# 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 $(\alpha, \beta)$ are unknown parameters.

   (a) Derive an expression for the partial effect of $x_i$ on $p(x_i)$ in this model.

   **Solution:** We have

   $$\frac{d}{dx}p(x) = \frac{\partial}{\partial x}F(\alpha + \beta x) = F'(\alpha + \beta x)\beta$$

   so all that remains is to calculate $F'$. By the quotient rule,

   $$F'(z) = \frac{d}{dz}\left(\frac{e^z}{1 + e^z}\right) = \frac{e^z(1 + e^z) - e^z e^z}{(1 + e^z)^2} = \frac{e^z}{(1 + e^z)^2}$$

   Therefore,

   $$\frac{d}{dx}p(x) = \left\{\frac{\exp(\alpha + \beta x)}{[1 + \exp(\alpha + \beta x)]^2}\right\}\beta$$

   (b) Write out the log-likelihood function $\ell_N(\alpha, \beta)$ for this model, simplifying your result as far as possible.

   **Solution:** The likelihood of a single observation is given by

   $$L_i(\alpha, \beta) = f(y_i|x_i, \alpha, \beta) = F(\alpha + \beta x_i)^{y_i}\left[1 - F(\alpha + \beta x_i)\right]^{1 - y_i}$$

   and the corresponding log-likelihood is

   $$\ell_i(\alpha, \beta) = \log L_i(\alpha, \beta) = y_i \log\left[F(\alpha + \beta x_i)\right] + (1 - y_i)\log\left[1 - F(\alpha + \beta x_i)\right].$$

Substituting the definition of $F$ and simplifying, we obtain

$$\begin{aligned}
\ell_i(\alpha, \beta) &= y_i \log \left[ \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right] + (1 - y_i) \log \left[ 1 - \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right] \\
&= y_i(\alpha + \beta x_i) - y_i \log\left[1 + \exp(\alpha + \beta x_i)\right] + (1 - y_i) \log(1) \\
&\quad - (1 - y_i) \log\left[1 + \exp(\alpha + \beta x_i)\right] \\
&= y_i(\alpha + \beta x_i) - \log\left[1 + \exp(\alpha + \beta x_i)\right]
\end{aligned}$$

Because our observations are iid, the log-likelihood function equals the sum of the likelihoods of each observation. Hence,

$$\ell_N(\alpha, \beta) = \sum_{i=1}^{N} \left\{ y_i(\alpha + \beta x_i) - \log\left[1 + \exp(\alpha + \beta x_i)\right] \right\}$$

(c) Using your answer to the preceding part, derive the first-order conditions for the maximum likelihood estimators of $\alpha$ and $\beta$. Simplify your results as far as possible.

**Solution:** Differentiating,

$$\begin{aligned}
\frac{\partial \ell_N}{\partial \alpha} &= \sum_{i=1}^{N} \frac{\partial}{\partial \alpha} \ell_i(\alpha, \beta) = \sum_{i=1}^{N} \frac{\partial}{\partial \alpha} \left\{ y_i(\alpha + \beta x_i) - \log\left[1 + \exp(\alpha + \beta x_i)\right] \right\} \\
&= \sum_{i=1}^{N} \left[ y_i - \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)} \right] = \sum_{i=1}^{N} \left[ y_i - F(\alpha + \beta x_i) \right]
\end{aligned}$$

and similarly

$$\begin{aligned}
\frac{\partial \ell_N}{\partial \beta} &= \sum_{i=1}^{N} \frac{\partial}{\partial \beta} \ell_i(\alpha, \beta) = \sum_{i=1}^{N} \frac{\partial}{\partial \beta} \left\{ y_i(\alpha + \beta x_i) - \log\left[1 + \exp(\alpha + \beta x_i)\right] \right\} \\
&= \sum_{i=1}^{N} \left[ y_i x_i - \frac{\exp(\alpha + \beta x_i) x_i}{1 + \exp(\alpha + \beta x_i)} \right] = \sum_{i=1}^{N} \left[ y_i - F(\alpha + \beta x_i) \right] x_i
\end{aligned}$$

Therefore, the first-order conditions are

$$\sum_{i=1}^{N} \left[ y_i - F(\widehat{\alpha} + \widehat{\beta} x_i) \right] \begin{bmatrix} 1 \\ x_i \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}.$$

2. This question concerns the Probit regression model $\mathbb{P}(y = 1|\mathbf{x}) = \Phi(\mathbf{x}'\boldsymbol{\beta})$ where $\Phi$ is the standard normal CDF.

