SciencesPo
DEPARTMENT OF ECONOMICS

ScPoEconometrics

Differences-in-Differences and Regression Discontinuity Design

Florian Oswald, Gustave Kenedi and Pierre Villedieu
SciencesPo Paris
2020-04-21

Recap from last week

- Applied inference tools to regression analysis
- *Standard error* of regression coefficients
- *Statistical significance* of regression coefficients

Recap from last week

- Applied inference tools to regression analysis
- *Standard error* of regression coefficients
- *Statistical significance* of regression coefficients

Today: Program Evaluation

Differences-in-differences

- Exploits changes in policy over time
- Find the appropriate control group
- *Empircal application:* impact of the ***minimum wage*** on ***employment***

Recap from last week

- Applied inference tools to regression analysis
- *Standard error* of regression coefficients
- *Statistical significance* of regression coefficients

Today: Program Evaluation

Differences-in-differences

- Exploits changes in policy over time
- Find the appropriate control group
- *Empircal application:* impact of the ***minimum wage*** on ***employment***

Regression Discontinuity Design

- Life is full of random rules
- Exploits knowledge of assignment rule
- *Empirical application:* effect of ***alchol consumption*** on ***mortality***

Program Evaluation

- Multiple regression often does not provide causal estimates because of ***selection on unobservables***.

Program Evaluation

- Multiple regression often does not provide causal estimates because of ***selection on unobservables***.

- RCTs are one way to solve this problem but they are often impossible to do.

Program Evaluation

- Multiple regression often does not provide causal estimates because of ***selection on unobservables***.
- RCTs are one way to solve this problem but they are often impossible to do.
- Four main causal evaluation methods used in economics:
    - ***instrumental variables (IV)***,
    - ***propensity-score matching***,
    - ***differences-in-differences (DiD)***, and
    - ***regression discontinuity designs (RDD)***.

Program Evaluation

- Multiple regression often does not provide causal estimates because of ***selection on unobservables***.
- RCTs are one way to solve this problem but they are often impossible to do.
- Four main causal evaluation methods used in economics:
    - ***instrumental variables (IV)***,
    - ***propensity-score matching***,
    - ***differences-in-differences (DiD)***, and
    - ***regression discontinuity designs (RDD)***.
- These methods are used to identify **causal relationships** between treatments and outcomes.

Program Evaluation

- Multiple regression often does not provide causal estimates because of ***selection on unobservables***.
- RCTs are one way to solve this problem but they are often impossible to do.
- Four main causal evaluation methods used in economics:
  - ***instrumental variables (IV)***,
  - ***propensity-score matching***,
  - ***differences-in-differences (DiD)***, and
  - ***regression discontinuity designs (RDD)***.
- These methods are used to identify **causal relationships** between treatments and outcomes.
- In this lecture, we will cover two popular and rigorous program evaluation methods:
  1. **differences-in-differences**,
  2. **regression discontinuity designs**.

Differences-in-Differences

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

DiD Requirements:

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

DiD Requirements:

- 2 time periods: before and after treatment.

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

DiD Requirements:

- 2 time periods: before and after treatment.
- 2 groups:

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

DiD Requirements:

- 2 time periods: before and after treatment.
- 2 groups:
  - *control group:* never receives treatment,

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

DiD Requirements:

- 2 time periods: before and after treatment.
- 2 groups:
  - *control group:* never receives treatment,
  - *treatment group:* initially untreated and then fully treated.

Differences-in-Differences (DiD)

- Usual starting point: subjects are not randomly allocated to treatment ⚠️

DiD Requirements:

- 2 time periods: before and after treatment.
- 2 groups:
  - *control group:* never receives treatment,
  - *treatment group:* initially untreated and then fully treated.
- Under certain assumptions, control group can be used as the counterfactual for treatment group

An Example: Minimum Wage and Employment

An Example: Minimum Wage and Employment

- Imagine you are interested in assessing the **causal** impact of increasing the minimum wage on (un)employment.

An Example: Minimum Wage and Employment

- Imagine you are interested in assessing the **causal** impact of increasing the minimum wage on (un)employment.
- Why is this not that straightforward? What should the control group be?

An Example: Minimum Wage and Employment

- Imagine you are interested in assessing the **causal** impact of increasing the minimum wage on (un)employment.
- Why is this not that straightforward? What should the control group be?
- Seminal 1994 paper by prominent labor economists David Card and Alan Krueger entitled "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania"

An Example: Minimum Wage and Employment

- Imagine you are interested in assessing the **causal** impact of increasing the minimum wage on (un)employment.
- Why is this not that straightforward? What should the control group be?
- Seminal 1994 paper by prominent labor economists David Card and Alan Krueger entitled "Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania"
- Estimates the effect of an increase in the minimum wage on the employment rate in the fast-food industry. Why this industry?

