Solutions to the Examination for the Master's Course Methods of Econometrics, winter semester 2023/24

Problem 1

Person	1	2	3	4	5	6	7	8	9	10
Cost PT [€]	2.00	2.00	2.00	0.00	0.00	2.00	2.00	2.00	2.00	0.00
Cost car [€]	2.00	2.00	2.00	2.00	2.00	3.00	2.00	2.00	3.00	2.00
Time PT [min]	30	40	50	40	50	60	30	40	50	40
Time car [min]	30	30	30	30	30	30	30	30	30	25
Daytime	day	night	day	night	day	night	day	night	night	night
Gender	m	m	m	m	m	m	f	f	f	f
Decision	PT	PT	car	PT	PT	car	PT	car	PT	car

(a) Stated choice/preverence because it is asked "under which circumstances they would ...", so it is a *hypothetical* situation

(b) Characteristics: Costs C_i and times T_i; socioecomomic variable: gender G (0=male, 1=female); external variable: daytime D (0=day, 1=night).

(c) Given: binomial Logit model with

$$V_i = \beta_1 \delta_{i1} + \beta_2 C_i + \beta_3 T_i + \beta_4 D + \beta_5 GD,$$

The characteristics T_i and C_i are modelled in a generic way since the same parameters β_2 and β_3 are defined for the costs and times for both alternatives, respectively.

(d) Meaning of the parameters:

- β_2: cost appraisal/sensitivity, < 0 expected
- β_3: travel time sensitivity, < 0 expected
- β_4: additional PT preference (for men) in the night compared to day (no clear sign)
- β_5: additional PT preference in the night for woman compared to men (generally negative because of security)
- *note* (not required): $\beta_4 + \beta_5$: additional PT preference for women in the night compared to day

(Not required): The intercept β_1 means the ad-hoc preference PT over cars for males at daylight if the trip costs zero and can be performed instantaneously (*"free beaming"*).

Hint: Be specific! Something like β_4: parameter for the night or β_5: "influence of the weather" will not do!

(e) Parameter values as given in the problem statement:

$$\hat\beta_1 = 3.12, \quad \hat\beta_2 = -1.08, \quad \hat\beta_3 = -0.199, \quad \hat\beta_4 = 0.39, \quad \hat\beta_5 = -2.42$$

For the first person, we have

$$C_1 = C_2 = 2,\ T_1 = T_2 = 30,\ D = 0, G = 0.$$

Inserting this in the model in (c), we have

$$V_1 = -5.01, \quad V_2 = -8.13$$

Still easier: One could also calculate directly the utility difference $V_1 - V_2 = \hat{\beta}_1 = 3.12$).

Using the general equation for the binary Logit model (MNL for 2 alternatives), we have

$$P_1 = \frac{e^{V_1}}{e^{V_1} + e^{V_2}} = \frac{1}{1 + e^{V_2 - V_1}} = 0.958, \quad P_2 = 1 - P_1 = 0.0423.$$

(f) Property sums as given by the data and estimate for the model with $\hat{\boldsymbol{\beta}} = \mathbf{0}$ resulting in $P_1 = P_2 = 1/2$:

- Factor 1:
  $X_1^{\text{data}} = \sum_n \sum_i y_{ni} \delta_{i1} = \sum_n y_{n1}$,
  $X_1^{\text{MNL}}(\hat{\boldsymbol{\beta}} = \mathbf{0}) = \sum_n \sum_i y_{ni} P_{ni} \delta_{i1} = \sum_n y_{n1}/2 = N_1/2$,
  so X_1 = #realized/modelled decisions for PT:
  $$X_1^{\text{data}} = N_1 = 6, \quad X_1^{\text{MNL}}(\hat{\boldsymbol{\beta}} = \mathbf{0}) = N/2 = 5$$
- Factor 2: X_2= # realized/modelled total cost:
  $$X_2^{\text{data}} = 17, \quad X_2^{\text{MNL}}(\hat{\boldsymbol{\beta}} = \mathbf{0}) = 36/2 = 18$$
- Factor 4: X_4= # realized/modelled # PT decisions in the night:
  $$X_4^{\text{data}} = 3, \quad X_4^{\text{MNL}}(\hat{\boldsymbol{\beta}} = \mathbf{0}) = 6/2 = 3$$
- Factor 5: X_5= # realized/modelled # PT decisions by women in the night:
  $$X_5^{\text{data}} = 1, \quad X_5^{\text{MNL}}(\hat{\boldsymbol{\beta}} = \mathbf{0}) = 3/2 = 1.5.$$

Note: There was a printing error in the exam formulation (the δ_{i1} selectors were missing at the $\beta_4 D$ and $\beta_5 GD$ terms of the utility specification) changing the values of the factors 4 and 5. Of course, the changed values got full marks, if correct.

