ECON 6511: Advanced Applied Econometrics
Homework 7 Solutions

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 $\beta_0 = 0.5$.
   Answer: If *spread* is zero, there is no favorite, and the probability that the team we (arbitrarily) label the favorite should have a 50% chance of winning.

   (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.
   Answer: The linear probability model estimated by OLS is:

   $$favwin = 0.577 + 0.0194 spread$$

   Using the usual standard error, the t statistic for $H_0 : \beta_0 = .5$ is $\frac{.577-.5}{.028} = 2.75$, which leads to rejecting H_0 against a two-sided alternative at the 1% level (critical value $\approx$ 2.58). Using the robust standard error reduces the significance but nevertheless leads to strong rejection of H_0 at the 2% level against a two-sided alternative: $t = \frac{.577-.5}{.032} \approx 2.41$ (critical value ≈ 2.33).

   (c) Is *spread* statistically significant? What is the estimated probability that the favored team wins when $spread = 10$?
   Answer: As we expect, *spread* is very statistically significant using either standard error, with a t statistic greater than eight. If spread $= 10$ the estimated probability that the favored team wins is $.577 + .0194(10) = .771$.

   (d) Now, estimate a probit model for $P(favwin = 1|spread)$. Interpret and test the null hypothesis that the intercept is zero. [*Hint*: Remeber $\Phi(0) = 0.5$.]
   Answer: $P(favwin = 1|spread) = \Phi(-0.0106 + 0.0925 spread)$. If $\beta_0 = 0$ then $P(favwin = 1|spread) = \Phi(\beta_1 spread)$ and $P(favwin = 1|spread) = \Phi(\beta_1 \cdot 0) = 0.5$ This is the analog of testing whether the intercept is 0.5 in the LPM. From the table, the t statistic for testing $H_0 : \beta_0 = 0$ is only about -0.102, so we do not reject H_0.

   (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])".

Answer: When $spread = 10$ the predicted response probability from the estimated probit model is $\Phi(-.0106 + .0925(10)) = \Phi(.9144) \approx .820$. This is somewhat above the estimate for the LPM.

(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.

Answer: When $favhome$, $fav25$, and $und25$ are added to the probit model, the value of the log-likelihood becomes -262.64. Therefore, the likelihood ratio statistic is $2[-262.64 - (-263.56)] = 2(263.56 - 262.64) = 1.84$. The p-value from the χ^2_3 distribution is about .61, so $favhome$, $fav25$, and $und25$ are jointly very insignificant. Once spread is controlled for, these other factors have no additional power for predicting the outcome.

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?

Answer: The probit estimates from approve on white are:

$$P(approve = 1|white) = \Phi(0.547 + 0.784 white)$$

As there is only one explanatory variable that takes on just two values, there are only two different predicted values: the estimated probabilities of loan approval for white and nonwhite applicants. Rounded to three decimal places these are 0.708 for nonwhites and 0.908 for whites. Without rounding errors, these are identical to the fitted values from the linear probability model. This must always be the case when the independent variables in a binary response model are mutually exclusive and exhaustive binary variables. Then, the predicted probabilities, whether we use the LPM, probit, or logit models, are simply the cell frequencies. (In other words, 0.708 is the proportion of loans approved for nonwhites and 0.908 is the proportion approved for whites.)

(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?

Answer: With the set of controls added, the probit estimate on *white* becomes about .520 (se $\approx$ .097). Therefore, there is still very strong evidence of discrimination against nonwhites.

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

Answer: When we use logit instead of probit, the coefficient (standard error) on *white* becomes .938 (.173).

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

Answer: The average marginal effect (standard error) is 0.0863868 (.015954) for probit and (0.0828195) (0.0150873) for logit.

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 $y82$.

Answer: The Poisson regression model is:

$$\begin{aligned}\lambda(kids|X) = &-3.060 - 0.048educ + 0.204age - 0.0022age^2 + 0.350black + 0.088east \\ &+ 0.142northcen + 0.080west - 0.015farm - 0.057othrural + 0.031town \\ &+ 0.074smcity + 0.093y74 - 0.029y76 - 0.016y78 - 0.020y80 - 0.193y82 \\ &- 0.214y84\end{aligned}$$

The coefficient on y82 means that, other factors in the model fixed, a woman?s fertility was about 19.3% lower in 1982 than in 1972.

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

Answer: Because the coefficient on black is so large, we obtain the estimated proportionate difference as exp(.36) - 1 $\approx$ .433, so a black woman has 43.3% more children than a comparable nonblack woman. (Notice also that black is very statistically significant.)

(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 $n-k-1$.

Answer: From the above table, $\hat{\sigma}^2 \approx 0.89$, which shows that there is actually underdispersion in the estimated model.