

Political Science 239 - Section Notes VII

Rocio Titunik

Fall 2006

1 How to set-up a Socrates account

If you want to have a Socrates account to be able to use a computer remotely for the midterm, you should go to <http://socrates.berkeley.edu/Acct/apply.html> and download the "Application for Computing Services". Complete this application form and submit it *personally* to the User and Account Services office, which is located at

address: 2195 Hearst Avenue, Room 111

hours: 10:00a.m.to 4:00p.m. weekdays

telephone: 510-642-7355

email: accounts@berkeley.edu

.Accounts are set up within 2 working days, so if you do it tomorrow you should have your account up and running for the midterm.

Alternatively, you can use the computer Lab in 750 Barrows. You should check in with Gilbert if you are planning to do so. He'll help you get a key if you don't have one already.

2 Bootstrapping the Variance-Covariance Matrix of OLS estimators

Consider the classical OLS model

$$Y = X\beta + \epsilon$$

where the outcome Y is a $(n \times 1)$ vector, the $(n \times k)$ matrix X is the matrix of covariates and assume all classical assumptions hold.

The OLS estimator is

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

Note that since $Y = \beta X + \epsilon$ we can write

$$\begin{aligned}\hat{\beta} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \epsilon) \\ &= (X^T X)^{-1} (X^T X) \beta + (X^T X)^{-1} X^T \epsilon \\ &= \beta + (X^T X)^{-1} X^T \epsilon\end{aligned}$$

so we have

$$(\hat{\beta} - \beta) = (X^T X)^{-1} X^T \epsilon$$

Under the classical assumptions $\mathbb{E}(\hat{\beta} | X) = \beta$, the variance of $\hat{\beta}$ is

$$\begin{aligned}
\text{Var}(\widehat{\beta} | X) &= \mathbb{E} \left\{ (\widehat{\beta} - \beta) (\widehat{\beta} - \beta)^T | X \right\} \\
&= \mathbb{E} \left\{ \left((X^T X)^{-1} X^T \epsilon \right) \left((X^T X)^{-1} X^T \epsilon \right)^T | X \right\} \\
&= \mathbb{E} \left\{ (X^T X)^{-1} X^T \epsilon \epsilon^T X (X^T X)^{-1} | X \right\} \\
&= (X^T X)^{-1} X^T \mathbb{E} \{ \epsilon \epsilon^T | X \} X (X^T X)^{-1} \\
&= (X^T X)^{-1} X^T (\sigma^2 I) X (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1} (X^T I X) (X^T X)^{-1} \\
&= \sigma^2 (X^T X)^{-1}
\end{aligned}$$

We can estimate this variance-covariance matrix by

$$\widehat{\text{Var}}(\widehat{\beta} | X) = \widehat{\sigma}^2 (X^T X)^{-1} = \frac{\widehat{\epsilon \epsilon^T}}{n - k} (X^T X)^{-1}$$

But we can also use bootstrapping. Say you've bootstrapped B times and you have B ($k \times 1$) vectors $\widehat{\beta}_1^*, \widehat{\beta}_2^*, \dots, \widehat{\beta}_B^*$. The bootstrapped variance-covariance matrix of $\widehat{\beta}$ is simply:

$$\frac{1}{B} \sum_{b=1}^B \left[\widehat{\beta}_b^* - \overline{\widehat{\beta}}^* \right] \left[\widehat{\beta}_b^* - \overline{\widehat{\beta}}^* \right]^T$$

where

$$\overline{\widehat{\beta}}^* = \frac{1}{B} \sum_{b=1}^B \widehat{\beta}_b^*$$

Note a very important point. When you estimate $\text{Var}(\widehat{\beta} | X)$ by $\frac{\widehat{\epsilon \epsilon^T}}{n - k} (X^T X)^{-1}$ you are relying on the classical assumptions. When you estimate $\text{Var}(\widehat{\beta} | X)$ by $\frac{1}{B} \sum_{b=1}^B \left[\widehat{\beta}_b^* - \overline{\widehat{\beta}}^* \right] \left[\widehat{\beta}_b^* - \overline{\widehat{\beta}}^* \right]^T$ you are not. This means that if the classical assumptions fail, $\frac{1}{B} \sum_{b=1}^B \left[\widehat{\beta}_b^* - \overline{\widehat{\beta}}^* \right] \left[\widehat{\beta}_b^* - \overline{\widehat{\beta}}^* \right]^T$ would still give you the right answer (as long as you have the necessary conditions for bootstrapping to work, namely, independence).

