

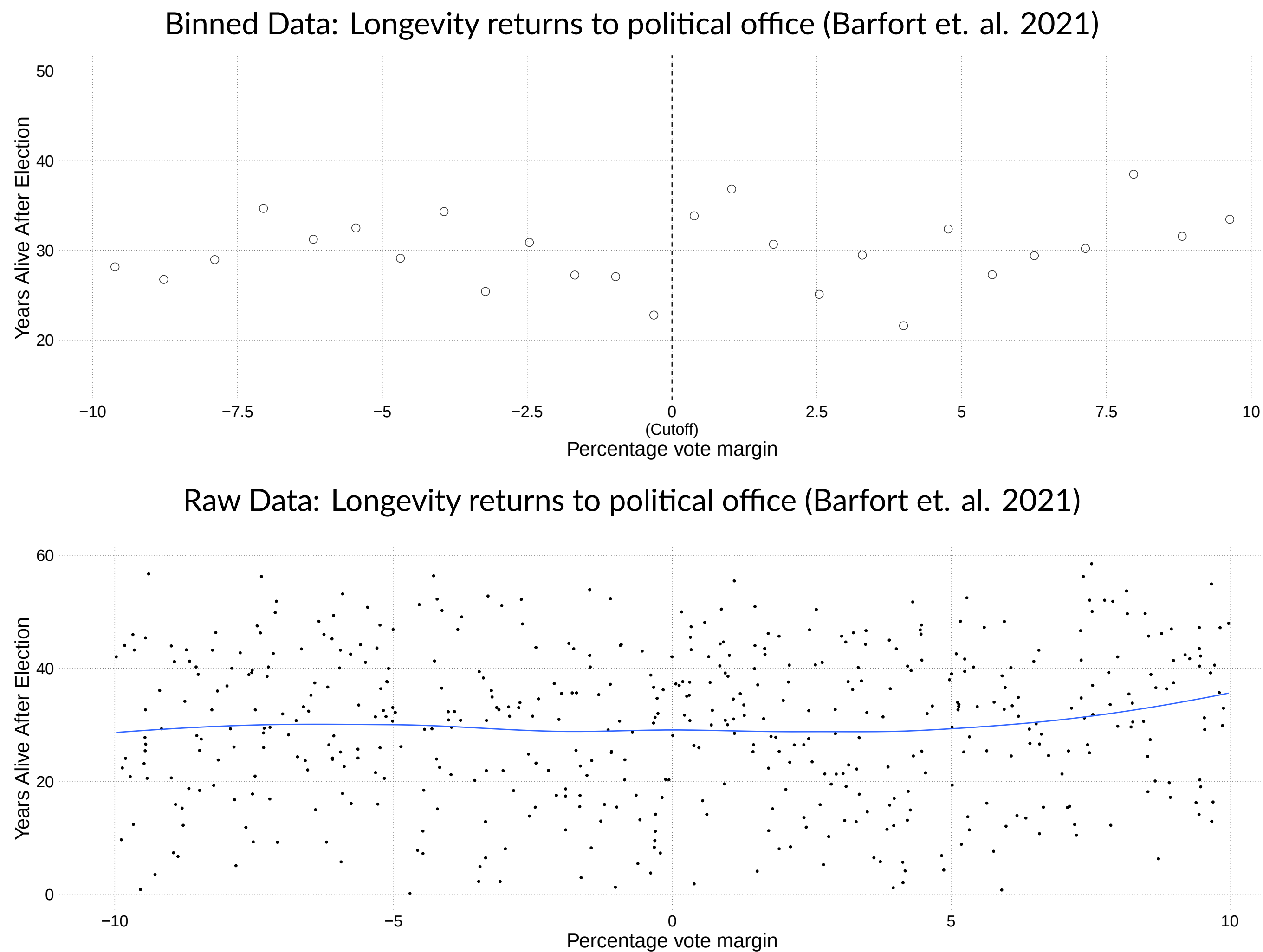


## Abstract

Regression Discontinuity Designs (RDD) are widely popular in the social sciences as they can recover local average treatment effects under a few assumptions. However, concerns about **natural variation** and **noise** in the data have led to skepticism of their results. We propose using **Change Point Analysis (CPA)** as a tool within the RDD framework to explore variation in the running variable and increase the credibility of RDD findings. By incorporating CPA, researchers may efficiently identify and evaluate discontinuities in the raw data without the need for pre-specification. The ability of CPA to correctly identify the theory-driven discontinuity should lend