Political Science 686/Pollins Regression Class
Mid-Term Examination

Study Guide

You should be thoroughly familiar and comfortable with the following:

- How regression coefficients (like correlation coefficients) summarize relationships between the variation and covariation between two or more variables.

- How the total variation in our dependent variable can be partitioned between "explained" and "unexplained" variation, and how R^2 is nothing more than a summary of this whole relationship.

- Regression Assumptions. How they keep OLS "ordinary", and exactly when and why we invoke each assumption in our derivations of **b**, Var (**b**), and the proof that the OLS estimator is unbiased.

- What an "estimator" is, the properties of estimators we examine to compare their performance, why those properties matter.

- Understand the estimating equation for $\sigma_{b_k}^2$ (trivariate case). How to the terms in this formula help us see directly the causes of sharp/sloppy parameter estimates?

- Be sure that you understand the general concept of "standard error". Know to a certainty that the standard error of the parameters is *not* the same thing as the standard error of the estimate. Know exactly what those are; i.e., what they show us we are deviating *from*.

- Know how to add, subtract, multiply, transpose, and invert matrices.

- Know how to fully and completely interpret regression coefficients. Show me that you know how to get the numbers to tell you their *substantively* rich story. And that you know how to listen to them.

- Hypothesis testing. How do you know whether X and Y are related? How can you make knowledge claims like "X is related to Y"? When would you use a one-tailed vs. a two-tailed test? How do you choose a level of significance?

- How all the many different *scalar* representations of regression equations (bivariate, trivariate, multivariate...) are captured directly and completely by the simple equation $\mathbf{y} = \mathbf{Xb} + \mathbf{e}$.

- How all of the above points are not isolated factoids. Rather, they hang together, they rely upon and work with each other.

Yes, you *can* do this.

Study Guide for 686 Final "Reality Check"

You should be thoroughly familiar and comfortable with the following:

- All Study Guide points listed on the mid-term guide.

- Precisely where and how OLS estimates are hit by *each* of the pathologies we discussed (Heteroscedasticity, Autocorrelation, Collinearity, Measurement Error, Specification Error).

- Be able to 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. Note that you must understand both **b** and Var(**b**).

- Steps taken in the diagnosis and GLS treatment of Heteroscedasticity; how and why each step is taken. (i.e., the technique of Weighted Least Squares) Be able to identify two diagnostic techniques for this problem.

- Steps taken in the diagnosis and GLS treatment of Autocorrelation; how and why each step is taken (i.e., the technique known as "Single Iteration Cochrane-Orcutt"). Be able to identify two diagnostic techniques for this problem.

- Techniques for identifying "unusual cases". Why "stargazing" is a good place to start, but not a sufficient means for identifying outliers. The logic underlying "deletion diagnostics". Know two techniques beyond stargazing for identifying unusual cases.

- Steps taken in the diagnosis and management of Collinearity; how and why each step is taken. Be able to describe best diagnostic procedures, and to sketch at least two approaches to managing this problem.

- Know the consequences of measurement error (both systematic and random) for OLS estimates. Know how and when these consequences are tied to the violation of specific regression assumptions so that you can clearly identify to location of the problem.

- Know the consequences of Specification Error (omitted variables and superfluous variables). Know whether, how and when these consequences are tied to the violation of specific regression assumptions so that you can clearly identify to location of the problem.

- Know the difference between R^2, adjusted R^2 and the standard error of the estimate. Know how each is a measure of the "fit" between our model and the data. Know the boundaries and the units that each is expressed in.

Yes, you *can* do this.