3 Bootstrapping with a lagged outcome as covariate

Let α and ρ be two unknown parameters, with $|\rho| < 1$. For $t = 1, 2, \dots, T$ we have the model

$$Y_t = \alpha + \rho Y_{t-1} + \epsilon_t$$

with y_0 fixed and ϵ_t IID with mean zero and variance σ^2 . The y_{t-1} is what we call a lag term, since it is the value of the outcome for the previous period. To estimate α and ρ by OLS, we construct the following vectors:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ Y_T \end{pmatrix}, \quad X = \begin{pmatrix} 1 & Y_0 \\ 1 & Y_1 \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & Y_{T-1} \end{pmatrix}, \quad \beta = \begin{pmatrix} \alpha \\ \rho \end{pmatrix}, \quad \epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \cdot \\ \cdot \\ \epsilon_T \end{pmatrix}$$

As usual, the OLS estimator for β is $\hat{\beta} = (X^T X)^{-1} X^T Y$ and we can use bootstrapping to estimate the variance-covariance matrix of $\hat{\beta}$. The problem is that non-parametric bootstrapping does not work in this case, because the observations are not independent. So we can't stack Y and X in a matrix and sample rows with replacement, because this rows are not independent. But we can use parametric bootstrapping because the ϵ_t are IID. This is, if the model is correctly specified, the OLS residuals $\hat{\epsilon}_t$ (which are unbiased estimators of the disturbances ϵ_t) will be independent. So while (Y, X) is not a good candidate for bootstrapping, $\hat{\epsilon}$ is.

The only trick is that we need to maintain the structure of the model in every bootstrap sample. The algorithm works as follows. First estimate the model by OLS, and get full sample estimators of α and ρ (I'll refer to those as $\hat{\alpha}^{FS}$ and $\hat{\rho}^{FS}$ and I will define $\hat{\beta}^{FS} \equiv \begin{pmatrix} \hat{\alpha}^{FS} \\ \hat{\rho}^{FS} \end{pmatrix}$), as well as the residuals $\hat{\epsilon}^{FS} = Y - X\hat{\beta}^{FS}$. Next, we fix Y_0 and also fix $\hat{\beta}^{FS}$ and $\hat{\epsilon}^{FS}$. Now resample the $\hat{\epsilon}^{FS}$ with replacement to get a sample $\hat{\epsilon}_1^*, \hat{\epsilon}_2^*, \dots, \hat{\epsilon}_T^*$. The problem is that if we do

$$Y^* = X\hat{\beta}^{FS} + \hat{\epsilon}^*$$

we will break the lagged structure $Y_t = \alpha + \rho Y_{t-1} + \epsilon_t$. So we need to generate the Y_t^* 's one at the time, using $\hat{\alpha}^{FS}$, $\hat{\rho}^{FS}$ and the $\hat{\epsilon}_t^*$'s. This is how you do it:

$$\begin{aligned} Y_1^* &= \hat{\alpha}^{FS} + \hat{\rho}^{FS} Y_0 + \hat{\epsilon}_1^*, \\ Y_2^* &= \hat{\alpha}^{FS} + \hat{\rho}^{FS} Y_1^* + \hat{\epsilon}_2^*, \\ Y_3^* &= \hat{\alpha}^{FS} + \hat{\rho}^{FS} Y_2^* + \hat{\epsilon}_3^*, \\ &\cdot \\ &\cdot \\ Y_T^* &= \hat{\alpha}^{FS} + \hat{\rho}^{FS} Y_{T-1}^* + \hat{\epsilon}_T^*, \end{aligned}$$

Note the difference between the first line and the rest: because you need to fix the initial condition, Y_0 is not obtained through bootstrapping. So this first line looks like standard parametric bootstrapping. But now in order to preserve the model in our bootstrap sample, once we've fixed Y_0 and obtained Y_1^* in the standard way, we "feed" Y_1^* to the second equation to generate Y_2^* and we do this recursively until we generate the last Y_T^* . So we have the following bootstrap dataset:

$$Y^* = \begin{pmatrix} Y_1^* \\ Y_2^* \\ \cdot \\ \cdot \\ Y_T^* \end{pmatrix}, \quad X^* = \begin{pmatrix} 1 & Y_0 \\ 1 & Y_1^* \\ \cdot & \cdot \\ \cdot & \cdot \\ 1 & Y_{T-1}^* \end{pmatrix}, \quad \hat{\epsilon}^* = \begin{pmatrix} \hat{\epsilon}_1^* \\ \hat{\epsilon}_2^* \\ \cdot \\ \cdot \\ \hat{\epsilon}_T^* \end{pmatrix}$$

And then we compute the bootstrap estimator $\hat{\beta}^*$ as

$$\hat{\beta}^* = (X^{*T} X^*)^{-1} X^{*T} Y^*$$

It should be clear that the main difference between this method and the standard parametric bootstrapping method is that we don't let X fixed in the bootstrap samples. The reason is that the second column of X is the lagged outcome, and therefore when we generate a bootstrap outcome

we need to regenerate X . That's why we use X^* instead of X . You can repeat this B times to get B estimators: $\widehat{\beta}_1^*, \widehat{\beta}_2^*, \dots, \widehat{\beta}_B^*$. Then proceed as usual to calculate confidence intervals, p-values, standard errors, etc. Remember, in every bootstrap round you resample from the same residuals $\widehat{\epsilon}^{FS}$, and calculate new Y^*, X^* and $\widehat{\beta}^*$.