Part 1: Analytic Exercises

1) Explain why when there is full compliance to an instrument, the LATE is equal to the ATE.

ANSWER: Under full compliance, all individuals are compliers. The LATE is the ATE for the complier sub-group. Hence, when all individuals are compliers, the LATE is equal to the ATE. (10 points)

2) A researcher has a binary treatment D and a binary instrument Z. She reports that the estimate of the slope coefficient from the regression of D on Z has a value of 0.4, and that the estimate of the slope coefficient from the regression of Y on Z has a value of 3.

a) What is the Wald IV estimate? (10 points)

ANSWER: Wald IV is

$$\beta(Z) = \frac{Cov(Y,Z)}{Cov(Z,D)} = \frac{Cov(Y,Z)/V(Z)}{Cov(Z,D)/V(Z)} = \frac{3}{0.4} = 7.5$$

b) What is the estimate for the intent-to-treat (ITT) estimator? (10 points)

ANSWER: ITT is the regression of Y on Z, $ITT = 3$.

c) Explain why the IV estimate might be more relevant to policy than the ITT. (10 points)

ANSWER: If the policy is centered on the treatment D itself, then we would like to know the effect of the treatment, rather than of the instrument.

d) Explain why the ITT might be more relevant to policy than the IV. (10 points)

ANSWER: If the policy is centered on the instrument Z itself, then we might want to know how agents respond to the offer of the instrument.

3) A researcher has a binary treatment $D \in \{0, 1\}$ and a binary observed covariate $W \in \{1, 2\}$. She reports estimates of the following conditional means: $\hat{E}(Y|D = 1, W = 1) = 1$, $\hat{E}(Y|D = 1, W = 2) = 2$, $\hat{E}(Y|D = 0, W = 1) = 3$, $\hat{E}(Y|D = 0, W = 2) = 4$. She also reports that the sample fraction with $W = 1$ is 0.5, and $W = 2$ is 0.5.

a) What is the OLS estimate of the ATE (i.e the estimate of the slope from the regression of Y on D)? (10 points)

ANSWER: OLS estimate is

$$\hat{E}(Y|D = 1) - \hat{E}(Y|D = 0) = (0.5 * 1 + 0.5 * 2) - (0.5 * 3 + 0.5 * 4) = -2$$

b) Assuming the CIA holds, what is the estimate for the matching estimator for the ATE, matching on exact W values? (10 points)

ANSWER: Matching is

$$\hat{E}(Y|D = 1, W = 1) - \hat{E}(Y|D = 0, W = 1) = 1 - 3 = -2$$

$$\hat{E}(Y|D = 1, W = 2) - \hat{E}(Y|D = 0, W = 2) = 2 - 4 = -2$$

ATE estimate is then

$$0.5 * (-2) + 0.5 * (-2) = -2$$

4) A researcher is considering a matching estimator based on a conditional independence assumption $Y_1, Y_0 \perp D|W$. The researcher's dataset has 20,000 observations, and W takes on 2,000 unique values.

a) Describe the problems with the matching estimator based on matching on all 2,000 values. (10 points)

ANSWER: There may be a sparse cells problem, where there is no overlapping support for at least some W cells. Even if there is overlapping support, small numbers of observation per cell may lead to an imprecise estimate.

b) Describe at least one potential solution to these problems, and how this solution balances bias versus variance. (10 points)

ANSWER: One solution is to aggregate some cells so that there are less than 500 values to match on. Another solution is to match based on the propensity score, and this may lead to some reduction in dimensionality. Both solutions might introduce bias but reduce the variance of the estimator.

5) Assume U_V is distributed Uniform $[-3, 3]$, and $V = Z + U_V$ with $Z \in \{0, 1\}$:

a) Show the range of U_V values for the complier, defier, always taker, and never taker groups. (10 points)

ANSWER: Compliers are such that $V = 1 + U_V > 0$ and $V = U_V < 0$ (ignore ties). Then we have for compliers, $U_V > -1$ and $U_V < 0$, so complier range is $U_V \in (-1, 0)$. Always takers are then $U_V \in (0, 3)$ and never takers are then $U_V \in (-1, -3)$.

b) Compute the fraction of the population in each group. (10 points)

ANSWER: Given the assumption that $U_V \in U(-3, 3)$, the compliers are then 1/6 of the population, always takers 3/6, and never takers 2/6.

6) A researcher conducts an experiment where the treatment D is randomized.

a) Explain one limitation of the experiment in estimating the full distribution of treatment effects of D (10 points)

ANSWER: Randomized experiments only identify the ATE. If there is heterogeneity in treatment effects, then the ATE may not be represenative of the range of possible treatment effects.

b) Describe a balance test and why researchers routinely use them. (10 points)

ANSWER: A balance test is a test for proper randomization where the researcher tests for whether variables which should be unaffected by the treatment are equally balanced in the treatment and control samples. The null hypothesis for the test is typically formed in terms of differences in means: $E(X|D = 1) - E(X|D = 0)$, where X is a variable believed to be unaffected by the treatment. Researchers routinely use these tests because they provide a simple check to make sure the experiment was properly conducted and independence of potential outcomes is likely to hold.

Part 2: Data Exercise

7) Use the data sample from practicedata.dta, which contains 4,567 observations on Y, D, X, Z.

a) Compute the OLS estimator for the effect of the treatment on Y (i.e. regression of Y on D). Ignore X (10 points)

b) Compute the Wald IV estimator using Z as an instrument. Ignore X. (10 points)

c) Compute a matching estimator for the ATE assuming the CIA holds: $Y_1, Y_0 \perp D|[X, Z]$. (20 points)

d) Compute a two step control function estimator for the ATE assuming joint Normality: U_0, U_1, U_V distributed joint Normal. Assume treatment is determined $D = 1\{\delta_0 + \delta_1 Z + U_V > 0\}$. Ignore X. (20 points)

Show all of your results in one Table, reporting 4 numbers for a), b), c), and d). You do not need to show standard errors, just the estimates. You MUST include STATA code showing all of your work leading to your estimates (we want to give you partial credit!).