credibility to the design, while finding other discontinuities may signal noisy data or other threats to identification. We demonstrate the robustness of CPA in detecting meaningful breakpoints through empirical applications and simulations.

## Natural Variation and Unclear Discontinuities

Social science data contains natural variation and noise which makes identifying discontinuities unclear. Given this natural variation, the logical rival hypothesis should first be “an uninterrupted curvilinear regression” (Thistlethwaite & Campbell 1960).



- Winning a close election adds 5 - 10 years to politician's life

## Change Point Analysis

**Change point analysis (CPA)** is a method used to identify areas in the data where the mean or variance change. In context of the running variable in a regression discontinuity, CPA can identify points in the data where a discontinuity is likely to have occurred.

## Reasons to use Bayesian Change Point Analysis

- Detect changes and discontinuity in the running variable **without pre-specification**
- Probabilistic framework** to calculate the posterior mean, variance, and the probability of a discontinuity at each point
- Flexibility** to handle various types of data

## Algorithm for Bayesian Change Point Analysis

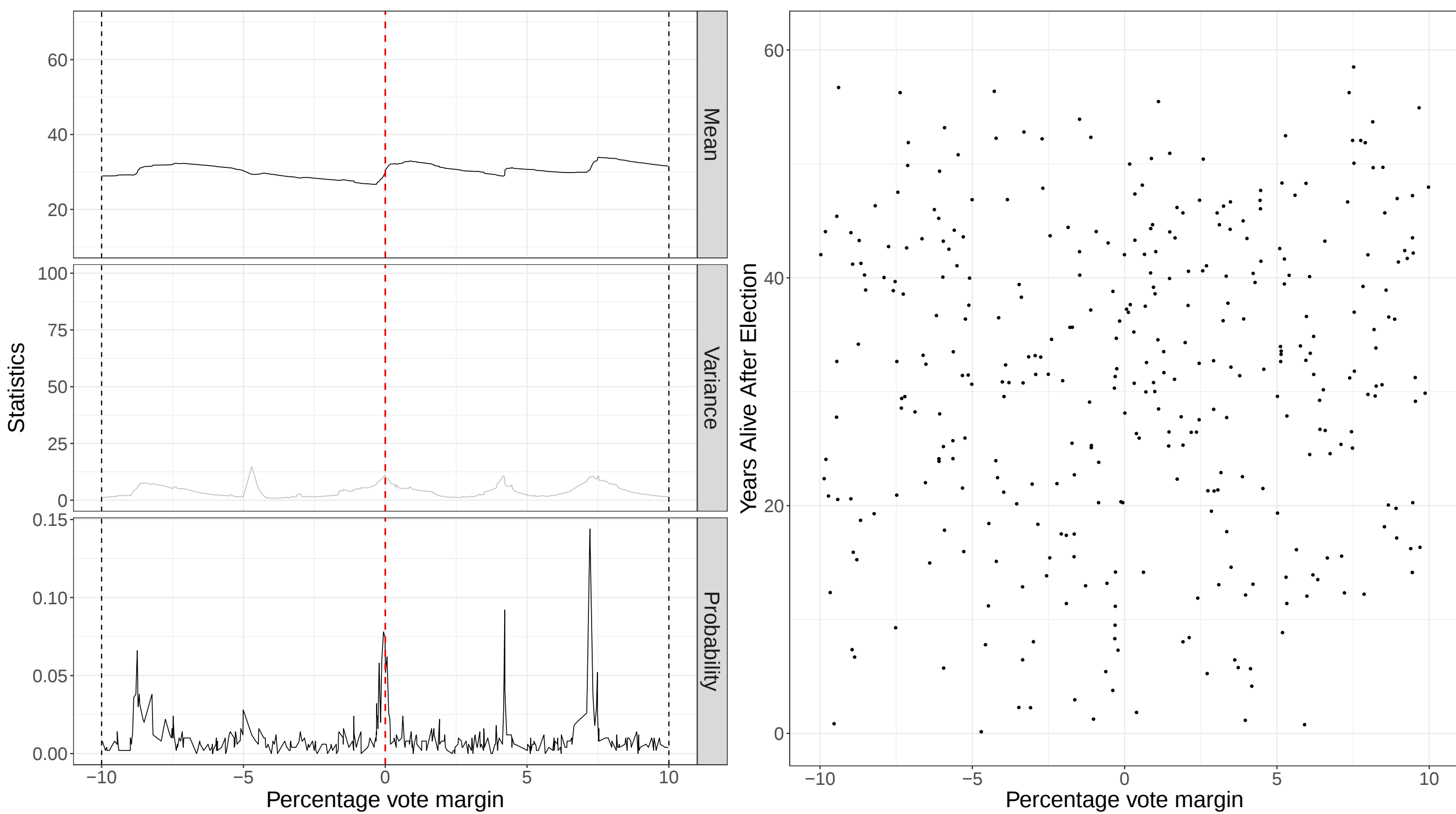
Estimates the likelihood of a change occurring at a specific point ( $U_i$ ) given the data  $X$ , and other estimated changes in the data ( $U_j$ ).

