Empirical Project II

ECON 2390

November 27, 2024

1 Question 1

1.1 Part (a)

Table 1 presents the results of the following regression equation estimated using ordinary least squares (OLS):

$$Y_m = \beta_0 + \beta_1 PANwin_m + \varepsilon_m, \tag{1.1}$$

Table 1: OLS Estimation

(Intercept)	43.626***
	(11.519)
PANwin	0.904
	(13.485)
Num.Obs.	152
R2	0.000
R2 Adj.	−0.007

The results reveal no statistically significant differences in overall homicide rates between municipalities where PAN candidates won the elections and those who lost during the lame-duck period.

However, these findings are unlikely to represent the causal effect of a PAN mayor taking office, as election outcomes are not random. For instance, municipalities that elected PAN candidates may have experienced higher homicide rates before the elections. They might have chosen PAN candidates hoping to improve public safety or due to unobservable factors influencing both violence and voting behavior. For example, municipalities with lower education levels might experience higher levels of violence and also be more likely to elect right-wing politicians.

In either scenario, if PAN mayors did reduce homicide rates during the lame-duck period, the OLS estimates would obscure this effect due to pre-election differences. Consequently, the estimates would underestimate the true causal impact.

1.2 Part (b)

A key identifying assumption in the sharp RDD framework is the continuity of the conditional means of the potential outcomes at the cutoff. Formally, this requires $\mathbb{E}(Y(1) \mid X)$ and $\mathbb{E}(Y(0) \mid X)$ to be continuous at $X = c$. This assumption ensures that any observed discontinuity in the outcome variable at the cutoff arises solely from the change in treatment status and not from other factors coinciding with the threshold.

In Dell's study, X represents the margin of PAN victory in municipal elections, and $c = 0.5$ is the cutoff separating victory from defeat. The identifying assumption implies that homicide rates, conditional on treatment status (whether a PAN mayor takes office), should not change discontinuously at the electoral threshold. In essence, the only source of discontinuity in homicide rates at the cutoff should be the change in the electoral winner's identity, as determined by the treatment assignment.

Continuity might be violated by manipulation or selective sorting around the cutoff. The absence of such manipulation requires the probability density function (pdf) of X, denoted $f(X)$, to be continuous at $X = c$. This ensures that the estimated difference in the conditional means at the cutoff reflects the treatment effect alone rather than the influence of strategic sorting around the threshold.

In the context of Dell's study, this implies that the pdf of the PAN victory margin must be smooth at the electoral victory cutoff. If the margin were manipulated—such as through electoral fraud, strategic behavior, or administrative interference—this could introduce bias. For instance, if municipalities with higher pre-existing homicide rates were more likely to manipulate election outcomes in favor of PAN candidates, the resulting discontinuity would confound the treatment effect estimate. Ensuring continuity of $f(X)$ at the cutoff is therefore critical to attributing any observed discontinuity in homicide rates solely to the treatment effect of a PAN mayor taking office.

1.3 Part (c)

Figure 1 illustrates the results of the McCrary test for density discontinuity:

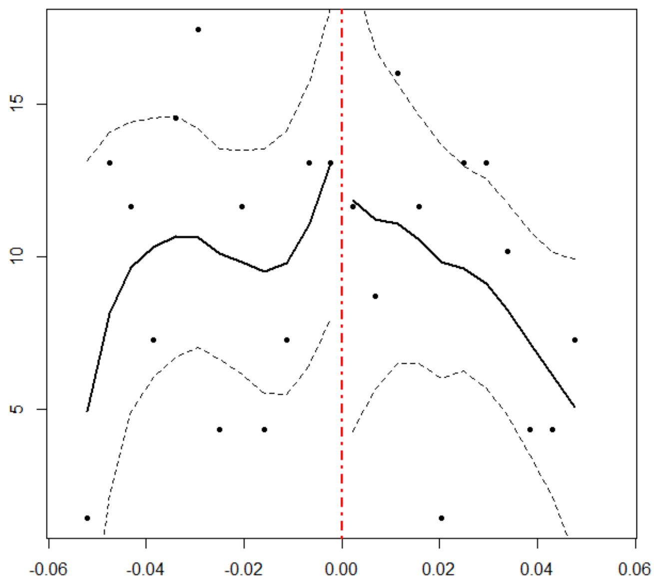
The graph illustrates the predicted distribution of the running variable, the PAN victory margin. It demonstrates that the probability density function of the running variable is continuous at the cutoff. This continuity reduces concerns that differences in homicide rates at the cutoff could be attributed to manipulation of the running variable. Although the test results do not eliminate the possibility of other factors changing discontinuously at the electoral victory cutoff and influencing homicide rates, they support the validity of the key identifying assumption: the continuity of the conditional expectations of counterfactual outcomes at the cutoff.

