International School of Economics at TSU
Econometrics II
Lasha Chochua

Problem Set 5

Instructions: You are encouraged to solve the problems before the recitation. Additionally, you are encouraged to work in groups. It is **not mandatory** to submit solutions unless stated otherwise. However, if you would like to share your solution, I would be happy to review it.

Problem 1: Let z_1 be a vector of variables, let z_2 be a continuous variable, and let d_1 be a dummy variable.

a. In the model

$$\mathbb{P}(y = 1 \mid \mathbf{z}_1, z_2) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 z_2^2),$$

find the partial effect of z_2 on the response probability. How would you estimate this partial effect?

b. In the model

$$\mathbb{P}(y = 1 \mid \mathbf{z}_1, z_2, d_1) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1),$$

find the partial effect of z_2. How would you measure the effect of d_1 on the response probability? How would you estimate these effects?

c. Describe how you would obtain the standard errors of the estimated partial effects from parts a and b.

Solution:

a. If $\mathbb{P}(y_i = 1 \mid \mathbf{z}_1, z_2) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 z_2^2)$ then

$$\frac{\partial \mathbb{P}(y = 1 \mid \mathbf{z}_1, z_2)}{\partial z_2} = (\gamma_1 + 2\gamma_2 z_2) \cdot \phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 z_2^2);$$

for given $\mathbf{z}$, the partial effect is estimated as

$$(\hat{\gamma}_1 + 2\hat{\gamma}_2 z_2) \cdot \phi(\mathbf{z}_1\hat{\delta}_1 + \hat{\gamma}_1 z_2 + \hat{\gamma}_2 z_2^2),$$

where, of course, the estimates are the probit estimates.

b. In the model

$$\mathbb{P}(y_i = 1 \mid \mathbf{z}_1, z_2, d_1) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1),$$

the partial effect of z_2 is

$$\frac{\partial \mathbb{P}(y = 1 \mid \mathbf{z}_1, z_2, d_1)}{\partial z_2} = (\gamma_1 + \gamma_3 d_1) \cdot \phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 d_1 + \gamma_3 z_2 d_1).$$

The effect of d_1 is measured as the difference in the probabilities at $d_1 = 1$ and $d_1 = 0$:

$$\mathbb{P}(y = 1 \mid \mathbf{z}, d_1 = 1) - \mathbb{P}(y = 1 \mid \mathbf{z}, d_1 = 0) = \Phi(\mathbf{z}_1\delta_1 + \gamma_2 + (\gamma_1 + \gamma_3)z_2) - \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2).$$

Again, to estimate these effects at given $\mathbf{z}$ and — in the first case — d_1, we just replace the parameters with their probit estimates, and use average or other interesting values of $\mathbf{z}$.

c. If the estimated partial effect is for particular values of $(\mathbf{z}_1, z_2, d_1)$, for example,

$$(\hat{\gamma}_1 + \hat{\gamma}_3 d_1^0) \cdot \phi(\mathbf{z}_1\hat{\delta}_1 + \hat{\gamma}_1 z_2^0 + \hat{\gamma}_2 d_1^0 + \hat{\gamma}_3 z_2^0 d_1^0),$$

then we can apply the delta method from Chapter 3 (and referred to in Part III). Thus, we would require the full variance matrix of the probit estimates as well as the gradient of the expression of interest, such as

$$(\gamma_1 + 2\gamma_2 z_2) \cdot \phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 + \gamma_2 z_2^2),$$

with respect to all probit parameters.

If we are interested in the average partial effect (APE) of d_1 going from zero to one then we estimate it as

$$N^{-1}\sum_{i=1}^{N}\left[\Phi(\mathbf{z}_{1i}\hat{\delta}_1 + (\hat{\gamma}_1 + \hat{\gamma}_3)z_{2i} + \hat{\gamma}_2) - \Phi(\mathbf{z}_{1i}\hat{\delta}_1 + \hat{\gamma}_1 z_{2i})\right],$$

that is, we estimate the effect for each unit i and then average these across all i. If we want a standard error for this, we would use the extension of the delta method.

