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

Notice: For pedagogic purposes and for future students, the answers given here are more elaborate than the ones that would already have got full marks.

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 $V_i = 0$.*

   Yes. Since only total utility differences matter, you can add any constant c to all utilities, particularly $c = -V_1$ to make vanish V_1. Generally (for use also for the following questions), the choice probability can be expressed in terms of the deterministic utilities V_i and the random utilities ϵ_i making up the total utilities $U_i = V_i + \epsilon_i$ as

$$\begin{aligned} P_i &= \text{Prob}(U_i > U_j \text{ for all } j \neq i) \\ &= \text{Prob}(V_i + \epsilon_i > V_j + \epsilon_j \text{ for all } j \neq i) \\ &= \text{Prob}(V_i - V_j + \epsilon_i - \epsilon_j > 0 \text{ for all } j \neq i) \end{aligned} \tag{1}$$

2. *Without changing the choice probabilities, you can multiply all V_i with a common nonzero factor.*

   No. If you only multiply the deterministic utilities V_i with a common factor but not the random utilities ϵ_i, you will change the inequality (1) defining the general choice probabilities

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 .*

   Yes. If they are independent, subtracting ϵ_1 from all utilities is just a special case of Question 1.

4. *If there are I alternatives, only $I - 1$ deterministic and random utilities can be defined.*

   Yes. You *can* define I deterministic and random utilities. However, only $I - 1$ of them are independent.

5. *You can multiply* both *deterministic and random utilities for all alternatives with a common positive factor without changing the choice probabilities.*

   Yes. Multiplying all terms in the inequality of (1) with a common *positive* factor will not change the inequality, so the choice probability remains unchanged.

   Notice: This is valid even when applying a strictly monotonously increasing function with definition range $I\!R$ (such as e^x) to all $V_i + \epsilon_i$. However, you may not multiply with a *negative* factor or zero or apply a function that is not strictly monotonously increasing.

6. *The expectation value of the random utilities must be =0.*

   No. If they were $\neq 0$ as is the case for the Gumbel-distributed random utilities of the Logit model, you could just subtract the expectation as in Point 1 without changing the choice probabilities

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.*

   Yes. From (1), it follows for the binomial case

$$\begin{aligned} P_1 &= P(V_1 - V_2 > \epsilon_2 - \epsilon_1) \\ &= F_{\epsilon_2 - \epsilon_1}(V_1 - V_2), \end{aligned}$$

   i.e., the choice probability is just the distribution function $F(\epsilon)$ for the random utility difference $\epsilon = \epsilon_2 - \epsilon_1$. Possible correlations only appear in this distribution function. For the MNL, we have the well-known analytic expression $P_i = e^{V_i} / \sum_j e^{V_j}$

Problem 2 (45 points)

(a) The reason is the *(no-)response bias.* Unlike people recruited by telephone or with a personal one-off link, the interviewer has no control over the interviewed people. This is problematic since the propensity to respond to the survey may be correlated in an unknown and systematic way with the personal preferences, i.e., with the result

Notice: The socioeconomic variables obtained during the survey may be used to partially compensate for the response bias but unknown factors remain

(b) In the standard discrete-choice models, the alternative set must be exclusive and complete.

- Exclusivity, i.e., at most one alternative may be chosen excludes multi-modal trips (e.g., bicycle-train). Solution: define as alternatives the mainly used mode.
- Completeness, i.e., at least one alternative must be chosen: Particularly, in a revealed-choice setting, there are other options, e.g., motorcycles or working in homeoffice. Solution: Two further alternatives 5: other means of transport; 6: did not travel.

(c) Advantages and disadvantages of a Revealed-Preferences (RP) with respect to a Stated-Preferences (SP) design

+ More realistic since, in contrast to the hypothetical SP decisions, actually performed choices are queried,
+ less biased since it is harder to actually lie (e.g., pretending to have done the trip by bike while actually having used a car) than give "socially desired" answers in a hypothetical context,
- the characteristics of the not chosen alternatives must be obtained separately (people who only use the car will not know how long a bus trip will take) while, in SP, the choice set defines everything,
- less efficient use of the previous interviewee's time: in SP, one could dynamically go to the "tipping point" of one's decisions (and ask about several situations) while, in RP, only one option may be viable, (e.g. for a long commute in a region without public transport), so no information can be derived in obtaining it.

