POL-GA 1251
Quantitative Political Analysis II
Homework 8
(due May 1)

We are going to work with the following paper,

> Mullainathan, Sendhil, & Ebonya Washington. 2009. "Sticking with Your Vote: Cognitive Dissonance and Political Attitudes." *American Economic Journal: Applied Economics.* 1(1): 86-111.

Read to the bottom of page 88, and then stop and answer question 1 below. I want you to develop your imagination for coming up with placebo tests and so I want you to think of your own tests before you read what they do. Then continue reading and complete the rest of the assignment.

Links to the paper, all replication materials, and also the ANES data are posted alongside the assignment on our course page. Using these materials, carry out the following:

1. *Identification*: Mullainathan and Washington discuss (p. 88) how simple comparisons of the approval ratings by voting versus non-voting partisans may fail to identify the causal effect of voting *per se* on approval ratings. Why is this the case? How does their identification strategy (the one using voting eligibles versus ineligibles) get around this problem? Discuss some possible threats to their identification strategy and propose "placebo tests" that you would want to see to assess these threats.
   (4 points)

2. *Replication*: Replicate Tables 3 and 7 from the paper. The estimates won't be identical because the current ANES dataset includes some updates and corrections since the version that Mullainathan and Washington used. But the estimates should be close. Some other hints for those trying to work with the authors' Stata .do files:

   - To convert variable names from upper case to lower case in Stata, you can use the `renvars` function. Type `findit renvars` and install.
   - Variables are named differently in the current data, which are updated from the data used in the paper: use `vcf0303x` for `vcf0303a`.
   - Change all the file paths that appear in the `.do` file to avoid errors.
   - If you want to use the `outreg` function, type `ssc install outreg` to install it. You might prefer to use `estout` instead though. (I find that `estout` makes nicer tables.)
   (10 points)

3. *Quantile regression*: Can we say whether effects on the "feeling thermometer" (Table 3) are concentrated at the lower end of the scale, in which case those who would otherwise have a rather low opinion tend nonetheless to maintain a moderate level of support or at the upper end of the scale, meaning that those who would otherwise take a moderate position are induced to be strong partisans; or are the effects occurring somewhere in the middle? Is there no pattern along these lines? To assess this, run quantile regressions using the same

specification as for Table 3, but estimating effects for the position of .25, .5, and .75 quantiles. Assess in terms of both the magnitude of coefficients and the statistical significance. In Stata, use the `bsqreg` command to get quantile effect estimates with bootstrapped standard errors. In `R`, you can use the `boot` option for `summary.rq` in the `quantreg` package to do the same. (6 points)

4. *Index construction*: Mullainathan and Washington use placebo regressions on measures of knowledge, exposure, and interest in politics to assess confounding due to acquisition of information about politics. They find no reason to reject the null hypothesis for each of these regressions, providing support for their identification strategy. But is this just because the measures are too noisy to detect confounding effects? To assess this, we will try to use some index methods:

   - First, you need to deal with missing data in the outcomes. What we will do is to simply impute values for the missing outcomes using predicted values from the model specification that you are using to estimate effects. (In principle you would want to use either multiple imputation or bootstrapping to account for the imputation uncertainty, but we will ignore that for now.)
   - Once all the missing data is addressed, compute the correlation matrix for the outcomes. Discuss. Check to see if any variables need to be reverse coded before being incorporated into the index (e.g., if there are any variables that are negatively correlated with the rest.)
   - Now, let's try a few different approaches:
     - Sum score: just add up the different outcome measures to create a sum score index.
     - Principal component factor scores: conduct a principal components (or factor analysis) analysis using the routines in Stata or R. Present a "scree plot" which gives an indication of the number of dimensions of meaningful orthogonal variation is contained in the variables. Then, use the factor score for the first principal component as your index. (The Stata and R principal components scoring functions have methods for estimating the factor scores.)
     - Finally, construct an inverse-covariance weighted (ICW) average along the lines of Anderson (2008). You can find code to do it in Stata or R here:
       `https://github.com/cdsamii/make_index`
       (Before using the code, make sure that you reverse code any variables so that all move in the same direction.)
   - Estimate the placebo regression on each index. What do you find? Does this change your conclusion about the potential for political activity to threaten their identification strategy?
     (10 points)