International School of Economics at TSU

Econometrics 2

Lasha Chochua

Problem Set 10

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?

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

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 we have a distribution with the following pdf (called a gamma distribution)*

$$f(x \mid a) = \frac{a^5}{(4)!} x^4 e^{-ax}.$$

Furthermore, Suppose we have independent data $x_1, x_2, ..., x_m$ drawn from this distribution. Find the maximum likelihood estimate (MLE) for a.

Problem 5: *In this problem we will use maximum likelihood estimates to develop Gauss' method of least squares for fitting lines to data.*

Bivariate data means data of the form

$$(x_1, y_1),\ (x_2, y_2),\ \ldots,\ (x_n, y_n).$$

For bivariate data the simple linear regression model assumes that, for some values of the parameters a and b, we have

$$y_i = ax_i + b + \text{ random measurement error.}$$

The model assumes the measurement errors are independent and identically distributed and follow a $N(0, \sigma^2)$ *distribution. (The values* x_i *may or may not be random.)*

It turns out that, under some assumptions about random variation of measurement error, one way to find a "best" line is by solving a maximum likelihood problem.

The goal is to find the values of the model parameters a *and* b *that give the MLE for this model. To guide you, we note that the model says that*

$$y_i \sim N(ax_i + b, \sigma^2).$$

Also remember that you know the density function for this distribution.

(a) *For a general datum* (x_1, y_1) *give the likelihood and log likelihood functions (these will be functions of* y_1*,* x_1*,* a*,* b*, and* σ*.)*

(b) *Consider the data (1,8), (3,2), (5,1). Assume that* $\sigma = 3$ *is a known constant and find the maximum likelihood estimate for* a *and* b*.*

Note: *We gave you a specific value of* σ*, to avoid the distraction of one more symbol. If you look at your calculations, you should see that the value of* σ *plays no role in finding the MLE for* a *and* b*. We get the same answer no matter what the value.*

Problem 6:

(a) *Suppose we have data 1.2, 2.1, 1.3, 10.5, 5 which we know is drawn independently from a uniform*(a, b) *distribution. Give the maximum likelihood estimate for the parameters* a *and* b*.*

(b) *Suppose we have data* $x_1, x_2, \ldots, x_n$ *which we know is drawn independently from a uniform*(a, b) *distribution. Give the maximum likelihood estimate for the parameters* a *and* b*.*