1.4 Part (d)

Figure 2 presents four scatterplots illustrating the relationship between the PAN win margin and the homicide rate.

The different specifications influence the plots in several ways. Higher-order polynomials make the plots more volatile and amplify the observed jump at the cutoff. Similarly, the specified number of bins affects the shape of the curves. Despite these variations, all plots—including those not reported here—consistently show a discontinuous increase in homicide rates at the

Figure 1: McCrary Test: Density of Victory Margin



cutoff. This suggests that PAN mayors taking office might have positively impacted homicide rates during the lame-duck period.

1.5 Part (e)

Table 2 presents the **rdrobust** and OLS estimators, along with their corresponding standard errors.

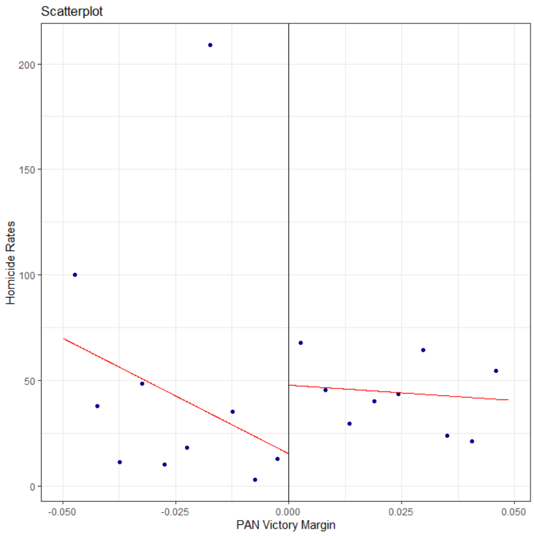
Table 2: Causal Effect Estimates for Different Specifications

Method	Estimate	Standard.Error
Non-parametric (Linear)	88.71	47.72
Non-parametric (Quadratic)	100.48	59.43
OLS (Linear)	76.40	25.88
OLS (Quadratic)	73.87	24.16

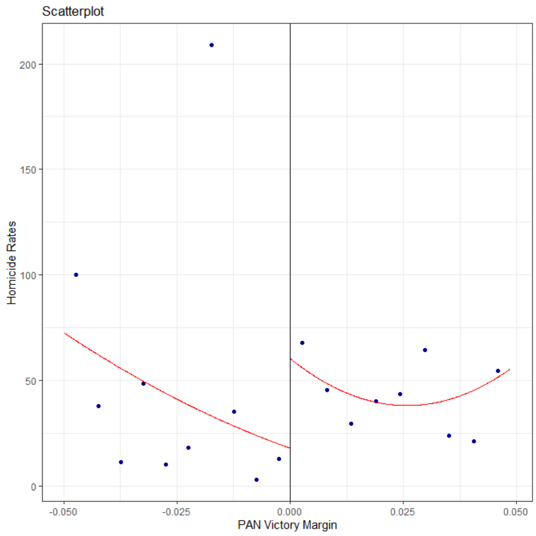
Using the OLS estimator with a linear polynomial, a narrow PAN victory is associated with an increase in homicide rates of approximately 76 cases per 100,000 people. The quadratic polynomial OLS specification yields a slightly lower estimate, predicting an increase of around 74 cases. In contrast, the **rdrobust** estimator, which employs local polynomial regression, suggests a larger increase: approximately 89 cases per 100,000 people with a linear polynomial and 100 cases with a quadratic polynomial. The OLS estimates exhibit greater precision, as reflected in their smaller standard errors compared to the **rdrobust** estimates.