(d) *Classify each of the variables (i) to (viii) as one of the following:*

- (a) alternative-specific constant: none
- (b) characteristic: (iv),(v)
- (c) socioeconomic variable: (i), (ii), (vii), (viii)
- (d) external variable: (vi)
- (e) endogenous variable: (iii)

(e) The gender g cannot be formulated generically as $V_i^{\mathrm{g}} = \beta_0 g$ since this does not distinguish between alternatives, so the value of the dummy $g = 0$ or $=1$ just drops out since only utility differences matter.

Hint: Generally, socioeconomic or external variables must be formulated in an alternative-specific way (as done in problem part (f)) or by interactions with characteristics, or both (problem part (g))

(f) $\beta_1 < 0$ and $\beta_2 < 0$ since, in bad weather ($w = 1$ instead of $=0$), the relative preference to the reference alternative "car" drops for both pedestrians and cycling.

Utility increase of car driving relativ to cycling if the weather turns bad:

$$\Delta V = (V_4 - V_2)_{\text{bad weather}} - (V_4 - V_2)_{\text{nice weather}} = -\beta_2 > 0$$

(g) Interaction of the weather variable with the travel time T_i reflects the fact that the weather-related utility differences increase with travel time:

$$V_i^{\text{w}} = \beta_i T_i w$$

Hint 1: All four β_i are independent and well-defined since utility differences (the only thing that matters) cannot be written in terms of parameter differences: For example, the difference $V_{ni}^{\text{w}} - V_{n4}^{\text{w}} = w(\beta_i T_{ni} - \beta_4 T_{n4})$ does not allow to express β_4 in terms of $\beta_1, ..., \beta_3$ for all decisions/persons n

Hint 2: A generic formulation $V_i^{\text{w}} = \beta w T_i$ (with a common β) is technically possible (since the interaction with T_i distinguishes between alternatives) but not realistic: Then, the weather would just globally change the travel time sensitivity for all modes

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) The "none" entry for the speed limit of the 8^{th} data tuple is plausibly set to the average desired speed of all the drivers on this road (which plausibly can only be a section of a German Autobahn), for example 130 km/h. The use of macroscopic variables is justified since the endogenous variable (traffic flow/road capacity) is macroscopic

(b) According to the problem statement, there is a maximum inside the definition range of the speeds since, for zero speed limit $x_1 = 0$, the flow/capacity is zero, and for very high limits x_1, the capacity reduces due to the traffic disturbances caused by the speed differences. However, a linear function cannot have a maximum inside its definition interval

(c) – β_0: Intercept without meaning (would be the capacity for an Autobahn with zero trucks and a speed limit of zero which is clearly outside of the applicability range of this model

– β_1 and β_2: Slope (> 0) and curvature $\beta_2 < 0$ of the parabola-shaped dependence of the capacity with the speed. The maximum is at a speed limit $-\beta_1/(2\beta_2$ (equal to 100 km/h for the numerical values below)

– $\beta_3 < 0$: capacity difference of city streets with respect to freeways: Since streets have traffic lights and other perturbations that do not exist on freeways, the difference is negative

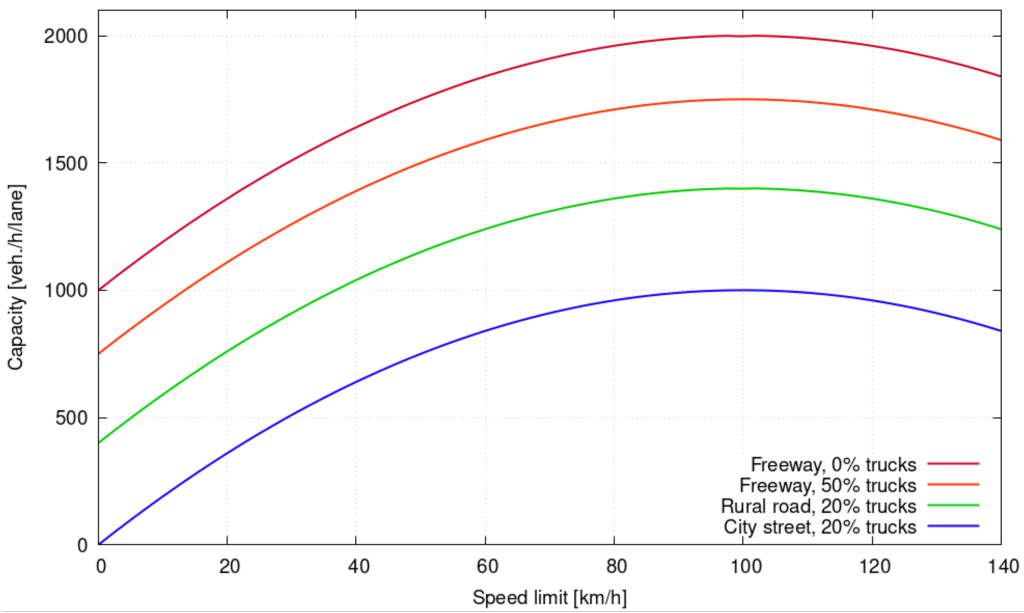
– $\beta_4 < 0$: capacity difference of extraurban roads with respect to freeways. Also this difference is negative since roads have more curves and worse visibility than freeways

