Limited Dependent Variables & Selection: PS #1

Francis DiTraglia

HT 2021

This problem set is due on *Monday of HT Week 6 at noon*. You do not have to submit solution to questions 1–2; they will be discussed in class but will not be marked.

Question #1 will not be marked; you do not have to submit a solution.

1. Let $y \sim$ Poisson(θ).

   (a) Using steps similar to the derivation of $\mathbb{E}[y]$ from the lecture slides, show that $\mathbb{E}[y(y-1)] = \theta^2$.

   Solution:

   $$\begin{aligned}\mathbb{E}\left[y(y-1)\right] &= \sum_{y=0}^{\infty} y(y-1)\left(\frac{e^{-\theta}\theta^y}{y!}\right) = \sum_{y=2}^{\infty} y(y-1)\left(\frac{e^{-\theta}\theta^y}{y!}\right)\\ &= \theta^2 \sum_{y=2}^{\infty} \frac{e^{-\theta}\theta^{y-2}}{(y-2)!} = \theta^2 \sum_{y=0}^{\infty} \frac{e^{-\theta}\theta^y}{y!} = \theta^2\end{aligned}$$

   The first equality is the definition of $\mathbb{E}[y(y-1)]$ for a Poisson RV. The second uses the fact that $y(y-1) = 0$ for $y = 0$ and $y = 1$ so the first two terms of the infinite sum are zero. The third factors θ^2 out of the infinite sum (we can always do this provided that the sum converges) and cancels $y(y-1)$ from $y!$ in the denominator. The fourth shifts the index of summation, and the final recognizes that the infinite sum is now a Poisson pmf summed over all possible values of y and hence equals one.

   (b) Use your answer to the preceding part, along with the result $\mathbb{E}[y] = \theta$, to show that $\text{Var}(y) = \theta$.

   Solution: Recall that $\text{Var}(y) = \mathbb{E}(y^2) - \mathbb{E}(y)^2$. Hence,

   $$\begin{aligned}\mathbb{E}\left[y(y-1)\right] &= \mathbb{E}(y^2) - \mathbb{E}(y)\\ &= \mathbb{E}(y^2) - \mathbb{E}(y)^2 + \left[\mathbb{E}(y)^2 - \mathbb{E}(y)\right]\\ &= \text{Var}(y) + \left[\mathbb{E}(y)^2 - \mathbb{E}(y)\right]\end{aligned}$$

and solving for $\text{Var}(y)$,

$$\text{Var}(y) = \mathbb{E}\left[y(y-1)\right] + \mathbb{E}(y) - \mathbb{E}(y)^2.$$

From the preceding part we know that $\mathbb{E}\left[y(y-1)\right] = \theta$ and from the lecture slides we know that $\mathbb{E}(y) = \theta$. Therefore, $\text{Var}(y) = \theta^2 + \theta - \theta^2 = \theta^2$.

Question # 2 will not be marked; you do not have to submit a solution.

2. Suppose that we observe count data $y_1, \ldots, y_N \sim$ iid p_o and our model $f(y_i|\theta)$ is a Poisson(θ) probability mass function. Show that $\widehat{K} = s_y^2/(\bar{y})^2$ where we define $s_y^2 = \frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2$ and $\bar{y} = \frac{1}{N}\sum_{i=1}^N y_i$.

Solution: Because θ is a scalar, by definition

$$\widehat{K} \equiv \frac{1}{N}\sum_{i=1}^N \left[\frac{d}{d\theta}\log f(y_i|\widehat{\theta})\right]^2$$

Here $\log f(y_i|\theta) = y_i \log(\theta) - \theta - \log(y_i!)$ and, as derived in the lecture slides, $\widehat{\theta} = \bar{y}$. Differentiating with respect to θ and substituting into the expression for $\widehat{K}$ given above, we have

$$\begin{aligned}
\widehat{K} &= \frac{1}{N}\sum_{i=1}^N \left[y_i/\bar{y} - 1\right]^2 = \frac{1}{N}\sum_{i=1}^N \left[y_i^2/(\bar{y})^2 - 2y_i/\bar{y} + 1\right] \\
&= \frac{1}{(\bar{y})^2}\left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - \frac{2}{\bar{y}}\left[\frac{1}{N}\sum_{i=1}^N y_i\right] + \left[\frac{1}{N}\sum_{i=1}^N 1\right] \\
&= \frac{1}{(\bar{y})^2}\left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - \frac{2}{\bar{y}}\cdot\bar{y} + 1 = \frac{1}{(\bar{y})^2}\left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - 1 \\
&= \frac{1}{(\bar{y})^2}\left\{\left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - (\bar{y})^2\right\}.
\end{aligned}$$

It remains to show that the term in the curly braces equals s_y^2. Expanding,

$$\begin{aligned}
s_y^2 &\equiv \frac{1}{N}\sum_{i=1}^N (y_i - \bar{y})^2 = \frac{1}{N}\sum_{i=1}^N \left(y_i^2 - 2y_i\bar{y} + \bar{y}^2\right) \\
&= \left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - 2\bar{y}\left[\frac{1}{N}\sum_{i=1}^N y_i\right] + \bar{y}^2\left[\frac{1}{N}\sum_{i=1}^N 1\right] \\
&= \left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - 2(\bar{y})^2 + (\bar{y})^2 \\
&= \left[\frac{1}{N}\sum_{i=1}^N y_i^2\right] - (\bar{y})^2.
\end{aligned}$$