The key difference between the two approaches lies in their underlying assumptions. The OLS method relies on a parametric specification, assuming that the relationship between the outcome (homicide rates) and the explanatory variable (PAN victory margin) can be captured

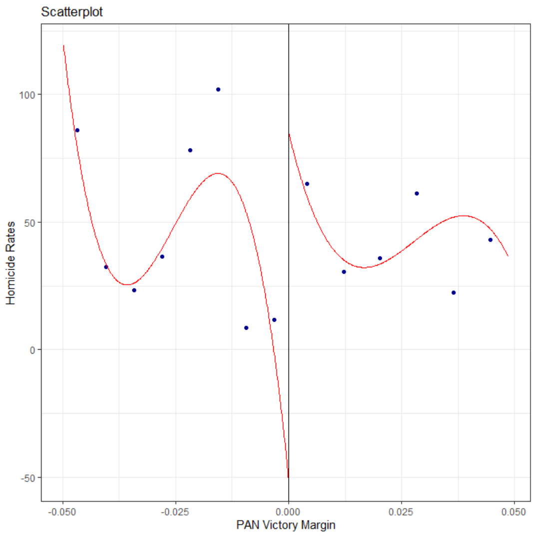
Figure 2: PAN Victory Margin and Homicide Rate



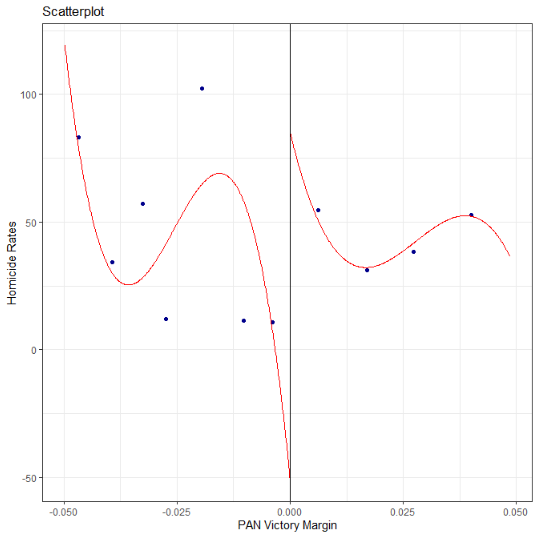
(a) Default bins, linear polynomial



(b) Default bins, quadratic polynomial



(c) **es** bins, cubic polynomial



(d) **qs** bins, cubic polynomial

by a global quadratic function. In contrast, the **rdrobust** estimator focuses on local behavior, fitting a polynomial only within a specified bandwidth around the cutoff, thus minimizing the influence of distant observations. This nonparametric approach avoids strong assumptions about the global functional form of the relationship.

While the OLS and **rdrobust** methods are built on different econometric principles, their consistent results lend credibility to the estimated treatment effect, suggesting that the choice of method does not critically alter the inference drawn near the cutoff.

1.6 Part (f)

1.6.1 Donut-hole Approach

Table 3 presents the **rdrobust** estimates using the "Donut-hole" approach, which excludes observations within a narrow range around the cutoff. The estimated effect of a PAN victory on the homicide rate appears sensitive to the exclusion fraction. For instance, excluding observations within 10% of the optimal bandwidth increases the estimated effect by approximately 24 cases per 100,000 people, while excluding observations within 20% significantly reduces the effect, rendering it statistically insignificant.

The motivation behind this approach is to assess the presence of selective sorting near the cutoff. In the absence of sorting, removing observations in a narrow band around the cutoff should not substantially alter the estimated treatment effect. However, while the Donut-hole method provides a useful robustness check, it has limitations. Excluding observations near the cutoff can increase the estimator's bias and variance by relying more heavily on observations farther from the threshold. Consequently, the approach should be interpreted with caution. Indeed, the results in Table 3 suggest the potential sensitivity of the estimates to the choice of exclusion fraction, particularly for larger exclusions, casting doubt on the robustness of the estimates.

Table 3: Donut-hole Approach

Exclusion Fraction	Estimate	Std Error
0.05	88.71	47.72
0.10	112.91	61.48
0.20	14.94	61.36

1.6.2 Bandwidth Selection

Figure 3 illustrates the estimated effect of a PAN mayor taking office on homicide rates across various bandwidth choices. The figure also includes robust bias-corrected confidence intervals for each estimate.

