First Assignment

Malgorzata, Kevin, Johannes

December, 2018

First Problem: (Simulation: Latent Variable, Probit Model):

Consider the following JOINT distributions of random variables $X_0 \& \varepsilon_j^*$ (You must simulate it as a joint (multivariate) normal distributions!).

$$\begin{pmatrix} X_0 \\ \varepsilon_1^0 \end{pmatrix} \sim N\left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 2^2 & 0 \\ 0 & 1 \end{pmatrix}\right], \begin{pmatrix} X_0 \\ \varepsilon_2^* \end{pmatrix} \sim N\left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 2^2 & 0 \\ 0 & 2^2 \end{pmatrix}\right], \begin{pmatrix} X_0 \\ \varepsilon_3^* \end{pmatrix} \sim N\left[\begin{pmatrix} 10 \\ 0 \end{pmatrix}, \begin{pmatrix} 2^2 & 3 \\ 3 & 2^2 \end{pmatrix}\right]$$

with

$$\beta_0 = -30$$
$$\beta_1 = 4$$

$$Y_j^* = \beta_0 + \beta_1 * X_1 + \varepsilon_j$$

$$Y_j = \begin{cases} 1 \text{ if } Y_j^* > 0 \\ 0, \text{ otherwise} \end{cases}$$

$$j \in \{1, 2, 3\}$$

a) [1P] Please simulate the three datasets (i.e.: $Y_j, X_1, \varepsilon_j^*; j \in \{1,2,3\}$) with a sample size of 30000 observations. For each of the three datasets, estimate a probit model of Y $_\text{i}$ on x_1 and save the estimate for $\hat{\beta}_1$

b) [1.5P] Repeat a 400 times while saving all the different estimate in vectors for all three models (based on the three different populations). Plot the kernel density estimates for $\hat{\beta}_1$ based on the three populations next to each other. Describe shortly the choices you faced and made estimating the density functions.

i. [1.5P] Now concentrate on the kernel density plot based on the first population (i.e.: $Y_1, X_1, \varepsilon_1^*$) . Does the distribution of $\overline{\beta}_1$ conform to your expectations? More specifically, explain which distribution you would expect (and why) and whether the plotted density conform to that expectation (no formal tests necessary).

ii. [1P] Compare the distributions of $\widehat{\beta_1}$ from the population $j = 1$ to the $\widehat{\beta_1}$ from the population $j = 2$ and $j = 3,$ respectively. Why do the means of the distributions differ?

iii. [1P] Compute the mean of $\widehat{\beta_1}$ from the population $j = 2$. Can you explain, why the distribution of this particular $\overline{\beta_1}$ concentrate approximately around this value?

Simulate again all three populations as you did in a. But this time, estimate and save the average marginal probability effect of x_1 . Repeat the estimation 400 times with a sample size of each iteration equaling 30000.

c) [1P] Analogously to b), plot the kernel density estimates for all three AMPE's. Determine the values around which the sample distributions are concentrated.

i. [1.5P] Calculate the relative difference (in percent) between $\hat{E}[AMPE|j = 1]$ and $\hat{E}[AMPE|j = 3]$. Which estimate would you use to ascertain the effect of the variable x_1?

ii. [1.5P] Calculate the relative difference (in percent) between $\hat{E}[AMPE|j = 1]$ and $\hat{E}[AMPE|j = 2]$.

iii. [1.5P] Based on the results in b-ii) would you expect the differences observed c-i) and c-ii)? Please provide a detailed explanation for the observed results

Problem 2: (Marginal effects estimation & Interpretation):

a) [0.5P] Can you learn anything from the estimated coefficients? Explain shortly.

b) [0.5P] Are the S.E. valid, or do you need to adjust them for heteroscedasticity? Explain.

c) [2P] Re-estimate the model from a) but this time include age in addition to 1d1. You see that the estimated coefficient of 1d1 changes. Explain why? Additionally, show that your explanation is supported by the data.

d) Finally, estimate the model from a) but include ldl squared next to ldl as a control variable.

i.) [1P]Based on the estimated coefficients from a) and d) draw the two resulting marhinal probability effects of ldl as a function of ldl $\in [1; 15]$ next to each other.

ii.) [0.5P] Are any of the marginal probability effects linear in ldl? Explain why.

iii.) [1P] What is the advantage of the marginal probability effect based on the estimation in d) over the one based on a)? Explain shortly.

iv.) [1.5P] Calculate and properly interpret both marginal probability effects for the mean value of ldl in the sample.

v. [1P] Are any of the effects computed in iv), ceteris paribus effects? Explain shortly.