# Exercise 9

Version 1 (April 5, 2022)

Mikkel Plagborg-Møller

Due in lecture on Tuesday, April 19

*The problem set should be submitted in paper form. You may turn in one problem set per group of at most 3 students. You may discuss the exercises with any of your classmates. Please attach your code to the problem set. Use any software you like, although you learn the most by avoiding canned statistics/econometrics packages. If you find any errors or require clarification, please let me know right away.*

## Question 1

Consider the same data set and probit specification as in Question 1 of Exercise 8.

i) Estimate the marginal effect of years of education on the probability that a married woman enters the labor force, for an individual whose covariate values equal the sample averages of the covariates (treat education as a continuous variable).

ii) Estimate the average partial effect of years of education on the probability that a married woman enters the labor force (treat education as a continuous variable).

iii) Estimate the average partial effect (on the probability that a married woman enters the labor force) of going from having 0 kids less than 6 years old to having 1 kid less than 6 years old.

iv) Calculate standard errors for the estimates in (i)–(iii). You may use any method that is valid for i.i.d. data.

## Question 2

The data set `heating.csv` is available on the course website. The data consist of observations on 246 households making choices on heating systems. The first five variables are fixed installation costs of each heating system in dollars, the second five are annual operating cost of each heating system in dollars, followed by household choice. The order of the choices is: heat pump, gas central, electric central, gas room, and electric room. The household choice coding follows the same order (heat pump is 0, gas central is 1, etc). This is a different sample from the one used in the lecture slides on discrete choice models.

We will focus on a discrete choice model where the utility of household $i$ from choice $j$ is

$$U_{ij} = \alpha_j + \beta_1 IC_{ij} + \beta_2 OC_{ij} + u_{ij}.$$

$IC_{ij}$ is the fixed (installation) cost for alternative $j$ for household $i$ and $OC_{ij}$ is the variable (operating) cost for alternative $j$ for household $i$. Let $p_{ij}(OC_{ij}, IC_{ij})$ be the probability that household $i$ chooses heating system $j$ conditional on $OC_{ij}$ and $IC_{ij}$. It can be shown that if $u_{ij}$ are independent and have a type I extreme value distribution, then

$$p_{ij}(OC_{ij}, IC_{ij}) = \frac{e^{\alpha_j + \beta_1 IC_{ij} + \beta_2 OC_{ij}}}{\sum_{k=0}^{m} e^{\alpha_k + \beta_1 IC_{ik} + \beta_2 OC_{ik}}}.$$

Normalize $\alpha_0 = 0$ so that there are six free parameters, $\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$, $\beta_1$, and $\beta_2$.

i) Let $Y_{ij}$ be an indicator variable that takes value one if household $i$ chooses alternative $j$, and value zero otherwise. Write the log likelihood of the sample in terms of the variables $Y_{ij}$, $IC_{ij}$, $OC_{ij}$, and the parameters of the model.

ii) Derive the score of a single observation in terms of the variables $Y_{ij}$, $IC_{ij}$, $OC_{ij}$, and the probabilities $p_{ij}$.

iii) Calculate the maximum likelihood estimates of the coefficients $\alpha_1$, $\alpha_2$, $\alpha_3$, $\alpha_4$, $\beta_1$, and $\beta_2$. Use a vector of zeros as starting values. Calculate standard errors using the estimated information matrix. When carrying out the numerical optimization, try to exploit the analytical gradient of the log likelihood as implied by your derivation in (ii).

iv) Conduct a Likelihood Ratio test of the null hypothesis $H_0\colon \{\alpha_j = 0 \text{ for all } j = 1, 2, 3, 4\}$ against the alternative $H_1\colon \{\alpha_j \neq 0 \text{ for some } j = 1, 2, 3, 4\}$. Use a 5% significance level. Provide a brief economic interpretation of the result.

v) The market share of choice $j$ is defined as $E[p_{ij}(OC_{ij}, IC_{ij})]$, averaging over the population of individuals $i$. Predict the change in market shares for all heating choices if the variable costs of gas central and gas room go up by 10 percent for all households. Use the nonparametric bootstrap to compute standard errors for the estimated changes in market shares. Use at least 1000 bootstrap iterations and the estimates from part (iii) as optimization starting values in each bootstrap iteration.

## Question 3

We will use the same data set `nls.csv` as in Question 2 of Exercise 7. In the following, you are encouraged to use available software packages for quantile regression, see for example Roger Koenker's website: `http://www.econ.uiuc.edu/~roger/research/rq/rq.html`

Consider a quantile regression of wages on a constant, years of education, experience, and experience squared.

i) Plot the estimated quantile regression coefficient on education as a function of the quantile, using the grid $[0.05, 0.06, 0.07, \ldots, 0.94, 0.95]$ of quantiles. Include a (pointwise) 95% bootstrap confidence interval for each estimated coefficient in your plot.

## Question 4

Consider the quantile IV model and estimator of Chernozhukov & Hansen (*Journal of Econometrics*, 2008).

i) If the instrument vector $Z$ is nearly independent of the endogenous covariate vector $X$, we would intuitively expect identification to be weak. Discuss why the quantile IV estimator's objective function is likely to be nearly flat around the true parameters in this case. A brief discussion with simple mathematical expressions suffices, but try to be as precise as possible.

ii) Consider the case of a scalar endogenous covariate $X$. How could you construct an analogue of the Anderson-Rubin weak IV robust confidence region in the quantile IV model?