(g) Value of time:

$$\text{VoT} = \hat{\beta}_3/\hat{\beta}_2 = 0.184\,€/\text{min} \quad \text{or} \quad \text{VoT}[€/\text{h}] = 60\,\hat{\beta}_3/\hat{\beta}_2 = 11.06\,€/\text{h}.$$

The additional willingness of woman to pay (WtP) for a car in the night compare to men is given in utility units by $-\hat{\beta}_5$. In €, it is given by

$$\text{WtP} = \frac{-\hat{\beta}_5}{-\hat{\beta}_2} = 2.24\,€.$$

(h) Likelihood-ratio test for the full model (c) with 5 parameters compared to the restrained model without daytime and gender effects (3 parameters):

1. H_0: Both model have the same explanative power.
2. Test function $T = 2(\tilde{L}_{\text{full}} - \tilde{L}_{\text{restr}}) \sim \chi^2(5-3)|_{H_0}$.
3. Realisation: $t_{\text{data}} = 2(-4.26 + 6.27 = 4.02$.
4. Decision: H_0 rejected at $\alpha = 5\,\%$ if (cf. the χ^2 quantile table on the last page)
   $$t_{\text{data}} > \chi^2_{2,\,0.95} = 5.99 \;\Rightarrow\; \text{cannot be rejected.}$$

Remark (not required): Not even the trivial model with no parameters ($\hat{\boldsymbol{\beta}} = \mathbf{0}$) is significantly worse than the full model:

$$T \sim \chi^2(5)|_{H_0}, \quad t_{\text{data}} = 2(-4.26 + 6.93 = 5.34, \quad \chi^2_{5,\,0.95} = 11.07 \;\Rightarrow\; \text{no rejection.}$$

Hint: Variance-based tests such as the χ^2, likelihood-ratio or F tests are always testing for small deviations/variances, hence they are one-sided tests with the quantile $t_{1-\alpha}$ the relevant one.

Problem 2

Given model for the energy consumption on aroad segment of length L:

$$y = \beta_1 L + \beta_2 T + \beta_3 h + \beta_4 L v^2 + \beta_5 L \vartheta + \epsilon, \quad \epsilon \sim \text{ i.i.d. } N(0, \sigma^2).$$

(a) All factors are proportional to L, T, or h. Therefore, an intercept would be the energy demand for a segment of length $L = 0$ and altitude increase $h = 0$ covered in the travel time $T = 0$ which, obviously, needs no energy.

(b)
- β_1: constant energy demand per length, i.e. (see the problem statement) a resistance force on level roads not depending on speed, i.e., β_1 is proportional to the constant friction force. > 0 expected.

  Note (not required): more precisely, we have the friction force $F = mg\mu$ allowing to estimate the friction coefficient $\mu = \beta_1/(mg)$, e.g., $\mu = 0.015$ for a car weighting $1\,400$ kg. > 0 expected.
- β_2: energy demand per time (power) if $L = 0$ and $h = 0$, i.e., β_2 is equal to the standby power P_0 (ventilation, AC, electrical appliances, radio) for a standing but active car. > 0 expected.
- β_3: According to the problem statement, the energy for climbing is the gravitational force times h, so β_3 is equal to the gravitational force mg, i.e., (*not required*) the average mass of the tested vehicles can be estimated as $m = \beta_3/g = 1\,600$ kg. > 0 expected.
- β_4 is proportional to the force (energy contribution per length) proportional to v^2, i.e., relates to the wind drag. > 0 expected.

  Note (not required): more precisely, the wind drag force is given by $F_w = 1/2 c_d \rho_{\text{air}} A v^2$ allowing to estimate the wind drag coefficient $c_d = 2\beta_4/(\rho_{\text{air}} A)$ with the air density $\rho_{\text{air}} = 1.3\,\text{kg/m}$ at sea level and the frontal area $A = 2\,\text{m}^2$, typically. With these values, we obtain $c_d = 0.31$ which is a typical value for modern cars.
- β_5 is proportional to the speed variance and characterizes the additional energy losses by the braking maneuvers (note that β_5 is smaller for electrical vehicles which can recuperate part of the kinetic energy instead of transforming it to heat at the brakes).

