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.

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?

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 $cigs$?

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.

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.