PLSC 503: Problem Set 8

Thad Dunning

Department of Political Science

Yale University

Spring 2010

Due April 20, 2010

1 Theoretical/Conceptual Exercise

1. **The bootstrap**. An analyst assumes the following regression model:

$$Y = X\beta + \epsilon, \tag{1}$$

where Y is an $n \times 1$ vector of observable random variables. Here, X is a fixed $n \times p$ matrix with a vector of 1's as the first column, and ϵ is mean-zero vector of i.i.d. random variables with $\text{var}(\epsilon_i) = \sigma^2$. The OLS estimator for this model is $\hat{\beta} = (X'X)^{-1}X'Y$. The residuals from the OLS fit are $e = Y - X\hat{\beta}$.

Suppose the analyst uses the procedure described by Freedman (2009, Chapter 8) to bootstrap the regression model. In particular, for the kth bootstrap replicate, she samples at random with replacement from the vector e to produce an $n \times 1$ vector of bootstrap errors, $\epsilon_{(k)} = \{\epsilon_{(k)1}, ..., \epsilon_{(k)n}\}'$. For each bootstrap replicate, she then constructs $Y_{(k)} = X\hat{\beta} + \epsilon_{(k)}$ and fits the OLS estimator, $\hat{\beta}_{(k)} = (X'X)^{-1}X'Y_{(k)}$. There are 100 bootstrap replicates. Finally, let

$$\begin{aligned}
\hat{\epsilon}_{(k)} &= Y_{(k)} - X\hat{\beta}_{(k)} \\
s_k^2 &= \frac{\hat{\epsilon}'_{(k)}\hat{\epsilon}_{(k)}}{n-p} \\
\hat{\beta}_{\text{ave}} &= \frac{1}{100}\sum_{k=1}^{100}\hat{\beta}_{(k)} \\
V &= \frac{1}{100}\sum_{k=1}^{100}[\hat{\beta}_{(k)} - \hat{\beta}_{\text{ave}}][\hat{\beta}_{(k)} - \hat{\beta}_{\text{ave}}]'.
\end{aligned}$$

Say whether the following statements are true or false, and most importantly, explain your answers:

(a) $E(\epsilon_{(k)}) = 0_{n\times 1}$.

(b) $E(\hat{\beta}_{(k)}) = \hat{\beta}$.

(c) $E(s_k^2) = \sigma^2$.

(d) $E(s_k^2) = \frac{1}{n} e'e$.

(e) $E(s_k^2) = \frac{1}{n-p} e'e$.

(f) $E(V) = \sigma^2 (X'X)^{-1}$

(g) The square roots of the diagonal elements of V are the bootstrap standard errors.

(h) The sample SD of the $\hat{\beta}_{(k)}$'s is a good approximation to the SE of $\hat{\beta}$.

(i) $\hat{\epsilon}_{(k)} \perp X$ for all k.

(j) The bootstrap can provide evidence that the original data were produced according to equation (1).

(k) The bootstrap can provide evidence that $E(\hat{\beta}) = \beta$ if the original data were produced according to equation (1), with i.i.d. errors, $E(\epsilon_i) = 0$ and $\text{var}(\epsilon_i) = \sigma^2$.