Institutional Details

- In the US, there is a national minimum wage, but states can depart from it.

Institutional Details

- In the US, there is a national minimum wage, but states can depart from it.
- April 1, 1992: New Jersey minimum wage increases from $4.25 to $5.05 per hour.

Institutional Details

- In the US, there is a national minimum wage, but states can depart from it.
- April 1, 1992: New Jersey minimum wage increases from \$4.25 to \$5.05 per hour.
- Neighboring Pennsylvania did not change its minimum wage level.

Institutional Details

- In the US, there is a national minimum wage, but states can depart from it.
- April 1, 1992: New Jersey minimum wage increases from \$4.25 to \$5.05 per hour.
- Neighboring Pennsylvania did not change its minimum wage level.

Institutional Details

- In the US, there is a national minimum wage, but states can depart from it.
- April 1, 1992: New Jersey minimum wage increases from \$4.25 to \$5.05 per hour.
- Neighboring Pennsylvania did not change its minimum wage level.

Pennsylvania and New Jersey are ***very similar***: similar institutions, similar habits, similar consumers, similar incomes, similar weather, etc.

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania
- Timing:

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania
- Timing:
    - Survey before NJ MW increase: Feb/March 1992

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania
- Timing:
    - Survey before NJ MW increase: Feb/March 1992
    - Survey after NJ MW increase: Nov/Dec 1992

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania
- Timing:
  - Survey before NJ MW increase: Feb/March 1992
  - Survey after NJ MW increase: Nov/Dec 1992
- What comparisons do you think they did?

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania
- Timing:
  - Survey before NJ MW increase: Feb/March 1992
  - Survey after NJ MW increase: Nov/Dec 1992
- What comparisons do you think they did?

Let's take a closer at their data

```
  # install package that contains the cleaned data
remotes::install_github("b-rodrigues/diffindiff")
  # load package
library(diffindiff)
  # load data
ck1994 <- njmin
```

Card and Krueger (1994): Methodology

- Surveyed 410 fast-food establishments in New Jersey (NJ) and eastern Pennsylvania
- Timing:
  - Survey before NJ MW increase: Feb/March 1992
  - Survey after NJ MW increase: Nov/Dec 1992
- What comparisons do you think they did?

Let's take a closer at their data

```r
  # install package that contains the cleaned data
remotes::install_github("b-rodrigues/diffindiff")
  # load package
library(diffindiff)
  # load data
ck1994 <- njmin
```

```r
ck1994 %>%
  select(sheet,chain,state,observation,empft,emppt) %>
  head()
```

```
## # A tibble: 6 x 6
##   sheet chain  state        observation   empft emppt
##   <chr> <chr>  <chr>        <chr>         <dbl> <dbl>
## 1 46    bk     Pennsylvania February 1992  30    15
## 2 49    kfc    Pennsylvania February 1992   6.5   6.5
## 3 506   kfc    Pennsylvania February 1992   3     7
## 4 56    wendys Pennsylvania February 1992  20    20
## 5 61    wendys Pennsylvania February 1992   6    26
## 6 62    wendys Pennsylvania February 1992   0    31
```

Card and Krueger DiD: Tabular Results

Average Employment Per Store Before and After the Rise in NJ Minimum Wage

Variables	Pennsylvania	New Jersey
FTE employment before	23.33	20.44
FTE employment after	21.17	21.03
Change in mean FTE employment	-2.17	0.59

Card and Krueger DiD: Tabular Results

Average Employment Per Store Before and After the Rise in NJ Minimum Wage

Variables	Pennsylvania	New Jersey
FTE employment before	23.33	20.44
FTE employment after	21.17	21.03
Change in mean FTE employment	-2.17	0.59

- Employment in PA *decreased* while it *increased* in NJ!

Card and Krueger DiD: Tabular Results

Average Employment Per Store Before and After the Rise in NJ Minimum Wage

Variables	Pennsylvania	New Jersey
FTE employment before	23.33	20.44
FTE employment after	21.17	21.03
Change in mean FTE employment	-2.17	0.59

- Employment in PA *decreased* while it *increased* in NJ!