– $\beta_5 < 0$: dependence of the capacity on the truck percentage: More trucks means a lower flow, so $\beta_5 < 0$

Remark: One can define a *passenger-car equivalent* (pce) of trucks by

$$\text{pce} = \frac{\text{capacity (100\,\% cars)}}{\text{capacity (100\,\% trucks)}} = \frac{y(x_3 = 0)}{y(x_3 = 100)} = \frac{y(x_3 = 0)}{y(x_3 = 0) + 100\beta_5}$$

Remark 2: For illustrative purposes, here is a plot of typical outcomes of this estimated model:



(d) Similarly as in discrete-choice situations, a third dummy β_6 times a freeway dummy will not be independent since we already have three constants to distinguish between the three roads: β_0, β_3, and β_4: The value of β_6 can be absorbed into β_0: simultaneously increasing β_3, β_4, and β_6 by a constant c is equivalent to decreasing β_0 by c

(e) This will model a multiplicative effect of trucks. For example, 100 % of trucks would half the capacity compared to no trucks instead of reducing the capacity by a constant amount.

Remark: In this approach, we would also have a cosntant $\text{pce} = (1 + 100\beta_5)^{-1}$

(f) Estimated capacities in veh/h/lane:

   (i) City, speed limit 40 km/h, 20 % trucks: $\hat{y} = \hat{\beta}_0 + 40\hat{\beta}_1 + 1\,600\hat{\beta}_2 + \hat{\beta}_3 + 20\hat{\beta}_5 = 640$

   (ii) Freeway (120 km/h) on Sunday: $\hat{y} = \hat{\beta}_0 + 120\hat{\beta}_1 + 14\,400\hat{\beta}_2 = 1\,960$.

(g) In significance tests, null hypotheses can only be rejected, not supported. Hence, it is not a good idea to test for a positive linear speed dependence since this is expected anyway and surely cannot be rejected at an error probability of $\alpha = 5\,\%$ if the estimated value is positive.

   Test of a negative linear speed dependence:

   (i) H_0: $\beta_1 < 0$

   (ii) Test statistic $T = \frac{\hat{\beta}_1}{\sqrt{\hat{V}_{11}}} \sim T(4)$ since we have $n = 10$ data sets and $m = 6$ parameters to estimate

   (iii) Realisation $t = \sqrt{20} = 4.472$

   (iv) Decision: H_0 rejected if $t > t^{(4)}_{1-\alpha} = t^{(4)}_{0.95} = 2.132$ which is the case

(h) In order to calculate the confidence interval for the estimate $\hat{y}$ itself, one needs the variances and covariances of all relevant parameter estimates. For a freeway with no trucks and a speed limit of 100 km/h, we have (in veh/h/lane)

$$
\begin{aligned}
\hat{y} &= \hat{\beta}_0 + 100\hat{\beta}_1 + 10\,000\hat{\beta}_2 = 2\,000, \\
\hat{V}(\hat{y}) &= \hat{V}_{00} + 10^4\hat{V}_{11} + 10^8\hat{V}_{22} + 2(100\hat{V}_{01} + 10\,000\hat{V}_{02} + 10^6\hat{V}_{12} = 438\,100
\end{aligned}
$$

Hence, the α- confidence interval (CI) reads

$$\mathrm{CI}_\alpha = [\hat{y} - \Delta\hat{y}_\alpha, \hat{y} + \Delta\hat{y}_\alpha]$$

with

$$\Delta\hat{y}_\alpha = t^{(4)}_{1-\alpha/2}\sqrt{\hat{V}(\hat{y})}$$

so, with $\alpha = 5\,\%$, we have

$$\Delta\hat{y} = 2.776\sqrt{438\,100} = 1\,838, \quad \mathrm{CI} = [162, 3838].$$

Obviously, the number of data sets (just $n = 10$ for six parameters to estimate) is not sufficient for a meaningful estimate