ECNS561, Fall 2011

Take-home Graded Practice Opportunity 6

Due date: December 6, 2011, 5:00 p.m.

Conceptual practice with Gauss-Markov assumption violations

1. Explain how serial autocorrelation may occur due to an omitted variable. Provide an example serial autocorrelation due to variable omission.
2. In non-technical terms (i.e., in a manner that individuals without an econometrics education) explain the concept of stationarity. Why is it useful? Why do we need it?
3. In the following three scenarios an OLS regression is estimated. In each scenario, are any Gauss-Markov assumptions violated? If so, which assumptions are violated and why?
   (a) A model of stock prices as a function of changing market conditions.
   (b) A model of expenditures on housing as a function of income.
   (c) A model of median income on per-capita level of public education spending.
4. Recall that the GLS estimator can be written as: $\tilde{\beta}_{GLS} = (\boldsymbol{X'V^{-1}X})^{-1}\boldsymbol{X'V^{-1}y}$. Show that this estimator minimizes the weighted sum of squared errors:

$$(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})'\boldsymbol{V}^{-1}(\boldsymbol{y} - \boldsymbol{X}\tilde{\boldsymbol{\beta}})$$

Applied analysis

You are examining crop insurance policies sold in the United States in 2009. Specifically, you are focused on policies sold in U.S. counties for canola. A dataset describing this information as well as other related characteristics is located on the course website ("data_hw6.csv") or at this web address:
http://www.montana.edu/bekkerman/classes/ecns561f11/data_hw6.csv

The dataset contains the following variables:

Variable	Description
crop_yr	Year
fips	FIPS county identifier
county_name	County name
state	State name
crop_name	Crop name
ins_name	Crop insurance policy name
cov_lvl	Level of insurance coverage (percentage)
policies_sold	Number of policies sold
net_acres	Number of acres insured
liability	Total liability (maximum potential production loss if 100% loss occurred)
total_prem	Total premiums paid by farmers to be insured
indem_amt	Amount of indemnities paid
loss_ratio	Ratio of indemnities to premiums

You would like to empirically analyze factors affecting the number of crop insurance policies sold. Your goals are the following:

1. Organize and clean the data as necessary. Provide a brief overview of your steps.
2. Manipulate the data so that you have the sum of values for each variable across the two different insurance plans.
3. Describe the model that you would seek to estimate using the dataset. Justify each variable in your model with an explanation of how the variable is economically relevant and what it will explain.
4. Provide visual and statistical summary statistics of the relevant variables. What are the potential economic implications? (*Hint: compare the mean and the median. What do they imply?*)
5. Estimate the OLS model.
6. Empirically diagnose the data for potential problems related to outliers. Describe whether you need to deal with outliers and if so, how you dealt with them.
7. Without empirically testing for it, is heteroskedasticity a potential problem in this model? If so, what might be the source?
8. Empirically test for heteroskedasticity.
9. In class (and in the course notes), we discussed the procedure for estimating feasible GLS estimators. Use this procedure to estimate the heteroskedasticity-adjusted estimator.

SAS coding hints:

(a) In PROC REG, you can output residuals using the code line: `output out=resid_data r=e_hat;` (to be placed after the `model ...;` line).

(b) In PROC REG, you can output predicted values using the code line: `output out=pred_data p=y_hat;` (to be placed after the `model ...;` line).

(c) In PROC REG, you can assign weights as follows:

```
proc reg data=myData;
weight weightVariable;
model y = x1 x2 ...;
run;
```

10. Estimate the OLS estimator, but adjust the standard errors using White's heteroskedasticity-robust method. How do these results differ from the FGLS estimator?