DiD Estimate

Differences-in-differences causal estimate: $0.59 - (-2.17) = 2.76$

Card and Krueger DiD: Tabular Results

Average Employment Per Store Before and After the Rise in NJ Minimum Wage

Variables	Pennsylvania	New Jersey
FTE employment before	23.33	20.44
FTE employment after	21.17	21.03
Change in mean FTE employment	-2.17	0.59

- Employment in PA *decreased* while it *increased* in NJ!

DiD Estimate

Differences-in-differences causal estimate: $0.59 - (-2.17) = 2.76$

Card and Krueger DiD: Tabular Results

Average Employment Per Store Before and After the Rise in NJ Minimum Wage

Variables	Pennsylvania	New Jersey
FTE employment before	23.33	20.44
FTE employment after	21.17	21.03
Change in mean FTE employment	-2.17	0.59

- Employment in PA *decreased* while it *increased* in NJ!

DiD Estimate

Differences-in-differences causal estimate: $0.59 - (-2.17) = 2.76$

Yes the essence of differences-in-differences is *that* simple! 😀

Card and Krueger DiD: Tabular Results

Average Employment Per Store Before and After the Rise in NJ Minimum Wage

Variables	Pennsylvania	New Jersey
FTE employment before	23.33	20.44
FTE employment after	21.17	21.03
Change in mean FTE employment	-2.17	0.59

- Employment in PA ***decreased*** while it ***increased*** in NJ!

DiD Estimate

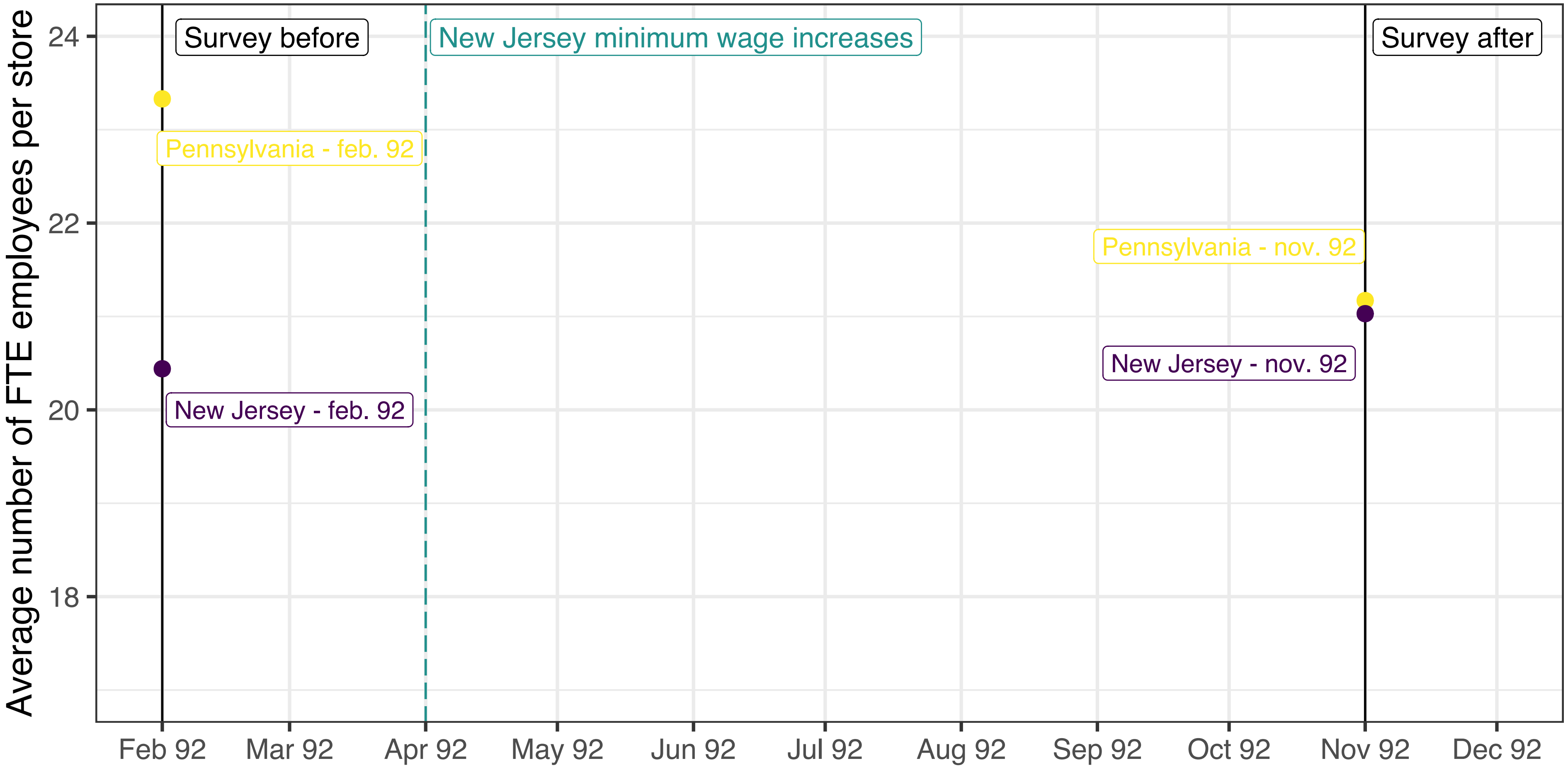
Differences-in-differences causal estimate: $0.59 - (-2.17) = 2.76$