3. Let $\widehat{\boldsymbol{\beta}}$ be the conditional maximum likelihood estimator of $\boldsymbol{\beta}_o$ in a Poisson regression model with conditional mean function $\mathbb{E}(y_i|\mathbf{x}_i) = \exp(\mathbf{x}_i'\boldsymbol{\beta}_o)$, based on a sample of iid observations $(y_1, \mathbf{x}_1), \ldots, (y_N, \mathbf{x}_N)$.

   (a) Derive the first-order conditions for $\widehat{\boldsymbol{\beta}}$.

Solution: The log-likelihood of the i^{th} observation is given by

$$\begin{aligned}\ell_i(\boldsymbol{\beta}) \equiv \log f(y_i|\mathbf{x}_i, \boldsymbol{\beta}) &= y_i \log\left[\exp\left\{\mathbf{x}_i'\boldsymbol{\beta}\right\}\right] - \exp(\mathbf{x}_i\boldsymbol{\beta}) - \log\left(y_i!\right)\\ &= y_i\mathbf{x}_i'\boldsymbol{\beta} - \exp(\mathbf{x}_i'\boldsymbol{\beta}) - \log\left(y_i!\right)\end{aligned}$$

and hence the score vector is

$$\mathbf{s}_i(\boldsymbol{\beta}) \equiv \frac{\partial \ell_i(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} = y_i\mathbf{x}_i - \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)\mathbf{x}_i = \mathbf{x}_i\left[y_i - \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)\right].$$

Therefore, $\widehat{\boldsymbol{\beta}}$ solves the first order condition

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i\left[y_i - \exp\left(\mathbf{x}_i'\boldsymbol{\beta}\right)\right].$$

In other words,

$$\frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_i\left[y_i - \exp\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right)\right] = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i\widehat{u}_i = \mathbf{0}.$$

Notice that we are free to include or exclude the $1/N$ factor since multiplying both sides by N gives

$$\sum_{i=1}^{N} \mathbf{x}_i\left[y_i - \exp\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right)\right] = \sum_{i=1}^{N}\mathbf{x}_i\widehat{u}_i = \mathbf{0}.$$

   (b) Using your answer to the previous part show that, so long as $\mathbf{x}_i$ includes a constant, the residuals $\widehat{u}_i \equiv y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})$ sum to zero, as in OLS regression.

Solution: The first order conditions derived in the preceding part are a *collection* of equations: one for each regressor x_j. If $\mathbf{x}$ contains a constant, then one of the x_j is simply equal to one. Substituting, the first-order condition for this regressor is

$$\frac{1}{N}\sum_{i=1}^{N} 1 \cdot \left[y_i - \exp\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right)\right] = \frac{1}{N}\sum_{i=1}^{N}\widehat{u}_i = 0.$$

Multiplying through by N gives $\sum_{i=1}^{N}\widehat{u}_i = 0$.

(c) Using your answer to the preceding part, show that $\left[\frac{1}{N}\sum_{i=1}^N \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})\right] = \bar{y}$, where $\bar{y}$ is the sample mean of y, so that $\bar{y}\widehat{\beta}_j$ equals the estimated average partial effect of x_j in this model.

Solution: Since $\widehat{u}_i \equiv y_i - \exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}})$, we have $\exp(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}) = y_i - \widehat{u}_i$. Hence,

$$\frac{1}{N}\sum_{i=1}^N \exp\left(\mathbf{x}_i'\widehat{\boldsymbol{\beta}}\right) = \frac{1}{N}\sum_{i=1}^N (y_i - \widehat{u}_i) = \frac{1}{N}\sum_{i=1}^N y_i - \frac{1}{N}\sum_{i=1}^N \widehat{u}_i = \bar{y} - 0 = \bar{y}.$$

(d) Explain why multiplying the estimated coefficients from this model by $\bar{y}$ makes them roughly comparable to the corresponding OLS estimates from the model $y_i = \mathbf{x}_i'\boldsymbol{\theta} + \varepsilon_i$.

Solution: The result of the preceding part implies that the estimated average partial effect of x_j in a Poisson regression model equals $\bar{y}\widehat{\beta}_j$. In a linear regression model, the partial effects do not vary with $\mathbf{x}$. Hence the estimated average partial effect of x_j is simply $\widehat{\theta}_j$. In other words: the estimated *coefficients* in a linear regression are APEs, while the estimated coefficients in a Poisson regression must be rescaled by $\bar{y}$ to convert them to APEs. After carrying out this conversion we are comparing apples-to-apples, albeit from different models. Accordingly we should expect $\widehat{\theta}_j$ and $\bar{y}\widehat{\beta}_j$ to be more comparable in magnitude that $\widehat{\theta}_j$ and $\widehat{\beta}_j$.

4. 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.