Methods in Transportation Econometrics and Statistics (Master)

Winter semester 2021/22, Tutorial No. 6

Problem 6.1: Estimating heteroskedastic data: vehicle heading

For present or novel driver assistance systems (e.g., responsive navigation, intersection assistant), it is crucial to know the actual vehicle heading (e.g. 0 degrees: North, 90 degrees: West) and the yaw rate (change of heading over time). There are three methods to measure the heading by car sensors:

- $\hat{y}_1$: Integrating the steering angle over time using steering wheel transmission and speed information,
- $\hat{y}_2$: directly by a yaw-rate sensor (as in smartphones),
- $\hat{y}_3$: evaluating the change of the GPS positions.

For a true change of heading by 30 degrees, all three methods give an unbiased estimate $E(\hat{y}_i) = 30$ but they have known heteroskedastic errors of $\sigma_1 = 2\,\text{deg}$, $\sigma_2 = 2\,\text{deg}$ and $\sigma_3 = 4\,\text{deg}$. The question is how to minimize the error by making use of all the three data sources.

(a) Argue why it is plausible to assume independence for the estimation errors.

(b) Give the variance of the arithmetic means $\overline{\hat{y}}$.

(c) Is it possible to obtain a better result by ignoring the "worse" sources or weighting them less? For this purpose, calculate the variance of the weighted mean using the weights $w_1 = w_2 = w$, $w_3 = 1 - 2w$ (why this expression for w_3) and minimize the variance by varying w.

(d) Expert question: Give the optimal ("efficient") weighting for n unbiased independent estimates $\hat{y}_i$ with generally different variances σ_i^2. *Hint:* use the technique of Lagrange multipliers.

Problem 6.2: Survey in the audience

In a stated-choice survey on the favoured transport mode for the route to the university (binary: alternative 1: self-powered, i.e., foot/bike; alternative 2: motorized, i.e., public transport, motorcycle or car), $n = 19$ participants were asked to give their favoiurite means of transport in different hypothetical situations. Attributes are the complex travel time T_i and the ad-hoc costs C_i. The result was as follows:

Hypothetical situation	Alternative 1: self-powered	Alternative 2: motorized	Chosen 1	chosen 2
1	30 min	30 min, free	3	16
2	30 min	40 min, free	6	13
3	30 min	50 min, free	19	0
4	30 min	30 min + 1 €	17	2
5	30 min	30 min + 2 €	19	0
6	20 min	30 min, free	14	5
7	10 min	30 min, free	19	0

The deterministic utilities for the two alternatives are specified as follows:

$$\begin{aligned} V_1 &= \beta_1 + \beta_2 T_1, \\ V_2 &= \beta_2 T_2 + \beta_3 C_2. \end{aligned} \tag{1}$$

(a) Give the meaning of the three parameters β_j, $j = 1, 2, 3$ and also, if applicable, the expected sign of the numerical value. Why there is no alternative-specific constant for the second alternative?

(b) The maximum-likelihood estimation for the i.i.d. Probit model (random utility components $\epsilon_i \sim i.i.d.N(0, 1)$) give following result (estimator ± standard error):

$$\hat{\beta}_1^{\mathrm{P}} = -1.90 \pm 0.46, \quad \hat{\beta}_2^{\mathrm{P}} = -0.229 \pm 0.041, \quad \hat{\beta}_3^{\mathrm{P}} = -3.67 \pm 0.71.$$

Test which of the factors are significant (the corresponding parameters are significantly different from zero at a level $\alpha = 5\,\%$). *Hint:* Test the null hypothesis $\beta_j = 0$ assuming an asymptotic normal distribution for the estimation errors.

(c) Give the choice probabilities predicted by the Probit model for the first three situations.

(d) Compare the Probit choice probabilities with that of the Logit model (same deterministic utilities, parameters estimated as $\hat{\beta}_1^{\mathrm{L}} = -2.42$, $\hat{\beta}_2^{\mathrm{L}} = -0.283$, $\hat{\beta}_3^{\mathrm{L}} = -4.59$).

(e) Multiplying the probit parameters by $\pi/\sqrt{6}$ nearly results in the logit estimators. Discuss if this is an incidental or systematic result by observing that (i) the density function of the Gumbel distributed logit random utilities (RUs) has a similar shape as the Gaussian distributed Probit RUs, (ii) the standard deviation of the Logit RUs is $= \pi/\sqrt{6}$ while that of the Probit model is =1.

(f) Show that the parameter ratios such as the value of time (VOT) $\hat{\beta}_2/\hat{\beta}_3$ are essentially the same in both models (discuss).

Problem 6.3: Questionnaire design for a conjoint analysis

In a stated-choice survey, the factors influencing the choice of the transport modes "public transport" and "car" for the daily commute to work shell be determined. Other modes are not relevant/available. The factors are the travel-time differences $T_{\text{car}} - T_{\text{PT}}$ with possible values $\{-20\,\text{min}, 0, 20\,\text{min}\}$ and the ad-hoc cost differences $C_{\text{car}} - C_{\text{PT}}$ with the values $\{-1€, 1€\}$.

(a) Give the questionnaire for the *full factorial design*.

(b) Check whether a questionnaire specified by the choice sets $(-20\,\text{min}, -1€)$, $(20\,\text{min}, -1€)$, and $(20\,\text{min}, 1€)$ satisfies the criteria for orthogonal design.