Problem 2: Consider the probit model

$$\mathbb{P}(y=1 \mid \mathbf{z}, q) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 q),$$

where q is independent of $\mathbf{z}$ and distributed as Normal$(0,1)$; the vector $\mathbf{z}$ is observed but the scalar q is not.

a. Find the partial effect of z_2 on the response probability, namely,

$$\frac{\partial \mathbb{P}(y=1 \mid \mathbf{z}, q)}{\partial z_2}.$$

b. Show that $\mathbb{P}(y=1 \mid \mathbf{z}) = \Phi[\mathbf{z}_1\delta_1/(1+\gamma_1^2 z_2^2)^{1/2}]$.

c. Define $\rho_1 \equiv \gamma_1^2$. How would you test $H_0 : \rho_1 = 0$?

d. If you have reason to believe $\rho_1 > 0$, how would you estimate δ_1 along with ρ_1?

Solution:

a. If $\mathbb{P}(y=1 \mid \mathbf{z}, q) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 q)$ then

$$\frac{\partial \mathbb{P}(y=1 \mid \mathbf{z}, q)}{\partial z_2} = \gamma_1 q \cdot \phi(\mathbf{z}_1\delta_1 + \gamma_1 z_2 q),$$

assuming that z_2 is not functionally related to $\mathbf{z}_1$.

b. Write $y^* = \mathbf{z}_1\delta_1 + r$, where $r = \gamma_1 z_2 q + e$, and e is independent of $(\mathbf{z}, q)$ with a standard normal distribution. Because q is assumed independent of $\mathbf{z}$,

$$q \mid \mathbf{z} \sim \text{Normal}(0, \gamma_1^2 z_2^2 + 1);$$

this follows because $\mathbb{E}(r \mid \mathbf{z}) = \gamma_1 z_2 \mathbb{E}(q \mid \mathbf{z}) + \mathbb{E}(e \mid \mathbf{z}) = 0$. Also,

$$\begin{aligned}\text{Var}(r \mid \mathbf{z}) &= \gamma_1^2 z_2^2 \text{Var}(q \mid \mathbf{z}) + \text{Var}(e \mid \mathbf{z}) + 2\gamma_1 z_2 \text{Cov}(q, e \mid \mathbf{z}) \\ &= \gamma_1^2 z_2^2 + 1\end{aligned}$$

because $\text{Cov}(q, e \mid \mathbf{z}) = 0$ by independence between e and $(\mathbf{z}, q)$. Thus, $r/\sqrt{\gamma_1^2 z_2^2 + 1}$ has a standard normal distribution independent of $\mathbf{z}$. It follows that

$$\mathbb{P}(y=1 \mid \mathbf{z}) = \Phi\left(\frac{\mathbf{z}_1\delta_1}{\sqrt{\gamma_1^2 z_2^2 + 1}}\right). \tag{1}$$

c. Because $\mathbb{P}(y = 1 \mid \mathbf{z})$ depends only on γ_1^2, this is what we can estimate along with δ_1. (For example, $\gamma_1 = -2$ and $\gamma_1 = 2$ give exactly the same model for $\mathbb{P}(y = 1 \mid \mathbf{z})$.) This is why we define $\rho_1 = \gamma_1^2$. Testing $H_0 : \rho_1 = 0$ is most easily done using the score or LM test because, under H_0, we have a standard probit model.

Let $\hat{\delta}_1$ denote the probit estimates under the null that $\rho_1 = 0$. Define $\hat{\phi}_i = \phi(\mathbf{z}_{i1}\hat{\delta}_1)$,
$\hat{\Phi}_i = \Phi(\mathbf{z}_{i1}\hat{\delta}_1)$,
$\hat{u}_i = y_i - \hat{\Phi}_i$, and
$\tilde{u}_i \equiv \hat{u}_i / \sqrt{\hat{\Phi}_i(1 - \hat{\Phi}_i)}$ (the standardized residuals).

