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:

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

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

5. What do we need to normalize $Var(u_i)$ to for the parameters to be identified? Why does Stata not output a constant term?

In a probit model, we need to normalize the variance of the error term $Var(u_i)$ to 1 for the parameters to be identified. This is because the probit model assumes that the error term follows a standard normal distribution, which has a mean of 0 and a variance of 1. Otherwise we can

multiply parameters and cut-off point to get different models explain the same probabilities, making the model un-identifiable.

Stata does not report a constant term because the in ordered probit model, the cut-off points are estimated instead of the intercept. The cut-off points represent the thresholds that separate the different categories of the dependent variable.

6. Give an intuition why the estimated parameters for the probit model are similar to those of the ordered probit model.

Both the probit and ordered probit models estimate the effect of covariates on an unobserved latent variable (inclination to have children). Although the outcome variable differs (binary vs. ordered categories), the underlying structure is similar, explaining why the coefficients in the two models are close.

7. What is the probability to have three children for an individual who is 35 years old with no high school graduation?

$$Y^* = \hat{\beta}_{age} \times \frac{35}{100} + \hat{\beta}_{age^2} \times \frac{35}{100}^2$$
$$= 0.26945 \times 0.35 + (-0.3453101) \times 12.25$$
$$= -4.134732225$$

Since the prodicted value is negative, which is much lower than the cut-off point for having three children, we can conclude that the probability of having three children for an individual who is 35 years old with no high school graduation is 0.

8. What is the expected number of children predicted given household characteristics?

$\Pr(0) = \Phi(cut1 - Y^*) = \Phi(4.436445 - (-4.13)) = \Phi(8.566445) \approx 1$
$\Pr(1) = \Phi(cut2 - Y^*) - \Phi(cut1 - Y^*) = \Phi(5.178941 - (-4.13) - 1 = \Phi(9.308941)) - 1 \approx 1 - 1 = 0$
$\Pr(2) = \Phi(cut3 - Y^*) - \Phi(cut2 - Y^*) = \Phi(6.082635 - (-4.13)) - 1 = \Phi(10.212635) - 1 \approx 1 - 1 = 0$
$\Pr(3) = \Phi(cut4 - Y^*) - \Phi(cut3 - Y^*) = \Phi(6.83505 - (-4.13)) = \Phi(10.96505) - 1 \approx 1 - 1 = 0$
$\Pr(4+) = 1 - \Phi(cut4 - Y^*) = 1 - \Phi(10.96505) \approx 1 - 1 = 0$
To sum up, the expected number of children predicted given this household characteristics is:

$$E[enf] = 0 \cdot \Pr(0) + 1 \cdot \Pr(1) + 2 \cdot \Pr(2) + 3 \cdot \Pr(3) + 4 + \cdot \Pr(4+) = 0 + 0 + 0 + 0 + 0 = 0$$

Part II — Theory

Problem 2

Let *grad* be a dummy variable for whether a student-athlete at a large university graduates in five years. Let *hsGPA* and *SAT* be high school grade point average and SAT score, respectively. Let study be the number of hours spent per week in an organized study hall. Suppose that, using data on 420 student-athletes, the following logit model is obtained:

$$Pr(grad = 1|hsGPA, SAT, study) = \Lambda(-1.17 + 0.24hsGPA + 0.00058SAT + 0.073study)$$

where $\Lambda(x) = \frac{exp(z)}{1+exp(z)}$ is the logit function. Holding hsGP A fixed at 3.0 and SAT fixed at 1,200, compute the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week.

Firstly, we compute the graduation probability for someone who spent 10 hours per week in study hall:

$$
\begin{aligned}
Pr(grad = 1|hsGPA = 3.0, SAT = 1200, study = 10) &= \Lambda(-1.17 + 0.24 \cdot 3.0 + 0.00058 \cdot 1200 + 0.073 \cdot 10) \\
&= \Lambda(-1.17 + 0.72 + 0.696 + 0.73) \\
&= \Lambda(0.976) \\
&= \frac{exp(0.976)}{1 + exp(0.976)} \\
&\approx \frac{2.651}{1 + 2.651} \\
&\approx 0.7266
\end{aligned}
$$

Then we compute the graduation probability for someone who spent 5 hours per week in study hall:

$$
\begin{aligned}
Pr(grad = 1|hsGPA = 3.0, SAT = 1200, study = 5) &= \Lambda(-1.17 + 0.24 \cdot 3.0 + 0.00058 \cdot 1200 + 0.073 \cdot 5) \\
&= \Lambda(-1.17 + 0.72 + 0.696 + 0.365) \\
&\approx 0.6479
\end{aligned}
$$

Finally, we can compute the difference in graduation probability:

$$0.7266 - 0.6479 = 0.0787$$

This means that the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week is approximately 0.0787, or 7.87 percentage points.

