Solutions to the
online exam for the Masters Course
Methods of Econometrics, winter semester 2020/21

Problem 1 (20 points)

An econometrist wants to investigate the influencing factors for the amount of bicycle usage y: bicycle trips per day and per person in Germany and Austriay using a regression model. Please mark all of the following statements that are true

- The country is a nominal or qualitative variable and one can model it using a dummy variable $x_1 = 1$ for Germany and x_1=0 for Austria ✓ [This is just the normal dummy for a dichotomic (binary) qualitative variable]

- While it is somewhat unorthodox, you could use the dummy $x_1 = 42.105$ for Germany and $x_1 = 0$ for Austria ✓ [The dummy needs not to be 1 or 0. You can use any nonzero value provided you set the contribution=0 for the reference value. Only the parameter changes accordingly such that the product parameter*dummy is fixed]

- Since Germany is a much bigger country, you cannot use $x_1 = 0$ for Germany and $x_1 = 1$ for Austria ✗ [You can use a single dummy variable whenever there is a binary variable, regardless of context]

- It is possible to include the influence of the season by setting $x_2 = 1, 2, \ldots, 12$ for January, February, ... December ✗ [Same reasoning as for the age: non-monotonous dependence, hence no direct linear modelling possible

- It is possible to use the binary dummy $x_2 = 1$ for winter, and $x_2 = 0$ for the other seasons ✓ [This is just the usual dummy for a dichotomic variable]

- When chosing $x_2 = 0$ for winter and =1 for the other seasons instead, the assocoated parameter $beta_2$ would change its sign ✓ [The sign of β_2 is changed because now the difference when going from summer to winter is no longer $+\beta_2$ but $-\beta_2$ while the contribution itself, of course, must be unchanged.]

- The proximity [Nähe] to cities is a categorial variable and can be modelled by $x_3 = 0$ inside city, $x_3 = 1$ near a city, $x_3 = 2$: far away from a city ✗ ✓ [There is a natural order (the proximity), so x3 is a categorial variable. If the steps can be considered to be equal, it can be used directly (ordinal number as x3=solution). However, treating it as a qualitative variable is always possible. So I also considered also the answer "no" to be correct]

- Since the age is a metric variable, you can include its influence on y by the contribution $\beta_4 x_4$ with x_4 the age in years ✗ [The age is a metric variable but its influence is not monotonous (first increasing, then decreasing). Hence a direct linear use is not possible]

- If the effect of the gender is modelled by the contribution $\beta_5 x_5$ with $x_5 = 0$ (male) and =1 (female) and $\beta_5 < 0$ is significant, this means that male persons use the bicycle more than female persons (watch out!) ✗ [The age is a metric variable but its influence is not monotonous (first increasing, then decreasing). Hence a direct linear use is not possible]

- If $\beta_5 < 0$ is significant, men use the bike more often than women if all other conditions are the same ✓ [if $\beta_5 < 0$, then female ($x_5 = 1$) have a negative contribution while the contribution for men is zero, so men use the bike more often if ceteris paribus (all other factors are equal)]

Problem 2 (14 points)

In an epidemic spread model, we have three sub-models:
Model A: the number of new infections in week n is proportional to the product of the number of new infections and the percentage [Anteil] of susceptible [infizierbare] persons in the past week $n-1$
Model B: the percentage of susceptible persons in week $n-1$ is given by 100
Model C: the number of Covid-19 deaths in week n is proportional to the number of new infections in week $n-3$
Please tick [whlen Sie aus] all of the following correct statements :

- The exogenous variables of Model A are the number of infections and the percentage of susceptible persons in week $n-1$ ✓ [Is in the text]
- Model A is chained to itself (time evolution) ✓ [The number of infections in week n depends on the number in week $n-1 \Rightarrow$ chained time evolution]
- Model A is linear ✗ [The input of Model A is the *product* of the number of infections and the number of susceptible persons. This can be called bilinear but not linear]
- Models B and C are linear ✓ [In Model B, the exogenous variables/factors "weekly new infections") enter as a sum ("cumulated number") with prefactor $1/n$ ($n =$ population number) $\Rightarrow$ linear. In Model C, the exogenous variable new infections three weeks ago enters "proportional". Since no other exogenous variables are mentioned, the proportionality factor is constant and this model linear as well.]
- All models A, B and C are stochastic ✗ [These are macroscopic models (output percentages or numbers) which all are deterministic]
- Model B is macroscopic while the models A and C are microscopic ✗ [All three models are macroscopic]
- The exogenous variables of Model B for week n are all the $n-1$ infection numbers of the past weeks ✓ [Stands in the text]
- Models B and C are chained to Model A. ✗ [Only Model C is chained to model A, not B: The percentage of susceptible persons influences the new infection but not the number of deaths which is determined by the infections alone.]
- Model C is chained to Model B ✗ [Model C is only chained to Model A]
- There is a feedback link from B to A ✓ [The percentage of susceptible persons calculated in B is fed back as an input to Model A]
- There is a feedback link from C to A ✗ [Model C (the number of deaths) depends on A but the number of deaths do not incluence the infection process A]
- Model C is not needed to model the infection spread ✓ [Model C is not needed to model the infection spread: correct Model C only describes the consequences (deaths) but does not couple back to A or B. Hence, it is not needed to describe the infection spread (#infections) alone. This is also true when including ICUs [Intensive car units, Intensivstationen] coming at their limits: Then, the case fatality rate would increase making model C depend nonlinearly on the number of infections but still no feedback.]