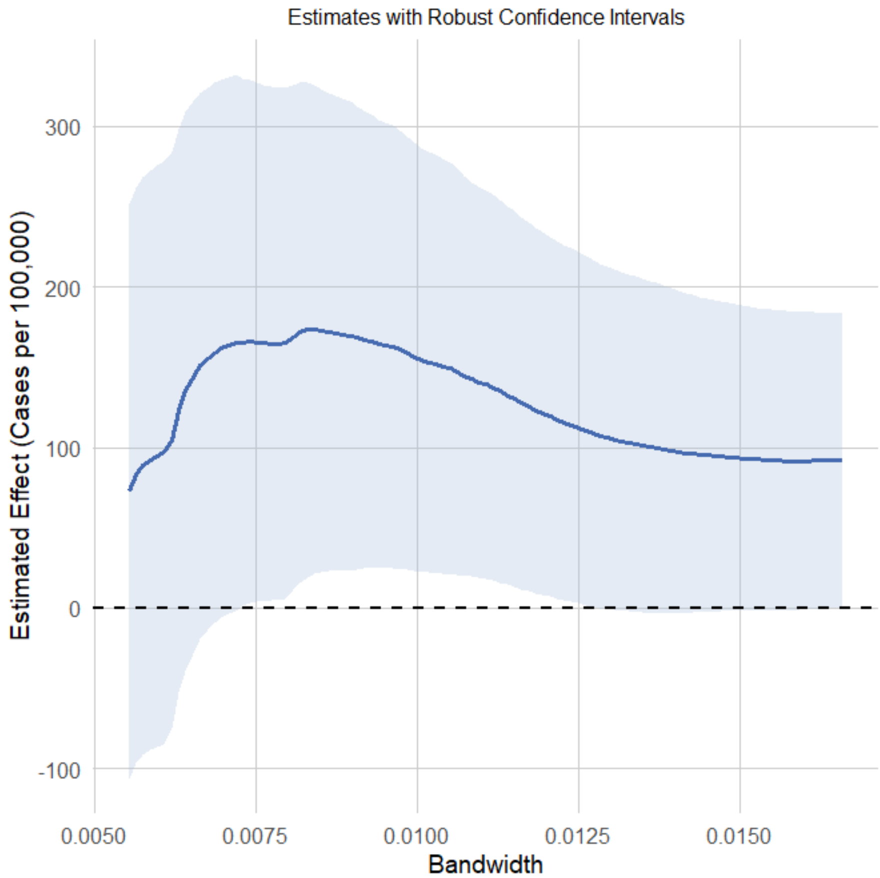
The figure reveals a hump-shaped relationship between the bandwidth and the estimated effect. Notably, the estimator is statistically significant only for a subset of the bandwidths examined, including the optimal bandwidth used in the baseline specification, as indicated by the robust bias-corrected confidence intervals. Overall, the positive effect of a PAN victory on homicide rates appears robust to bandwidth choice, particularly within a narrow range around the optimal bandwidth.

1.7 Part (g)

Many recent papers have exploited the RD design to identify the causal effects of specific characteristics of winning candidates on post-election outcomes. Focusing on close elections mitigates concerns that unobservable factors simultaneously affect both the odds of the winning candidate possessing this characteristic and the outcomes of interest.

Marshall discusses a potential and largely overlooked threat to identification. He argues that the characteristic of interest (denoted X) might systematically influence candidates' popularity.

Figure 3: PAN Victory and Homicide Rates: Different Bandwidths



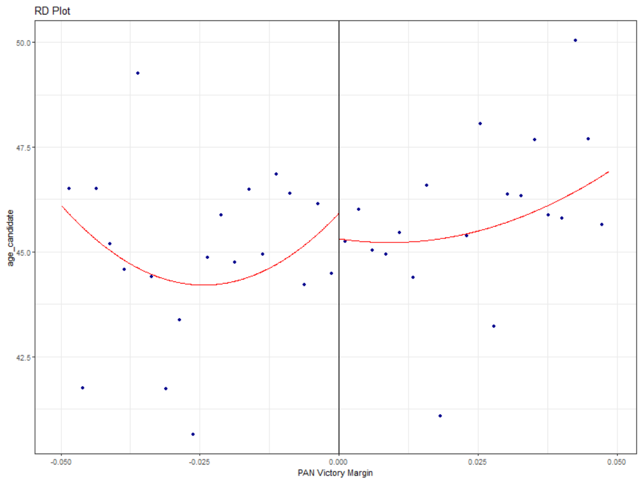
The fact that the elections were close implies the existence of other factors (denoted Z), which the author refers to as *compensating differentials*, that affect the candidates' popularity in the opposite direction. For instance, if X increases a candidate's popularity while Z decreases it, then in close elections, winners who possess X might also be more likely to possess Z. This selection pattern would hold even if X and Z are generally uncorrelated. If Z also affects the outcome of interest, Y, the standard RD estimates, which do not account for Z, would be biased. The resulting bias is similar to the one arising in the standard case of the omitted variables when the omitted variable is negatively correlated with the main explanatory variable. Specifically, Marshall demonstrates that if Z influences popularity and Y in the same (opposite) direction, the RD estimate would exhibit a downward (upward) bias.