(a) Derive the first order conditions for the maximum likelihood estimator $\widehat{\boldsymbol{\beta}}$ based on an iid sample $(y_1, \mathbf{x}), \ldots, (y_N, \mathbf{x}_N)$.

**Solution:** The Probit likelihood for a single observation is given by

$$L_i(\boldsymbol{\beta}) = \Phi(\mathbf{x}_i'\boldsymbol{\beta})^{y_i} \left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right]^{1-y_i}$$

and hence the corresponding log-likelihood is

$$\ell_i(\boldsymbol{\beta}) \equiv \log L_i(\boldsymbol{\beta}) = y_i \log \Phi(\mathbf{x}_i'\boldsymbol{\beta}) + (1 - y_i) \log\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right]$$

while the score vector is

$$\begin{aligned}
\mathbf{s}_i \equiv \frac{\partial}{\partial \boldsymbol{\beta}} \ell_i(\boldsymbol{\beta}) &= y_i \left[\frac{\varphi(\mathbf{x}_i'\boldsymbol{\beta})}{\Phi(\mathbf{x}_i'\boldsymbol{\beta})}\right] \mathbf{x}_i - (1 - y_i) \left[\frac{\varphi(\mathbf{x}_i'\boldsymbol{\beta})}{1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})}\right] \mathbf{x}_i \\
&= \frac{\varphi(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i}{\Phi(\mathbf{x}_i'\boldsymbol{\beta})\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right]} \left\{\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right] y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta})(1 - y_i)\right\} \\
&= \frac{\varphi(\mathbf{x}_i'\boldsymbol{\beta})\mathbf{x}_i \left[y_i - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right]}{\Phi(\mathbf{x}_i'\boldsymbol{\beta})\left[1 - \Phi(\mathbf{x}_i'\boldsymbol{\beta})\right]}
\end{aligned}$$

Because $\Phi$ lacks a closed-form, this expression cannot be simplified further. The first-order conditions are simply $\sum_{i=1}^N \mathbf{s}_i = \mathbf{0}$.

(b) Suppose that $y = \mathbb{1}\left\{\mathbf{x}'\boldsymbol{\beta} + u > 0\right\}$ 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.

**Solution:** First note that

$$\mathbb{P}(y = 1|\mathbf{x}) = \mathbb{P}(\mathbf{x}_i'\boldsymbol{\beta} + u > 0) = \mathbb{P}(-u < \mathbf{x}'\boldsymbol{\beta}).$$

Now, since $u$ is independent of $\mathbf{x}$, so is $-u$. Moreover, by the symmetry of the normal distribution $-u \sim \mathcal{N}(0, 1)$. Therefore $\mathbb{P}(-u < \mathbf{x}'\boldsymbol{\beta}) = \Phi(\mathbf{x}'\boldsymbol{\beta})$.

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?

   **Solution:** Because all of the attributes vary across alternatives, this is a conditional logit model.

   (b) Show that the partial effects for this model are given by

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

**Solution:** By the quotient rule,

$$\begin{aligned}\frac{\partial P_{ni}}{\partial \mathbf{x}_{nk}} &= \frac{\partial}{\partial \mathbf{x}_{nk}}\left[\frac{\exp(V_{ni})}{\sum_{j=1}^{J}\exp(V_{nj})}\right]\\ &= \frac{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{ni}) - \exp(V_{ni})\left[\sum_{j=1}^{J}\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{nj})\right]}{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]^2}\end{aligned}$$

Now, because $V_{nj}$ only contains $j$-specific attributes $\partial \exp(V_{nj})/\partial \mathbf{x}_{nk} = \mathbf{0}$ for any $k \neq j$. Hence, the preceding simplifies to

$$\begin{aligned}\frac{\partial P_{ni}}{\partial \mathbf{x}_{nk}} &= \frac{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{ni}) - \exp(V_{ni})\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{nk})}{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]^2}\\ &= \frac{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{ni})}{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]^2} - \frac{\exp(V_{ni})\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{nk})}{\left[\sum_{j=1}^{J}\exp(V_{nj})\right]^2}\\ &= \frac{\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{ni})}{\sum_{j=1}^{J}\exp(V_{nj})} - \frac{P_{ni}\left[\frac{\partial}{\partial \mathbf{x}_{nk}}\exp(V_{nk})\right]}{\sum_{j=1}^{J}\exp(V_{nj})}\\ &= P_{ni}\left(\frac{\partial V_{ni}}{\partial \mathbf{x}_{nk}}\right) - P_{ni}P_{nk}\left(\frac{\partial V_{nk}}{\partial \mathbf{x}_{nk}}\right)\end{aligned}$$

