Previous RDD Papers

Rafael Alcantara*

January 31, 2024

Contents

1 Previous RDD papers

It is worth noting that most of the papers write the DGPs in terms of $E[Y_1|X] = \mu_1(x)$ and $E[Y_0|X] = \mu_0(x)$. While this makes sense from the perspective that most methods consist of approximating each function and taking their differences at $x = c$, I rewrite the DGPs in terms of a prognostic and treatment function, $\mu(x)$ and $\tau(x)$ respectively, as this makes more sense from a BCF perspective.

*rafaelca10@al.insper.edu.br

1.1 Without W

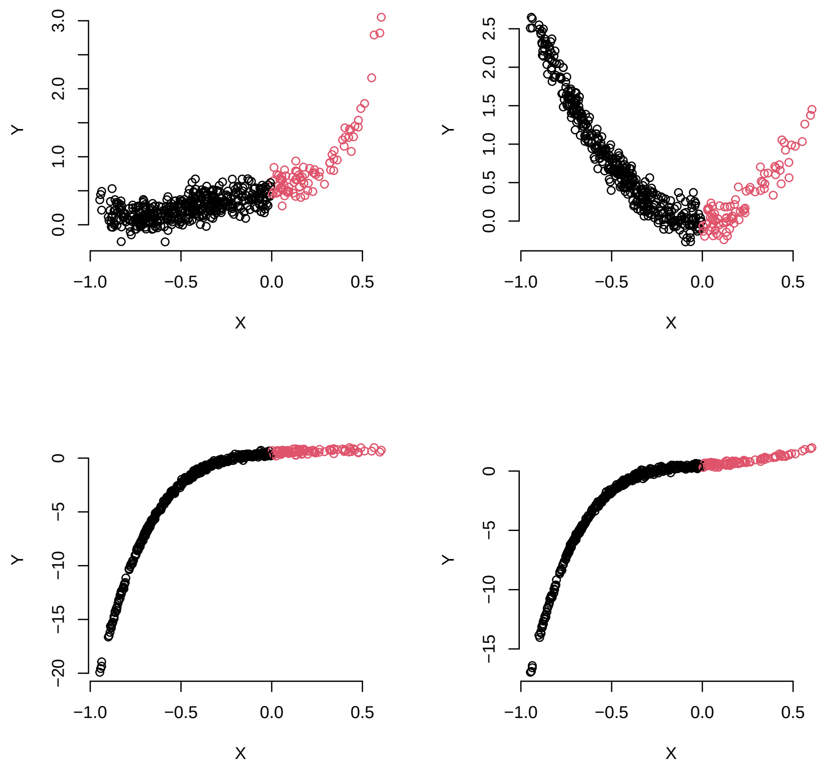
1.1.1 IK2012

Imbens and Kalyanaraman (2012) consider 4 different DGPs. For every DGP, they consider $n = 500$ and sample X from $X \sim 2\mathcal{B}(2,4) - 1$, $Z = \mathbf{1}(X \geq 0)$ and $\varepsilon \sim \mathcal{N}(0, 0.1295^2)$. For each DGP, the prognostic and treatment effect functions are defined as:

1. Based on Lee (2008)[1]:
    - $\mu(x) = 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5$
    - $\tau(x) = 0.04 - 0.43x - 10.18x^2 - 12.22x^3 - 12.53x^4 - 3.77x^5$
2. Quadratic on X:
    - $\mu(x) = 3x^2$
    - $\tau(x) = x^2$
3. Constant ATE:
    - $\mu(x) = 0.42 + 0.84x - 3x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5$
    - $\tau(x) = 0.1$
4. Constant ATE and curvature at the threshold is zero at both sides:
    - $\mu(x) = 0.42 + 0.84x + 7.99x^3 - 9.01x^4 + 3.56x^5$
    - $\tau(x) = 0.1$

This is what each DGP looks like:

[1] Every simulation based on "Lee data" refers to this



1.1.2 CCT2014

Calonico et al. (2014) repeat n, X and ε from Imbens and Kalyanaraman (2012). Their regression functions follow[2]:

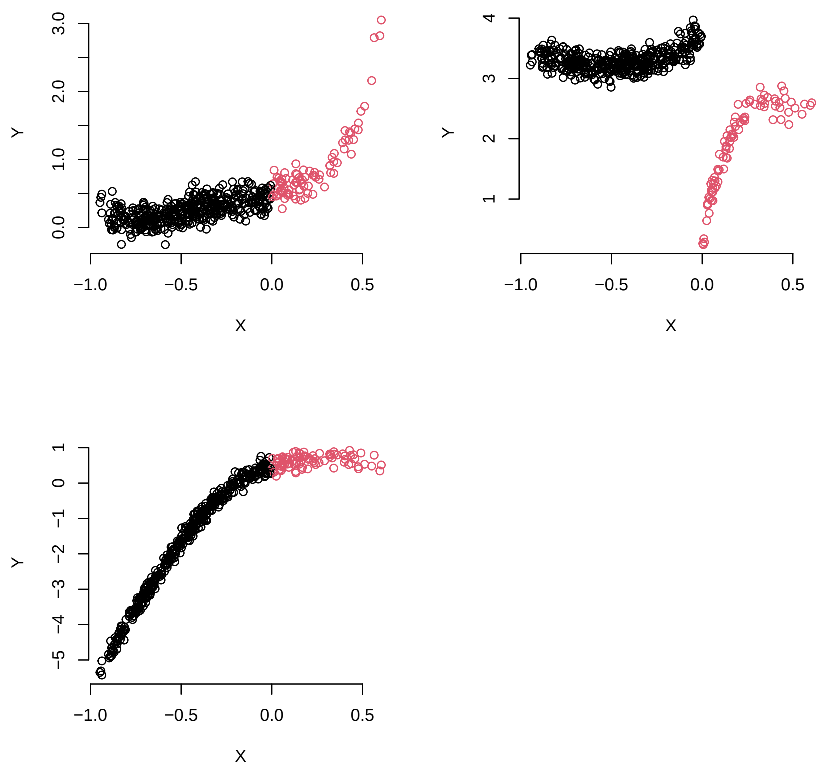
1. Lee data
2. Based on Ludwig and Miller (2007)[3]:
   - $\mu(x) = 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5$
   - $\tau(x) = -3.45 + 16.19x - 58.09x^2 + 72.85x^3 - 45.25x^4 + 9.8x^5$
3. Variation of Lee data

[2]Simulations detailed in supplemental material

[3]Every simulation based on "LM data" refers to this

- $\mu(x) = 0.48 + 1.27x - 3.59x^2 + 14.147x^3 + 23.694x^4 + 10.995x^5$
- $\tau(x) = 0.04 - 0.43x + 3.29x^2 - 16.544x^3 - 24.595x^4 - 7.435x^5$

This is what each DGP looks like:



1.1.3 BRBM2019

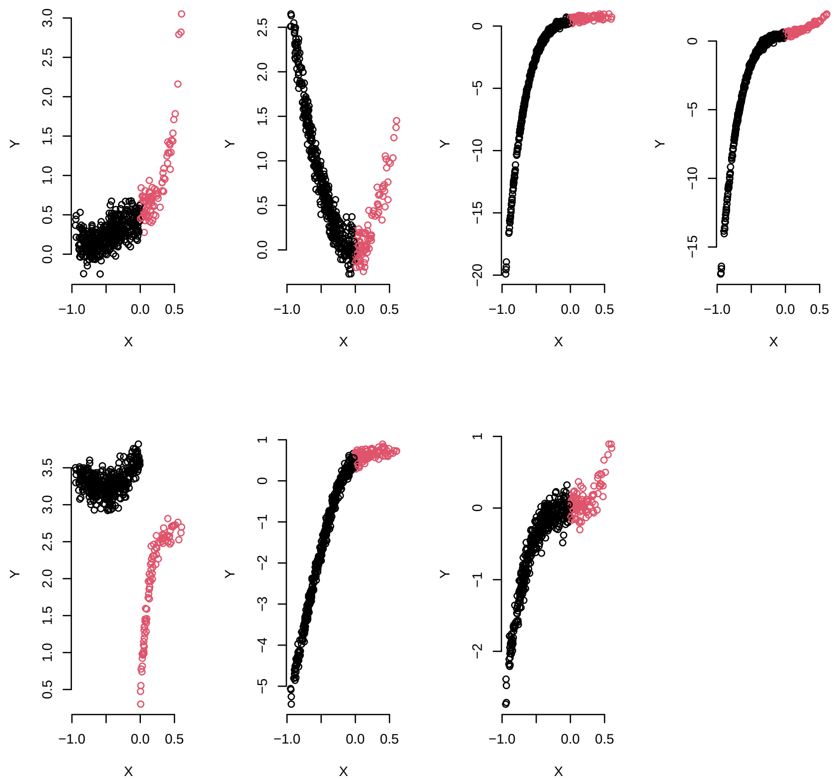
Branson et al. (2019) fit two different GP regressions to treated and untreated units. They generate 1000 samples of 7 different DGPs. In each case, n, X and ε are the same as Imbens and Kalyanaraman (2012). The regression functions are the same as Imbens and Kalyanaraman (2012) and Calonico et al. (2014) plus one additional setup:

1. Cubic functions:

    - $\mu(x) = 3x^3$

- $\tau(x) = x^3$

The DGPs look like this:



1.1.4 CCF2020

Calonico et al. (2020) consider a variation of the LM data with a different cutoff and higher error variance, but same parameters for μ and τ.

1.2 With W

1.2.1 CGS2023

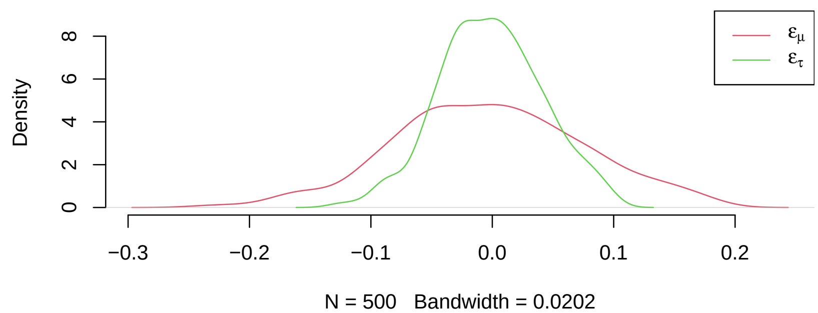
Chib et al. (2023) analyzes two DGPs. First, the classic Lee data with t-distributed erros instead of Gaussian errors. They also consider a setup with nonparametric errors as follows. μ, τ are an extension of the Lee data DGP that also includes W but in such a way that there still are no heterogeneous effects. For this DGP they also propose a more intricate error structure. The DGP is:

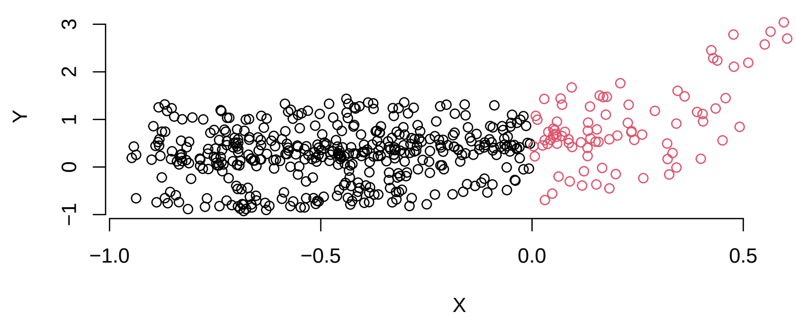
$$\begin{aligned}
\mu(x,w) &= 0.48 + 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 + h(w) + \varepsilon_\mu \\
\tau(x,w) &= 0.04 - 0.43x - 10.18x^2 - 12.22x^3 - 12.53x^4 - 3.77x^5 + \varepsilon_\tau \\
h(w) &= \frac{\sin(\pi w/2)}{1 + w^2(\text{sign}(w)+1)} \\
w &\sim U(-\pi, \pi) \\
\varepsilon_\mu &= \varepsilon_0 \\
\varepsilon_\tau &= \varepsilon_1 - \varepsilon_0.
\end{aligned} \tag{1}$$

The errors follow:

$$\begin{aligned}
F(\varepsilon_0) &= \sigma_0 F(\varepsilon) \\
F(\varepsilon_1) &= \sigma_1 F(\varepsilon) \\
F(\varepsilon) &= \frac{1}{3} \times \Phi(\varepsilon + 2.5) + \frac{1}{3} \times \Phi(\varepsilon) + \frac{1}{3} \times \Phi(\varepsilon - 2.5) \\
\sigma_0 &= 0.1295 \\
\sigma_1 &= 0.2.
\end{aligned} \tag{2}$$

This is what that second DGP looks like:





1.2.2 CCT2019

Calonico et al. (2019)[4] consider 4 variations of the Lee data by adding a pre-determined binary covariate (previous democratic share). Each model includes this covariate differently. For DGP 1, the covariate is irrelevant and the DGP is the same as the classic Lee data DGP. For the others, the covariate is relevant. For all three, X and W follow:

[4] `https://github.com/rdpackages-replication/CCFT_2019_RESTAT/blob/master/CCFT_2019_RESTAT_simuls.do`

$$
\begin{aligned}
w_r &= 0.49 + (1.06 - 0.45)x + (5.74 - 5.51)x^2 \\
&\quad + (17.14 - 20.60)x^3 + (19.75 - 13.32)x^4 + (7.47 - 10.95)x^5 + \varepsilon_w \\
w_l &= 0.49 + 1.06x + 5.74x^2 + 17.14x^3 + 19.75x^4 + 7.47x^5 + \varepsilon_w \\
y_r &= 0.38 + 0.63x - 2.85x^2 + 8.43x^3 - 10.24x^4 + 4.32x^5 + 0.28w_r + \varepsilon_y \\
y_l &= 0.36 + 0.96x + 5.47x^2 + 15.28x^3 + 15.87x^4 + 5.14x^5 + 0.22w_l + \varepsilon_y \\
\sigma_y &= 0.1295 \\
\sigma_w &= 0.13537.
\end{aligned}
\tag{3}
$$

This implies:

$$
\begin{aligned}
y &= \mu(x, \varepsilon_w) + \tau(x, \varepsilon_w)z + \varepsilon_y \\
\mu(x, \varepsilon_w) &= 0.47 + 1.19x + 6.73x^2 + 19.05x^3 + 20.21x^4 + 6.78x^5 + 0.22\varepsilon_w \\
\tau(x, \varepsilon_w) &= 0.049 - 0.36x - 0.87x^2 - 10.35x^3 - 27.85x^4 - 2.78x^5 + 0.06\varepsilon_w.
\end{aligned}
\tag{4}
$$

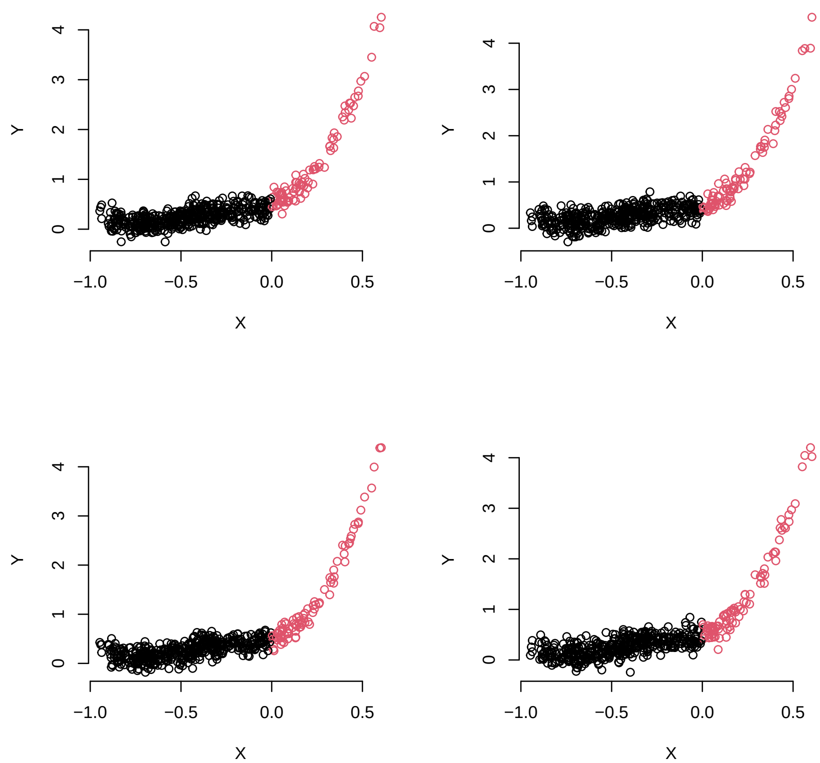
What differs from one DGP to the other is the joint distribution of $\varepsilon_y, \varepsilon_w$:

1. DGP 2: $\sigma_{yw} = 0.2692\sigma_y\sigma_w$
2. DGP 3: $\sigma_{yw} = 0$
3. DGP 4: $\sigma_{yw} = 0.5384\sigma_y\sigma_w$