Compensating differentials might pose an issue in Dell's study. To illustrate, consider the following scenario: suppose PAN candidates are generally more popular due to, for example, belonging to the ruling party. In municipalities with close elections, PAN candidates might be less competent than candidates from other parties (who garnered a significant share of votes despite facing generally more popular PAN candidates). If competence reduces homicide rates, the observed increase in homicide rates associated with the election of a PAN candidate could be partially explained by PAN candidates being less competent. Consequently, the causal effect of PAN policies or affiliation on homicide rates reported in Dell's study might be overstated.

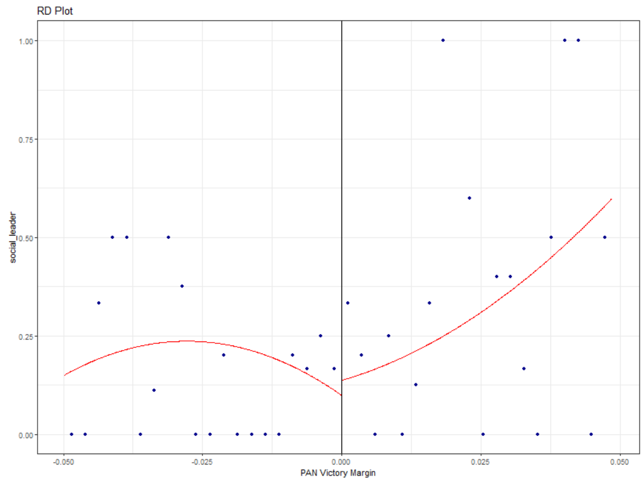
1.8 Part (h)

Figure 4 displays four scatterplots that examine the relationship between the PAN victory margin and the characteristics of winning candidates. These scatterplots allow for a comparison between PAN-affiliated winners and those from other parties. The analysis focuses on four key characteristics: candidates' age (Panel A), binary indicators for being a social leader and having a business background (Panels B and C, respectively), and years of education (Panel D). Each scatterplot is based on a specification with 20 bins and quadratic polynomial regressions.

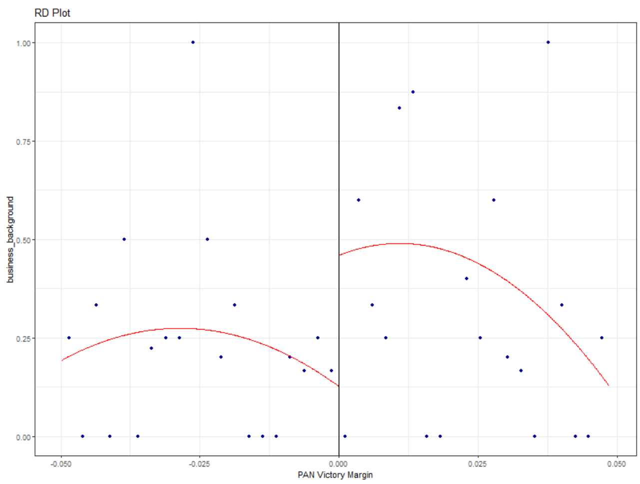
Figure 4: PAN Victory Margin and Candidate Characteristics



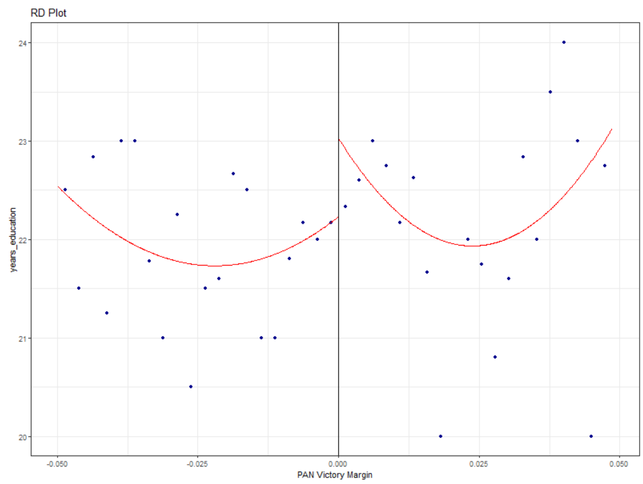
(a) Age



(b) Social Leadership



(c) Business Background



(d) Education

Panels A and B show little evidence of discontinuity at the electoral victory cutoff. On the other hand, Panels C and D indicate a discontinuous jump at the cutoff, suggesting that PAN winning candidates might be more likely to have a business background and have more years of education. Other characteristics not reported here that feature discontinuous jumps include campaign budget and years in local government.