Yes the essence of differences-in-differences is *that* simple! 😀

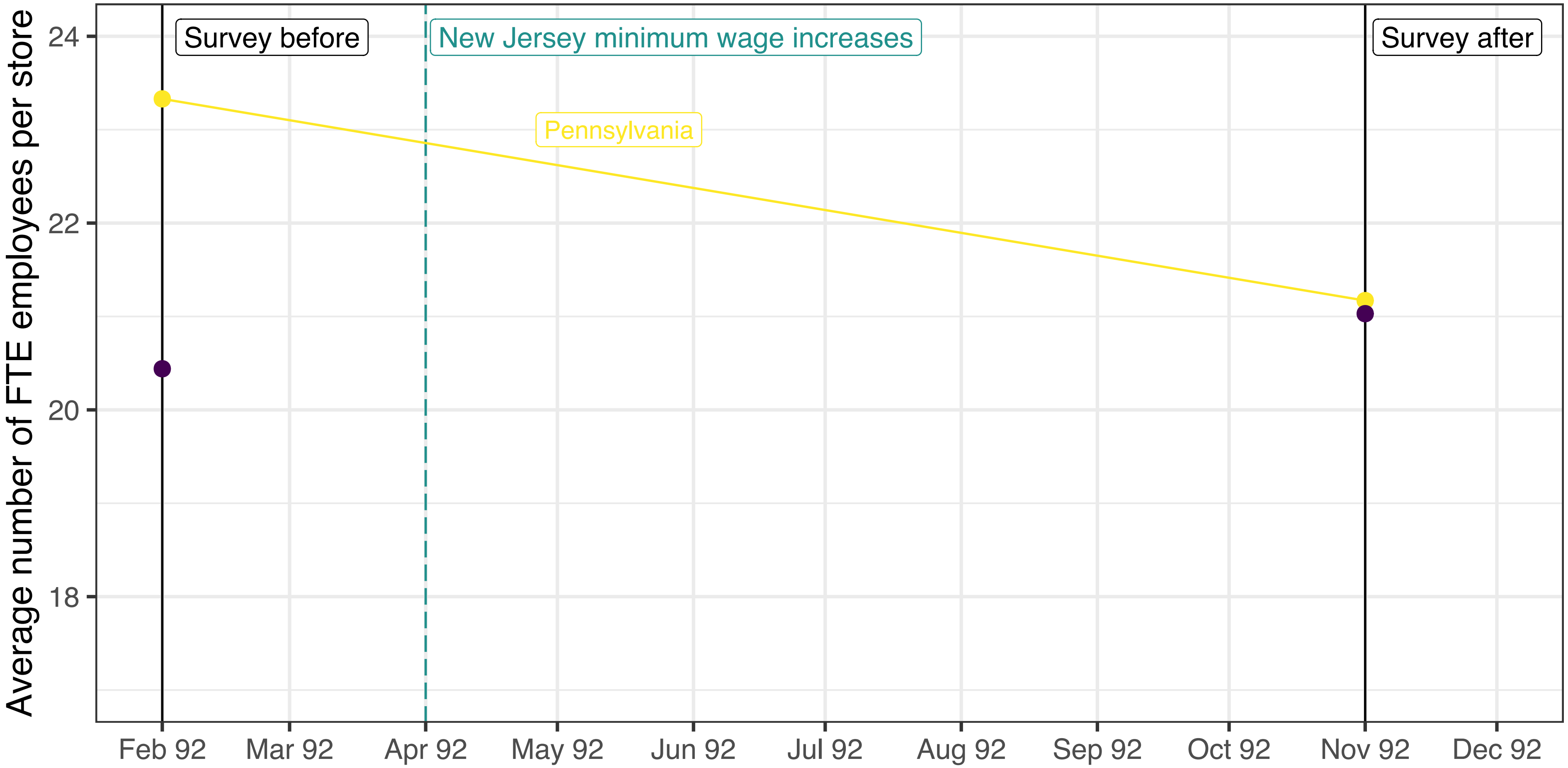
Let's look at these results graphically.

