

Political Science 239

Problem Set 8

Due date: Thursday, November 2nd, 2006

For this problem set you will use simulated data available in the file PSet8.RData. In this file, you will find the data frames "exercise1" and "exercise2". The former is the same data that you used in Problem Set 7. You can find the data on the usual website <http://are.berkeley.edu/~rocio/teaching.html>. You will also find the file "ProblemSet8_data.R", which has the code that was used to generate the data. You should take a look at this code to get an idea of what the true population parameters are.

Exercise 1 *Using the simulated random variables $Y, X_1, X_2,$ and X_3 in data frame "exercise1", use parametric bootstrapping to calculate 95% confidence intervals for the OLS estimators of $\alpha, \beta_1, \beta_2,$ and β_3 in the model $Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$. Also, use parametric bootstrapping to estimate the variance-covariance matrix of these OLS estimators (this is, you must calculate not only their variances but also their covariances).*

Exercise 2 *The data in the data frame "exercise2" was simulated using the following model:*

$$y_t = \alpha + \rho y_{t-1} + \epsilon_t$$

$$y_0 = 0$$

where

$$|\rho| < 1$$

and the shocks ϵ_t are iid with mean 0 and variance σ^2 . You are asked to calculate bootstrapped standard errors and bootstrapped bias for the OLS estimators of α and ρ (referred to as $\hat{\alpha}$ and $\hat{\rho}$). Explain why you cannot use non-parametric bootstrapping in this case. Can you use parametric bootstrapping? What is the crucial assumption for parametric bootstrapping to work in this kind of model? Does it hold in this case? If you decide that parametric bootstrapping works in this case, provide bootstrapped standard errors and bias for $\hat{\alpha}$ and $\hat{\rho}$.

Exercise 3 This question refers to the paper "Water for Life: The Impact of the Privatization of Water Services on Child Mortality", by S Galiani, P Gertler, and E Schargrodsky, published in the *Journal of Political Economy*, volume 113 (2005), pages 83–120 (<http://scholar.google.com> will take you to a place where you can download it using your Berkeley student id). Please write two pages addressing the following questions (and, of course, any other issues that you find relevant):

1. Describe and discuss the identification strategy of the paper. Do you find any weaknesses? What parts do you find convincing?
2. Explain the importance of Section II.B in the paper. If you could do it again, what would you do differently?
3. Perform the following thought experiment. Holding the estimation procedure in Section III constant and assuming that you have access to all existing data in Argentina, what data would you add to increase our confidence in the results? Now do the reverse. Holding the data constant, discuss what things you would change and/or add to the estimation to increase our confidence in the validity of the results.
4. The paper presents different pieces of evidence to convince the reader that the estimated effect is not spurious. If you had to defend the paper using only one of these pieces of evidence, which one would it be? Provide an argument to support your choice.
5. In sum, are you convinced that these estimates are capturing a causal effect?

Exercise 4 (DUE WEDNESDAY, NOVEMBER 8TH) *This question will help you check whether your bootstrapping code for Exercise 2 actually works. First, using the bootstrapping code that you did for Exercise 2, bootstrap a hypothesis test to test the null hypothesis that $\rho = 0$. Do you reject the null hypothesis? What is your bootstrapped p-value? Now generate data from the model in Exercise 2 and use Monte Carlo simulations to show that you can recover the correct test level for your bootstrapped tests. You should use the file `bs1mc1.R` as a reference.*