Panels A and B reveal little evidence of a discontinuity at the electoral victory cutoff. In contrast, Panels C and D show a noticeable jump at the cutoff, suggesting that PAN-winning candidates are more likely to have a business background and possess more years of education.

Additional characteristics not reported here that also exhibit discontinuities include campaign budget and years in local government.

These findings suggest that the observed positive effect of a PAN mayor taking office on homicide rates may be driven by systematic differences in candidate characteristics between PAN and other parties rather than by PAN-specific policies or affiliation. This casts some doubt on the validity of Dell's results.

Does controlling for these (and other observable) characteristics eliminate estimation bias? Marshall argues that *compensating differentials* do not introduce bias if either of the following conditions is satisfied: (1) the characteristic of interest (PAN affiliation in Dell's study) does not affect electoral outcomes, or (2) the compensating differentials do not influence the outcome of interest (homicide rates in Dell's study). If neither condition holds, then there *must* exist a compensating differential that makes elections closely contested. While controlling for observable covariates removes them from the set of compensating differentials, unobservable covariates would still confound the analysis, leading to biased estimates.

To ensure unbiased estimation, one must control for all compensating differentials, reducing differences in electoral outcomes solely to idiosyncratic shocks. However, this ideal scenario is rarely achievable in observational studies. Hence, controlling for candidate characteristics in the main regression is unlikely to fully resolve the bias in Dell's study.

1.9 Part (i)

Table 4 presents the corrected PCRD estimators, given three different factors on the OLS estimators.

Table 4: Corrected PCRD

OLS Factor	Adjusted Value
0.5x	103.13
1x	105.77
1.5x	108.42

Accounting for observable candidates' characteristics slightly increases the estimated positive effect of PAN mayors on homicide rates. This suggests that the compensating differentials captured by these characteristics help reduce homicide rates. The finding is plausible: in close elections, PAN-winning candidates are often more educated and more likely to have prior experience in local government or business. These traits collectively reflect competence, which may contribute to lower crime rates. Consequently, the positive effect of PAN policies or party affiliation on homicide rates could be even greater, suggesting that Dell's estimates likely underestimate the true effect.

2 Question 2

2.1 Part (a)

Table 5 presents the results for the 2SLS and Wald estimators, computed using a residualized instrumental variable (IV).

Table 5: Wald and 2SLS Estimators

Estimator	Value
Wald Estimator	7.48
2SLS Estimator	7.48

As expected, the two estimators are identical, which reflects the mechanics of the 2SLS estimation procedure. Without covariates, the 2SLS estimator is given by the ratio of covariances:

$$\frac{\mathrm{cov}(Z, Y)}{\mathrm{cov}(Z, D)}, \tag{2.1}$$

And this ratio is precisely the Wald estimator. However, since the covariate vector X is included in the 2SLS specification, we must explicitly partial out its influence when computing the Wald estimator. Therefore, the residualized IV is used instead of the raw IV to ensure consistency with the regression framework.

2.2 Part (b)

One can identify the conditional Local Average Treatment Effect (LATE) using the Wald estimator, conditioned on the covariate vector X. Formally, the conditional LATE is expressed as:

$$\theta_{\mathrm{LATE}|X} = \frac{\mathbb{E}\left[Y_i \mid Z_i = 1, X_i\right] - \mathbb{E}\left[Y_i \mid Z_i = 0, X_i\right]}{\mathbb{E}\left[D_i \mid Z_i = 1, X_i\right] - \mathbb{E}\left[D_i \mid Z_i = 0, X_i\right]}. \tag{2.2}$$

$\theta_{\mathrm{LATE}|X}$ represents the average treatment effect for compliers within a specific covariate group X.

