UPPSALA UNIVERSITET

Assignment 3: Panel data, count models, bootstrapping, and simulating QI I

Panel data and count data pose challenges to our regular OLS framework, as these data types violate the OLS assumptions. As OLS is often unsuitable for these types of data, other alternatives need to be considered. In this assignment we want you to first design a Monte Carlo Experiment to assess the effects of using different types of methods for handling (or not) panel data. Secondly, we want you to apply a range of different models on a (subset of a) real-life dataset with a count variable as the dependent variable, and to simulate the quantities of interest. In the motivation section, briefly discuss how panel and count data violate the OLS assumptions, what the upside of bootstrapping is, and how simulating QI:s can benefit your research. The assignment is due **April 24: 17.00**.

UPPSALA UNIVERSITET

Assignment 3: Part 1 I

1. Generate 100 unique intercepts, a_i, from the random uniform distribution $U(-1, 1)$, with 10 observations for each intercept and a unique ID for each intercept. In `R` this is done by `df <- data_frame(id = rep(1:100,10), a = rep(runif(100,-1,1),10)`
2. Generate two X variables where $X_1 \sim U(a, 3)$ and $X_2 \sim N(3, 2)$
3. Assume that the true DGP is $Y_i = a_i + \beta_1 X_1 + \beta_2 X_2 + u_i$ where a_i is the interecept for the individual observation, $\beta_1 = 1.5$, and $\beta_2 = 0.7$.
4. Draw a realization of the DGP and fit four models on the data
   - Regular OLS

UPPSALA UNIVERSITET

Assignment 3: Part 1 II

- OLS with Fixed Effects
- OLS with Random Effects
- OLS with bootstrapping (1000 bootstraps)

5. Present the regression results for X_1 and X_2
6. Run an MC analysis using the setup from above and investigate whether the estimate of β_1 is unbiased and effective. You may lower the number of bootstraps to 100 to reduce computational time. Note: this step may take some time to complete computationally. To keep track of progress, one could include cat(i,"\n"), where 'i' is the index in the loop, at the end of the loop to keep track of progress.
7. Discuss the results, and the advantages and suitability of each of the models used

UPPSALA UNIVERSITET

Assignment 3: Part 2 I

1. Load the country-month dataset provided for the exercise
2. Create a dummy variable for 'democracy', taking the value 1 if fvp_democracy$>$ 0.75
3. Assume that the model you wish to test is $fatalities = \beta_0 + \beta_1 ln(population) + \beta_2 ln(GDP_{cap}) + \beta_3 Democracy$
4. Fit the model using
   - OLS
   - OLS with logarithmic dependent variable (remember to take $log(Y + 1)$ since $log(0) = -\infty$
   - Possion regression
   - Negative binomial regression
   - Zero-inflated Negative binomial regression
5. Present the results in a regression table

UPPSALA UNIVERSITET

Assignment 3: Part 2 II

6. Compare the results with regards to direction and statistical significance and discuss the differences
7. Select two of the models and
   1. Obtain bootstrapped parameter estimates and standard errors for the coefficients. Discuss how they differ from the non-bootstrapped solution and why that might be.
   2. Use a simulated QI approach, choose freely between the average case and the observed value solution, to estimate the extent to which democracy affects the expected number of fatalities. Present the results.
   3. OPTIONAL: Repeat the above analysis but using bootstrapping instead of the simulated QI approach.