Econometrics--Econ 388

Spring 2005, Richard Butler

Final Exam

your name____

Section	Problem	Points Possible
I	1-20	3 points each
II	21	10 points
	22	15 points
	23	10 points
	24	5 points
	25	15 points
	26	5 points
III	27	20 points
IV	28	30 points
	29	30 points

I. Define or explain the following terms:

1. inverse of a matrix--

2. median--

3. standardized score (or Z-score)--

4. slope coefficient in a simple regression model-

5. type I vs. type II errors -

6. law of large numbers-

7. cointegration of two time series, w_t and v_t--

8. dummy variable trap-

9. LaGrange-Multiplier test-

10. method of moment estimators –

11. maximum likelihood estimation criterion -

12. logistic regression model -

13. “boundedness” problem in the probability model -

14. structural vs. reduced form equations -

15. Breusch-Pagan test-

16. dynamically complete models-

17. weak dependence (in time series) -

18. Durbin-Watson test -

19.omitted variable bias -

20. prediction variance for y_T (value of y_t next year) when the true model is $y_t = x_t\beta + \mu_t$ when the usual assumptions hold--

II. Some Concepts

21. Probability Theory: Suppose that a fair coin is relabeled so that heads are 1, and tails are 2, and that the fair coin is tossed twice. The outcome of the first toss is i (so i=1 or 2) and the outcome of the second toss is j (so j=1 or 2). From these experimental outcomes construct the following two random variables:

$X = i + j$
$Y = |i-j|$

a) Construct the joint distribution of random variables X and Y.

b) What is the conditional distribution of X given Y=0? What is the conditional distribution of of X given that Y=1?

c) Are X and Y independent?

22. Write Shazam programs to make the following tests or estimate the following models requested below, assuming that the sample variables A and B are endogeneous, and that the exogeneous variables are C, D, E, and F.

a. $A_i = \beta_0 + \beta_1 B_i + \beta_2 C_i + \beta_3 D_i + \mu_i$ Do a Hausman test for endogeneity of B on the right hand side of the equation.

b. For the same model as in (a), write out the Shazam code to test for overidentifying restrictions on the “extra” identifying variables, E and F.

c. For the same model as in (a), write out the Shazam code to estimate the model in (a) by two stage least squares.

23. Some Matrix Algebra problems: Let $A=\begin{bmatrix}1 & 0 & 1\\ 2 & 1 & 1\end{bmatrix}$; $B=\begin{bmatrix}1 & 0\\ 0 & 1\\ 1 & 0\end{bmatrix}$; $C=\begin{bmatrix}2 & 1\\ 1 & 1\end{bmatrix}$

a. compute ABC; CAB; BCA if they are conformable for multiplication.

b. Verify (by calculation for these matrices) that (ABC)' =C'B'A' where A' means the transpose of matrix A.

c. Verify that trace(BCA)=trace(ABC) = trace(CAB).

24. Let W_i=weight in pounds of the ith person, and H_i=height of the ith person in inches above five feet (so a six-foot person would have H=12). Suppose that a sample of 100 males yields the following sample regression function (ignoring the error term), which is statistically significant at the five percent level:

$W_i = 125 + 4.0\ H_i$ Sum of Squared Errors=980.0

a. If coach is 76 inches tall, what is his predicted weight?

Now suppose that a friend suggests adding Fi, the percent body fat, to the equation (where F=10 means 10 percent body fat). The body fat of our 100 males is measured, and the new model is estimated as follows:

$W_i = 120 + 4.1\ H_i + .3\ F_i$ Sum of Squared Errors=900.0

b. If coach has a percent body fat of 25%, what is his predicted weight based on the second equation?

c. Which equation do you prefer, if you only want to include statistically significant regressors? (Show your reasoning, including any necessary derivations).

25. Consider the population regression function $y_i = x_i\beta + \mu_i$ for semiannual (6 months) macro data i with vector of aggregate independent regressors x_i and μ_i the usual white noise error term: independently normally distributed with mean 0 and variance σ^2. But the actual data comes either as **annual data**:

$$\bar{y}_1 = y_1 + y_2; \qquad \bar{y}_2 = y_3 + y_4; \qquad \bar{y}_3 = y_5 + y_6$$

etc. with annual data for the independent variables defined analogously

$$\bar{x}_1 = x_1 + x_2; \qquad \bar{x}_2 = x_3 + x_4$$

etc (that is, you add up two semiannual observations to get an annual observation), or the alternative is that the data is given as a **moving average**:

$$y_1^* = \frac{y_1 + y_2}{2}; \qquad y_2^* = \frac{y_2 + y_3}{2}; \qquad y_3^* = \frac{y_3 + y_4}{2}$$

etc. with annual data for the independent variables defined analogously

$$x_1^* = \frac{x_1 + x_2}{2}; \qquad x_2^* = \frac{x_2 + x_3}{2}$$

etc (that is, you add up this period's semiannual value with next period's semiannual value and divide by two to get moving averages).

a. What is the mean, variance, and covariance ($E(\bar{\mu}_i\,\bar{\mu}_{i+1})$) of the error term if we use annual data (regress $\bar{y}$ on $\bar{x}$)?

b. What is the mean, variance, and covariance ($E(\mu_i^*\,\mu_{i+1}^*)$) of the error term if we use moving average data (regress y^* on x^*)?

c. Prove that the estimator using annual data is biased or unbiased (regress $\bar{y}$ on $\bar{x}$).

d. Prove that the estimator using moving average data is biased or unbiased (regress y^* on x^*).

e. Which would you prefer to use: annual data or moving average data? Why?

26. Prove, using matrix algebra for the multiple regression, that if you double the value of y_i for all observations i, then you double the estimated value of the coefficient vector, $\hat{\beta}$.

III. Some Bigger Proofs.

27. Prove the Gauss-Markov theorem (that OLS estimators are BLUE).

28. Given the usual regression model $Y = X\beta + \mu$ where the population error terms have a first order autoregressive process: $\mu_t = \rho\,\mu_{t-1} + e_t$ where e_t is a white noise error term, independently distributed with zero mean and variance σ^2, derive the variance-covariance matrix for μ.

29. Given the usual assumptions about the n by k matrix of instrumental variables, *Z*, prove that the instrumental variable estimator is consistent:

i.e., prove that for $\hat{\beta}_{IV} = (Z'X)^{-1}Z'Y$ that plim $\hat{\beta}_{IV} = \beta$. (You may assume that $\frac{Z'X}{n}$ is a positive definite matrix with finite elements for any value of *n*, the sample size).