Now, suppose that $i \neq k$. Then $\partial V_{ni}/\partial \mathbf{x}_{nk} = \mathbf{0}$, so we obtain

$$\frac{\partial P_{ni}}{\partial \mathbf{x}_{nk}} = -P_{ni}P_{nk}\left(\frac{\partial V_{nk}}{\partial \mathbf{x}_{nk}}\right) = -P_{ni}P_{nk}\boldsymbol{\beta}, \quad i \neq k.$$

If instead $i = k$, we obtain

$$\begin{aligned}\frac{\partial P_{ni}}{\partial \mathbf{x}_{ni}} &= P_{ni}\left(\frac{\partial V_{ni}}{\partial \mathbf{x}_{ni}}\right) - P_{ni}P_{ni}\left(\frac{\partial V_{ni}}{\partial \mathbf{x}_{ni}}\right)\\ &= P_{ni}(1 - P_{ni})\left(\frac{\partial V_{ni}}{\partial \mathbf{x}_{ni}}\right)\\ &= P_{ni}(1 - P_{ni})\boldsymbol{\beta}\end{aligned}$$

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 con-

ditional 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 $\gamma_1 v_2 + \gamma_2 v_2^2$.

**Solution:** By the Law of Iterated Expectations and Assumption (d*)

$$\mathbb{E}(u_1) = \mathbb{E}[\mathbb{E}(u_1|v_2)] = \mathbb{E}[\gamma_1 v_2 + \gamma_2(v_2^2 - 1)] = \gamma_1 \mathbb{E}(v_2) + \gamma_2[\mathbb{E}(v_2^2) - 1].$$

Since $z \sim \mathcal{N}(0, 1)$ by Assumption (c), it follows that

$$\mathbb{E}(u_1) = \gamma_1 \times 0 + \gamma_2 \times (1 - 1) = 0.$$

If instead Assumption (d*) had taken the form $\gamma_1 v_2 + \gamma_2 v_2^2$, i.e. without subtracting one from the second term, we would have obtained $\mathbb{E}(u_1) = \gamma_2$, violating the part of Assumption (b) that imposes $\mathbb{E}(u_1) = 0$.

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

**Solution:** The argument is very similar to that given in the lecture slides, with a few minor modifications. We'll begin by adapting the logic of Lemma 1 from the slides. Since step 1 of the lemma only used Assumption (b), it remains true that $u_1$ and $\mathbf{x}$ are conditionally independent given $v_2$. Note that only the *final* part of step 2 uses Assumption (d). Thus everything before this point continues to apply, in particular

$$\mathbb{E}(y_1|\mathbf{x}, v_2) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \mathbb{E}(u_1|v_2).$$

Now, substituting (d*) for $\mathbb{E}(u_1|v_2)$, we obtain

$$\mathbb{E}(y_1|\mathbf{x}, v_2) = \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 v_2 + \gamma_2(v_2^2 - 1).$$

The only change in step 3 is that we now have a *different* expression from step 2, namely the preceding equality. Substituting this, we obtain

$$\begin{aligned}\mathbb{E}(y_1|\mathbf{x}, y_2) &= \mathbb{E}_{v_2|(\mathbf{x},y_2)}\left[\mathbb{E}(y_1|\mathbf{x}, v_2)\right] = \mathbb{E}\left[\mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1 v_2 + \gamma_2(v_2^2 - 1)\middle|\, \mathbf{x}, y_2\right] \\ &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\mathbb{E}[v_2|\mathbf{x}, y_2] + \gamma_2\mathbb{E}\left[(v_2^2 - 1)|\mathbf{x}, y_2\right]\end{aligned}$$

and evaluating this expression at $y_2 = 1$, we see that

$$\begin{aligned}\mathbb{E}(y_1|\mathbf{x}, y_2 = 1) &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\mathbb{E}[v_2|\mathbf{x}, y_2 = 1] + \gamma_2\mathbb{E}\left[(v_2^2 - 1)|\mathbf{x}, y_2 = 1\right] \\ &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\mathbb{E}[v_2|\mathbf{x}, y_2 = 1] + \gamma_2\left\{\mathbb{E}(v_2^2|\mathbf{x}, y_2 = 1) - 1\right\} \\ &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\lambda(\mathbf{x}'\boldsymbol{\delta}_2) + \gamma_2\left\{\mathbb{E}(v_2^2|\mathbf{x}, y_2 = 1) - 1\right\}\end{aligned}$$

