ECON 6511: Advanced Applied Econometrics
Homework 7
Due in class March 7, 2018

1. (Wooldridge, Chapter 17, Problem 1) Use the data in PNTSPRD.dta for this exercise.

   (a) The variable $favwin$ is a binary variable if the team favored by the Las Vegas point spread wins. A linear probability model to estimate the probability that the favored team wins is:

   $$P(favwin = 1|spread) = \beta_0 + \beta_1 spread$$

   Explain why, if the spread incorporates all relevant information, we expect $\beta_0 = 0.5$.

   (b) Estimate the model from part (a) by OLS. Test $H_0 : \beta_0 = 0.5$ against a two-sided alternative. Use both the usual and heteroskedasticity-robust (command "robust") standard errors.

   (c) Is *spread* statistically significant? What is the estimated probability that the favored team wins when $spread = 10$?

   (d) Now, estimate a probit model for $P(favwin = 1|spread)$. Interpret and test the null hypothesis that the intercept is zero. [*Hint*: Remeber $\Phi(0) = 0.5$.]

   (e) Use the probit model to estimate the probability that the favored team wins when $spread = 10$. Compare this with the LPM estimate from part (c). In Stata, to obtain $G(x\hat{\beta})$, where $G()$ is the normal CDF, you can use the command: "display normprob([values])" where we replace "[values]" with $\hat{\beta}_0 + \hat{\beta}_1 x_1 + \ldots \hat{\beta}_k x_k$. In this example, we want to use our estimates of β_0 and β_1 and $spread = 10$. For logit, we would use the command "display invlogit([values])".

   (f) Add the variables $favhome$, $fav25$, and $und25$ to the probit model and test joint significance of these variables using the likelihood ratio test at the 5% level. (How many *df* (or restrictions) are in the chi-square distribution?) Interpret this result, focusing on the question of whether the spread incorporates all observable information prior to a game.

2. (Wooldridge, Chapter 17, Problem 2) Use the data in LOANAPP.dta for this exercise.

   (a) Estimate a probit model of *approve* on *white*. Find the estimated probability of loan approval for both whites and nonwhites (using approach from 1(e) above). How do these compare with the linear probability estimates?

   (b) Now, add the variables *hrat*, *obrat*, *loanprc*, *unem*, *male*, *married*, *dep*, *sch*, *cosign*, *chist*, *pubrec*, *mortlat*1, *mortlat*2, and *vr* to the probit model. Is there statistically significant evidence of discrimination against nonwhites?

(c) Estimate the model from part (b) by logit. Compare the coefficient on *white* to the probit estimate.

(d) Compute the average partial effect (APE) to estimate the sizes of the discrimination effects for probit and logit. Interpret these results.

3. (Wooldridge, Chapter 17, Problem 5) Using the data in FERTIL1.dta.

(a) Estimate a Poisson regression model for *kids*, using the same variables in Slide 4 from Lecture 3. Interpret the coefficient on $y82$.

(b) What is the estimated percentage difference in fertility between a black woman and a nonblack woman, holding other factors fixed?

(c) Obtain $\hat{\sigma}^2$ using formula $\frac{\sum_{i=1}^{n} \frac{\hat{u}^2}{\hat{y}_i}}{n-k-1}$. Is there evidence of over- or underdispersion? To compute this, use the approach from class discussed in class: First obtain $\hat{y}$ and use this to obtain $\hat{u}^2$. Then create the ratio $\frac{\hat{u}^2}{\hat{y}}$. Use the command "collapse (sum) ratio" to sum the ratio across all observations (this will leave only one observation in your dataset equal to the sum). Finally, divide the sum by $n - k - 1$.