ECON2390, Applied Econometrics

Problem set 2

1. **Regression Discontinuity Design.** This question is based on the paper "Trafficking Networks and the Mexican Drug War" by Melissa Dell (2015, AER). Dell studies the effects of close elections of PAN mayors[1] on drug-related violence. For more context and details about the empirical strategy, read the Introduction and the Econometric Framework (page 1748) section.

   Use the dataset *Dell.dta* in Canvas to answer the following questions. This data corresponds to the municipalities with mayoral elections involving a mayoral PAN candidate. For this problem set, we will focus exclusively on the effect of a PAN mayor taking office on the overall homicide rate during the lame duck period.

   (a) Consider the following regression:

   $$Y_m = \beta_0 + \beta_1 \cdot \text{PANwin}_m + \epsilon_m$$

   where Y_m is the overall homicide rate in municipality m and PANwin_m is a dummy equal to one if a PAN mayor taking office. Estimate β_1 by OLS and interpret the result. Discuss why an OLS regression may not be sufficient to capture the causal effect. In what way might these estimates be biased?

   (b) Identify and explain the key assumptions necessary for the regression discontinuity design (RDD) to be valid in Dell's analysis. Describe how these assumptions apply specifically to the setting in her study.

   (c) Use the McCrary test to empirically evaluate whether the running variable in the regression discontinuity design shows evidence of manipulation. Present the plot, interpret the results, and discuss how this test informs us about the continuity of the conditional expectations of counterfactual outcomes in the running variable.

   (d) Use the `rdplot` command to create binned scatterplots illustrating the relationship between the PAN win margin and the homicide rate Y_m.

      i. Start with the default numbers of bins and a linear polynomial specification to observe the basic trend.

      ii. Try higher-order polynomials (quadratic and cubic)

      iii. Try different procedures for selecting the number bins: `es`, `qs`, 10, and 20

      Describe the patterns you observe and discuss how the different specifications affect the plot.

   (e) Estimate the effect of a PAN mayor on the homicide rate using both `rdrobust` and a parametric (OLS) approach.

      i. For `rdrobust`, use a triangular kernel and fit polynomials of degree 1 and 2.

      ii. For the parametric model, use the same polynomial degrees (1 and 2) and estimate the regression within the optimal bandwidth selected by `rdrobust`.

[1] The National Action Party (Partido Acción Nacional, PAN) is a center-right political party in Mexico. Crucially, during President Felipe Calderón's administration (2006–2012), PAN led a major anti-drug trafficking campaign, deploying thousands of federal troops to combat organized crime.

Compare and interpret the results from each approach. Discuss the differences between `rdrobust` and OLS, and any differences in the estimates, precision, or assumptions between the two methods.

(f) Perform robustness checks on the main RDD result (polynomial of degree 1 with triangular kernel) using `rdrobust` as follows:

i. Apply a 'donut hole' approach by re-estimating the model after excluding observations within a narrow range around the threshold (specifically, those within 5%, 10%, and 20% of the optimal bandwidth on either side). Discuss how this method assesses the robustness of your results.

ii. Use a loop to re-estimate the model with bandwidths ranging from 0.50 to 1.50 times the optimal bandwidth (in increments of 0.01). Plot the estimated coefficients and their confidence intervals across these bandwidths. Interpret the results.

For the next three questions, consider the article "Can Close Election Regression Discontinuity Designs Identify Effects of Winning Politician Characteristics?" by John Marshall (2024, AJPS) and the *Candidates.dta* dataset containing candidate-level covariates for the candidates in our main sample.

(g) Explain Marshall's concern about compensating differentials and discuss its implications for the causal interpretation of regression discontinuity designs in Dell's setting.

(h) Using the *Candidates.dta* dataset, test for discontinuities in candidate characteristics. What do your findings imply about the validity of Dell's results? Additionally, would controlling for these covariates in the main regression address potential biases in the estimates?

(i) Using equation (9) from Marshall's paper (page 507), implement the suggested approach to bound the effect of a PAN mayor on the homicide rate. Since we lack identification strategies for the compensating differentials, to capture plausible values of $\hat{\gamma}_k$, estimate a separate OLS regression of Y_m on each potential compensating differential to obtain its corresponding coefficient. Then, for each compensating differential, multiply the OLS coefficient by 0.5, 1, and 1.5 to generate a grid of plausible values for $\hat{\gamma}_k$. Using these values, compute the corrected PCRD estimator.

