Final Exam

Answer five of the following questions. You must answer question 7. The questions are weighted equally. You have 2.5 hours. You may use a calculator. Brevity is recommended.

1. Answer true or false and state why.
   a. When the disturbances are serially correlated and the regressors include a lagged endogenous variable, the OLS coefficient estimates will be biased and inconsistent.
   b. Under the null hypothesis, the Lagrange multiplier (LM), likelihood ratio (LR), and Wald (W) tests always agree.
   c. When collinearity is known to be present, then OLS-based inferences are meaningless.
   d. A proper instrument for instrumental variables estimation is one which is uncorrelated with the regressors but correlated with the disturbances.
   e. When the RHS variables move smoothly, the appropriate critical value for the Durbin-Watson test is closer to $d_U$ than $d_L$.
   f. The power of a test is determined by the variability of the explanatory variables.
   g. Suppose the disturbances are non-normal, then the usual t-statistics will still be asymptotically appropriate in large samples.
   h. When the model has right-hand side variables correlated with the disturbances, we can correct any difficulties through the use of generalized least squares.

2. Consider the following multivariate regression model (for $t = 1, 2, ..., n$):
   \[ y_t = x_t' \beta + u_t \]
   where $x_t$ is a $k \times 1$ vector of observations on the $k$ independent variables for period $t$, $y_t$ is the value of the dependent variable and $\beta$ is the $k \times 1$ vector of unknown coefficients. The disturbance term satisfies the classical stochastic assumptions. Suppose we seek to predict the value of the dependent variable given by
   \[ y_* = x_*' \beta + u_* \]
   for some observation $*$ outside the sample. We treat $x_*$ as given for prediction purposes.
   a. Show that $\hat{y} = x_*' \hat{\beta}$ is the minimum mean-squared error predictor of $y_*$.
   b. Determine the mean and variance of the least-squares based predictor $\hat{y}_* = x_*' \hat{\beta}$ and the prediction error $y_* - \hat{y}_*$.
   c. Suppose $u_t$ i.i.d. $N(0, \sigma^2)$, then what is the distribution of (support your answer!)
   \[ (y_* - \hat{y}_*)/\{\sigma^2[1 + x_*'(X'X)^{-1}x_*]\}^{1/2}? \]
   d. What is the distribution of this statistic if we replace $\sigma^2$ by $s^2$? How might this statistic be used to obtain a prediction interval for $y_*$?

3. Consider the following model of demand and supply for laptop computers:
   \[ Q_i = \alpha_1 + \alpha_2 P_i + \alpha_3 Y_i + u_i^d \quad \text{(Demand)} \]
   and
   \[ Q_i = \beta_1 + \beta_2 P_i + u_i^s \quad \text{(Supply)} \]
   where $Q_i$ is the quantity sold/purchased, $P_i$ is the per unit price, and $Y_i$ is the income level for the customer involved in sale $i$. We assume that both error terms have the usual ideal properties. We consider $Y_i$ to be an exogenous variable.
a. Show that the disturbance term for both equations is correlated with $P_i$. What problems will this introduce for the OLS estimation of each equation?
b. With an appropriate choice of instruments, discuss the properties of instrumental variables estimates of the supply equation.
c. Discuss how you might test whether OLS or IV estimates should be used for the supply equation.
d. Discuss the feasibility of estimating the demand equation by instrumental variables.
e. Suppose we rewrite the supply equation as $Q_i = \beta_1 + \beta_2 P_i + \beta_3 C_i + u_i^s$ where $C_i$ measures the per unit cost of production, then how would the IV estimation of the Demand equation be impacted?

4. Consider the following multivariate regression model:

$$y = X_1 \beta_1 + x_k \beta_k + u$$

where $X_1$ is an $n \times (k - 1)$ matrix of observations of the first $k - 1$ independent and $x_2$ is an $n \times 1$ vector of observation of the $k^{th}$ variable, $y$ is the vector of observations on the dependent variable, and $\beta_1$ and $\beta_k$ are the corresponding coefficients. The disturbance term satisfies the ideal stochastic assumptions.

a. Suppose that the last variable is collinear with the first $k - 1$. What is meant by collinearity in this model? Near extreme collinearity?
b. Describe the properties, including possible optimality, of the OLS estimator of $\beta_k$ under near extreme collinearity.
c. In the case of near extreme collinearity, describe the distribution of

$$\frac{(\hat{\beta}_k - \beta_k)}{(s^2[(X'X)^{-1}]_{kk})^{1/2}}$$

under both the null hypothesis that $\beta_k = \beta_k^0$ and the alternative $\beta_k = \beta_k^1$. What happens to the power as the collinearity becomes extreme?
d. How might one test for the possible collinearity discussed above. Be specific.
e. Discuss possible solutions, if any, to the collinearity problem.