The gradient of the mean function in (1) with respect to δ_1, evaluated under the null estimates, is simply $\hat{\phi}_i \mathbf{z}_{i1}$. The only other quantity needed is the gradient with respect to ρ_1 evaluated at the null estimates. But the partial derivative of (1) with respect to ρ_1 is, for each i,

$$-(\mathbf{z}_{i1}\delta_1)(z_{i2}^2/2)(\rho_1 z_{i2}^2 + 1)^{-3/2}\, \phi\left(\frac{\mathbf{z}_{i1}\delta_1}{\sqrt{\gamma_1^2 z_2^2 + 1}}\right).$$

When we evaluate this at $\rho_1 = 0$ and $\hat{\delta}_1$, we get $-(\mathbf{z}_{i1}\hat{\delta}_1)(z_{i2}^2/2)\hat{\phi}_i$. Then, the score statistic can be obtained as NR_u^2 from the regression

$$\tilde{u}_i \text{ on } \frac{\hat{\phi}_i \mathbf{z}_{i1}}{\sqrt{\hat{\Phi}_i(1 - \hat{\Phi}_i)}}, \quad \frac{(\mathbf{z}_{i1}\hat{\delta}_1) z_{i2}^2 \hat{\phi}_i}{\sqrt{\hat{\Phi}_i(1 - \hat{\Phi}_i)}};$$

under H_0, $NR_u^2 \overset{a}{\sim} \chi_1^2$.

d. The model can be estimated by MLE using the formulation with ρ_1 in place of γ_1^2. It is not a standard probit estimation but a kind of "heteroskedastic probit."

Problem 3: Consider taking a large random sample of workers at a given point in time.
Let $sick_i = 1$ if person i called in sick during the last 90 days, and zero otherwise. Let $\mathbf{z}_i$ be a vector of individual and employer characteristics. Let $cigs_i$ be the number of cigarettes individual i smokes per day (on average).

a. Explain the underlying experiment of interest when we want to examine the effects of cigarette smoking on workdays lost.

b. Why might $cigs_i$ be correlated with unobservables affecting $sick_i$?

c. One way to write the model of interest is

$$\mathbb{P}(sick = 1 \mid \mathbf{z}, cigs, q_1) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 cigs + q_1),$$

where $\mathbf{z}_1$ is a subset of $\mathbf{z}$ and q_1 is an unobservable variable that is possibly correlated with $cigs$. What happens if q_1 is ignored and you estimate the probit of $sick$ on $\mathbf{z}_1$, $cigs$?

d. Can $cigs$ have a conditional normal distribution in the population? Explain.

e. Explain how to test whether $cigs$ is exogenous. Does this test rely on $cigs$ having a conditional normal distribution?

f. Suppose that some of the workers live in states that recently implemented no-smoking laws in the workplace. Does the presence of the new laws suggest a good IV candidate for $cigs$?

Solution:

a. What we would like to know is that, if we exogenously change the number of cigarettes that someone smokes per day, what effect would this have on the probability of missing work over a three-month period? In other words, we want to infer causality, not just find a correlation between missing work and cigarette smoking.

b. Since people choose whether and how much to smoke, we certainly cannot treat the data as coming from the experiment we have in mind in part a. (That is, we cannot randomly assign people a daily cigarette consumption.) It is possible that smokers are less healthy to begin with, or have other attributes that cause them to miss work more often. Or, it could go the other way: cigarette consumption may be related to personality traits that make people harder workers. In any case, $cigs$ might be correlated with the unobservables in the equation.

c. If we start with the model

$$\mathbb{P}(y = 1 \mid \mathbf{z}, cigs, q_1) = \Phi(\mathbf{z}_1\delta_1 + \gamma_1 cigs + q_1), \tag{2}$$

but ignore q_1 when it is correlated with $cigs$, we will not consistently estimate anything of interest, whether the model is linear or nonlinear. Thus, we would not be estimating a causal effect. If q_1 is independent of $cigs$, the probit ignoring q_1 does estimate the average partial effect of another cigarette.

d. No. There are many people in the working population who do not smoke. Thus, the distribution (conditional or unconditional) of $cigs$ piles up at zero. Also, since $cigs$ takes on integer values, it cannot be normally distributed. But it is really the pile up at zero that is the most serious issue.

e. Use the Rivers-Vuong test. Obtain the residuals, $\hat{r}_2$, from the regression of $cigs$ on $\mathbf{z}$. Then, estimate the probit of y on $\mathbf{z}_1, cigs, \hat{r}_2$ and use a standard t-test on $\hat{r}_2$. This does not rely on normality of r_2 (or $cigs$). It does, of course, rely on the probit model being correct for y under H_0.

