ECOM40006/90013 Econometrics 3

Week 12 Extras

Question 1: The Poisson Regression Model

Binary variables are not the only type of discrete dependent variable that we would like to measure. A commonly seen source of data comes from observations on *count data*, which can take values $y_i = 0, 1, 2, \ldots$ and so on. For these kinds of models, we have the *Poisson regression model.* The model is, unsurprisingly, built on the *Poisson* distribution[1]

$$\mathbf{P}(y_i = k) = \frac{e^{-\lambda}\lambda^k}{k!}, \qquad \lambda > 0.$$

In this question we'll explore some of the various miscellaneous derivations associated with the Poisson distribution and the Poisson regression model.

(a) The first property of the Poisson distribution that we're going to investigate is the fact that $\mathbb{E}(y_i) = \lambda$. To get you started, here is the first line:

$$\mathbb{E}(y_i) = \sum_{k=0}^{\infty} k \times \mathbf{P}(y_i = k) = \sum_{k=0}^{\infty} k \times \frac{e^{-\lambda}\lambda^k}{k!}.$$

(b) Show that the variance of y_i is also λ, i.e. $\mathbb{E}(y_i) = \mathrm{Var}(y_i) = \lambda$. Below are some hints:

- $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, right?
- However, calculating $\mathbb{E}(X^2)$ is a gigantic pain. Perhaps you might want to try calculating $\mathbb{E}[X(X-1)]$ first. You might think that it seems really weird to calculate this, but note that doing so means (i) you don't have to calculate $\mathbb{E}(X^2)$ directly and (ii) you might potentially be able to reuse your steps from (a) above...

(c) Now let's look at the Poisson regression model. The model relies on the following *link function*

$$\mathbb{E}(y_i|x_i) = \lambda_i = \exp(x_i'\beta).$$

(i.) Explain how using an exponential function for λ_i may avoid conditional mean interpretation issues relative to using a linear model like $\lambda_i = x_i'\beta$.

(ii.) Derive an expression for the log-likelihood.

(iii.) Derive an expression for the first-order conditions associated with the log-likelihood for the Poisson regression model. *Note that the intuition from the binary response models derivations will come in handy.*

[1] Did you know that *Poisson* means 'fish' in French? Sounds fishy, if you ask me.

(iv.) Obtain the Hessian associated with the Poisson regression model. *Again, the binary response models derivations may help. The derivations here aren't as bad, fortunately.*

(d) Consider the marginal probability effect of a change in a regressor x_{ik}:

$$\frac{\partial \mathbf{P}(y_i = j|x_i)}{\partial x_{ik}} = \frac{\partial f(y_i|x_i)}{\partial x_{ik}} = f(y_i|x_i)\tau(y_i).$$

Verify this result, and show specifically that $\tau(y_i) = (y_i - \lambda_i)\beta_k$.

One of the primary uses of econometrics lies in interpreting regression models. The Poisson regression model is no different! Consider the following example based on Cameron and Trivedi's *Microeconometrics: Methods and Applications.* Let y_i represent the yearly number of visits to a doctor for individual i, and x_i the number of years of education for that individual. Then consider the model

$$y_i = \beta_0 + \beta_1 x_1 + \text{controls} + u_i,$$

where controls represent other regressors that may also have an effect on the number of yearly visits (but we won't consider them here since the only goal is to just get intepretation sorted).

(e) Based on the partial of λ_i with respect to a regressor x_{ik}, obtain an interpretation for an arbitrary coefficient estimate β_k in a Poisson regression.

(f) Suppose a Poisson regression is conducted on the model above and it is found that $\hat{\beta}_1 = 0.03$. How would you interpret this coefficient?

(g) Further suppose in the Poisson regression model that an *average marginal effect* is calculated:

$$\underbrace{\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \lambda_i}{\partial x_{ik}}}_{AME_k} = \left[\frac{1}{n}\sum_{i=1}^{n} \exp(x_i'\beta)\right]\beta_k.$$

In the same regression, it is found that the estimated value of $AME_1 = 0.202$. How would you interpret this?

Question 2: Binary Response (Interpretations and Approximation)

This exercise makes use of the supplementary R code `Week12ExtraCode.R`*, which is provided alongside this exercise. Note that the code will save five plots in PDF format so you might like to set a working directory first before running the code.*

One of the discussions made in lectures concerns the idea of comparing the coefficients across the logit, probit and linear probability models. In this exercise, we'll use our trusty Monte Carlo simulation methods to take a look at how these kinds of approximations compare under various conditions.

For starters, the model considered will be a classic linear regression on a *binary* variable y_i:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where u_i is an i.i.d. standard normal disturbance term. Note that the choice of distribution for u_i will also have an impact on whether certain model estimates can be considered to be consistent. For this simulation, we will be considering the following values

$$\beta_0 = -0.1, \qquad \beta_1 = 0.3,$$

for 1000 repeated samples of $n = 25$, $n = 50$ and $n = 100$ observations each.

(a) Obtain density estimates for the estimated coefficients $\hat{\beta}_1$ from the linear probability model. Compared to the true value of β_1, how well does the linear probability model perform?

(b) Repeat part (a), but for the probit regression model instead.

(c) Obtain the density estimates of $\hat{\beta}_1$ using a logit link function. Based on the assumptions made in this model, do you think that a logit model is appropriate?

(d) A suggested method in which one can make the coefficients comparable between logit and probit models is to divide the logit estimates by $\pi/\sqrt{3}$. Do this for the logit estimates in (c) above, and calculate the difference between these and probit estimates in (b). How well does the approximation perform?

(e) Another suggested approximation concerns the linear probability model: dividing the linear probability model estimates by 4 allows one to obtain an approximation for the same coefficient if a logit model were to be used instead. Do this and calculate the difference between these and the actual logit estimates in (c). How well does the approximation perform?