Problem 3

1. Are all parameters identifiable ? What is the usual normalization?

No, in multinomial logit model, the probabilities of the different outcomes are not independent from each other because they are set to sum to 1, meaning that if we know the probabilities of all but one outcome, we can determine the probability of the last outcome. This means that we need to normalize one of the parameters to make the model identifiable. The usual normalization is to set the coefficient of one of the categories to zero.

2. Compute $\mathbb{P}(y = l | x = k)$

For $x = 1$, the inactive group:

$$\mathbb{P}(y = 1 | x = 1) = \frac{150}{450} = \frac{1}{3}$$
$$\mathbb{P}(y = 2 | x = 1) = \frac{200}{450} = \frac{4}{9}$$
$$\mathbb{P}(y = 3 | x = 1) = \frac{100}{450} = \frac{2}{9}$$

For $x = 2$, the employee group:

$$\mathbb{P}(y = 1 | x = 2) = \frac{100}{450} = \frac{2}{9}$$
$$\mathbb{P}(y = 2 | x = 2) = \frac{150}{450} = \frac{1}{3}$$
$$\mathbb{P}(y = 3 | x = 2) = \frac{200}{450} = \frac{4}{9}$$

For $x = 3$, the independent group:

$$\mathbb{P}(y = 1 | x = 3) = \frac{20}{100} = \frac{1}{5}$$
$$\mathbb{P}(y = 2 | x = 3) = \frac{50}{100} = \frac{1}{2}$$
$$\mathbb{P}(y = 3 | x = 3) = \frac{30}{100} = \frac{3}{10}$$

3. Write the log-likelihood of the sample

In the multinomial logit model:

$$P(y_i = l \mid x_i) = \frac{\exp(V_{il})}{\sum_{m=1}^{3} \exp(V_{im})}$$

where $V_{il} = \sum_k b_{lk} x_{ik}$
Substituting the multinomial logit probability into the log-likelihood:

$$\log \mathcal{L} = \sum_{i=1}^{n} \log \left(\frac{\exp(V_{iy_i})}{\sum_{m=1}^{3} \exp(V_{im})} \right)$$

simplifying gives:

$$\log\left(\frac{\exp(V_{iy_i})}{\sum_{m=1}^{3}\exp(V_{im})}\right) = \log(\exp(V_{iy_i})) - \log\left(\sum_{m=1}^{3}\exp(V_{im})\right)$$

further simplifying gives:

$$= V_{iy_i} - \log\left(\sum_{m=1}^{3}\exp(V_{im})\right)$$

Finally, we can write the log-likelihood as:

$$\log\mathcal{L} = \sum_{i=1}^{n}\left(V_{iy_i} - \log\left(\sum_{m=1}^{3}\exp(V_{im})\right)\right)$$

4. Compute the estimator of $\gamma_{lk} = \frac{expb_{lk}}{expb_{1k}+expb_{2k}+expb_{3k}}$ for l and by maximizing the log-likelihood under the identifying constraint on $\gamma_{1k} + \gamma_{2k} + \gamma_{3k} = 1$

Empirically, γ_{lk} can be esetimated by the observed proportions of the sample in each labor market status, which is the same as the probabilities we computed in part 2.

Labor Market Status	γ_{1k} (Abstain)	γ_{2k} (Mr. John)	γ_{3k} (Ms. Does)
Inactive	$\frac{1}{3}$	$\frac{4}{9}$	$\frac{2}{9}$
Employee	$\frac{2}{9}$	$\frac{1}{3}$	$\frac{4}{9}$
Independent	$\frac{1}{5}$	$\frac{1}{2}$	$\frac{3}{10}$

5. Note $\hat{\gamma}_k = (\hat{\gamma}_{1k}, \hat{\gamma}_{2k}, \hat{\gamma}_{3k})$. Show that $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ are asymptotically independent.

