Assoc. Prof. A. Sevtap Kestel, Ph.D.
Department of Empirical Research and Econometrics

Assignment 3

Problem 1 (Wooldridge, Problem C7.2, Page 259)

Use the data in WAGE2.RAW for this exercise.

(i) Estimate the model

$$log(wage) = \beta_0 + \beta_1 educ + \beta_2 exper + \beta_3 tenure + \beta_4 married + \beta_5 black + \beta_6 south + \beta_7 urban + u$$

and report the results in the usual form. Holding other factors fixed, what is the approximate difference in monthly salary between blacks and nonblacks? Is this difference statistically significant?

(ii) Add the variables $exper^2$ and $tenure^2$ to the equation and show that they are jointly insignificant at even the 20% level.

(iii) Extend the original model to allow the return to education to depend on race and test whether the return to education does depend on race.

(iv) Again, start with the original model, but now allow wages to differ across four groups of people: married and black, married and nonblack, single and black, and single and nonblack. What is the estimated wage differential between married blacks and married nonblacks?

Problem 2 (Wooldridge, Problem C7.13, Page 263)

Use the data in APPLE.RAW to answer this question.

(i) Define a binary variable as $ecobuy = 1$ if $ecolbs > 0$ and $ecobuy = 0$ if $ecolbs = 0$. In other words, *ecobuy* indicates whether, at the prices given, a family would buy any ecologically friendly apples. What fraction of families claim they would buy ecolabeled apples?

(ii) Estimate the linear probability model

$$ecobuy = \beta_0 + \beta_1 ecoprc + \beta_2 regprc + \beta_3 faminc + \beta_4 hhsize + \beta_5 educ + \beta_6 age + u$$

and report the results in the usual form. Carefully interpret the coefficients on the price variables.

(iii) Are the non-price variables jointly significant in the LPM? (Use the usual F statistic, even though it is not valid when there is heteroskedasticity.) Which explanatory variable other than the price variables seems to have the most important effect on the decision to buy ecolabeled apples? Does this make sense to you?

(iv) In the model from part (ii), replace $faminc$ with $\log(faminc)$. Which model fits the data better, using $faminc$ or $\log(faminc)$? Interpret the coefficient on $\log(faminc)$.

(v) In the estimation in part (iv), how many estimated probabilities are negative? How many are bigger than one? Should you be concerned?

(vi) For the estimation in part (iv), compute the percent correctly predicted for each outcome, *ecobuy* $= 0$ and *ecobuy* $= 1$. Which outcome is best predicted by the model?

Problem 3 (Wooldridge, Problem C8.4, Page 296)

Use VOTE1.RAW for this exercise.

(i) Estimate a model with $voteA$ as the dependent variable and $prtystrA$, $democA$, $log(expendA)$, and $log(expendB)$ as independent variables. Obtain the OLS residuals, $\hat{u}_i$, and regress these on all of the independent variables. Explain why you obtain $R^2 = 0$.

(ii) Now, compute the Breusch-Pagan test for heteroskedasticity. Use the F statistic version and report the p-value.

(iii) Compute the special case of the White test for heteroskedasticity, again using the F statistic form. How strong is the evidence for heteroskedasticity now?

Problem 4 (Wooldridge, Problem 17.2, Page 614)

Let *grad* be a dummy variable for whether a student-athlete at a large university graduates in five years. Let $hsGPA$ and SAT be high school grade point average and SAT score. Let *study* be the number of hours spent per week in organized study hall. Suppose that, using data on 420 student-athletes, the following logit model is obtained:

$$\hat{P}(grad = 1|hsGPA, SAT, study) = \Lambda(-1.17 + .24hsGPA + .00058SAT + .073study),$$

where $\Lambda(z) = exp(z)/[1 + exp(z)]$ is the logit function. Holding $hsGPA$ fixed at 3.0 and SAT fixed at 1200, compute the estimated difference in the graduation probability for someone who spent 10 hours per week in study hall and someone who spent 5 hours per week.

Problem 5 (Wooldridge, Problem C17.1, Page 615)

Use the data in PNTSPRD.RAW for this exercise.

(i) 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 $\beta_0 = .5$.

(ii) Estimate the model from part (i) by OLS. Test $H_0 : \beta_0 = .5$ against a two-sided alternative. Use both the usual and heteroskedasticity-robust standard errors.

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

(iv) Now, estimate a probit model for $P(favwin = 1|spread)$. Interpret and test the null hypothesis that the intercept is zero. [Hint: Remember that $\Phi(0) = .5$.]

(v) Use the probit model to estimate the probability that the favored team wins when $spread = 10$. Compare this with the LPM estimate from part (iii).

(vi) Add the variables $favhome$, $fav25$, and $und25$ to the probit model and test joint significance of these variables using the likelihood ratio test. (How many *df* 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.

Problem 6 (Wooldridge, Problem C17.5, Page 617)

Refer to Table 13.1 in Chapter 13. There, we used the data in FERTIL1.RAW to estimate a linear model for $kids$, the number of children ever born to a woman.

(i) Estimate a Poisson regression model for $kids$, using the same variables in Table 13.1. Interpret the coefficient on $y82$.

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

(iii) Obtain $\hat{\sigma}$. Is there evidence of over- or underdispersion?

(iv) Compute the fitted values from the Poisson regression and the R-squared as the squared correlation between $kids_i$ and $\widehat{kids}_i$. Compare this with the R-squared for the linear regression model.