Prof. Dr. Kai Carstensen
Prof. Dr. Uwe Jensen

Online Examination in Econometrics I

(Winter Term 2020/21)

Examination regulation

April 12, 2021, 08:00

Preliminary remarks:

1. Please read these instructions carefully!
2. You are permitted to use any auxiliary tools.
3. Write your name and enrollment (matriculation) number on every sheet of paper!
4. Don't use a pencil!
5. **Round your solutions to 4 decimal places.**
6. For all tests use a significance level of 5%, if nothing else is specified.
7. You have 60 minutes in total to answer the exam questions.

Good luck!

Problem 1 (25 points)

Consider the following model explaining an employee's probability (in percentage points) of being retired

$$ret_i = \beta_0 + \beta_1 \cdot age_i + \beta_2 \cdot lninc_i + u_i \quad i = 1, \ldots, N. \tag{1}$$

where *ret* is a dummy variable that is 100 for a retired person and zero otherwise, *age* is given in years and *lninc* is the log of net household income in USD. Assume that we estimated the following parameters for a random sample of $N = 100$ current or former employees:

$$ret_i = -\underset{(1.5562)}{68.2} + \underset{(1.0221)}{2.64} \cdot age_i + \underset{(0.0344)}{0.07} \cdot lninc_i + u_i \quad i = 1, \ldots, N. \tag{2}$$

Standard errors are given in parentheses.

1. **(4P)** Test the significance of β_1 and β_2 separately.
2. **(7P)** Assume that the estimation using de-meaned data yields $\hat{\boldsymbol{\beta}} = \begin{pmatrix} 2.64 \\ 0.07 \end{pmatrix}$ again and $\widehat{Avar}(\boldsymbol{\beta}) = \begin{pmatrix} 1.0446 & 0.0324 \\ 0.0324 & 0.0012 \end{pmatrix}$. Perform a Wald test to test significance of β_1, β_2 jointly.
3. **(4P)** Why are the two separate tests in 1. not sufficient to test joint significance of β_1 and β_2 at a given level α? **Hint**: Recall that $\Pr(H_0 \text{ is rejected} | H_0 \text{ is correct}) = \alpha$. Assume that both decisions are independent and calculate $\Pr(\text{at least one } H_0 \text{ is rejected} | \text{both } H_0 \text{ are correct})$.

Now, assume that we estimated the following extended model

$$\begin{aligned} ret_i = &- 62.28 + 2.12 \cdot age_i - 0.48 \cdot female_i - 0.08 \cdot lninc_i \\ &+ 1.4 \cdot educ_i + 7.6 \cdot house_i + u_i \quad i = 1, \ldots, N \end{aligned} \tag{3}$$

with the following additional regressors: *female* is a gender dummy that is 1 for a female and zero otherwise, *educ* measures education in years, and *house* is a dummy that is 1 for a person that lives in their own house and zero otherwise.

4. **(3P)** Briefly explain why the sign of *lninc* is different after including additional regressors.
5. **(4P)** Calculate the retirement probabilities
   - of a 62 years old female earning 66,000 USD, having 18 years of education and living in her own house,
   - and of a 20 years old male earning 11,000 USD, having 14 years of education and renting his place.

   Are these probabilities meaningful? If not, formulate a requirement that our regression method should satisfy to deal with probabilities.
6. **(3P)** By checking the following Figure 1, do you think homoskedasticity would be a valid assumption? Why or why not?

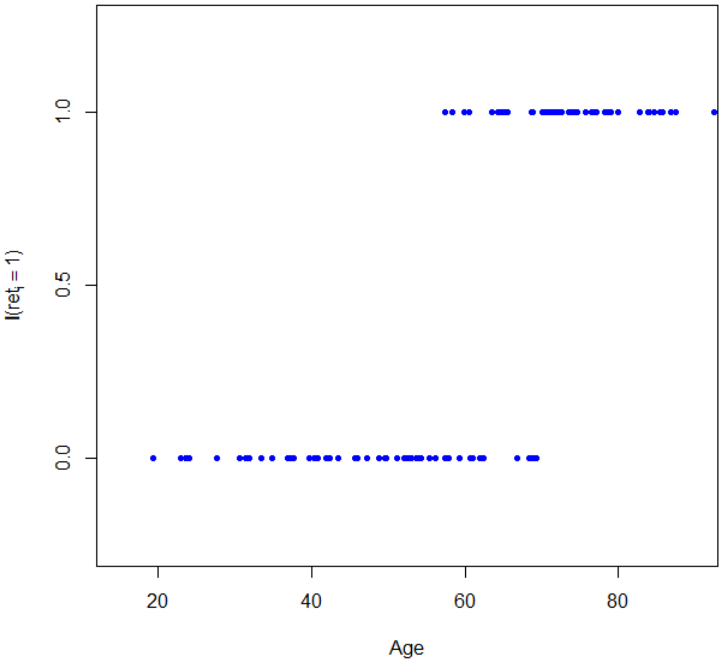

Figure 1: Observed retirement ages

Problem 2 (20 Points)

1. Suppose you are given the following model for a random sample of $N = 100$ students:

$$y_{1i} = \beta_1 x_{1i} + e_{1i} \tag{4}$$
$$y_{2i} = \beta_2 x_{2i} + e_{2i} \tag{5}$$

where y_{1i} denotes the reading skill (on a scale from 0 to 10), y_{2i} the computer skill (on a scale from 0 to 10), x_{1i} the number of books read annually and x_{2i} the average number of hours spent in front of the PC, all for student i. All variables are demeaned.