Problem 3 (14 points)

In summer, vacation will probably be possible but under different conditions as usual. Therefore, a travelling agency plans to model demand shifts of their customers using the MNL. The agency offers following categories: the Alpes, North Sea, city trips, and low mountain ranges [Mittelgebirge] in Germany. Please tick all correct statements

- Both deterministic and random utilities can be exponentiated, $V_i \to \exp(V_i), \epsilon_i \to \exp(\epsilon_i)$, without changing the choice probabilities. ✓ [Since exponentiation is a positive strictly monotonously increasing function, the *argument* of the maximum of $(V_i + \epsilon_i)$, hence the probabilities, are not changed]
- Both deterministic and random utilities can be multiplied by 22 without changing the choice probabilities ✗ [Multiplication by a negative number makes a minimum out of a maximum and vice versa]
- The alternatives are sufficiently different to allow using the MNL ✗ ✓ [Both answers gain points.Yes: North Sea, cities and Alps/low mountain ranges: obvious; Furthermore, the destination Alps and low mountain ranges differ in their location and also the usual age group. No: It can be argued that the Alps and the low mountain ranges have many commonalities, e.g., you can hike in both of them. Unfortunately, a justification/discussion could not be given]
- For two customers, the calibrated choice model predicts customer 1: P1=0.2, P2=0.3, P3=0.1,P4=0.4, customer 2: P1=0.6, P2=0.15, P3=0.05, P4=0.2. Does this model satisfy the IIA condition? ✓ [IIA (independence of irrelevant alternatives) means the relative preference P_1/P_2 of two alternatives does not depend on other alternatives. Here, alternative 1 increased its attractivity and as consequence, the other three loose in popularity by 50%, so P_2/P_3, P_2/P_4, and P_3/P_4 remain constant]
- The alternative set is complete ✗ [At least, the no-travel option is missing]
- The alternatives are exclusive ✓ [Four distinct and unique destinations are offered]

Problem 4 (36 points)

In a binary choice situation bicycle (alternative $i = 1$) vs. public transport (PT, alternative 2), the utilities depend on the total travel times T_i (minutes), the ad-hoc costs C_i (Euro), and the gender g_n of the decision maker ($g_n = 0$ if male and =1, otherwise). The model is given by

$$\begin{aligned} V_1 &= \beta_0 + \beta_1 g_n + \beta_2 C_1 + \beta_3 T_1, \\ V_2 &= \beta_2 C_2 + \beta_4 T_2 \end{aligned}$$

The calibrated parameters for a Logit model are

$$\hat{\beta}_0 = 1.3, \quad \hat{\beta}_1 = 0.6, \quad \hat{\beta}_2 = -1, \quad \hat{\beta}_3 = -0.2, \quad \hat{\beta}_4 = -0.1.$$

- In a stated-choice situation, 15 out of 40 female and 32 out of 60 male decision makers chose the bicycle. Give following property sums that a calibrated Logit model needs to reproduce:

  Factor of β_0: Property sum [47] all persons chosing the bike

  Factor of β_1: Property sum [15] all females chosing the bike

- Give the probability that a male with semester ticket will chose the bicycle if, for both modes, the total travel time is 30 minutes [...] Here, the individual numerical value of the following formula has to be entered (the parameter values were different for each student, so are the numerical values):

  $$V_1 = \beta_0 + \beta_3 * 30, \quad V_2 = \beta 4 * 30, \quad P_1 = \exp(V1)/(\exp(V1) + \exp(V2))$$

- For the same travel times, calculate the bike riding probability for a woman who needs to pay 2 euro to use the public transport [...] The individual numerical value of the following formula has to be entered:

  $$V_1 = \beta_0 + \beta_1 + \beta_3 * 30, \quad V2 = \beta_2 * 2 + \beta_4 * 30, \quad P_1 = \exp(V1)/(\exp(V1) + \exp(V2))$$

- Give the implied value of time (VOT) in Euro per hour for bicycle riding [...] The numerical value of $60 * \beta_3/\beta_2$ (notice that β_3/β_2 gives the implied VOT in Euro per minute)

Following variances and covariances are given:

$$V(\hat{\beta}_1) = 0.08, \quad V(\hat{\beta}_2) = 0.16, \quad V(\hat{\beta}_3) = 0.01, \quad \text{and} \quad \text{Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right) = -0.02$$

Test following null hypotheses at the $\alpha = 5\%$ level

- H_{01}: There is no gender difference in the preference for riding the bike. $t =$ [...], critical value for t: [...], H_{01} rejected? [...]

