ECONOMICS 582 - ELEMENTS OF ECONOMETRICS I

UNIVERSITY OF TENNESSEE, FALL 2019

Problem Set 3

Due 10/4/2019, 6:00PM

Submit your typed solutions as well as one R or Python script (specifically, one .r or .py file) through the Canvas assignment portal. Show all work. Group work is not permitted on this assignment, although you may bring questions to the lab. Be sure to include your name on each file that you turn in.

Unlike problem sets 1 and 2, almost all of your work will be in R (or Python, if you choose), and you will use your typed solution document to discuss and interpret results.

The Sample1_Data.dta file on Canvas is a partial replication file for the study "Ever failed, try again, succeed better: Results from a randomized educational intervention on grit," by Alan, Boneva, and Ertac, published in *The Quarterly Journal of Economics*. The article is also posted to Canvas.[1]

1. Replicate coefficient values (not standard errors or permutation p values) for Table VI, Sample 1 results. You may use matrix algebra or packages such as `lm()`. The outcome for "Math grade" is `mathgrade2` in the data, and the outcome for "Verbal grade" is `verbalgrade2`. As described in notes to Table VI, controls for this least-squares regression include `grit` (the treatment), `male`, `raven`, `mathscore1`, `verbalscore1`, `belief_survey1`, `grit_survey1`, and `csize`. For many of these controls, you will need to replace missing values with variable means in order to replicate Table VI results. Report coefficient results for `grit` *and other control variables*, control means for each outcome, and observation counts in a table in your typed solution document. Ideally, your code should report these values (or better yet, format and save the table as you wish to report it) so that you do not have to hunt for them and transfer them manually. Automate!

2. For the `grit` treatment effect estimate, follow the residual regression procedure described by Hansen in section 3.18. For each outcome:

    1. Regress the outcome on controls other than `grit`, including the constant. Store residuals, $\mathbf{e}_1$.
    2. Regress `grit` on controls and a constant. Store residuals $\mathbf{e}_2$.
    3. Regress $\mathbf{e}_1$ on $\mathbf{e}_2$.

    After step 3, use logical assertion statements to confirm that the coefficient on `grit` from question 1 is identical to the coefficient on $\mathbf{e}_2$ from question 2, step 3, and that residuals from this procedure are the same as they are in question 1. In your typed solution document, include a table reporting coefficient estimates for $\mathbf{e}_2$ from step 3. In your own words, (1) describe why these reported coefficient should be identical to the least squares coefficients from question 1, and (2) describe the meaning behind step 2 in this context, where the `grit` treatment was randomly assigned.

3. Compute leverage values h_{ii} and prediction errors $\tilde{e}_i$ (see Hansen equation 3.45). Compute the difference between least squares residuals and $\tilde{e}_i$, and for each of the two outcomes, plot this difference against h_{ii} (see the `plot()` command, and be sure that your code saves the plot and that you do not have to do this manually). Include these plots in your typed solution, along with a brief answer to the following: Does the pattern of $\hat{e}_i - \tilde{e}_i$ against h_{ii} make sense? Why or why not?

[1] You will not need the full replication file for this assignment, but if you are interested, it is here: https://doi.org/10.7910/DVN/SAVGAL.

4. For both outcomes, compute R^2 according to the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \hat{e}_i^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

   Confirm, using a logical assertion statement in your code, that your computation of R^2 is the same as the value returned from `lm()`.

5. Add a randomly generated variable `bananas` to the control matrix using the `runif()` function, and re-estimate the models for math and verbal grades by least squares. In a table, report coefficients on `grit` from the original model and the model with `bananas`, along with R^2 values. Discuss any similarities or differences.