$$\frac{P(U_i = 1|X, U_j, j \neq i)}{P(U_i = 0|X, U_j, j \neq i)} = \frac{\int_0^\gamma p^b(1-p)^{n-b-1} dp \left[ \int_w^{w_0} \frac{w^{b/2}}{(W_1+B_0w)^{(n-1)/2}} dw \right]}{\int_0^\gamma p^{b-1}(1-p)^{n-b} dp \left[ \int_w^{w_0} \frac{w^{(b-1)/2}}{(W_0+B_1w)^{(n-1)/2}} dw \right]}$$

where  $U_i$  represents a change existing at location  $i$ ,  $X$  represents the observed data, and  $b$  are blocks/regimes in which location  $i$  falls within. The right side of equation first calculates the likelihood of blocks ( $b$ ) of locations ( $i$ ) changing to a new level. The second integral estimates the fit of having a change occur at the specified spots given the within ( $W$ ) and between ( $B$ ) sum of squares and a weighted ratio ( $w$ ) of variances in the data and a prior ( $\frac{\sigma^2}{\sigma^2 + \sigma_0^2}$ ).

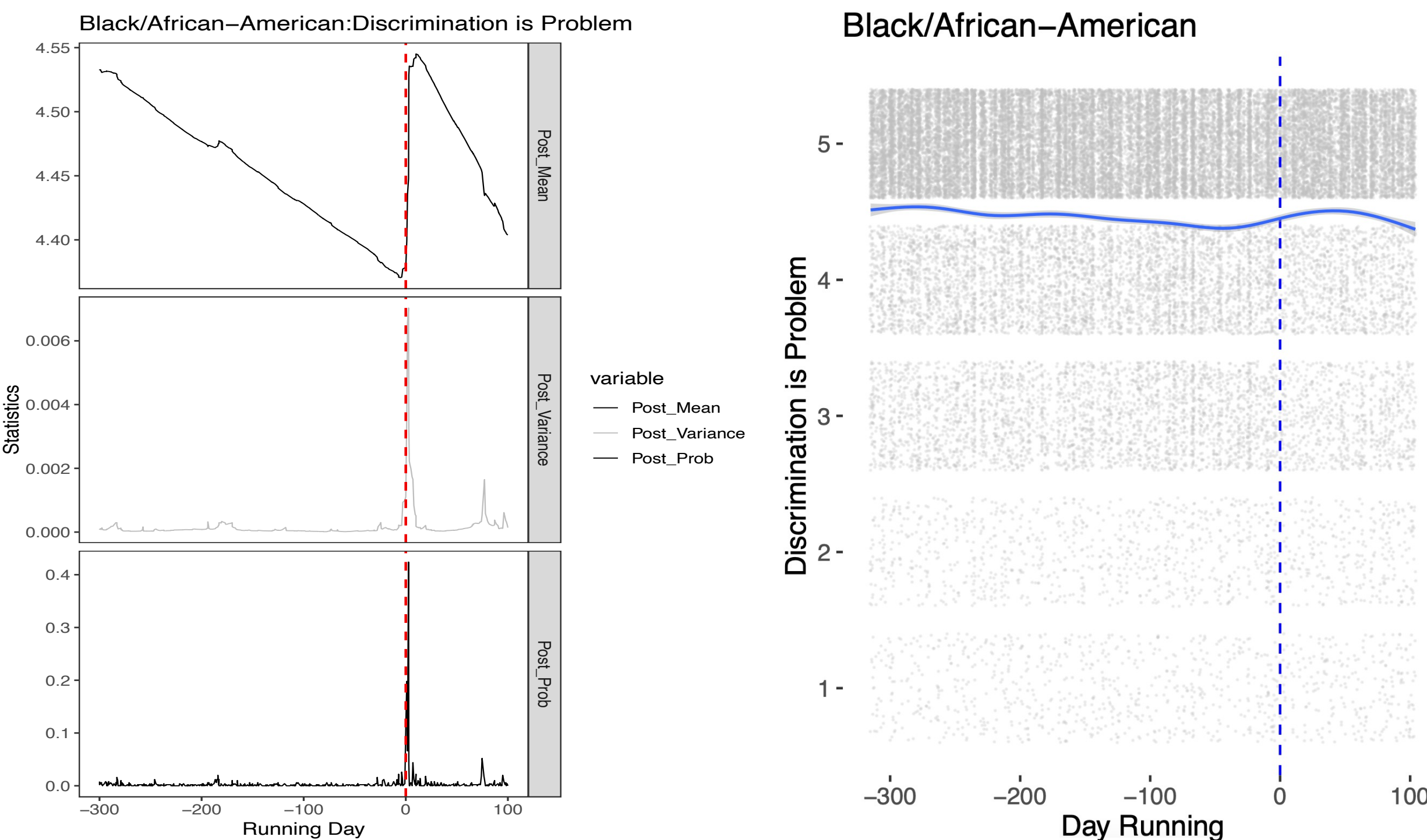
Outcomes: Posterior Mean, Posterior Variance, Posterior Probability of a Change Occurring

## Application 1: Barfort et. al. (2021)



- Evidence of posterior mean changes at theoretically-derived discontinuity
- Evidence of mean changes at other points along the running variable
- Variance change throughout the running variable
- Multiple discontinuities throughout
- Noisy Data

