İKT 452-554

Homework 2

Question (1) Using all observations of "mus03data" (Cameron and Trivedi's book MUS), obtain the OLS estimates of the following model:

$ltotexp = \beta_0 + \beta_1 suppins + \beta_2 phylim + \beta_3 actlim + \beta_4 totchr + \beta_5 age + \beta_6 female + \beta_7 income + u$

a) Compute standard errors in three ways: default, heteroskedastic, and cluster-robust, where clustering is on the number of chronic problems. Use the "estimates" command to produce a table with three sets of coefficient estimates and standard errors.

b) Do you observe any differences between the three sets of estimates? Comment on the reasons for the differences.

c) Compare your results to those obtained using the first 500 observations of the dataset (as we did in Quiz 2).

d) In the model with robust standard errors, test at 5% the joint significance of the demographic variables age, income, and female.

e) In the model with robust standard errors, test the hypothesis that being male has the same impact on medical expenditures as being ten years older.

f) Re-estimate the model with robust standard errors, this time under the constraint that the coefficients of "phylim" and "actlim" are the same.

g) Use the "mfx" command to compute the marginal effects of the explanatory variables.

h) Show by example that the square of the correlation coefficient is the same as the R-squared from the regression.

i) Create graphs that show the probability density functions of $totexp$ and $ltotexp$. Comment on the graphs. Do the distributions look symmetric or skewed?

Question (2)

Let $X \sim N(0,1)$. Generate ten thousand random values for X. Draw a plot of the probability density function of X you just created.

(Hints : Modify and use the sample program below, copied from Ch.4 of Cameron and Trivedi MUS. Pay attention not to misspell the quotations in `x'. "kdensity" command generates a kernel estimate of the pdf of X. Example: kdensity x.)

```
. * Simulation using postfile
. set seed 10101
. postfile sim_mem xmean using simresults, replace
. forvalues i = 1/10000 {
  2.      drop _all
  3.      quietly set obs 30
  4.      tempvar x
  5.      generate `x' = runiform()
  6.      quietly  summarize `x'
  7.      post sim_mem (r(mean))
  8. }
. postclose sim_mem
```

Question (3)

Inspect the variables in "loanapp.dta" (one of the datasets used by Wooldridge). Using this dataset, estimate a model that explains "approve", a binary dependent variable that shows whether a mortgage loan to an individual is approved or not.

a) Consider the regression $approve = \beta_0 + \beta_1 white + other\ factors$. If there is discrimination against non-whites, what do you expect the sign of β_1 to be?

b) Regress *approve* on *white* and report the results in the usual form. Interpret the coefficient on *white*. Is it statistically significant? Is it practically large?

c) As controls, add the variables *hrat*, *obrat*, *loanprc*, *unem*, *male*, *married*, *dep*, *sch*, *cosign*, *chist*, *pubrec*, *mortlat1*, *mortlat2*, and *vr*. What happens to the coefficient on *white*? Is there still evidence of discrimination against nonwhites?

d) Now allow the effect of race to interact with the variable measuring other obligations as a percent of income (*obrat*). Is the interaction term significant?

e) Using the model from part (d), what is the effect of being white on the probability of approval when *obrat* = 32, which is roughly the mean value in the sample? Obtain a 95% confidence interval for this effect.