Hint: As always when explaining parameters: be specific: Describe the parameters, not the factors. For example β_2: "time sensitivity" or β_2: "related to time" will not give any points in the future!

(c) $n = 25$ data points, LSE estimate (expectation $\pm$ standard deviation)

$$\hat\beta_1 = 210 \pm 50, \quad \hat\beta_2 = 2\,100 \pm 700, \quad \hat\beta_3 = 16\,000 \pm 2\,000, \quad \hat\beta_4 = 0.4 \pm 0.1, \quad \hat\beta_5 = 0.5 \pm 0.3.$$

speed $v = L/T = 25\,\text{m/s}$, needed energy

$$y = 1000\hat\beta_1 + 40\hat\beta_2 + 30\hat\beta_3 + 1000 * 25^2 \hat\beta_4 + 0 = 1\,024\,000\,\text{Ws}$$

*Hint: 3 %*1000 m = 30 m, not 0.3 m!*

(d) As above, we obtain

$$y = 1000\hat\beta_1 + 40\hat\beta_2 - 60\hat\beta_3 + 1000 * 25^2 \hat\beta_4 + 0 = -416\,000\,\text{Ws}$$

This energy is negative because the gravitational downhill force is greater than the friction force, the wind drag, and the force to generate the standby power together. In internal combustion vehicles (gasoline, Diesel) this gravitational energy is lost and the model breaks down.

Note (not required): In battery-electric vehicles, the gravitational energy can be partially recuperated in battery-electric vehicles. The model remains valid for full recuperation; otherwise, it needs to be modified.

(e) Since the value of the speed in km/h is 3.6 times the value in m/s, the parameter $\hat{\beta}_4$ and $\hat{\beta}_5$ need to be multiplied by $1/3.6^2$ (speed squared and speed variance in $(\text{km/h})^2$ have 3.6^2 the value of speed squared and speed variance in $(\text{m/s})^2$ but the energy is, of course, unchanged).

(f) Why it is a bad idea to add an additional factor $\beta_6 vT$? The associated factor $vT = L/T * T = L$ is identical to the factor of β_1, so we would have multi-colinearity violating one of the data Gauß-Markow assumptions.

(g) $H_{01} : \beta_2 = 0$:

$$T = \frac{\hat{\beta}_2}{\sqrt{\hat{V}_{22}}} \sim T(25-5) = T(20)|_{H_0}, \quad t_{\text{data}} = \frac{2\,100}{700} = 3.$$

H_0 is rejected if

$$|t_{\text{data}}| > t_{0.975}^{(20)} = 2.086 \;\Rightarrow\; \text{rejected.}$$

$H_{02} : \beta_2 < 1\,000$:

$$T = \frac{\hat{\beta}_2 - 1\,000}{\sqrt{\hat{V}_{22}}} \sim T(25-5) = T(20)|_{H_0}, \quad t_{\text{data}} = \frac{1\,100}{700} = 1.57$$

H_0 is rejected if

$$t_{\text{data}} > t_{0.95}^{(20)} = 1.725 \;\Rightarrow\; \text{not rejected.}$$

Notice: In the second test, we have t_{data}, not $|t_{\text{data}}|$, and the critical value is given by the 95^{th} percentile, not the 97.5^{th} percentile because the errors over which to distribute the error probability $\alpha = 5\,\%$ can only be on the right-hand side, not on both.

(h) Expectation and variance of the parameter $\beta_4' = \beta_4 + \gamma\beta_5$ of the offline model, with constant $\gamma = 0.4$

Expectation:

$$\hat{\beta}_4' = \hat{\beta}_4 + \gamma\hat{\beta}_5 = 0.6,$$

Variance:

$$V(\hat{\beta}_4') = V(\hat{\beta}_4) + \gamma^2 V(\hat{\beta}_5) + 2\gamma V_{45} = 0.1^2 + \gamma^2 0.3^{-} 2\gamma V_{45} = 0.0196.$$

Problem 3 (30 points)

Given is an e-bike with following main components:

- Iron/steel (motor, chain, gear shift, etc): 7 kg (CO_2-neutral recycling: 30 %)
- Aluminum (frame, wheels): 12 kg (recycling 50 %)
- Battery: 5 kg (recycling 0 %)
- Other materials (rubber, plastic etc): 1 kg (recycling 20 %).

