ECON 4848 - Applied Econometrics
Professor Carlos Martins-Filho
Final Examination
Date: 12.16.2009

IMPORTANT INSTRUCTIONS:

1. Your answers should be on just one side of a sheet of paper. Use a new sheet of paper to start the answer of each question.

2. Question 5 involves the use of the Stata data set file located on the course webpage:

http://spot.colorado.edu/~martinsc/ECON_4848.html

You will have to write and run a Stata `do` file to answer some of the items in Question 5. You DO NOT need to submit the output of your `do` file as part of your answers.

QUESTIONS:

Question 1. The following table (from Stock and Watson, 2003) contains estimates of the demand for economic journals. The regressand is the logarithm of the number of subscriptions ($log(Y_i)$) in United States Libraries for the year 2000. There are $n = 180$ observations. Four models (1)-(4) were estimated:

Regressor	(1)	(2)	(3)	(4)
$log(pricepercitation)$	-0.533**(0.034)	-0.408**(0.044)	-0.961**(0.160)	-0.899**(0.145)
$log(pricepercitation)^2$			0.017(0.025)	
$log(pricepercitation)^3$			0.0037(0.0055)	
$log(Age)$		0.0424**(0.119)	0.373**(0.118)	0.374**(0.118)
$log(Age) \times log(pricepercitation)$			0.056**(0.052)	0.141**(0.040)
$log(characters/1,000,000$		0.206*(0.098)	0.235(0.098)	0.229(0.096)
Intercept	4.77**(0.055)	3.21**(0.38)	3.41**(0.38)	3.43**(0.38)
F Statistic (p value)			0.25 (0.779)	
s^2	.75	.70	.69	.68
Adjusted R^2	0.555	0.607	0.622	0.626

The row with F-value contains the value of the F-statistic for the hypothesis that the parameters associated with $log(pricepercitation)^2$ and $log(pricepercitation)^3$ are both zero. Standard errors are given in parenthesis to the right of the parameter estimates and the p-value for the F-test is given to the right of the F-statistic. Parameter estimates are statistically different from zero at the *5% level or **1% level. Age is the age of the journal (years since first publication), $pricepercitation$ is the subscription price divided by the number of citations received, $characters/1,000,000$ is a way to standardize the number of characters published per year.

a) Using these results find justification for the following conclusions:

- Demand is less elastic for older than newer journals
- The evidence supports a linear function of log of price per citation.
- Demand is greater for journals with more characters, holding price and age constant.

b) Is there an assumption of normality of $logY_i$ given the regressors? How can you tell?

c) Can you give an interpretation for the intercept in model (4)?

Question 2: The variable *smokes* is a binary variable equal to 1 if a person smokes and 0 otherwise. The following model has been estimated by OLS.

$$\widehat{smokes} = 0.656 - 0.069 log(price\, of\, cigarettes) + 0.012 log(income) - 0.029 education$$

$$+0.020 Age - 0.00026 Age^2 - 0.101 rest - 0.026 white$$

white is a dummy variables which is 1 if the individual is white and zero otherwise. *rest* is a dummy variable which is 1 if the individual lives in a state where there are restrictions for smoking in restaurants. The meaning of the other regressors should be evident. Estimated standard deviations for each parameter estimate are, in the order they appear in the regression: 0.855, 0.204, 0.026, 0.006, 0.006, 0.00006, 0.039, 0.052. The heterocedastic robust versions are, in order, 0.856, 0.207, 0.026, 0.006, 0.005, 0.00006, 0.038, 0.050.

a) Holding other regressors fixed, if education increases by by four years, what happens to the estimated probability of smoking?

b) At what point does another year of age reduce the probability of smoking?

c) Interpret the coefficient on the variable *rest*.

d) Individual number 206 in the sample has the following chraacteristics: $price\, of\, cigarettes = 67.44$, $income = 6500$, $education = 16$, $Age = 77$, $rest = 0$, $white = 0$ and $smokes = 0$. Compute the predicted probability of smoking for this person and comment on the result.

Question 3: An economist wants to test whether or not personal computer ownership, represented by a binary variable, PC, impacts college GPA, $colGPA$ of students. With data on SAT (Scholastic Aptitude Test – a test taken prior to admission to colleges and universities) scores, and high school GPA, hsGPA, she assumes the model

$$colGPA_i = \beta_0 + \beta_1 hsGPA_i + \beta_2 SAT_i + \beta_3 PC_i + u_i \;\; i = 1, ..., n$$

a) She proposes the following test: $H_0 : \beta_3 = 0$ and $H_A : \beta_3 \neq 0$ based on the following test statistic $t = \frac{\hat{\beta}_3}{\hat{\sigma}_{\hat{\beta}_3}}$, where $\hat{\beta}_3$ is an OLS estimator, and $\hat{\sigma}_{\hat{\beta}_3}$ is its estimated standard deviation. She postulates that under the null hypothesis t has a Student's t distribution with $n - 4$ degrees of freedom. Give a complete list of assumptions under which her test statistic has the postulated distribution.

b) Do you believe that it is reasonable to assume that PC_i is correlated with u_i in this model? Why or why not?

c) Irrespective of your answer in b), what are the consequences of PC_i being correlated with u_i for the test proposed in a)? For the OLS estimator of β_3?

Question 4: Consider the following regression model to explain monthly beer consumption:

$$beer = \beta_0 + \beta_1 income + \beta_2 price + \beta_3 education + \beta_4 female + u$$

with $E(u|income, price, education, female) = 0$ and $E(u^2|income, price, education, female) = \sigma^2 income^2$, and female a dummy variable that takes on the value 1 if the individual is a female.

a) Describe, in detail, how you would obtain the Feasible Generalized Leas Squares estimator for $\beta_0, \cdots, \beta_4$.

b) In this case, is β_{FGLS} identical to β_{GLS}? Explain.

c) If you had the choice of using the FGLS estimator you described in a) or the OLS estimator with robustly estimated standard errors to test the hypothesis that $\beta_2 = 0$ against the alternative that it is different from zero, which would you use? Explain.

Question 5: Use the data set *vote1.dta* for this question.

a) Estimate the following regression model by OLS.

$$voteA = \beta_0 + \beta_1 prtystrA + \beta_2 democA + \beta_3 log(expendA) + \beta_4 log(expendB) + u$$

a) Obtain OLS residuals $\hat{u}$ and regress these residuals on all regressors of the model. What is the R^2 of this regression? Why does it have this value?

b) Using White's test for homocedasticity indicate whether or not there is support for an assumption of constant variance in this model.

c) Using heterocedastic robust standard errors test the hypothesis that β_1 and β_2 are equal to zero, against the alternative that either one is not at the 5% level of significance.