Liming Lin
Professor Moshe Buchinsky
Econometrics II
Problem Set 8
Apr. 25th, 2025

Part I — Empirical

Problem 1

1. According to our estimation, what is the impact of degree level on enf?

The probit model estimates how education level affects the probability of having children. The coefficients correspond to changes in an unobserved latent variable Y*, which reflects an individual's underlying propensity to have children. Compared with the reference gorup (no high school graduation), the high school graduation (**dipl=3**) has a negetive impacts on the inclination to have children (Y^*) and thus enf. The coefficient for **dipl=3** is also negative and the absolute value is larger than the coefficient for **dipl=2** (high school graduation). This indicates that the probability of having children decreases as education level increases.

2. How would you test that the parameter for **_Idipl_3** is significantly different from the parameter for **_Idipl_2** ? Compute the t - $statistic$ and give the result

We can use a t-test to compare the coefficients of **_Idipl_3** and **_Idipl_2**. The null hypothesis is that the two coefficients are equal, while the alternative hypothesis is that they are not equal. The t-statistic can be calculated as follows:

$$t = \frac{\hat{\beta}_{Idipl1_2} - \hat{\beta}_{Idipl1_3}}{\sqrt{Var(\hat{\beta}_{Idipl1_2}) + Var(\hat{\beta}_{Idipl1_3}) - 2Cov(\hat{\beta}_{Idipl1_2}, \hat{\beta}_{Idipl1_3})}}$$

$$= \frac{-0.1324138 - (-0.2885005)}{\sqrt{0.000559 + 0.000348 - 2*0.000104}} \approx 5.9$$

Since the t-statistic is greater than the critical value from the t-distribution, we reject the null hypothesis and conclude that the coefficients are significantly different.

3. What is the impact of age on enf? At which age is the probability to have children the highest?

Again since we estimate a probit model here, the coefficients on age and age sqaured indicate the effects on the laten variable Y^* which in turn affects enf. In this case, by holding education level constant, with one year increase in age, Y^* increases by $0.002996346 - 0.00766951 \cdot age$.

To find the age at which the probability of having children is highest, we can set the derivative of

the latent variable with respect to age to zero and solve for age:

$$\frac{dY^*}{dage} = 0.002996346 - 0.00766951 \cdot age = 0$$

$$age^* = \frac{0.002996346}{0.00766951} \approx 0.39$$

Then we muliply the age by 100 to get the actual age. The probability of having children is highest at age 39.

4. Use the variance covariance matrix to compute an estimator of the standard error of this age

Since we calculate the age^* by

$$\text{Age}^* = -\frac{\hat{\beta}_{age}}{2\hat{\beta}_{age2}}$$

Since the two coefficients are estimated and Age^* is a function of them, we can use the delta method to compute the standard error of Age^*.
First, we define the funciton:

$$g(\hat{\beta}_{age}, \hat{\beta}_{age2}) = -\frac{\hat{\beta}_{age}}{2\hat{\beta}_{age2}}$$

Then we can compute the gradient of g:

$$\frac{\partial g}{\partial \hat{\beta}_{age}} = -\frac{1}{2\hat{\beta}_{age2}} \quad \text{and} \quad \frac{\partial g}{\partial \hat{\beta}_{age2}} = \frac{\hat{\beta}_{age}}{2\hat{\beta}_{age2}^2}$$

Thus:

$$\nabla g = \begin{pmatrix} -\frac{1}{2\hat{\beta}_{age2}} \\ \frac{\hat{\beta}_{age}}{\hat{\beta}_{age2}^2} \end{pmatrix} = \begin{pmatrix} -\frac{1}{2(-0.3834755)} \\ \frac{0.2996346}{2(-0.3834755)^2} \end{pmatrix} = \begin{pmatrix} 1.303986 \\ 1.018814 \end{pmatrix}$$

Then we construct the variance-covariance matrix of the two coefficients:

$$\text{Var}(\beta) = \begin{pmatrix} \text{Var}(\hat{\beta}_{age}) & \text{Cov}(\hat{\beta}_{age}, \hat{\beta}_{age2}) \\ \text{Cov}(\hat{\beta}_{age}, \hat{\beta}_{age_2}) & \text{Var}(\hat{\beta}_{age2}) \end{pmatrix} = \begin{pmatrix} 0.000041 & -0.000051 \\ -0.000051 & 0.000065 \end{pmatrix}$$

Lastly, we can compute the standard error of Age^* using the delta method:

$$\text{Var}(\text{Age}^*) \approx \nabla g' \, \text{Var}(\beta) \, \nabla g \approx 0.0000016755$$

And take the square root to get the standard error: 0.001294