5. Consider the following cross-sectional model of income:

$$y_i = \alpha + \beta x_i + u_i$$

where $y_i$ is income for individual $i$, $x_i$ is a variable that measures ability, say intelligence, and $u_i$ is an unobservable disturbance with is assumed to be $N(0, \sigma^2)$.

a. Assume the $x_i$ are nonstochastic and nonconstant, then what will be the distribution of the OLS estimates of $\alpha$ and $\beta$.
b. Alternatively, and more realistically, assume that $x_i$ are stochastic, what can be said about the distributions of the OLS estimates in small samples and large samples? Be explicit in any additional assumptions introduced.

Suppose that $x_i$ is unobservable but a proxy $x_i^*$, say I.Q., is observable but measures our ability variable with error

$$x_i^* = x_i + \varepsilon_i$$

where $\varepsilon_i$ is the measurement error, which is uncorrelated with $x_i$, and the model which is estimated on the basis of observable variables is

$$y_i = \alpha + \beta x_i^* + u_i^*$$

(c) Show that $x_i^*$ will be correlated with $u_i^*$. What problems will this cause for the OLS estimates and inferences of $\alpha$ and $\beta$.

d. Outline a procedure for estimating $\alpha$ and $\beta$ which will avoid the difficulties introduced.
in (iii). Be explicit about any additional information, variables, or assumptions that are needed to carry out this procedure.

6. Suppose we mistakenly estimate the following model:

\[ y_i = \alpha + \beta x_i + u_i^* \]  

(i)

when the true model has the form

\[ y_i = (\alpha + \beta x_i) \cdot \exp(u_i) \]  

(ii)

where \( x_i \) is nonstochastic and nonconstant and \( u_i \sim i.i.d. \).

a. Determine the mean and variance of \( u_i^* \) in (i). (Hint: Remember that \( \alpha + \beta x_i \) in (i) is the expectation of \( y_i \) and \( u_i^* \) is the residual).

b. What are the implications for the properties fo the OLS estimates of (i) under this form of misspecification?

c. What are the implications for OLS-based inferences based on (i)?

d. How might one test for the possibility of this form of misspecification?

e. Show how the estimation/inference problem can be corrected or avoided.

7. Consider the aggregate production model:

\[ \ln X_t = \beta_0 + \beta_1 \ln L_t + \beta_2 \ln K_t + u_t \]

where \( X_t \) is an index of U.S. GNP in constant dollars, \( L_t \) is a labor input index, \( K_t \) is a capital input index, and \( u_t \) is a disturbance term.

a. A regression of this equation from annual data for 1929-1948 yields:

\[ \ln X_t = -4.0576 + 1.6167 \ln L_t + 0.2197 \ln K_t \]

\[ \text{with } R^2 = .9759, s = .04573, d.f. = 17. \]

Construct a 95% confidence interval for the capital coefficient. What is the relationship of this interval to an hypothesis test?

b. A regression of the equation for the sample 1949-1967 yields:

\[ \ln X_t = -1.9564 + 0.8336 \ln L_t + 0.6631 \ln K_t \]

\[ \text{with } R^2 = .9904, s = .02185, d.f. = 16. \]

Test the hypothesis that the variances \( \sigma^2 \) are the same in both sub-samples (given the coefficients do not differ).

c. A regression over the complete sample 1929-1967 yields:

\[ \ln X_t = -3.8766 + 1.4106 \ln L_t + 0.4162 \ln K_t \]

\[ \text{with } R^2 = .9937, s = .037555, d.f. = 36. \]

Test the hypothesis that all the coefficients are the same in both samples (given the variances are the same).

d. What are the implications for the OLS estimates over the complete sample if we reject the hypothesis tested in (b)? If we reject the hypothesis tested in (c)?

e. Suppose the true model is given by

\[ \ln X_t = \beta_0 + \beta_1 \ln L_t + \beta_2 \ln K_t + \beta_3 E_t + u_t \]

where \( E_t \) is a measure of energy input. Discuss how this can or cannot explain the results in (b), (c), and (d).