Limited Dependent Variables & Selection: PS #2

Francis DiTraglia

HT 2021

This problem set is due on *Monday of HT Week 7 at noon*. You do not have to submit solutions to questions 3–4; they will be discussed in the class but will not be marked.

1. Let $(y_1, x_1), \cdots, (y_N, x_N)$ be a collection of iid observations where $y_i \in \{0, 1\}$ and x_i is continuously distributed. Suppose that $p(x_i) \equiv \text{Prob}(y_i = 1|x_i) = F(\alpha + \beta x_i)$ where $F(z) = e^z/(1 + e^z)$ and (α, β) are unknown parameters.
   (a) Derive an expression for the partial effect of x_i on $p(x_i)$ in this model.
   (b) Write out the log-likelihood function $\ell_N(\alpha, \beta)$ for this model, simplifying your result as far as possible.
   (c) Using your answer to the preceding part, derive the first-order conditions for the maximum likelihood estimators of α and β. Simplify your results as far as possible.

2. This question concerns the Probit regression model $\mathbb{P}(y = 1|\mathbf{x}) = \Phi(\mathbf{x}'\boldsymbol{\beta})$ where Φ is the standard normal CDF.
   (a) Derive the first order conditions for the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ based on an iid sample $(y_1, \mathbf{x}), \dots, (y_N, \mathbf{x}_N)$.
   (b) Suppose that $y = \mathbb{1}\{\mathbf{x}'\boldsymbol{\beta} + u > 0\}$ where $u \sim \mathcal{N}(0, 1)$ independently of $\mathbf{x}$ and $\mathbb{1}(\cdot)$ is the indicator function. Show that this model is in fact *exactly equivalent* to the Probit regression model.

Question #3 will not be marked; you do not have to submit a solution.

3. Consider a logit-Family model with $P_{ni} = \exp(V_{ni})/\sum_{j=1}^{J}\exp(V_{nj})$ and $V_{nj} = \mathbf{x}_{nj}'\boldsymbol{\beta}$.
   (a) What *variety* of Logit-family model is this? How can you tell?
   (b) Show that the partial effects for this model are given by

$$\frac{\partial P_{ni}}{\partial \mathbf{x}_{ni}} = P_{ni}(1 - P_{ni})\boldsymbol{\beta}, \quad \text{and} \quad \frac{\partial P_{ni}}{\partial \mathbf{x}_{nk}} = -P_{ni}P_{nk}\boldsymbol{\beta} \quad \text{for } i \neq k$$

Question #4 will not be marked; you do not have to submit a solution.

4. *This question is adapted from Wooldridge (2010).* Consider the Heckman selection model from the lecture slides. Assumption (d) of this model states that the conditional mean of u_1 given v_2 is linear: $\mathbb{E}(u_1|v_2) = \gamma_1 v_2$. In this question, you will explore the consequences of replacing Assumption (d) with a *quadratic* conditional mean function, in particular

$$\text{Assumption (d*)} \quad \mathbb{E}(u_1|v_2) = \gamma_1 v_2 + \gamma_2(v_2^2 - 1).$$

In your answers to the following parts, assume that all assumptions other than (d) of the Heckman Selection model continue to apply.

(a) Show that Assumption (c) and (d*) imply $\mathbb{E}(u_1) = 0$. Using your answer, explain why the RHS of Assumption (d*) does *not* take the form $\gamma_1 v_2 + \gamma_2 v_2^2$.

(b) Let a be a constant, $z \sim N(0,1)$ and $\lambda(\cdot)$ be the inverse Mills ratio defined in the lecture slides. It can be shown that:

$$\text{Var}(z|z > -a) = 1 - \lambda(a)\left[\lambda(a) + a\right].$$

Use this result to prove that

$$\mathbb{E}(y_1|\mathbf{x}, y_2 = 1) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\lambda(\mathbf{x}'\boldsymbol{\delta}_2) - \gamma_2\lambda(\mathbf{x}'\boldsymbol{\delta}_2)\mathbf{x}'\boldsymbol{\delta}_2.$$

Hint: $\mathbb{E}(v_2^2|v_2 > -a) = \text{Var}(v_2|v_2 > -a) + \left[\mathbb{E}(v_2|v_2 > -a)\right]^2$.

(c) Using the expression for $\mathbb{E}(y_1|\mathbf{x}, y_2 = 1)$ from the preceding part, explain how to carry out the Heckman Two-step procedure under assumption (d*).

(d) Consider a "naïve" OLS regression of y_1 on $\mathbf{x}_1$ for the subset of individuals with $y_2 = 1$. Without actually running the naïve regression, explain how you could use the estimates from your Heckman Two-step procedure in the preceding part to determine whether or not the naïve OLS of β_1 would be biased.