Discuss what the computed equation (9) reveals about the potential influence of compensating differentials and their implications for the interpretation of Dell's results.

2. **Instrumental Variables with Controls.** In this question, you will explore the properties of the Two-Stage Least Squares (2SLS) estimator when incorporating covariates, focusing on its performance in large samples. Consider the following data-generating process:

   The covariates $X \in \{(0,0),(1,0),(0,1),(1,1)\}$ are equally likely. Further, assume that the probability of Z (the instrument) given X is:

$$\mathbb{P}(Z=1 \mid X=(x_1,x_2)) = \begin{cases} 2/5 & \text{if } x_1=0 \text{ and } x_2=0, \\ 1/5 & \text{if } x_1=1 \text{ and } x_2=0, \\ 3/5 & \text{if } x_1=0 \text{ and } x_2=1, \\ 4/5 & \text{if } x_1=1 \text{ and } x_2=1. \end{cases}$$

   The conditional group shares (Never-Takers $G=\text{NT}$, Compliers $G=\text{CP}$, and Always-Takers $G=\text{AT}$) are given by:

$$\begin{aligned} &\mathbb{P}(G=\text{NT} \mid X) = 1/3, \\ &\mathbb{P}(G=\text{CP} \mid X) = 1/6 + x_1/6, \\ &\mathbb{P}(G=\text{AT} \mid X) = 1 - \mathbb{P}(G=\text{CP} \mid X) - \mathbb{P}(G=\text{NT} \mid X). \end{aligned}$$

   The potential outcomes for Y are:

$$Y_i(0)=\epsilon_i, \quad Y_i(1) = \begin{cases} 2+0.5x_1+1.5x_2+\nu_i & \text{if } G_i=\text{NT}, \\ 6+2x_1+x_2+\nu_i & \text{if } G_i=\text{CP}, \\ 10+x_1+\nu_i & \text{if } G_i=\text{AT}, \end{cases}$$

   where ϵ_i and ν_i are independent standard normal random variables. As always, $Y_i = D_iY_i(1) + (1-D_i)Y_i(0)$.

   (a) Generate a dataset with at least 10,000 observations. Consider the following outcome equation:

$$Y_i = \beta_0 + \beta_1 D_i + \beta_2 X_{1,i} + \beta_3 X_{2,i} + \beta_4 X_{1,i} \times X_{2,i} + e_i$$

   with corresponding first stage:

$$D_i = \alpha_0 + \alpha_1 Z_i + \alpha_2 X_{1,i} + \alpha_3 X_{2,i} + \alpha_4 X_{1,i} \times X_{2,i} + u_i.$$

   Calculate the 2SLS estimator. Then, calculate the Wald estimator using the residuals of Z from a regression of Z on X_1, $X2$, and $X_1 \times X_2$, instead of the usual Z. Are the 2SLS and residualized Wald estimators equal?

   (b) Explain how a researcher observing (Y_i, D_i, Z_i, X_i) would identify the conditional-on-X_i LATEs, $\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i, G = \text{CP}]$. Compare your proposed estimator to the true conditional LATEs in your data.

(c) Angrist and Imbens (1995) show that under the standard IV assumptions (Independence, Exclusion, First Stage, and Monotonicity) and a fully saturated model for covariates, the 2SLS estimator identifies a convex weighted average of the conditional LATEs:

$$\beta_{2sls} = \mathbb{E}\left[\frac{V\{\mathbb{E}[D_i \mid X_i, Z_i] \mid X\}}{\mathbb{E}[V\{\mathbb{E}[D_i \mid X_i, Z_i] \mid X\}]}\mathbb{E}[Y_i(1) - Y_i(0) \mid X_i, G = \text{CP}]\right].$$

Using your dataset, verify this relationship empirically. Which group gets more weight? Discuss how the data-generating process would need to change for the 2SLS estimator to identify the unconditional LATE $\mathbb{E}[Y_i(1) - Y_i(0) \mid G = \text{CP}]$.

(d) Suppose the interaction term $X_1 \times X_2$ is excluded from the set of controls, leaving only X_1 and X_2. Discuss whether the resulting 2SLS estimator can be interpreted as a convex-weighted average of treatment effects. Use your knowledge of the DGP to argue your case.

(e) Can a researcher observing (Y_i, D_i, Z_i, X_i) identify the unconditional LATE $\mathbb{E}[Y_i(1) - Y_i(0) \mid G = \text{CP}]$? If yes, compute the estimator and compare it to the true value in your data. If no, provide a clear argument as to why identification is not possible.