since $\mathbb{E}(v_2|\mathbf{x}, y_2 = 1) = \lambda(\mathbf{x}'\boldsymbol{\delta}_2)$ as we showed in Lemma 2 from the lecture slides. All that remains is to calculate $\mathbb{E}(v_2^2|\mathbf{x}, y_2 = 1)$. By the argument from step 1 of Lemma 2 with $v_2^2$ in place of $v_2$, we see that the distribution of $v_2^2$ given $(\mathbf{x}, y_2 = 1)$ is the same as that of $v_2^2$ conditional on $v_2^2 > c$, where we define $c \equiv -\mathbf{x}'\boldsymbol{\delta}_2$ as in the slides. Thus it suffices for us to derive an expression for $\mathbb{E}(v_2^2|v_2 > c)$ where $v_2 \sim N(0, 1)$. Now, recall the hint from the problem statement:

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

In the lecture slides we showed that $\mathbb{E}(v_2|v_2 > -a) = \lambda(a)$, and from the result in the problem statement we have $\mathrm{Var}(v_2|v_2 > -a) = 1-\lambda(a)[\lambda(a)+a]$ where $\lambda(a) \equiv \varphi(a)/\Phi(a)$ is the inverse Mills ratio. Substituting into the preceding equality and simplifying,

$$\begin{aligned}\mathbb{E}(v_2^2|v_2 > -a) &= 1 - \lambda(a)\left[\lambda(a) + a\right] + \left[\lambda(a)\right]^2 \\ &= 1 - \lambda(a)^2 - a\lambda(a) + \lambda(a)^2 \\ &= 1 - a\lambda(a)\end{aligned}$$

and taking $a = -c = \mathbf{x}'\boldsymbol{\delta}_2$, it follows that

$$\mathbb{E}(v_2^2|\mathbf{x}, y_2 = 1) = \mathbb{E}(v_2^2|v_2 > -a) = 1 - a\lambda(a) = 1 - (\mathbf{x}'\boldsymbol{\delta}_2)\cdot\lambda(\mathbf{x}'\boldsymbol{\delta}_2).$$

Finally, substituting this into our expression for $\mathbb{E}(y_1|\mathbf{x}, y_2 = 1)$ from above,

$$\begin{aligned}\mathbb{E}(y_1|\mathbf{x}, y_2 = 1) &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\lambda(\mathbf{x}'\boldsymbol{\delta}_2) + \gamma_2\left\{\mathbb{E}(v_2^2|\mathbf{x}, y_2 = 1) - 1\right\} \\ &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\lambda(\mathbf{x}'\boldsymbol{\delta}_2) + \gamma_2\left\{1 - (\mathbf{x}'\boldsymbol{\delta}_2)\cdot\lambda(\mathbf{x}'\boldsymbol{\delta}_2) - 1\right\} \\ &= \mathbf{x}_1'\boldsymbol{\beta}_1 + \gamma_1\lambda(\mathbf{x}'\boldsymbol{\delta}_2) - \gamma_2\lambda(\mathbf{x}'\boldsymbol{\delta}_2)\mathbf{x}'\boldsymbol{\delta}_2.\end{aligned}$$

(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*).

**Solution:** The first step is the same as in the lecture slides: run Probit on the full sample to estimate $\widehat{\delta}_2$ and then construct $\widehat{\lambda}_i \equiv \lambda(\mathbf{x}_i'\widehat{\delta}_2)$. In the second step, we run an OLS regression $y_{i1}$ on $\mathbf{x}_{i1}$, $\widehat{\lambda}_i$ *and* $\widehat{\lambda}_i\mathbf{x}_i'\boldsymbol{\delta}_2$ using the selected sample, i.e. the individuals with $y_{2i} = 1$. Compared to the procedure from class, this modified second step includes an extra regressor, namely $\widehat{\lambda}_i\mathbf{x}_i'\boldsymbol{\delta}_2$.

(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 $\beta_1$ would be biased.

**Solution:** The parameters $\gamma_1$ and $\gamma_2$ govern selection bias. If these are both zero, then the naïve regression does not suffer from selection bias. Thus, you could examine the estimates $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ of these estimates, and perhaps test the joint restriction that $\gamma_1 = \gamma_2 = 0$, to determine whether sample selection bias is present in a particular application.