Since each γ_k is estimated using a different and independent subset of the data, the estimators $\hat{\gamma}_1, \hat{\gamma}_2, \hat{\gamma}_3$ are asymptotically independent by the Central Limit Theorem.

6. Show that an estimator of the asymptotic variance-covariance matrix of $(\hat{\gamma}_{11}, \hat{\gamma}_{21})$ is

$$\frac{1}{450}\begin{bmatrix} \frac{15}{2} & \frac{9}{2} \\ \frac{9}{2} & \frac{27}{4} \end{bmatrix}^{-1}$$

We model choices between three outcomes (Abstain, Mr. John, Ms. Does) in the inactive group $k = 1$, with probabilities:

$$(\gamma_{11}, \gamma_{21}, \gamma_{31}), \quad \text{where } \gamma_{31} = 1 - \gamma_{11} - \gamma_{21}$$

and total sample size is 450 for the inactive group.
The multinomial likelihood is:

$$\mathcal{L}(\gamma) = \prod_{l=1}^{3}\gamma_{l1}^{n_l}$$

6

Taking logs:

$$\log \mathcal{L}(\gamma) = n_1 \log(\gamma_{11}) + n_2 \log(\gamma_{21}) + n_3 \log(1 - \gamma_{11} - \gamma_{21})$$

Partial derivatives of the log-likelihood with respect to γ_{11} and γ_{21} are:

$$\frac{\partial \log \mathcal{L}}{\partial \gamma_{11}} = \frac{n_1}{\gamma_{11}} - \frac{n_3}{1 - \gamma_{11} - \gamma_{21}}$$

$$\frac{\partial \log \mathcal{L}}{\partial \gamma_{21}} = \frac{n_2}{\gamma_{21}} - \frac{n_3}{1 - \gamma_{11} - \gamma_{21}}$$

Partial second derivatives:
- Second derivative with respect to γ_{11}:

$$\frac{\partial^2 \log \mathcal{L}}{\partial \gamma_{11}^2} = -\frac{n_1}{\gamma_{11}^2} - \frac{n_3}{(1 - \gamma_{11} - \gamma_{21})^2}$$

- Second derivative with respect to γ_{21}:

$$\frac{\partial^2 \log \mathcal{L}}{\partial \gamma_{21}^2} = -\frac{n_2}{\gamma_{21}^2} - \frac{n_3}{(1 - \gamma_{11} - \gamma_{21})^2}$$

- Cross derivative (mixed partial):

$$\frac{\partial^2 \log \mathcal{L}}{\partial \gamma_{11} \partial \gamma_{21}} = -\frac{n_3}{(1 - \gamma_{11} - \gamma_{21})^2}$$

The Fisher Information Matrix is the negative expected value of the second derivatives.
Replacing n_1, n_2, n_3 by their expected values ($n\gamma_{11}, n\gamma_{21}, n\gamma_{31}$):
Thus, the Information Matrix is:

$$\mathcal{I}(\gamma_{11}, \gamma_{21}) = n \times \begin{pmatrix} \frac{1}{\gamma_{11}} + \frac{1}{\gamma_{31}} & \frac{1}{\gamma_{31}} \\ \frac{1}{\gamma_{31}} & \frac{1}{\gamma_{21}} + \frac{1}{\gamma_{31}} \end{pmatrix} = 450 \times \begin{pmatrix} \frac{1}{\frac{1}{3}} + \frac{1}{\frac{2}{9}} & \frac{1}{\frac{2}{9}} \\ \frac{1}{\frac{2}{9}} & \frac{1}{\frac{4}{9}} + \frac{1}{\frac{2}{9}} \end{pmatrix} = 450 \times \begin{pmatrix} \frac{15}{2} & \frac{9}{2} \\ \frac{9}{2} & \frac{27}{4} \end{pmatrix}$$

The variance covariance matrix is the inverse of the Fisher Information Matrix:

$$\text{Var}(\hat{\gamma}) = \frac{1}{450} \begin{bmatrix} \frac{15}{2} & \frac{9}{2} \\ \frac{9}{2} & \frac{27}{4} \end{bmatrix}^{-1}$$