f. Assuming people will not immediately move out of their state of residence when the state implements no smoking laws in the workplace, and that state of residence is roughly independent of general health in the population, a dummy indicator for whether the person works in a state with a new law can be treated as exogenous and excluded from (2). (These situations are often called "natural experiments.") Further, *cigs* is likely to be correlated with the state law indicator because since people will not be able to smoke as much as they otherwise would. Thus, it seems to be a reasonable instrument for *cigs*.

Problem 4: Suppose individuals choose one of three commuting modes: **car**, **bus**, or **bike**.

Let:

- x_i = income of individual i
- $cost_{ij}$ = monetary cost of mode j for individual i
- $time_{ij}$ = travel time of mode j for individual i

The utility of individual i for mode j is:

$$U_{ij} = \beta_1 \cdot cost_{ij} + \beta_2 \cdot time_{ij} + \gamma_j x_i + \varepsilon_{ij}$$

Assume a multinomial logit structure with $\varepsilon_{ij} \sim$ i.i.d. Type I extreme value.

Normalize $\gamma_{\text{bike}} = 0$ for identification.

You estimate the model and obtain:

- $\hat{\beta}_1 = -0.8$, $\hat{\beta}_2 = -0.1$
- $\hat{\gamma}_{\text{car}} = 0.3$, $\hat{\gamma}_{\text{bus}} = 0.1$, $\hat{\gamma}_{\text{bike}} = 0$ (normalized)

For individual i:

- $x_i = 4$ (in tens of thousands of dollars)
- Costs: car = 6, bus = 3, bike = 0
- Times: car = 30, bus = 45, bike = 25

a. Compute the choice probabilities for this individual.

b. Compute the marginal effect of income on the probability of choosing car.

Solution:

a. Compute utilities (up to normalization)

Use:

$$V_{ij} = \beta_1 \cdot cost_{ij} + \beta_2 \cdot time_{ij} + \gamma_j x_i$$

- $V_{\text{car}} = -0.8 \cdot 6 + (-0.1) \cdot 30 + 0.3 \cdot 4 = -4.8 - 3 + 1.2 = -6.6$
- $V_{\text{bus}} = -0.8 \cdot 3 + (-0.1) \cdot 45 + 0.1 \cdot 4 = -2.4 - 4.5 + 0.4 = -6.5$
- $V_{\text{bike}} = -0.8 \cdot 0 + (-0.1) \cdot 25 + 0 = -2.5$

Exponentiate and compute denominator:

- $\exp(-6.6) \approx 0.00136$
- $\exp(-6.5) \approx 0.00150$
- $\exp(-2.5) \approx 0.0821$

$$D = 0.00136 + 0.00150 + 0.0821 \approx 0.085$$

Probabilities:

- $P_{\text{car}} \approx \frac{0.00136}{0.085} \approx 0.016$
- $P_{\text{bus}} \approx \frac{0.00150}{0.085} \approx 0.018$
- $P_{\text{bike}} \approx \frac{0.0821}{0.085} \approx 0.966$

b. Marginal effect of income on P_{car}

Marginal effect:

$$\frac{\partial P_j(x)}{\partial x} = P_j(x) \cdot \left(\gamma_j - \sum_{\ell} \gamma_\ell P_\ell(x) \right)$$

Compute weighted average:

$$\bar{\gamma} = 0.3 \cdot 0.016 + 0.1 \cdot 0.018 + 0 \cdot 0.966 \approx 0.0048 + 0.0018 = 0.0066$$

Then:

$$\frac{\partial P_{\text{car}}}{\partial x} = 0.016 \cdot (0.3 - 0.0066) = 0.016 \cdot 0.2934 \approx 0.0047$$

Problem 5: Consider the multinomial logit model derived from the random utility framework:

$$U_{ij} = X_i' \beta_j + \varepsilon_{ij}, \quad j = 1, \dots, J$$

where:

