POL-GA 1251
Quantitative Political Analysis II
Homework 2
Due Thursday, February 8.

I. (10 points) You will prove the consistency of OLS for estimating the PATE with experimental data. Suppose a simple random sample of units indexed by $i = 1, \ldots, N$ drawn from an infinitely large population. For each i in the sample a binary treatment, X_i, is randomly assigned, and a pre-treatment binary stratum indicator, Z_i, is recorded. By random assignment, of course, $X_i \perp\!\!\!\perp Z_i$. After treatment is assigned and the experiment runs its course, you record an outcome Y_i for each unit in the sample, where

$$Y_i = X_i Y_{1i} + (1 - X_i) Y_{0i},$$

with (Y_{1i}, Y_{0i}) being potential outcomes with respect to X_i as we have defined in class. Your analysis plan is such that you estimate a regression of the centered outcome on the centered treatment and stratum indicators, where centering refers to subtracting the sample mean. That is you use OLS to compute $\hat{\beta}_1$ and $\hat{\beta}_2$ as estimates of the coefficients in the following regression,

$$(Y_i - \bar{Y}) = \beta_1 (X_i - \bar{X}) + \beta_2 (Z_i - \bar{Z}) + e_i.$$

It is perfectly valid to have dropped the intercept, because a "centered" constant is just 0. Note as well that the slopes estimated from this regression are the same as those that would be estimated if we had taken the raw Y_i and regressed it on the raw X_i and Z_i (maybe make a little simulation to demonstrate this to yourself). Now define,

$$\mathrm{E}[X_i] = p \qquad \mathrm{E}[Z_i] = q \qquad Y_d = \begin{pmatrix} Y_1 - \bar{Y} \\ \vdots \\ Y_N - \bar{Y} \end{pmatrix} \qquad \mathbf{D}_d = \begin{pmatrix} X_1 - \bar{X} & Z_1 - \bar{Z} \\ \vdots & \vdots \\ X_N - \bar{X} & Z_N - \bar{Z} \end{pmatrix},$$

where $0 < p < 1$ and $0 < q < 1$. Complete the following:

1. Show that,

$$\frac{\mathbf{D}_d' \mathbf{D}_d}{N} \to \begin{pmatrix} p(1-p) & 0 \\ 0 & q(1-q) \end{pmatrix}, \text{ and so } \left(\frac{\mathbf{D}_d' \mathbf{D}_d}{N} \right)^{-1} \to \begin{pmatrix} \frac{1}{p(1-p)} & 0 \\ 0 & \frac{1}{q(1-q)} \end{pmatrix},$$

$$\text{and then, } \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \to \begin{pmatrix} \frac{\mathrm{Cov}(X_i, Y_i)}{\mathrm{Var}(X_i)} \\ \frac{\mathrm{Cov}(Z_i, Y_i)}{\mathrm{Var}(Z_i)} \end{pmatrix},$$

   as $N \to \infty$. (The only asymptotic result that you will need here is the law of large numbers. If you need a refresher, check out the Wikipedia entry!)

2. Show that $\frac{\mathrm{Cov}(X_i, Y_i)}{\mathrm{Var}(X_i)} = \mathrm{E}[Y_{1i} - Y_{0i}]$ to complete the proof.

II. (10 points) You will work with the following paper,

> Gerber, Alan S. & Gregory A. Huber. (2010). Partisanship, Political Control, and Economic Assessments. *American Journal of Political Science 54*(1): 153-173.

You can find the paper and replication materials on Huber's website:

> `http://huber.research.yale.edu/writings.html`

Perform the following, reporting your results in a professional-looking tables or graphs, with proper variable labels, standard error estimates, sample sizes, reasonable numbers of digits, and an explanatory caption for tables, and proper axis labeling and an explanatory caption for figures:

1. Provide an intuitive explanation of the assumptions necessary for the coefficient $B7$ in their regression model to identify the effect of a change in partisan control. What observable implications do Gerber and Huber propose to check on the validity of these assumptions?

2. Use OLS to recreate all of the analyses presented in Tables 3A, 3B, 4A, and 4B in the paper and provide an interpretation of the point estimates. Don't worry about using ordered probit or the precise meaning of the cluster robust standard errors, we will get to those issues in due course.

3. On p. 162, the authors interpret the results from column (1) in Table A in terms of semi-elasticities, writing that "respondents who are not affiliated with either party (e.g., Independents, whose partisanship score is 0) have an average decline in their anticipated holiday spending of 4.4 percentage points be tween the pre- and postelection surveys, or about \$34 given that the average preelection report of anticipated spending is about \$774. Among Strong Democrats (Party ID = 2), spending is instead predicted to increase by about 2.4 percentage points (\$19), while among Strong Republicans (Party ID = -2), it is predicted to decline by about 11.2 percentage points (\$87)." For column (4) on the same table, they write that "a strong Democrat's vacation spending is predicted to increase by about 20% (or \$286 above the average preelection plan of \$1,432), while a strong Republican's is expected to decline by about 11%, or \$162." Explain why these interpretations in terms of percentage changes are justified. That is, provide a proof for why it is correct to interpret the estimate of β from the following model,

$$\log(Y_i) = \alpha + \beta X_i + \varepsilon_i,$$

as a semi-elasticity in percentage terms.

4. Use the FWL theorem to construct scatter plots with regression lines that show the estimated coefficients on the party-ID variables for columns (2) and (4) form Table 4A. Demonstrate that you can use bivariate regressions of residualized variables to recover the same estimates for the coefficients that appear in the table.