(a) **(2P)** Assuming that Ω is unknown, briefly describe how you would estimate $\hat{\Omega}$.

(b) **(10P)** The following pieces of information are given:

$$\hat{\Omega} = \begin{pmatrix} 0.74 & -0.31 \\ -031 & 0.29 \end{pmatrix}$$
$$\sum_{i=1}^{N} \begin{pmatrix} x_{1i}^2 & x_{1i}x_{2i} \\ x_{1i}x_{2i} & x_{2i}^2 \end{pmatrix} = \begin{pmatrix} 15200 & 8600 \\ 8600 & 13100 \end{pmatrix}$$
$$\sum_{i=1}^{N} \begin{pmatrix} x_{1i}y_{1i} & x_{1i}y_{2i} \\ x_{2i}y_{1i} & x_{2i}y_{2i} \end{pmatrix} = \begin{pmatrix} 7250 & 7100 \\ 9600 & 11100 \end{pmatrix}$$

You can assume that SOLS.1 and SOLS.2 are fulfilled. Use the most efficient estimation method to estimate β_1 and β_2. Interpret your results.

(c) **(3P)** Why could it be problematic to only include one variable in each equation? Would your estimators still be consistent and unbiased, if you include a covariable that is unrelated to your response variable? Briefly explain. No formal proof is needed.

2. **(5P)** You are interested in a precise, unbiased estimate of β in the linear model $y = x\beta + u$, where x is a scalar random variable with zero mean and $\mathrm{E}(u|x) = 0$. Suppose the literature reports OLS estimates $\hat{\beta}_1$ and $\hat{\beta}_2$ which are based on two independent random samples of sizes N_1 and N_2 taken from the same population. However, you do not have access to the samples, all you know are the OLS estimates and the sample sizes.
Find a linear combination of $\hat{\beta}_1$ and $\hat{\beta}_2$ that is unbiased and has minimum asymptotic variance. How does it compare to the (infeasible) OLS estimator applied to the joint sample of size $N_1 + N_2$?

Problem 3 (15 points)

We want to fit the seminal Black-Scholes (1973) model to a sample of $i = 1, \ldots, N$ binary call options where the underlying asset and the terminal time T are the same for each option. At terminal time T, the price of a binary call option $C_{i,T}$ equals the option's payoff and depends on the relation between the underlying asset's price S_T and the option's so-called strike price K_i:

$$C_{i,T} = \begin{cases} 1 \text{ EUR} & \text{if } S_T > K_i \\ 0 \text{ EUR} & \text{if } S_T \leq K_i. \end{cases} \tag{6}$$

Currently, we are in time $t = 0$. The option price model predicts that the price of a binary call option is

$$\mathcal{C}_{i,0} = e^{-rT}\Phi\left(d_{i,0}\right), \qquad d_{i,0} = \frac{1}{\sigma\sqrt{T}}\left(\ln\left(\frac{S_0}{K_i}\right) + \left(r + \frac{\sigma^2}{2}\right)\cdot T\right) - \sigma\sqrt{T}, \tag{7}$$

where $\Phi(\cdot)$ denotes the cdf of the standard normal distribution. For each observed option, we know its current market price $C_{i,0}$, its strike price K_i, and the current price of the underlying asset S_0. We also know the risk-free rate r, and the terminal time T. Let us define the data vector $\boldsymbol{x}_i = (K_i, S_0, r, T)$ and assume that $\mathcal{C}_{i,0} = \mathrm{E}(e^{-rT}C_{i,T}|\boldsymbol{x}_i, \sigma)$. Note that the **only unknown parameter is** σ which we will estimate subsequently.

1. **(4P)** Suggest a way to calculate the unknown parameter σ separately for each observed option using its observed price $C_{i,0}$, our price formula in (7), and a numerical procedure. Explain briefly (no formulas, no derivations!). Why will these σ's in general differ between options?

2. **(11P)** Now, we want to estimate σ using all options of our sample and the objective function

$$q(\boldsymbol{w}_i, \sigma) = \frac{1}{2}\left[C_{i,0} - \mathrm{E}(e^{-rT}C_{i,T}|\boldsymbol{x}_i, \sigma)\right]^2 = \frac{1}{2}\left[C_{i,0} - \mathcal{C}_{i,0}\right]^2, \tag{8}$$

   where $\boldsymbol{w}_i = (C_{i,0}, \boldsymbol{x}_i)$. Which estimation approach is this? Derive the score and the Hessian of the objective function. To this end, derive $d'_{i,0} \equiv \frac{\partial}{\partial\sigma}d_{i,0}$ and use short-hand notation $d'_{i,0}$ subsequently. Furthermore, use the derivatives $\phi(x) \equiv \frac{\partial}{\partial x}\Phi(x)$, $\phi'(x) \equiv \frac{\partial}{\partial x}\phi(x)$, and $d''_{i,0} \equiv \frac{\partial}{\partial\sigma}d'_{i,0}$ without explicit calculation.