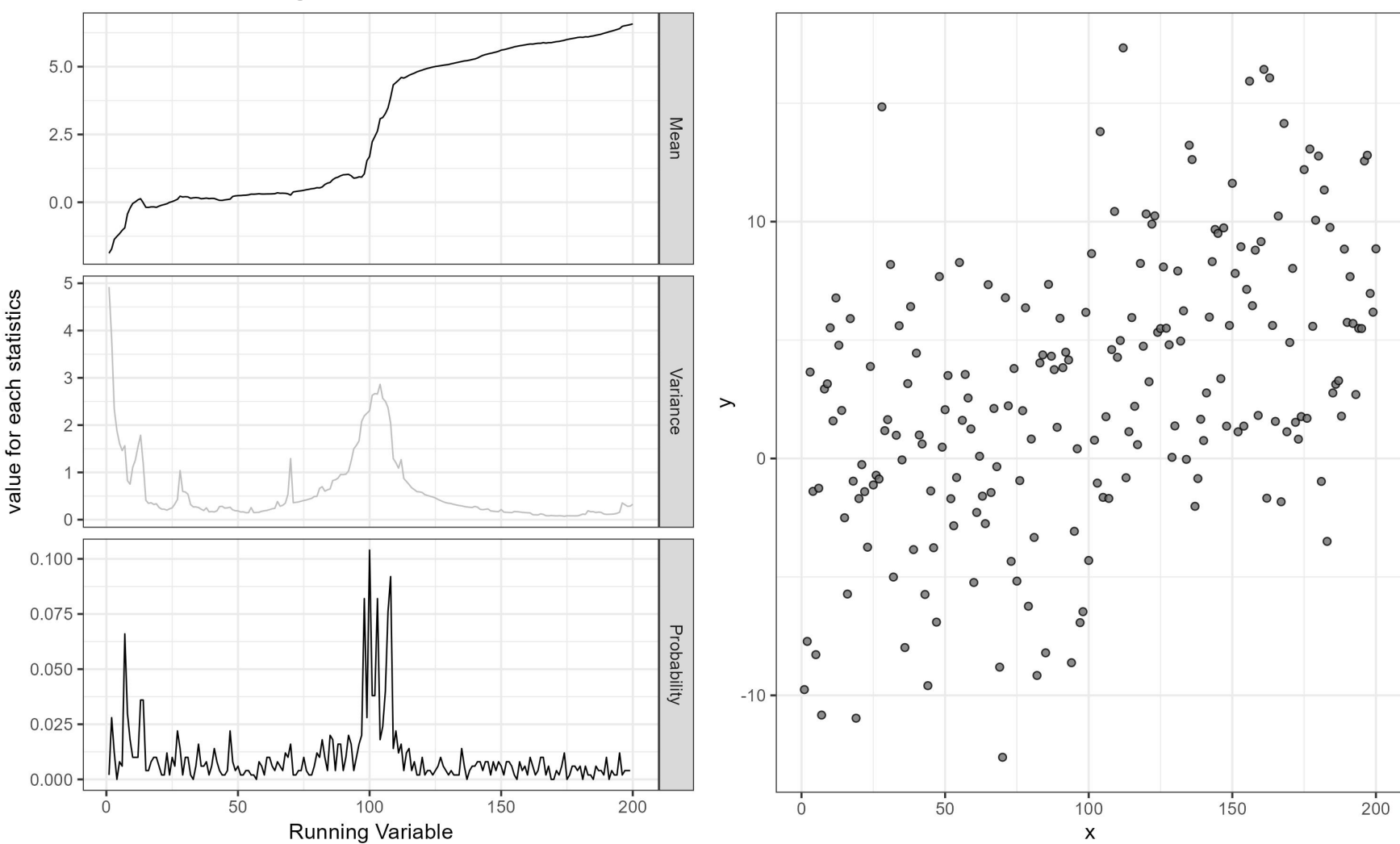
## Application 2: Opinion-Mobilizing Effect of George Floyd's Murder



- Reny and Newman 2021, Black Subgroup
- Clear evidence of change in posterior mean
- Pre-trend of decline before event
- Drastic increase in variance around event
- Variance is constant except the event
- Evidence that the discontinuity exists at the theoretically-derived location with little evidence elsewhere

## Illustrative Simulation: $\mu_{right} - \mu_{left} = 5, \sigma^2 = 5$

5 Unit Change with  $\sigma^2$  of 5 on each side of the discontinuity



## Steps to More Credible RDD Analyses

- Step 1. Visualize the Raw Data.** Plot the raw data with a curvilinear line without bins. Visually explore the running variable and set expectations.
- Step 2. Conduct a Change Point Analysis.** Identify whether a change has occurred. If so, does it match with where your theory would suggest the discontinuity lie? Identify whether there are mean, variance, or trend differences. Discuss the probability a change occurred. Are there other discontinuities?
- Step 3. Proceed with Regression Discontinuity.** Interpret results according to the exploratory evidence found in Steps 1 & 2.