Final Exam
Economics 446
Applied Econometrics and Economic Modeling
May 2, 2007
R. Sickles

You have three hours for this final exam. You may use a two-sided 81/2 x 11 piece of paper with notes, etc.

Problems (10 points each)

1. Consider the multiple regression model $y_i = \beta_1 + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \varepsilon_i$ for i=1,...,n. Explain the five key assumptions needed in order to establish the desirable properties of best linear unbiasedness of the ordinary least squares estimates of β_1, β_2, and β_3.

2. Answer the following questions about multicollinearity:

   a. Define what it is.
   b. Why does it occur?
   c. What are its consequences?
   d. How can it be detected?
   e. Is it true that multicollinearity is always a bad thing and nothing can be done about it? Give reasons for your answer together with illustrative examples.

3. Consider the following model of market equilibrium,

$$Q_i = \alpha_0 + \alpha_1 p_i + u_i i \quad \text{(demand equation)}$$
$$Q_i = \beta_0 + \beta_1 p_i + v_i \quad \text{(supply equation)}$$

   a. Show that p_i is an endogenous regressor in both equations, i.e., it is correlated with the error term.

   b. Show that OLS estimates of α_1 and β_1 are biased (this is usually called "endogeneity bias").

   c. Outline an alternative method for consistently estimating the coefficients.

4.   One of the more important implications of combining assets into portfolios is that pooling assets results in a diversification of risk. A prominent model that explains why different assets have different expected risk premia is the capital asset pricing model (CAPM). In terms of data on rates of return observed over time t, the CAPM may be expressed as:

$$R_{jt} - r_{ft} = \beta_j (R_{mt} - r_{ft}) + \varepsilon_t$$

where R_{jt} is the random return on asset j (at time t), r_{jt} is the risk-free rate of return (not a random variable because it is guaranteed), R_{mt} is the random return on the market as a whole, and ε_t is assumed to be an independently and identically distributed error term.

a. What data would you need to estimate the "beta" for different stocks in your portfolio? How would you interpret the stocks (aggressive-passive) with betas greater than or less than one?

b. Since the test of whether or not beta is greater or less than one is crucial in the CAPM, the correct standard error is crucial. How would you modify your testing procedure if you were told that the conditional variance of ε_t was serially correlated, following an AR(1) process in variances, instead of in levels. That is, large shocks are followed by large shocks and small shocks are followed by small shocks (this is call an ARCH (autoregressive conditionally heteroskedastic) process.

5. Results from estimating the relationship between peanut yield Y (in pounds per acre) and the application of nitrogen fertilizer N (in hundreds of pounds per acre) and phosphorus fertilizer P (in hundreds of pounds per acre) are found in the following table. The specified relationship is:

$$Y_t = \beta_1 + \beta_2 N_t + \beta_3 P_t + \beta_4 N_t^2 + \beta_5 P_t^2 + \beta_6 N_t P_t + e_t$$

a. Find and comment on the estimated functions describing the marginal response on yield to nitrogen when P=1, P=2, and P=3.
b. Find and comment on the estimated functions describing the marginal response of yield to phosphorus when N=1, N=2, and N=3.

Variable	Coefficient	Std. Error	t-statistic	Prob
C	1.385	1.264	1.10	0.2855
N	8.011	0.941	8.52	0.0000
P	4.800	0.941	5.10	0.0000
N2	-1.944	0.220	-8.85	0.0000
P2	-0.777	0.220	-3.54	0.0019

NP	-0.567	0.155	-3.65	0.0015

R-Squared = 0.892
Adjusted R-squared = 0.867
S.E. of regression = 0.538
Sum squared residuals = 6.079
Mean dependent variable = 12.04
Standard deviation of dependent variable = 1.473
F-statistic = 34.75
Prob (F-statistic) = 0.000
Dependent Variable: Y
Sample 1 27
Included observations: 27

6. Consider the distributed lag model:

$$y_t = \alpha + \beta_0 x_t + \beta_1 x_{t-1} + \beta_2 x_{t-2} + ... + \beta_n x_{t-n} + e_t \qquad t = n+1,...,T$$

where we assume that the error is distributed normally and independently with constant variance and mean zero.

a. Outline how you would estimate this model assuming the lag coefficients are distributed along a polynomial with order p<n
b. Outline how you would estimate this model assuming the lag coefficients are distributed along a geometric function.
c. What if any problems would either of these assumed lag distributions solve versus the unconstrained model lag model.

7. Labor economists study the determination of labor earnings using a "statistical earnings function." A simple example of such a regression, estimated using data for 31,093 men, is

$$\log Y_i = 7.58 + \underset{(0.00160)}{0.070} X_i + e_i.$$

Here Y denotes earnings and X is years of education; "log" denotes a natural logarithm. The value in parentheses is the estimated standard error.

a. Using your knowledge of logarithmic functional forms, explain the interpretation of the coefficient on education.
b. What does this equation predict would be the earnings of a hypothetical person with no education?
c. Obtain a 90% confidence interval for the rate of return to education.

8.   A suggested outline for writing an empirical report was provided to you. The outline was:
  a. Statement of the problem
  b. Review of the literature
  c. Formulation of the general model
  d. data sources and descriptions
  e. Model estimation and hypothesis testing
  f. Interpretation or results and conclusions
  g. Limitations of the study and possible extensions
  h. Acknowledgements
  i. References and tables

Assume you have written such a paper. Please spend a moment (15 minutes or so) to reflect on how these basic aspects of research can be used in your future job, in your future career in business, in your future in graduate school, and/or your understanding of the print and spoken media.

9.   A new online auction site specializes in selling automotive parts for classic cars. The founder of the company believes that the price received for a particular item increases with its age (i.e., the age of the car on which the item can be used in years) and with the number of bidders. Use the multiple regression output below to answer the following questions.

Summary measures

Multiple R	0.8391
R-Square	0.7041
Adj R-Square	0.6783
StErr of Estimate	148.828

Regression coefficients

	Coefficient	Std Err	t-value	p-value
Constant	-1242.986	331.204	-3.7529	0.0010
Age of part	75.017	10.647	7.0459	0.0000
Number of Bidders	13.973	10.443	1.3380	0.1940

a. Interpret each of the estimated regression coefficients. .
b. Is the founder of the company correct in believing that the price received for the item increases with its age and with the number of bidders? Explain your answer.
c. Identify and interpret the coefficient of determination (R^2) and the standard error of the estimate (s_e) for the model.
d. Would you recommend that this company examine any other factors to predict the selling price? If yes, what other factors would you want to consider?

10. Assume in a multiple regression model with quarterly observations that an AR(4) process is assumed wherein the residual is specified as:

$$e_t = \rho e_{t-4} + u_t$$

where u_t is serially uncorrelated, homoskedastic, and has zero mean. Outline how you would test for serial correlation in such a model and estimate the model if it were determined from your results that such serial correlation existed.