PROBLEM SET 3

Problem 1 (The Probit Model)

Consider the probit model

$$\begin{aligned} Y &= 1\left[Y^* > 0\right], \\ Y^* &= X\beta + U, \end{aligned}$$

where U and X are independent and U is normally distributed with mean zero and unit variance, with cumulative distribution function denoted

$$\Phi\left(c\right) \equiv \Pr\left\{U \leq c\right\}.$$

Y is a binary outcome variable that takes the value one if and only if $Y^* > 0$. $X = (X_1, ..., X_k)$ is a k-dimensional row vector and $\beta = (\beta_1, ..., \beta_k)'$ is a column vector of parameters to be estimated. A random sample of size n, $\{(y_i, x_i) : i = 1, ..., n\}$ is observed. Realizations of U and Y^* are not observed.

(a) Derive the probability that $Y = 1$ and the probability that $Y = 0$ conditional on $X = x$.

(b) Suppose that there is particular interest in the change in $\Pr\left(Y = 1|X = x\right)$ with respect to the k^{th} component of x, x_k. Derive the partial effect of x_k on $\Pr\left(Y = 1|X = x\right)$, assuming that X_k is continuously distributed.

(c) Derive the log-likelihood for maximum likelihood estimation of the parameter β.

(d) Suppose someone suggests that you estimate β using OLS instead of maximizing the likelihood in part (c). Provide two drawbacks with application of OLS with a binary outcome variable.

(e) What condition do you require for identification, and hence for consistency of the maximium likelihood estimator? (Hint: It is the same condition one would require with the linear probability model.)

Problem 2 (MLE and the Information Equality)

Let $y = (y_1, y_2, \ldots, y_n)$ be a random iid sample from the distribution $y_i \sim \mathcal{N}(\log\theta, 1)$, where $\theta > 0$ is an unknown scalar parameter.

(a) Write down the log-likelihood function $\log f(y|\theta) = \log\prod_{i=1}^{n} f(y_i|\theta)$ and calculate the maximum likelihood estimator $\hat{\theta}$. (Hint: remember that the pdf of a normally distributed random variable X with mean μ and variance σ^2 reads $\frac{1}{\sigma\sqrt{2\pi}}\exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$.)

(b) Derive the Hessian $H(y_i, \theta) = \frac{\partial^2 \log f(y_i|\theta)}{\partial\theta^2}$ for this model.

(c) Calculate the expected value of the Hessian $\mathbb{E}H(y_i, \theta_0)$ at the true parameter θ_0.

(d) Use your result in (c) to calculate the asymptotic variance $\text{AsyVar}(\sqrt{n}\hat{\theta})$, using our general result for the asymptotic variance of maximum likelihood estimators.

(e) Let $\hat{\mu} = \log\hat{\theta}$. Use your result in (d) to calculate the asymptotic variance $\text{AsyVar}(\sqrt{n}\hat{\mu})$.

Problem 3 (Linear Models, Applied)

Consider the probability of being arrested in a given year:

$$\begin{aligned} \mathrm{P}(arr86 = 1|\boldsymbol{X}) = G(&\beta_0 + \beta_1 pcnv + \beta_2 tottime + \beta_3 ptime86 + \beta_4 avgsen \\ &+\beta_5 inc86 + \beta_6 black + \beta_7 hispan + \beta_8 born60 + u) \end{aligned} \tag{1}$$

1) Read the grogger.txt dataset into R and generate a dummy variable *arr86* that equals one for individuals who was arrested in 1986 and zero otherwise.

2) Estimate equation (1) by OLS to obtain a linear probability model.

3) Visually inspect the residuals, and test for heteroskedasticity using a Breusch-Pagan test.

4) Correct for heteroskedasticity by computing robust standard errors. (What else could be done?)

5) What is the estimated effect of on the probability of arrest if *pcnv* increases from 0.25 to 0.75.

6) Test for joint significance of *avgsen* and *tottime* using a nonrobust and a robust test for the LPM.

7) Estimate (1) using a logit model.

8) Estimate (1) using a probit model.

9) Compare the *qualitative* results of the logit and probit models. Check that the logit coefficients are approximately 1.6 times the probit estimates.

10) Compute marginal effects of the probit model for a black man who were born in 1960 and has average values of *avgsen*, *tottime*, *inc86*, *ptime86* in the sample. What is the estimated effect on the probability of arrest if *pcnv* increases from 0.25 to 0.75?

11) In the probit model, find the average partial effect (APE) of *pcnv*.

12) Compare the marginal effects at the mean with the APE as well as the results from the LPM.

13) Compare the predictions of the three models:

    a) Generate a dummy equal to one if the model predicts the person to be arrested (i.e. if $\boldsymbol{X\beta} > 0$ in the logit/probit case and if $\boldsymbol{X\beta} > 0.5$ in the LPM)

    b) Generate a dummy equal to one if the model correctly predicts the person to be arrested or not arrested.

    c) Compare the predictions of the LPM and the probit model: which model has the highest share of correct predictions?

14) In the probit model, add the terms $pcnv^2$, $ptime86^2$, and $inc86^2$ to the model. Are these jointly significant? Describe the relationship between the probability of arrest and *pcnv*. In particular, at what point does the probability of conviction have a negative effect on the probability of arrest?