Limited Dependent Variables & Selection: PS #2

Francis DiTraglia

HT 2021

This problem set is due on *Friday in Week 3 of HT 2021*. You need only submit solutions to questions 1–4, as question #5 will not be marked. See the explanation immediately preceding question #5 for further information.

1. Suppose that we observe N iid draws $(y_i, \mathbf{x}_i)$ from a population of interest where $y_i \in \{0, 1\}$ and $\mathbf{x}_i$ is a $(k \times 1)$ vector of dummy variables indicating which of k mutually exclusive "bins" person i falls into. For example, suppose that $k = 2$ and we defined the bins to be "female" and "male." Then $\mathbf{x}_i' = \begin{bmatrix} 1 & 0 \end{bmatrix}$ would indicate that person i is female while $\mathbf{x}_1' = \begin{bmatrix} 0 & 1 \end{bmatrix}$ would indicate that person i is male. Note that $\mathbf{x}_i$ does not include an intercept to avoid the dummy variable trap. The following parts explore the results of fitting the linear probability model $\mathbb{P}(y_i|\mathbf{x}_i) = \mathbf{x}_i'\boldsymbol{\beta}$ by running an OLS regression of y_i on $\mathbf{x}_i$. Following the usual conventions, define

$$\mathbf{X}' = \begin{bmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_N \end{bmatrix}, \quad \mathbf{y}' = \begin{bmatrix} y_1 & y_2 & \cdots & y_N \end{bmatrix}$$

   (a) Let N_j denote the number of individuals in the sample who fall into category j. In other words, if $x_i^{(j)}$ is the jth element of $\mathbf{x}_i$, then $N_j \equiv \sum_{i=1}^N x_i^{(j)}$. Show that

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} N_1 & & & 0 \\ & N_2 & & \\ & & \ddots & \\ 0 & & & N_k \end{bmatrix}$$

   i.e. that $\mathbf{X}'\mathbf{X}$ is a $(k \times k)$ diagonal matrix with jth diagonal element N_j.

Solution: Expressed in summation form,

$$\mathbf{X}'\mathbf{X} = \begin{bmatrix} \mathbf{x}_1 & \cdots & \mathbf{x}_N \end{bmatrix} \begin{bmatrix} \mathbf{x}_1 \\ \vdots \\ \mathbf{x}_N \end{bmatrix} = \sum_{i=1}^N \mathbf{x}_i \mathbf{x}_i'$$

Consider an arbitrary element $\mathbf{x}_i\mathbf{x}_i'$ of the sum. Because the k dummy variables in $\mathbf{x}_i$ encode membership in k mutually exclusive categories, $x_i^{(j)}x_i^{(\ell)} = 0$ for any $j \neq \ell$. In other words, all of the off-diagonal elements of $\mathbf{x}_i\mathbf{x}_i'$ are zero.

Moreover, because each element of $\mathbf{x}_i$ is zero or one, the diagonal elements $x_i^{(j)}x_i^{(j)}$ simply equal $x_i^{(j)}$. Therefore, $\mathbf{x}_i\mathbf{x}_i = \text{diag}\{\mathbf{x}_i\}$ and we obtain

$$\mathbf{X}'\mathbf{X} = \sum_{i=1}^{N} \text{diag}\{\mathbf{x}_i\} = \text{diag}(N_1, \ldots, N_k).$$

(b) Substitute the preceding part into $\widehat{\boldsymbol{\beta}} \equiv (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ to obtain a simple, closed-form expression for $\widehat{\beta}_j$. Interpret your result.

Solution: We have defined the $(k \times 1)$ vector $\mathbf{x}_i'$ to be the ith *row* of $\mathbf{X}$. Now let $\boldsymbol{x}^{(j)}$ be the jth *column* of $\mathbf{X}$, i.e. the $(N \times 1)$ vector that stacks all N observations of $x_i^{(j)}$. Then we have

$$\mathbf{X} = \begin{bmatrix} \boldsymbol{x}^{(1)} & \cdots & \boldsymbol{x}^{(k)} \end{bmatrix}$$

and hence,

$$\widehat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y} = \begin{bmatrix} 1/N_1 & & & 0 \\ & 1/N_2 & & \\ & & \ddots & \\ 0 & & & 1/N_k \end{bmatrix} \begin{bmatrix} \boldsymbol{x}^{(1)\prime} \\ \vdots \\ \boldsymbol{x}^{(k)\prime} \end{bmatrix} \mathbf{y} = \begin{bmatrix} \mathbf{y}'\boldsymbol{x}^{(1)}/N_1 \\ \vdots \\ \mathbf{y}'\boldsymbol{x}^{(k)}/N_k \end{bmatrix}$$

Thus, we have shown that

$$\widehat{\beta}_j = \mathbf{y}'\boldsymbol{x}^{(j)}/N_j = \frac{1}{N_j}\sum_{i=1}^{N} x_i^{(j)} y_i = \frac{\#\text{of people in bin } j \text{ with } y = 1}{\#\text{of people in bin } j}$$

Hence $\widehat{\beta}_j$ is simply the sample analogue of $\mathbb{P}(y_i = 1 | i \text{ in bin } j)$.

(c) A critique of the LPM is that it can yield predicted probabilities that are greater than one or less than zero. Is this a problem in the present example?