  Point test of H_{01}: $\beta_1 = 0$:

  $$t = \frac{\hat{\beta}_1}{\sqrt{V(\hat{\beta}_1)}}, \quad t_c = z_{1-\alpha/2} = z_{0.975} = 1.96, \quad H_{01} \text{ rejected? yes if } |t| > t_c$$

  Notice that, in discrete-choice theory, we always assume the asymptotic limit where $t \sim N(0, 1)$. Depending on your individual parameter values, you enter "yes" or "no" in the last box (always act according to the "hints" appearing in the text box before you enter something into it!)

- H_{02}: The perceived costs are at least 0.2 utility units per Euro $t =$[...], critical value for t: [...] H_{02} rejected? [...]

  The *costs* are *at least* 0.2 utility units per Euro, i.e., the *deterministic utility increment* is *at most* minus 0.2 utility units per Euro, so we have an interval test of H_{02}: $\beta_2 < -0.2$:

  $$t = \frac{\hat{\beta}_2 - (-0.2)}{\sqrt{V(\hat{\beta}_2)}}, \quad t_c = z_{1-\alpha} = z_{0.95} = 1.64, \quad H_{02} \text{ rejected? yes if } t > t_c$$

  This test is never rejected since, for all parameter values, $t < 0$

- H_{03}: the value of time (VOT) β_3/β_2 for bicycle riding is at most 0.10 Euro/minute (6 Euro per hour). Test the equivalent hypothesis $\gamma = \hat{\beta}_3 - 0.1\hat{\beta}_2 \geq 0$: $\hat{\gamma} =$ [...], $V(\hat{\gamma}) =$ [...], $t =$ [...], critical value $t_c =$ [...], H_{03} rejected? [...]

  We have $\hat{\gamma} = \hat{\beta}_3 - 0.1\hat{\beta}_2$ and

  $$V(\hat{\gamma}) = V(\hat{\beta}_3) + 0.01 * V(\hat{\beta}_2) - 2 * 0.1 * \text{ Cov}\left(\hat{\beta}_2, \hat{\beta}_3\right) = 0.0156$$

  (The variances and covariances were the same for all), so

  $$t = \frac{\hat{\gamma}}{\sqrt{V(\hat{\gamma})}}, \quad t_c = -z_{0.95} = -1.64, \quad H_{03} \text{ rejected? yes if } t < t_c$$

  For all parameter values, this test is never rejected.

Problem 5 (16 points)

Given is a physical model for the required energy of electrical vehicles y [kWh/100 km]. This energy depends on the vehicle mass m, the speed V, the front area A, the air drag coefficient c_w, friction coefficient μ, and the power $P0$ needed for secondary consumers (radio, ventilation, air conditioning) as follows

$$y = a0 \left[\frac{P0}{V} + m\ g\ \text{ mu} + a1\ \text{ cw}\ A\ V^2 \right]$$

where g, $a0$ and $a1$ are known constants.

1. Assume one and the same vehicle and driver driving in different conditions. List all the exogenous variables that vary in this situation

   Variable 1 [...] Variable 2 [...]

   Again, follow the hint and enter just one of the symbols appearing in the formula such as P0, mu, a1, V, ... With the same vehicle and driver, quantities such as m (mass), cw (air drag coeficient) , mu (friction), A (vehicle front area), and a1 (prefactor of the air drag containing, e.g., the air density) are constant. What varies is the speed V and the standby power P0 (switching on and off radio, A/C, ventilation etc). So enter

   Variable 1 [V], Variable 2 [P0], or

   Variable 1 [P0], Variable 2 [V]

2. Assume now that the only exogenous variable is the speed V, so all factors of the linear regression model

   $$y = \sum_{j=0}^{2} \beta_j x_j + \epsilon$$

   are functions of V. Give the factors x_j and the associated parameters β_j in terms of $a0$, $a1$, $P0$, m, g, mu, and A

   Factor x_0 [a fixed function of V, e.g., V, V^2, 1, sin(V)] Parameter β_0 [example: P0*m/cw]

   Factor x_1 [a fixed function of V, e.g., V, V^2, 1, sin(V)] Parameter β_1 [example: P0*m/cw]

   Factor x_2 [a fixed function of V, e.g., V, V^2, 1, sin(V)] Parameter β_2 [example: P0*m/cw]

   The three factors are 1/V, 1, and V^2, and the associated β_j the rest of the corresponding term in the problem statement, so enter, e.g.,

Factor x_0: 1/V	Parameter β_0: a0*P0
Factor x_1: 1	Parameter β_1: a0*m*g*mu
Factor x_2: V^2	Parameter β_2: a0*a1*cw*A

   Of course, you can change the order without losing points, e.g., x_0: 1, x_1: 1/V