(a) Life-cycle repository with an energy need of $y_5^s = 1/15\,\text{kWh/km} * 30\,000\,\text{km} = 2\,000\,\text{kg}$:

$$\begin{aligned}\boldsymbol{y}^s = \boldsymbol{y}^s_{\text{prod}} + \boldsymbol{y}^s_{\text{drive}} + \boldsymbol{y}^s_{\text{rec}} &= \begin{pmatrix} 7\,\text{kg} \\ 12\,\text{kg} \\ 5\,\text{kg} \\ 1\,\text{kg} \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 5\,\text{kg} \\ 0 \\ 2\,000\,\text{kWh} \end{pmatrix} + \begin{pmatrix} -2.1\,\text{kg} \\ -6\,\text{kg} \\ 0 \\ -0.2\,\text{kg} \\ 0 \end{pmatrix} \\ &= \begin{pmatrix} 4.9\,\text{kg} \\ 6\,\text{kg} \\ 10\,\text{kg} \\ 0.8\,\text{kg} \\ 2\,000\,\text{kWh} \end{pmatrix}\end{aligned}$$

Note: The energy demand in the exam formulation was unrealistically low (0.1 kWh per 15 km instead of 1 kWh per 15 km) and it is now increased by a factor of 10 in the repository. Of course, the correspondingly different solutions got full marks if correct.

(b)

$$e_{CO_2} = e_1 = \boldsymbol{C}'\boldsymbol{y}^s = 1\,261\,\text{kg}$$

(c)
- Scenario I, charging in France at $C_5 = 50\,\text{g}\,CO_2/\text{kWh}$ instead of $C_5 = 400\,\text{g}\,CO_2/\text{kWh}$:

$$e_1 = \boldsymbol{C}'\boldsymbol{y}^s = 561\,\text{kg}$$

- Scenario II, Use a smaller battery/a durable battery: $y_3^s = 5\,\text{kg}$ instead of 10 kg:

$$e_1 = \boldsymbol{C}'\boldsymbol{y}^s = 1\,111\,\text{kg}$$

- Scenario III, tread more, so a total energy of $y_5^s = 1\,000\,\text{kWh}$ instead of 2 000 kWh:

$$e_1 = \boldsymbol{C}'\boldsymbol{y}^s = 861\,\text{kg}.$$

(d) Conventional bike with the same 30 000 km of riding: The rider needs additional food of the amount y_s in carbohydrates which, per kg of carbohydrates, leads to $C_5 = 3.5\,\text{kg} + 2.5\,\text{kg}$ of CO_2 emissions. Thus, we have

$$y_5^s = 30\,000\,\text{km} * 0.004\,\text{kg/km} = 120\,\text{kg carbohydrates}$$

and

$$C_5 = 3.5\,\text{kg/kg} + 2.5\,\text{kg/kg} = 6\ [\text{kg } CO_2 \text{ per kg of carbohydrates}]$$

Thus, without a battery and carbohydrates as "fuel"

$$\boldsymbol{y}^s = \begin{pmatrix} 1.4\,\text{kg} \\ 4\,\text{kg} \\ 0 \\ 0.8\,\text{kg} \\ 120\,\text{kg} \end{pmatrix}, \quad \boldsymbol{C} = \begin{pmatrix} 2 \\ 25 \\ 30 \\ 1 \\ 6 \end{pmatrix}.$$

The usual scalar product leads to a CO_2- emission of

$$e_1 = \boldsymbol{C}'\boldsymbol{y}^s = 824\,\mathrm{kg}$$

of which the greatest part $C_5 y_5^s = 720\,\mathrm{kg}$ are emitted by additional human breathing during driving. Remarkably, the ecological CO_2- footprint of conventional non-motorized bicycles over the lifetime is only a little better than that of e-bikes in Germany, and worse than e-bikes in France. The reason is that the "fuel combustion" of carbohydrates or other nutrients in the human body has a very low efficiency, and that making the food (in a way corresponding to the w2t emissions of gasoline or Diesel) is very CO_2 intensive as well. This overcompensates for the lighter bicycle and the missing CO_2-intensive battery.