International School of Economics at TSU
Econometrics II
Lasha Chochua

Problem Set 4

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: Consider the binary choice latent variable model

$$Y_i^* = X_i'\beta + e_i, \qquad Y_i = \mathbb{1}\{Y_i^* > 0\}, \qquad e_i \sim \mathcal{N}(0, \sigma^2). \tag{1}$$

Suppose $\sigma^2 \neq 1$, but you estimate the standard probit specification that imposes unit error variance.

a. Show that the probit response probability under (1) is $\Phi(X_i'\beta/\sigma)$, and conclude that what the probit MLE actually identifies is $\beta^* = \beta/\sigma$, not β itself.

b. Show that the marginal effect of X_{ij} on $\mathbb{P}(Y_i = 1 \mid X_i)$ is invariant to the normalization in the sense that the *true* marginal effect equals the marginal effect computed using the rescaled coefficient β^*.

Problem 2: Let $\hat{\beta}$ be the MLE for the logit model with i.i.d. data $\{(Y_i, X_i)\}_{i=1}^n$, $Y_i \in \{0, 1\}$. Define the score and the negative Hessian as

$$S_n(\beta) = \sum_{i=1}^n X_i(Y_i - \Lambda(X_i'\beta)), \qquad \mathcal{H}_n(\beta) = \sum_{i=1}^n X_i X_i' \, \Lambda(X_i'\beta)(1 - \Lambda(X_i'\beta)). \tag{2}$$

a. Derive the asymptotic distribution of $\hat{\beta}$ under correct specification, including an explicit derivation via a first-order Taylor expansion of the score around the pseudo-true value β_0.

b. Now suppose the model is misspecified: $\mathbb{P}(Y_i = 1 \mid X_i) = m(X_i)$ for some function m that does *not* equal $\Lambda(X_i'\beta)$ for any β. Show that the information matrix equality fails and that valid inference requires the full sandwich variance. Give an explicit expression for Ω under misspecification.

Problem 3: Consider a probit model with a single regressor: $\mathbb{P}(Y_i = 1 \mid X_i) = \Phi(\beta_0 + \beta_1 X_i)$, with $X_i \sim \mathcal{N}(0, 1)$ and $(\beta_0, \beta_1) = (0, 1)$. You wish to test $H_0 : \beta_1 = 1$ against $H_1 : \beta_1 \neq 1$.

a. Write down the Wald, likelihood ratio, and score test statistics. State explicitly what each requires you to compute.

b. Under H_0, all three statistics are asymptotically equivalent and converge to a χ^2_1 distribution. Briefly explain *why* they are asymptotically equivalent, and give one practical reason you might prefer one over the others.

Problem 4: A researcher fits both a linear probability model (LPM) and a logit model to the same binary outcome data and reports that the LPM coefficient on a regressor X_j is 0.04, while the logit coefficient is 0.18. The researcher concludes that the logit model finds "a much larger effect."

a. Explain why this comparison is misleading. What is the correct way to compare effect magnitudes across the two models?

b. Suppose the LPM produces fitted probabilities $\hat{P}_i$ in the range $[-0.05, 0.92]$. State two reasons why a binary choice researcher might still prefer logit despite the closeness of average marginal effects.

Problem 5: A researcher estimates a logit model for the probability that a household owns its home, with regressors *income* (in tens of thousands of dollars) and *college* (a dummy equal to 1 if the household head has a college degree). The estimated coefficients are

$$\hat{\beta}_{\text{income}} = 0.25, \qquad \hat{\beta}_{\text{college}} = 0.80, \qquad \hat{\beta}_0 = -1.50.$$

a. Derive the relationship between the logit coefficients and the **odds ratio** $\mathbb{P}(Y = 1 \mid X)/\mathbb{P}(Y = 0 \mid X)$. Then interpret $\hat{\beta}_{\text{income}}$ and $\hat{\beta}_{\text{college}}$ on the odds-ratio scale.

b. A student looks at part (a) and concludes: "So a college degree more than doubles the *probability* of homeownership." Explain carefully why this conclusion is wrong, and compute the correct marginal effect of *college* on the probability of homeownership for a household with *income* $= 5$ (i.e., \$50,000).