- U_{ij} is the latent utility individual i obtains from alternative j
- X_i is a vector of observed regressors (same across alternatives)
- β_j is an alternative-specific parameter vector
- $\varepsilon_{ij} \overset{\text{i.i.d.}}{\sim}$ Type I Extreme Value (Gumbel) distribution

The observed outcome is:

$$Y_i = \arg\max_j U_{ij}$$

To ensure identification, set $\beta_J = 0$ for the base category.

a. Show that the choice probabilities are:

$$P_j(X_i) = \frac{\exp(X_i'\beta_j)}{1 + \sum_{\ell=1}^{J-1} \exp(X_i'\beta_\ell)} \quad \text{for } j = 1, \dots, J-1$$

and

$$P_J(X_i) = \frac{1}{1 + \sum_{\ell=1}^{J-1} \exp(X_i'\beta_\ell)}$$

b. Explain why β_J is not identified. What is the economic intuition behind normalizing one coefficient vector?

c. Define the Independence of Irrelevant Alternatives (IIA) assumption and show how it arises in the multinomial logit model.

d. Discuss one scenario where the IIA assumption is likely to be violated and explain why.

Solution:

a. We want to derive the probability that individual i chooses alternative j:

$$P_j(X_i) = \mathbb{P}(U_{ij} > U_{i\ell} \ \forall \ell \neq j)$$

Since the errors are i.i.d. Type I extreme value, the probability that alternative j is chosen is:

$$P_j(X_i) = \frac{\exp(X_i'\beta_j)}{\sum_{\ell=1}^{J} \exp(X_i'\beta_\ell)}$$

Now apply the normalization $\beta_J = 0$:

- For $j = 1, \dots, J-1$:

$$P_j(X_i) = \frac{\exp(X_i'\beta_j)}{\sum_{\ell=1}^{J-1} \exp(X_i'\beta_\ell) + \exp(X_i' \cdot 0)} = \frac{\exp(X_i'\beta_j)}{1 + \sum_{\ell=1}^{J-1} \exp(X_i'\beta_\ell)}$$

- For $j = J$:

$$P_J(X_i) = \frac{\exp(0)}{1 + \sum_{\ell=1}^{J-1} \exp(X_i'\beta_\ell)} = \frac{1}{1 + \sum_{\ell=1}^{J-1} \exp(X_i'\beta_\ell)}$$

b The utility model is invariant to adding the same constant to all U_{ij}. That is:

$$U_{ij} = X_i'\beta_j + \varepsilon_{ij} \quad \text{and} \quad U_{ij} + c$$

yield the same choice probabilities, since ranking among U_{ij} remains unchanged.

Thus, only *differences* in utility matter — not absolute levels.

Therefore, we must impose a normalization, typically $\beta_J = 0$, to identify the model.

Economic intuition: Utility is ordinal. Only relative utilities determine choice, so we can interpret each β_j as *relative to the base category* J.

c. The **Independence of Irrelevant Alternatives (IIA)** states:

> The relative odds of choosing one alternative over another do not depend on the presence or attributes of other alternatives.

In the multinomial logit model:

$$\frac{P_j(X)}{P_k(X)} = \frac{\exp(X'\beta_j)}{\exp(X'\beta_k)} = \exp(X'(\beta_j - \beta_k))$$

This ratio **does not depend on** any other alternative $\ell \neq j, k$.

This strong property arises from the assumption that ε_{ij} are **i.i.d. Gumbel** — implying **independence of utilities** across choices.

d. Example: Red Bus / Blue Bus Problem

Suppose initial alternatives are: - Car with $P = 0.5$ - Bus with $P = 0.5$

Now split "Bus" into two indistinguishable options: - Red Bus - Blue Bus

The multinomial logit model implies:

$$P(\text{Car}) = P(\text{Red Bus}) = P(\text{Blue Bus}) = \frac{1}{3}$$

But this is **unreasonable**, as Red and Blue Bus are **close substitutes** — most of the original bus probability should split between them:

- $P(\text{Car}) \approx 0.5$
- $P(\text{Red Bus}) \approx 0.25$
- $P(\text{Blue Bus}) \approx 0.25$

Conclusion: IIA fails when some alternatives are *more similar* than others. In such cases, nested logit or multinomial probit are better suited.