In each case, $\varepsilon_y, \varepsilon_w$ are sampled jointly from a Gaussian with covariance:

$$
\Sigma = \begin{pmatrix} \sigma_y^2 & \sigma_{yw} \\ \sigma_{yw} & \sigma_w^2 \end{pmatrix}. \tag{5}
$$

This is what the data looks like:

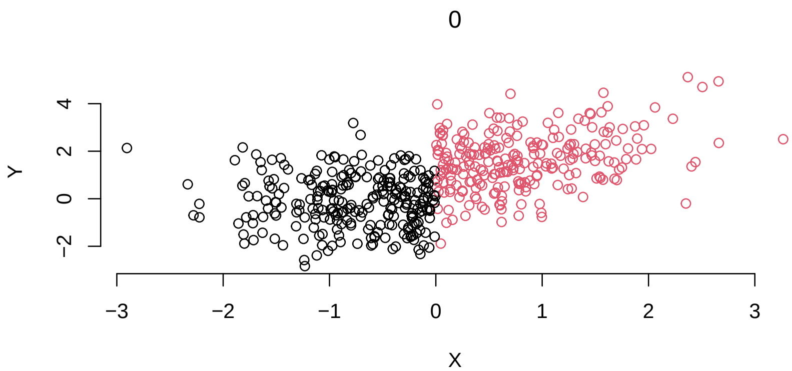


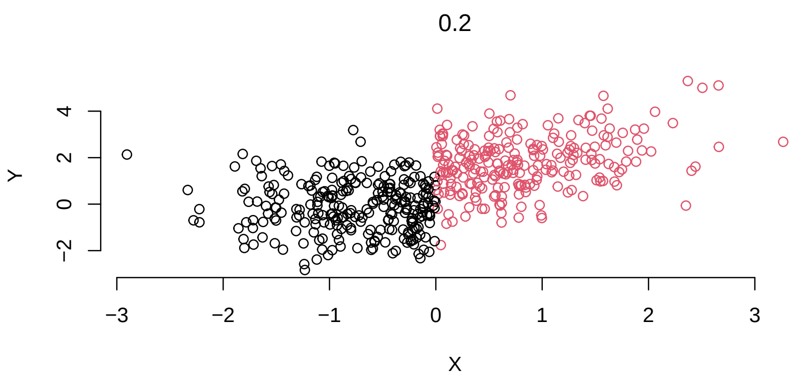
1.2.3 FH2019

Frölich and Huber (2019) analyze a setup where the additional covariates might be discontinuous at the cutoff. The DGP follows:

$$
\begin{aligned}
X, U_1, U_2, U_3 &\sim \mathcal{N}(0,1) \\
Z &= \mathbf{1}(X \geq 0) \\
W_1 &= \alpha Z + 0.5 U_1 \\
W_2 &= \alpha Z + 0.5 U_2 \\
\mu(x,w) &= \beta(w_1 + w_2) + \frac{\beta}{2}(w_1^2 + w_2^2) + 0.5x + 0.25x^2 \qquad (6) \\
\tau(x) &= 1 - 0.25x \\
y &= \mu(x,w) + \tau(x)z + u_3 \\
\alpha &\in \{0, 0.2\} \\
\beta &= 0.4.
\end{aligned}
$$

This is what that DGP looks like for both values of α:



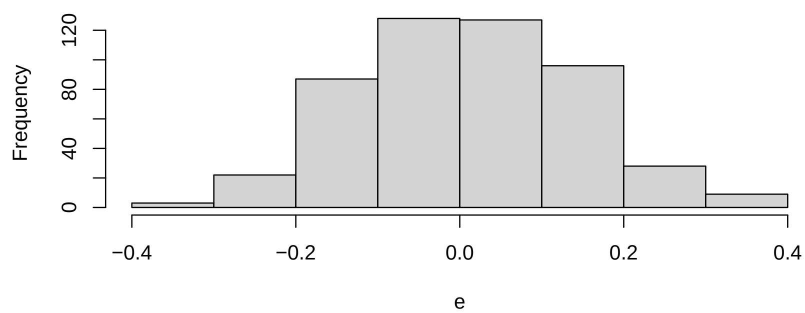


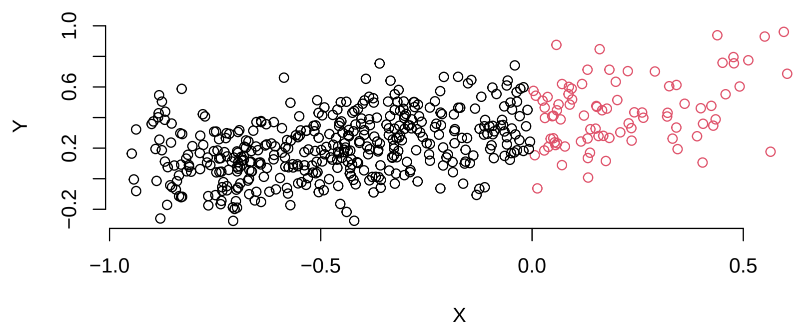
1.2.4 KR2023

Kreiß and Rothe (2021) also consider a polynomial DGP but W, ε are sampled jointly.

$$
\begin{aligned}
p &= 200 \\
X &\sim 2\mathcal{B}(2,4) - 1 \\
Z &= \mathbf{1}(X \geq 0) \\
(\varepsilon, W^T)^T &\sim \mathcal{N}(\mathbf{0}, \Sigma) \\
\Sigma &= \begin{pmatrix} \sigma_\varepsilon^2 & \nu^T \\ \nu & \sigma_W^2 I_p \end{pmatrix} \\
\mu(X,W) &= 0.36 + 0.96X + 5.47X^2 + 15.28X^3 + 15.87X^4 + 5.14X^5 + 0.22W^T\alpha \\
\tau(X,W) &= 0.02 - 0.34X - 8.31X^2 - 6.86X^3 - 26.11X^4 - 0.83X^5 + 0.06W^T\alpha \\
\sigma_\varepsilon &= 0.1295 \\
\sigma_W &= 0.1353 \\
\nu &\in \mathcal{R}^{200}, \quad \nu_k = \frac{0.8\sqrt{6}\sigma_\varepsilon^2}{\pi k} \\
\alpha &\in \mathcal{R}^{200}, \quad \alpha_k = \frac{2}{k^2} \\
Y &= \mu(X,W) + \tau(X,W)Z + \varepsilon.
\end{aligned}
\tag{7}
$$

This is what the errors and data looks like:





1.2.5 Reguly2021

Reguly (2021) fits a CART model to additional covariates and performs node-level polynomial regressions on X. He takes 1000 samples of each DGP and considers $n \in \{1000, 5000, 10000\}$. Importantly, he samples (X, W) only once so that the variation across MCMC samples comes only from ε and his treatment effect function includes only W and not X.

1. DGP 1:

    - $X \sim U(-1, 1)$
    - $W_1, W_2 \sim \text{Bernoulli}(0.5)$
    - $\mu(x) = 2x$
    - $\tau(w_1) = 2w_1 - 1$
    - $\varepsilon \sim \mathcal{N}(0, 1)$

2. DGP 2:

    - $X \sim U(-1,1)$
    - $W_1, W_2 \sim \text{Bernoulli}(0.5)$
    - $W_3, W_4 \sim U(-5,5)$
    - $\mu(x,w) = (4w_2 - 2)x$
    - $\tau(w_3) = 2w_3$
    - $\varepsilon \sim \mathcal{N}(0,1)$

3. DGP 3 (variation of Lee data):

    - $X \sim 2\mathcal{B}(2,4) - 1$
    - $W_1 \sim \text{Bernoulli}(0.5)$
    - $\mu(x,w_1) = 0.48 + w_1(1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5) + (1-w_1)(2.35x + 8.18x^2 + 22.21x^3 + 24.14x^4 + 8.33x^5)$
    - $\tau(x,w_1) = w_1(0.02 - 0.43x - 10.18x^2 - 12.22x^3 - 12.53x^4 - 3.77x^5) + (1-w_1)(0.07 - 1.14x - 11.08x^2 - 15.22x^3 - 14.13x^4 - 3.77x^5)$
    - $\varepsilon \sim \mathcal{N}(0,0.05)$

