Final

Duration: 180 Minutes
Name: ______________________________________
Student I.D.:________________________________

Questions
1:________________

2:________________

3:________________

4:________________

5:________________

6:________________

7:________________

8:________________

TOTAL:________________

Instructions: Show work and formulas used. Final answers with no work/formulas shown receive 0 credit.

1) Suppose

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \qquad cov(X) = \begin{bmatrix} 1 & -2 & 2 \\ -2 & 2 & 2 \\ 2 & 2 & 5 \end{bmatrix} \qquad mean(X) = \begin{bmatrix} 4 \\ 2 \\ 7 \end{bmatrix}$$

a) What is the expected value of $x_1 x_3$?
b) What is the variance of $(x_1 - x_3)$?
c) If instead x_1 and x_3 were independent of each other, what would the covariance matrix of X look like?
Be careful to write out the formulas and derivations you use. Marks will not be given for guesses.
[20pts]

2) State the practical consequences of multicollinearity and discuss situations in which multicollinearity is a problem.
[10pts]

3) List methods other than OLS to estimate β in $y = X\beta + \epsilon$ and briefly describe the conditions under which you would use these alternative estimation techniques.
[10pts]

3) State with reason whether the following statements are true, false or uncertain. Be precise.

a) If the residuals estimated from an OLS regression exhibit a systematic pattern, it means heteroskedasticity is present in the data.

b) In the presence of heteroskedasticity, OLS estimators are biased as well as inefficient.

c) A significant Durbin-Watson d does not necessarily mean there is autocorrelation.

d) The exclusion of an important variable(s) from a regression model may give a significant d value.

e) The use of dummy variables as explanatory variables invalidates the usual t and F tests. Special Maximum Likelihood methods must be used.

f) The problems of simultaneity and exogeneity are the same thing.

g) The problem of specification error is limited to measurement error, faulty omission of an irrelevant variable, faulty inclusion of a relevant variable, autocorrelation and heteroskedasticity.

h) Chow's Short test is identical to the Chow test for structural change except that it is easier to apply.

i) If we measure the explanatory variables with error, we face critical misspecification problems, but if we measure the dependent variable with error, we may not face any misspecification problems.

[40pts]

4) Tell me what the following SAS code is meant to do. Be sure to identify objects as you describe what the program does — i.e. identify what “final” is, what the “reg” procedure does, what “model” defines, what “output” does, etc.

```
proc reg data=final;
model sell= ask time lot rms bdrms rec; output out=pred p=yhat r=resid;
data new; set pred;
y2=yhat**2; y3=yhat**3;
el1=lag1(resid); el2=lag2(resid);
proc reg data=new;
model sell= ask time lot rms bdrms rec y2 y3;
test: test y2=0, y3=0;
proc reg data=new;
model resid= ask time lot rms bdrms rec el1 el2;
```

[20pts]

5) Consider the regression $y = X\beta + \epsilon$ where y is n by 1, and X is n by 3. Supose you believed this regression suffered from the problem of simultaneity.
a) Explain what the problem of simultaneity is.
b) If this problem causes bias of the OLS estimator, show this bias. If it doesn't cause bias, explain why.
c) What alternatives to OLS are available to "fix up" this problem?
[10pts]

6) Derive the matrix algebra form of the variance-covariance estimator of the OLS estimator $\hat{\beta}$. Be explicit about the assumptions you have to make, and tell me the number of columns and rows in the final formula you derive.
[10pts]

7) Suppose we were to model annual unemployment U as a function of W labour wage rate, K the cost of capital and C construction of new homes,

$$U = \beta_1 + \beta_2 W + \beta_3 K + \beta_4 C + \epsilon$$

but we do not observe the number of the unemployed, just whether or not there were more than 1,000,000 unemployed that year. What is this sort of estimation problem categorized as? What would the dependent variable be? Is OLS the best way to estimate the model? What would be the properties of OLS estimates of the parameters?
[10pts]

8) The Generalized Least Squares (GLS) estimator was designed to handle what problems? Describe one alternative to GLS that is also designed to handle the same problems, and discuss any advantages it may have over GLS.
[10pts]