

686 Final "Reality Check"

June 8, 1998

You are allotted **two hours** to complete this exam. Conditions are closed book, closed note. i.e., writing implement, hand calculator, and blank paper is all you may access. Pace yourself. Providing a respectable answer to all questions will likely win more total points than exhaustive answers to some, and no answer to others. Take care, and good luck.

1) Imagine that you have just executed an OLS regression on a model related to your own research, and the computer printed out: "Your Residuals Are Not White Noise". (Typical canned package. Either too helpful or not helpful enough.) Which of the following conditions could have generated that message? (List all that apply). For those that you list, describe *one* test you would employ to determine whether that was your problem.

- a) Non-uniform variance in the residuals.
- b) Collinearity
- c) An AR process in the residuals.
- d) Spherical residuals
- e) Homoscedasticity

2) Describe the difference between the OLS and GLS (or "Aitken") estimators. The fullest description will present the estimation formula for each, and describe the difference(s) between the two.

3. Your theory tells you that the percent vote for the Democratic Presidential candidate is a function of macroeconomic conditions, more specifically, higher unemployment favors the Democratic candidate, while higher inflation favors the Republican. You have quadrennial observations dating back to 1892. You estimate a model which predicts to %DEM as a function of the Inflation Rate (IR) and the Unemployment Rate (U). Your Results are:

$$\begin{aligned} \%DEM = 49.0 - 0.245IR + 0.339U \\ (0.082) \quad (0.126) \quad (\text{std. errors in parens}) \end{aligned}$$

N = 25 %DEM, IR, U coded as whole numbers 0-100.

- a) What are the values of student's t for your predictor variables? (See attached t table.)
- b) Are your predictor variables statistically significant at the .05 level?
- c) What would you predict regarding %DEM if the unemployment rate increased from 5% to 6%? What would you predict if the inflation rate increased from 8% to 9%? Are these predictions contingent upon each other, or independent of one another?

d) How would you assess the *substantive* significance of a model parameter? In other words, tell me how to *interpret* a model coefficient, and tell me how you would judge the size of the coefficient to be large or small .

4) What is the difference between the standard error of the *model* (also known as the standard error of the *estimate*) and the standard error of a *parameter*? In what units are each expressed?

5) What is the difference between R^2 and Adjusted R^2 ? Imagine I estimated two different models of some phenomenon called y . (note that I used the same observations on y while estimating both models). Which statistic would be appropriate to use in comparing the overall fit of these two models, R^2 , Adjusted R^2 , both, neither ? Why? What would your answer be if I told you that one model used observations on y , while the next model employed the log of y as the dependent variable?

6) We know that collinearity cannot be "cured", it can only be managed. We also know that the best way to manage the problem depends more upon the substance and theory underlying our model than it does on statistics. Construct or briefly describe substantive or theoretical conditions under which the following approaches would be most appropriate:

a) Factor-analyze the collinear variables, and use the resulting factor score for each case in place of those same variables.

b) A joint hypothesis test on the collinear variables.

c) Do nothing. Who cares? take the OLS estimates just as they are, and use the model as it is to predict y .

d) Drop one or more of the collinear variables, re-estimate that smaller model, and stop.

7) Consider the possibility of **measurement error** in our variables. Such errors may be made in our observations on y and/or in our observations on X_k . Such error may also be systematic or random. Discuss the various possibilities, and point out which are most damaging to our model estimates, and which are least damaging.

8) Statistically, what does it mean for a case to be an "outlier"? Identify one or two procedures or statistics you could employ to statistically identify outliers. Imagine you have so identified one or more outliers during your estimation of some model. What would you do next?

~ end ~