Table 6 presents the estimated conditional LATEs, groped by the covariate vectors. For comparison, it also provides the true conditional LATE values derived from the data-generating process (DGP). The comparison demonstrates that the conditional LATE estimator is unbiased, precisely estimating the true conditional LATE.

Table 6: Estimated and True Conditional LATE

x_1	x_2	Conditional LATE	True Conditional LATE
0	0	6.57	6
1	0	7.95	8
0	1	7.00	7
1	1	9.23	9

2.3 Part (c)

The equivalence between the convex weighted average of the conditional LATE, as derived in Angrist and Imbens (1995), and the standard 2SLS estimator is well-established. However, the two slightly differ in my calculations, potentially due to an error in one of the steps.

Table 7 presents the weights assigned to each conditional LATE. Higher weights are given to the covariate groups $(x_0, x_1) = (1, 0)$ and $(x_0, x_1) = (1, 1)$, where the variance of the conditional means is larger. A higher variance indicates greater differences in conditional expectations, reflecting a larger proportion of compliers. Consequently, the 2SLS estimator places greater weight on groups with more compliers. This outcome aligns with the purpose of the 2SLS estimator, which aims to estimate the average treatment effect for compliers.

Table 7: Weights

x_1	x_2	Weight
0	0	0.40
1	0	1.53
0	1	0.31
1	1	1.76

An important implication is that the 2SLS estimator would identify the unconditional LATE if all covariate groups had the same proportion of compliers. This condition would hold if the probability of being a complier were independent of X. However, in the specified DGP, this probability depends on x_1, leading to differing shares of compliers across covariate groups and, consequently, differing weights assigned by the 2SLS estimator.

2.4 Part (d)

No, the 2SLS estimator cannot be interpreted as a convex-weighted average of conditional LATEs when the interaction term is excluded from the set of controls. Let us break down this observation step by step.

The conditional LATE measures the average treatment effect among compliers, conditional on a fixed covariate vector X. This is equivalent to estimating the average effect of a change in Z on the outcome Y, specifically for compliers while holding X fixed. Achieving this requires isolating the effect of Z, which involves partialling out X.

In the given DGP, however, properly partialling out X necessitates including the interaction term $X_1 \times X_2$ as a control. The reason lies in the dependence of the probability distribution of Z on X_1, X_2, and $X_1 \times X_2$. Without controlling for the interaction term, the relationship between Z and X cannot be fully accounted for.

To illustrate, in a linear regression of Z on X_1, X_2, and $X_1 \times X_2$, the coefficient on the interaction term is 0.4 (or $0.8 - 0.4$). Excluding the interaction term would thus result in an incomplete specification of the covariate structure. Consequently, this omission would prevent holding X fixed, making it impossible to estimate the conditional LATE or any convex-weighted average of them accurately.

2.5 Part (e)

Observing (Y_i, D_i, Z_i, X_i), the unconditional LATE can be identified as a weighted average of the conditional LATEs, where the weights correspond to the relative sizes of the covariate groups.

Formally, by the Law of Iterated Expectations (LIE), we have:

$$\begin{aligned} \mathbb{E}\left[Y_i(1) \mid G = \mathrm{CP}\right] &= \mathbb{E}\left[\mathbb{E}\left[Y_i(1) \mid X_i = x, G = \mathrm{CP}\right]\right] \\ &= \sum_{x \in X} \mathbb{E}\left[Y_i(1) \mid X_i = x, G = \mathrm{CP}\right] \mathbb{P}(X_i = x) \end{aligned} \tag{2.3}$$

The same relationship holds for $\mathbb{E}\left[Y_i(0) \mid G = \mathrm{CP}\right]$, leading to:

$$\begin{aligned} \mathbb{E}\left[Y_i(1) - Y_i(0) \mid G\right] &= \sum_{x \in X} \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = x, G = \mathrm{CP}\right] \mathbb{P}(X_i = x) \\ &= \frac{1}{4} \sum_{x \in X} \mathbb{E}\left[Y_i(1) - Y_i(0) \mid X_i = x, G = \mathrm{CP}\right] \end{aligned} \tag{2.4}$$

Since the four covariate groups are of equal size, the unconditional LATE simplifies to the arithmetic mean of the conditional LATEs.

Table 8 presents the conditional LATE estimated from the data and compares it to the true unconditional LATE.

Table 8: Comparing Unconditional LATE

Unconditional LATE	True Unconditional LATE
7.69	7.50