Microeconometrics
Ethan Kaplan
Homework #2

Note : Please use STATA and hand in your a printed copy of your .do file in addition to your output.

1 Robust Standard Errors

Generate 10,000 data points, 100 each with

$$\epsilon_i \sim N(0, \sqrt{i}),\ i = 1, ..., 100$$

$$\mu \sim N(0, 5)$$

Generate:

$$X = \epsilon_i + \delta_X,\ \delta_X \sim N(12, 8)$$

Model A: Generate $Y = 5 + 10X + \epsilon_i$
Model B: Generate $Z = 5 + 10X + \mu$
Parameter Model: $Y = \alpha + \beta X + \eta$

1. Run OLS for model A. Do not use robust standard errors. Now, bootsrap (with 500 replications) your estimates for α and β (using non-robust OLS).Non-parametrically plot (using the kdensity command in Stata and the Epanechnikov kernel) the densities of $\hat{\alpha}$ and $\hat{\beta}$. Choose your own bandwidth. Also compute the standard deviation of both $\hat{\alpha}$ and $\hat{\beta}$. Compare these numbers to the OLS standard errors for $\hat{\alpha}$ and $\hat{\beta}$ on the full sample.
2. Now run OLS for model A with robust standard errors. Compare these numbers to the OLS standard errors for $\hat{\alpha}$ and $\hat{\beta}$ on the full sample.
3. Now divide your sample up into $j = 1, ..., 100$ where each the j groups do not intersect, the union of the j groups is the entire sample and each j group contains one of each i (from 1 to 100). Now run OLS for model A, clustering on j. Compare these numbers to the OLS standard errors for $\hat{\alpha}$ and $\hat{\beta}$ on the full sample as well as to your results in (2.).
4. Repeat (1.) for model B.
5. Repeat (2.) for model B.
6. Discuss your results.

2 Clustered Standard Errors

Generate 10,000 data points, 100 each (*i.e.* $j = 1, ..., 100$ and $i = 1, ..., 100$) with

$$\begin{aligned} \epsilon_{j1} &\sim N(0,5) \\ \epsilon_{ji} &= \epsilon_{ji-1} + \mu \end{aligned}$$

where

$$\mu \sim N(0,5)$$

Generate:

$$X = \epsilon_i + \delta_X,\ \delta_X \sim N(12,8)$$

Model A: Generate $Y = 5 + 10X + \epsilon_i$

1. Run OLS for model A. Report your results. Now, bootsrap (with 500 replications) your estimates for α and β (using non-robust OLS).Non-parametrically plot (using the kdensity command in Stata and the Epanechnikov kernel) the densities of $\hat{\alpha}$ and $\hat{\beta}$. Choose your own bandwidth. Also compute the standard deviation of both $\hat{\alpha}$ and $\hat{\beta}$. Compare these numbers to the OLS standard errors for $\hat{\alpha}$ and $\hat{\beta}$ on the full sample.

2. Now run OLS for model A, clustering on j. Compare your results to those in (1.).

3. Now run OLS for model A, clustering on i. Compare your results to those in (1.) and (2.). Which of (2.) and (3.) would you choose given your setup and why?

4. Pick 5 of the j at random (without replacement). Estimate $\hat{\alpha}$ and $\hat{\beta}$ for the sample including only observations in the five j groups using OLS and clustering on j. Save the standard errors for $\hat{\alpha}$ and $\hat{\beta}$. Repeat this 500 times. Compare the distribution standard errors to the standard errors you computed from (1.), (2.), and (3.), including the bootstrapped standard errors. Explain your results.

5. Run OLS on Model A with robust standard errors. Then, run OLS on Model A, clustering on the individual observation. Compare the results and make sense of them.

3 Poisson and Negative Binomial Distributions

1. Download the data from the website http://www-2.iies.su.se/~ekaplan/courses/appliedmicro.html by clicking on "Data for Homework II"

2. The data set is at the town level in the United States. Run OLS, Robust OLS, Poisson, Negative Binomial, and Robust Poisson regressions of numA00all (number of contributions to the republican party) on fox2000 (presence of Fox News in the cable system in the year 2000), noch2000 (number of channels in the year 2000), and popC2000 (census population in the year 2000). Report your coefficients. Compare your results? Are the means different? Standard Errors? Be careful!

3. For the Poisson model, compute the marginal effects for number of channels (noch2000) at the mean of the sample for the other variables for a large number of values of noch2000. Then run a non-parametric regression of the marginal effects of number of channels on noch2000. Compare your results here to their counterparts in your OLS regression.

4. Restructure the data set as a panel with two time periods: 1996 and 2000. Run an xtpoisson of numA* on noch*, popC*, and fox* with town fixed effects. Describe mathematically and in words what STATA is doing with the xtpoisson command and interpret your results.