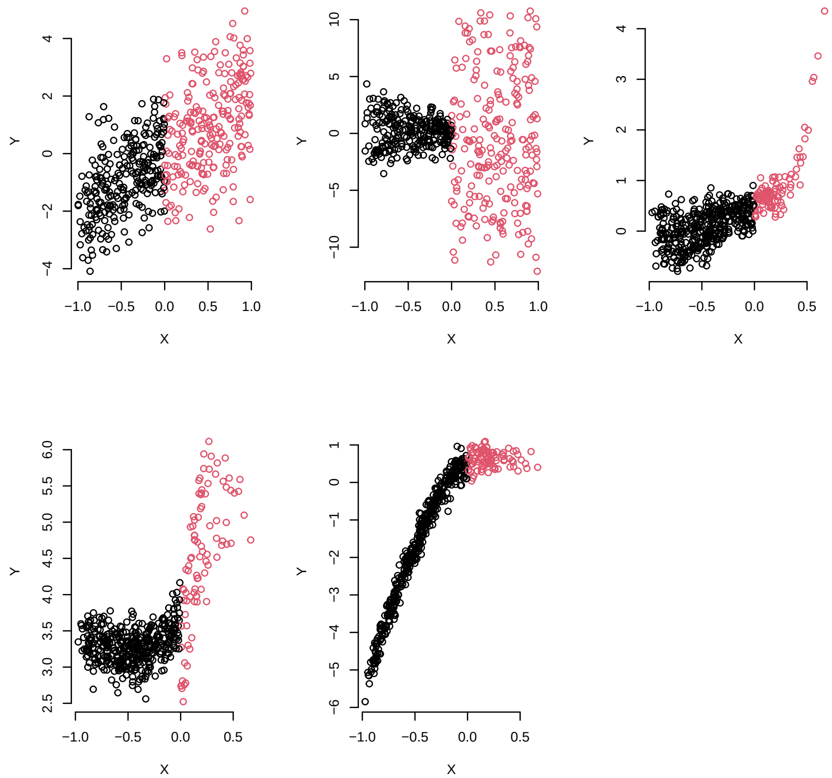
4. DGP 4 (variation of LM data)

    - $X \sim 2\mathcal{B}(2,4) - 1$
    - $W_1 \sim U(5,9)$
    - $\mu(x) = 3.71 + 2.30x + 3.28x^2 + 1.45x^3 + 0.23x^4 + 0.03x^5$
    - $\tau(w_1) = -0.45 + 16.19x - 58.09x^2 + 72.85x^3 - 45.25x^4 + 9.8x^5 - w_1$
    - $\varepsilon \sim \mathcal{N}(0,0.05)$

5. DGP 5: DGP 3 of Calonico et al. (2014)

This is what the DGPs look like[5]:

[5] DGP 3 and 4 look different from the paper. The former looks similar when using smaller sample sizes and the conversion from μ_0, μ_1 to μ, τ is correct. The latter had a typo in the paper so the one written here is a guess. I could not find the simulation codes for the paper to check this



Two things to note about this exercise. In DGP 1, we have $\tau(w = 1) = 1$, $\tau(w = 0) = -1$ and $E[\tau] = 0$. It might be interesting to include a case with zero ATE but some non-zero CATE to be estimated in our setup. In DGP 2, a similar thing happens but now with continuous CATE. The way τ is constructed is such that the continuous W introduces huge variability in it, this is something to keep in my mind when writing these simulations.

References

Branson, Z., Rischard, M., Bornn, L., and Miratrix, L. W. (2019). A nonparametric bayesian methodology for regression discontinuity designs. *Journal of Statistical Planning and Inference*, 202:14–30.

Calonico, S., Cattaneo, M. D., and Farrell, M. H. (2020). Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econometrics Journal*, 23(2):192–210.

Calonico, S., Cattaneo, M. D., Farrell, M. H., and Titiunik, R. (2019). Regression discontinuity designs using covariates. *Review of Economics and Statistics*, 101(3):442–451.

Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.

Chib, S., Greenberg, E., and Simoni, A. (2023). Nonparametric bayes analysis of the sharp and fuzzy regression discontinuity designs. *Econometric Theory*, 39(3):481–533.

Frölich, M. and Huber, M. (2019). Including covariates in the regression discontinuity design. *Journal of Business & Economic Statistics*, 37(4):736–748.

Imbens, G. and Kalyanaraman, K. (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of economic studies*, 79(3):933–959.

Kreiß, A. and Rothe, C. (2021). Inference in regression discontinuity designs with high-dimensional covariates. *arXiv preprint arXiv:2110.13725*.

Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics*, 142(2):675–697.

Ludwig, J. and Miller, D. L. (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly journal of economics*, 122(1):159–208.

Reguly, A. (2021). Heterogeneous treatment effects in regression discontinuity designs. *arXiv preprint arXiv:2106.11640*.