

# PLSC 503: Problem Set 6

Thad Dunning

Department of Political Science

Yale University

Spring 2010

Due February 23, 2010

# 1 Theoretical/Conceptual Exercises

1. **Multicollinearity.** Let  $Y_i = au_i + bv_i + \epsilon_i$  for  $i = 1, \dots, 100$ . The  $\epsilon_i$  are IID with mean 0 and variance 1. Here,  $u_i$  and  $v_i$  are fixed, not random. These two data variables have mean 0 and variance 1. The correlation between them is  $r$ . Let  $M = [u \ v]$  denote the (partitioned) design matrix.

(a) Show that the design matrix has rank 1 if  $r = 1$  or  $r = -1$ .

(b) Otherwise, let  $(M'M)^{-1}M'Y = (\hat{a} \ \hat{b})'$  be the OLS estimator for  $a$  and  $b$ . Is the OLS estimator biased or unbiased?

(c) Find the variance of  $\hat{a}$ ; of  $\hat{b}$ ; of  $\hat{a} + \hat{b}$ ; and of  $\hat{a} - \hat{b}$ . What happens if  $r = 0.99$ ?

(d) What are the implications of multicollinearity for drawing inferences about  $a$  and  $b$ ? What about their sum and their difference? What are the implications of exact collinearity?

(Note: *exact collinearity* here means,  $r = 1$  or  $r = -1$ ; *multicollinearity* means  $r \doteq 1$  or  $r \doteq -1$ ).

(e) True or false, and explain:

i. Multicollinearity leads to bias in the OLS estimator.

ii. Multicollinearity leads to bias in the estimated standard errors for the OLS estimates.

iii. Multicollinearity leads to big standard errors for some estimates.

2. **IVLS.** Work exercise 24 in the Discussion Questions in Chapter 9 of Freedman (2009).

Briefly explain why each statement is true or false.

## 2 Computer exercises

1. Work all of Lab 7 in Freedman (2009: pp. 301-2). If you are using Stata, ignore the prompts that relate to Matlab code.
2. In this exercise, you will use the data from the study of Miguel, Satyanath, and Sergenti (2004), which are available at: <http://elsa.berkeley.edu/emiguel/data.shtml>. (There should be a tilde between the first forward slash “/” and “emiguel”).
  - (a) Download the “Main Dataset” and “Results Do-file” from “Economic Shocks and Civil Conflict: An Instrumental Variables Approach,” along with any other files that may be of interest to you.
  - (b) Replicate Tables 1-4 in Miguel et al., using the code in “Results Do-file.” Turn in your results.
  - (c) Now regress `NCEP_g` on `gdp_g`, that is, regress current rainfall growth on lagged GDP growth. Turn in your output. Are your results surprising? Why or why not?
  - (d) What are some possible violations of the exclusion restriction in Miguel et al.’s study? Can you think of any possible ways to evaluate these potential violations empirically?