POL-GA 1251
Quantitative Political Analysis II
Homework 2

Due Wednesday, February 20 to Antonella or Ye.

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,

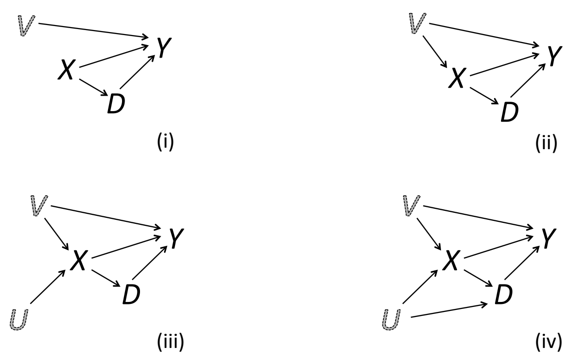
$$\frac{\mathbf{D}_d' \mathbf{D}_d}{N} \to \begin{pmatrix} p(1-p) & 0 \\ 0 & q(1-q) \end{pmatrix}, \text{ and so } \left(\frac{\mathbf{D}_d' \mathbf{D}_d}{N} \right)^{-1} \to \begin{pmatrix} \frac{1}{p(1-p)} & 0 \\ 0 & \frac{1}{q(1-q)} \end{pmatrix},$$

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

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. Consider the following four causal DAGs, where we assume that the variables X, D, and Y are observed, and then V and U are unobserved:



Recall from the theory of DAGs (cf. Morgan and Winship) that, for some variable W in a DAG, conditioning on W is equivalent to performing the following graph operations:

- If W is a collider, link all pairs of parents of W by drawing an undirected edge between them, connoting an induced dependency.
- For any ancestor of W, if this ancestor is itself a collider, link all pairs of parents of this ancestor with undirected edges to connote induced dependencies.
- Erase W from the graph and all edges connected with W.

Now answer the following questions:

1. For which of the graphs is the effect of D on Y non-parametrically identified by conditioning on X? (I.e., for which graphs does CIA hold with respect to conditioning on X?) Explain your reasoning. (5 points)

2. Now suppose that the following system of linear equations holds for each graph:

(i)	(ii)	(iii)	(iv)
$V = \varepsilon_v$	$V = \varepsilon_v$	$V = \varepsilon_v$	$V = \varepsilon_v$
		$U = \varepsilon_u$	$U = \varepsilon_u$
$X = \varepsilon_x$	$X = V + \varepsilon_x$	$X = U + V + \varepsilon_x$	$X = U + V + \varepsilon_x$
$D = X + \varepsilon_d$	$D = X + \varepsilon_d$	$D = X + \varepsilon_d$	$D = X + U + \varepsilon_d$
$Y = D + X + V + \varepsilon_y$	$Y = D + X + V + \varepsilon_y$	$Y = D + X + V + \varepsilon_y$	$Y = D + X + V + \varepsilon_y$

   where the $\varepsilon_{\cdot}$ terms are each independent draws from $\mathrm{N}(0,1)$. From these equations we see that the effect of a unit change in D on Y is defined as being equal to 1. Recall that you only observe Y, D, and X. Consider the OLS regression of Y on D and X. For each case, indicate whether this regression yields an unbiased estimate of the effect of D on Y. If not, is the estimate biased upward or downward? Why? (Hint: feel free to write some simulation code to check yourself.) (10 points)

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:

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 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) from 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.