Limited Dependent Variables & Selection: PS #1

Francis DiTraglia

HT 2020

1. Let $y \sim$ Poisson(θ).
   (a) Using steps similar to the derivation of $\mathbb{E}[y]$ from the lecture slides, show that $\mathbb{E}[y(y-1)] = \theta^2$.
   (b) Use your answer to the preceding part, along with the result $\mathbb{E}[y] = \theta$, to show that $\text{Var}(y) = \theta$.

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 $\bar{y} = \frac{1}{N}\sum_{i=1}^N y_i$.

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 $\widehat{\boldsymbol{\beta}}$.
   (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.
   (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.
   (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 $y_i = \mathbf{x}_i'\boldsymbol{\theta} + \varepsilon_i$.

4. *This question is adapted from Wooldridge (2010).* To answer it you will need to use the dataset `SMOKE.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/smoke.des`
   (a) Use a linear regression to predict *cigs*, the number of cigarettes smoked each day, using the regressors log(*cigpric*), log(*income*), *restaurn*, *white*, *educ*, *age*, and *age*^{2}. Interpret your findings. In particular: are cigarette prices and income statistically significant predictors? Does this depend on whether you use robust standard errors?

(b) Repeat the preceding part but estimate a *Poisson* regression with an exponential conditional mean function rather than a linear regression. Calculate the APEs for the Poisson model and compare them to the OLS estimates.

(c) If you calculated standard errors using the Poisson variance assumption, are cigarette prices and income statistically significant? Compare to your OLS results from above.

(d) Calculate $\widehat{\sigma}^2$. Does your estimate suggest evidence of overdispersion? If you use the Quasi-Poisson Variance assumption, how do your results compare to those of the preceding part?

(e) How do your answers to the preceding two parts change if you instead use the fully-robust "sandwich" standard errors?