Econ 120C **Ramu Ramanathan**
Fall 2003 **First Midterm answers**

I. Consider the model $Y_t = b_1 + b_2X_t + b_3S_t + u_t$, where Y = per-capita expenditure on health care, X is per-capita personal income, and S is the percent of seniors (that is, percent of population 65 years or over). Using data for the U.S. states and the District of Columbia (51 observations), the model was estimated by OLS, and the auxiliary regression for testing for heteroscedasticity is given next (it had R^2 = 0.417).

$$|\hat{u}_t| = 6.619 - 0.683X_t + 0.017X_t^2 + 0.063S_t - 0.003S_t^2$$

Ia. (3 points) Write down in symbolic terms the auxiliary equation that specifies how heteroscedasticity is determined. In it write down the null hypothesis of homoscedasticity. Note that your equation should not have any numerical values.

$s_t = a_1 + a_2X_t + a_3 X_t^2 + a_4S_t + a_5S_t^2$ $\qquad H_0$: $a_2 = a_3 = a_4 = a_5 = 0$

Ib. (5 points) Compute the test statistic, state its distribution under the null and d.f., carry out the test at the 1 percent level, and state your conclusion about homo-/heteroscedasticity in words.

$LM = nR^2 = 51 \times 0.417 = 21.267$. Under H_0 LM has the Chi-square distribution with 4 d.f. For a 1 percent test, the critical value is $LM^* = 13.2767$. Since $LM > LM^*$, we reject homoscedasticity and conclude that there is significant heteroscedasticity.

Ic. (4 points) Based on your results, are OLS estimators biased, inconsistent, or inefficient? What can you say about the validity of hypothesis tests based on OLS estimators?

OLS estimators are still unbiased and consistent but not BLUE or efficient. Also, hypothesis tests are invalid.

Id. (8 points) Describe step-by-step how the auxiliary equation can be used to obtain weighted least squares estimates of the bs (you can assume that there is no problem with negative error variance or standard deviation). Note: Your answer should not be in general symbolic terms. It should be quite specific to these models with numerical values indicated wherever they are known.

Compute $\hat{s}_t = 6.619 - 0.683X_t + 0.017X_t^2 + 0.063S_t - 0.003S_t^2$. Then define $w_t = 1/\hat{s}_t$. Next regress $w_t \times Y_t$ against, w_t, $w_t t \times X_t$, and $w_t \times S_t$, without a constant term.

II. Consider the double-log model of farm output

$$\ln(Q_t) = b_1 + b_2 \ln(K_t) + b_3 \ln(L_t) + b_4 \ln(A_t) + b_5 \ln(F_t) + b_6 \ln(S_t) + u_t$$

where *Q* is output, *K* is capital, *L* is labor, *A* is the acreage planted, *F* is the amount of fertilizers used, and *S* is the amount of seed planted. Using annual data for the years 1949 through 1988, the model was estimated and the Durbin–Watson statistic (d) was 1.409.

IIa. (4 points) Write down the auxiliary equation for the error term for using the DW statistic as the test statistic, state the null hypothesis you will test, and the alternative most common in economics.

$u_t = r\ u_{t-1} + e_t$ $\qquad H_0$: $r = 0$ $\qquad H_1$: $r > 0$

IIb. (4 points) Write down the critical values for the DW test, actually carry out the test, and state what the test result is.

We have, k¢ = 5 and n = 40. $d_L = 1.230$ and $d_U = 1.786$. Because $d_L < d < d_U$, the test is inconclusive.

Next you are asked to carry out an LM test for AR(2), that is, second-order autocorrelation.

IIc. (4 points) Write down the auxiliary equation for the error term and state the null hypothesis of no serial correlation.

$u_t = r_1 u_{t-1} + r_2 u_{t-2} + e_t$ $\qquad H_0$: $r_1 = r_2 = 0$

The auxiliary equation was estimated, after suppressing the 1949 and 1950 data, and the unadjusted R^2 was 0.687.

IId. (6 points) Compute the LM test statistic and state its distribution under the null and d.f. Write down the critical value for the 0.001 level and carry out the test. What do you conclude about serial correlation?

n-2 = 38 and $R^2 = 0.687$. Hence LM = 26.106. Under the null, LM has the Chi-square distribution with 2 d.f. LM = 13.816 < LM. Therefore we reject H_0 and conclude that there is significant second order serial correlation.*

IIe . (12 points) You want to use the generalized Cochrane–Orcutt procedure for taking account of AR(2) in the errors. Carefully describe the steps you should take to obtain the estimates. *Your answers should not be simply the index card copied but should refer specifically to the variables here.*

Step 1: Regress $ln(Q_t)$ against a constant, $ln(K_t)$, $ln(L_t)$, $ln(A_t)$, $ln(F_t)$, and $ln(S_t)$.

Step 2: Compute $\hat{u}_t = ln(Q_t) - \hat{b}_1 - \hat{b}_2 ln(K_t) - ..... - \hat{b}_6 ln(S_t)$.

Step 3: Regress $\hat{u}_t$ against $\hat{u}_{t-1}$ and $\hat{u}_{t-2}$ with no constant to obtain $\hat{r}_1$ and $\hat{r}_2$.

Step 4: Generate $Y_t^* = ln(Q_t) - \hat{r}_1 ln(Q_{t-1}) - \hat{r}_2 ln(Q_{t-2})$, $X_{t2}^* = ln(K_{t2}) - \hat{r}_1 ln(K_{t-1}) - \hat{r}_2 ln(K_{t-2})$, and similarly for the other explanatory variables.

Step 5: Regress Y_t^* against a constant and the X variables and obtain new estimates of the b_s.*

Step 6: Go back to Step 2 and iterate until the error sum of squares for Step 5 does not change by more than a certain percentage, say, 0.01 percent.