Solution: No. In this example our prediction $\widehat{y}_i$ for a person who falls into bin j is simply $\widehat{\beta}_j$. We see from the expression in the preceding part that this quantity is always between zero and one.

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 $(y_1, \mathbf{x}), \ldots, (y_N, \mathbf{x}_N)$.

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 Φ 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 $\mathbb{P}(-u < \mathbf{x}'\boldsymbol{\beta}) = \Phi(\mathbf{x}'\boldsymbol{\beta})$.

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

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

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

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

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

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:

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

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 $\text{Var}(v_2|v_2 > -a) = 1-\lambda(a)[\lambda(a)+a]$ where $\lambda(a) \equiv \varphi(a)/\Phi(a)$ is the inverse Mills ratio. Substituting into the preceding equality and simplifying,

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

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 β_1 would be biased.

> **Solution:** The parameters γ_1 and γ_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.

The following applied question will *not be marked*, but you encouraged to complete it nonetheless as it will build your understanding of the material from the lectures. Solving this problem will requires some of the R material from Lecture #6.

5. *This question is adapted from Wooldridge (2010).* To answer it you will need to use the dataset `BWGHT.RAW`, which can either be downloaded from the MIT Press website for the text, or loaded directly into R using the package `Wooldridge`. Documentation for the dataset is available in the R package or alternatively at `http://fmwww.bc.edu/ec-p/data/wooldridge/bwght.des`

   (a) Create a binary variable called *smokes* that equals one if a woman smokes during pregnancy, zero otherwise. Then estimate a probit regression that uses *motheduc*, *white*, and log(*faminc*) to predict *smokes*. Summarize your results.

   (b) Consider two white women with family income equal to the sample mean: Alice has 12 years of education while Beth has 16. What is the estimated difference in the probability of smoking during pregnancy for Alice compared to Beth?

   (c) Calculate the average partial effect of log(*faminc*) in your estimated model.

   (d) Calculate the pseudo-R-squared of your model.

> **Solution:** See the attached pdf document.

Solution to Question 5 from Problem Set #2

Francis J. DiTraglia

Birthweight Dataset

Mullahy (1997; ReStat)

```
# Load packages for robust standard errors (install them first!)
library(sandwich)
library(lmtest)

# Load the data from the wooldridge package (install it first!)
library(wooldridge)

# View the names of the columns
names(bwght)
```

```
##  [1] "faminc"   "cigtax"   "cigprice" "bwght"    "fatheduc" "motheduc"
##  [7] "parity"   "male"     "white"    "cigs"     "lbwght"   "bwghtlbs"
## [13] "packs"    "lfaminc"
```

Description of relevant variables: `bwght`

- `motheduc` mother's years of education
- `white` equals one if white
- `faminc` 1988 family income in $1000s
- `lfaminc` log of faminc
- `cigs` number of cigarettes smoked per day while pregnant
- `packs` number of packes of smoked per day while pregnant

Part (a) - Define the outcome variable `smokes`

```
# Check that cigs and packs agree about who smoked during pregnancy
any(bwght$cigs == 0 & bwght$packs > 0)
```

```
## [1] FALSE
```

```
any(bwght$packs == 0 & bwght$cigs > 0)
```

```
## [1] FALSE
```

```
# Since they agree, we can use either to define smokes
bwght$smokes <- ifelse(bwght$cigs > 0, yes = 1, no = 0)
```

Part (a) - Fit a Probit Regression

```
smoking_model <- smokes ~ motheduc + white + lfaminc
```

```r
probit <- glm(smoking_model, family = binomial(link = 'probit'), data = bwght)
coeftest(probit)
```

```
##
## z test of coefficients:
##
##               Estimate Std. Error z value  Pr(>|z|)
## (Intercept)  1.126273   0.250374  4.4984 6.848e-06 ***
## motheduc    -0.145060   0.020697 -7.0087 2.405e-12 ***
## white        0.189677   0.110680  1.7137 0.0865758 .
## lfaminc     -0.166911   0.049842 -3.3488 0.0008115 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Part (b) - Predictions for Alice and Beth

```r
mean_faminc <- mean(bwght$faminc)
Alice <- c(motheduc = 12, white = 1, lfaminc = log(mean_faminc))
Beth <- c(motheduc = 16, white = 1, lfaminc = log(mean_faminc))
predict_me <- data.frame(rbind(Alice, Beth))
predictions <- predict(probit, newdata = predict_me, type = 'response')
predictions
```

```
##      Alice       Beth
## 0.16183185 0.05853452
```

```r
diff(predictions)
```

```
##       Beth
## -0.1032973
```

Part (c) - Average partial effect of `lfaminc`

```r
# Average of g(x'beta_hat) where g is the std. normal density: dnorm
# (predict defaults to the scale of x'beta_hat)
mean(dnorm(predict(probit))) * coef(probit)[4]
```

```
##     lfaminc
## -0.03614676
```

Part (d) - Pseudo R-squared

```r
# Fit model with only an intercept
model0 <- smokes ~ 1
probit0 <- glm(model0, family = binomial(link = 'probit'), data = bwght)

# Pseudo R-squared
1 - logLik(probit) / logLik(probit0)
```

```
## 'log Lik.' 0.07838101 (df=4)
```