DiD Graphically

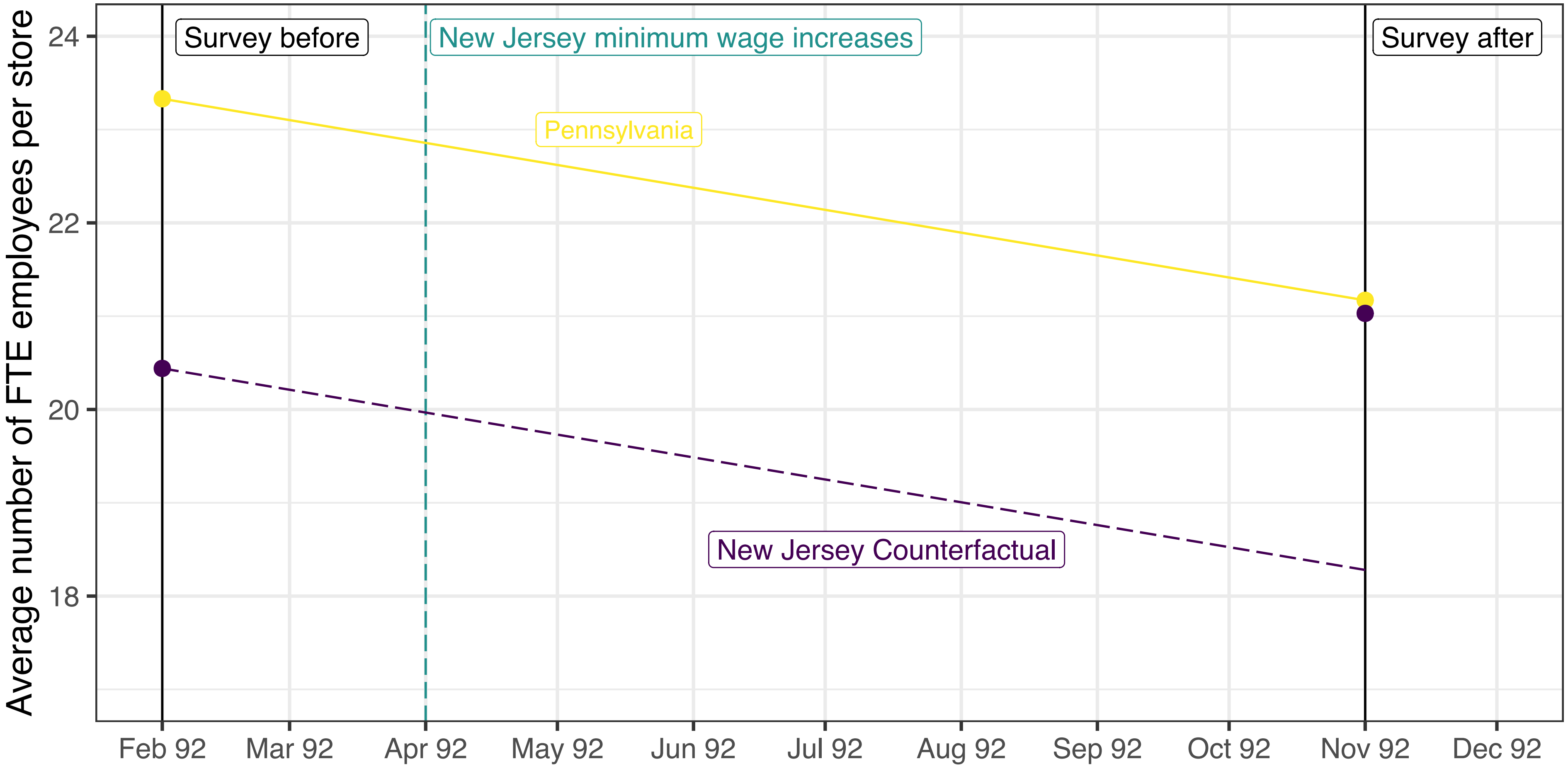
DiD Graphically



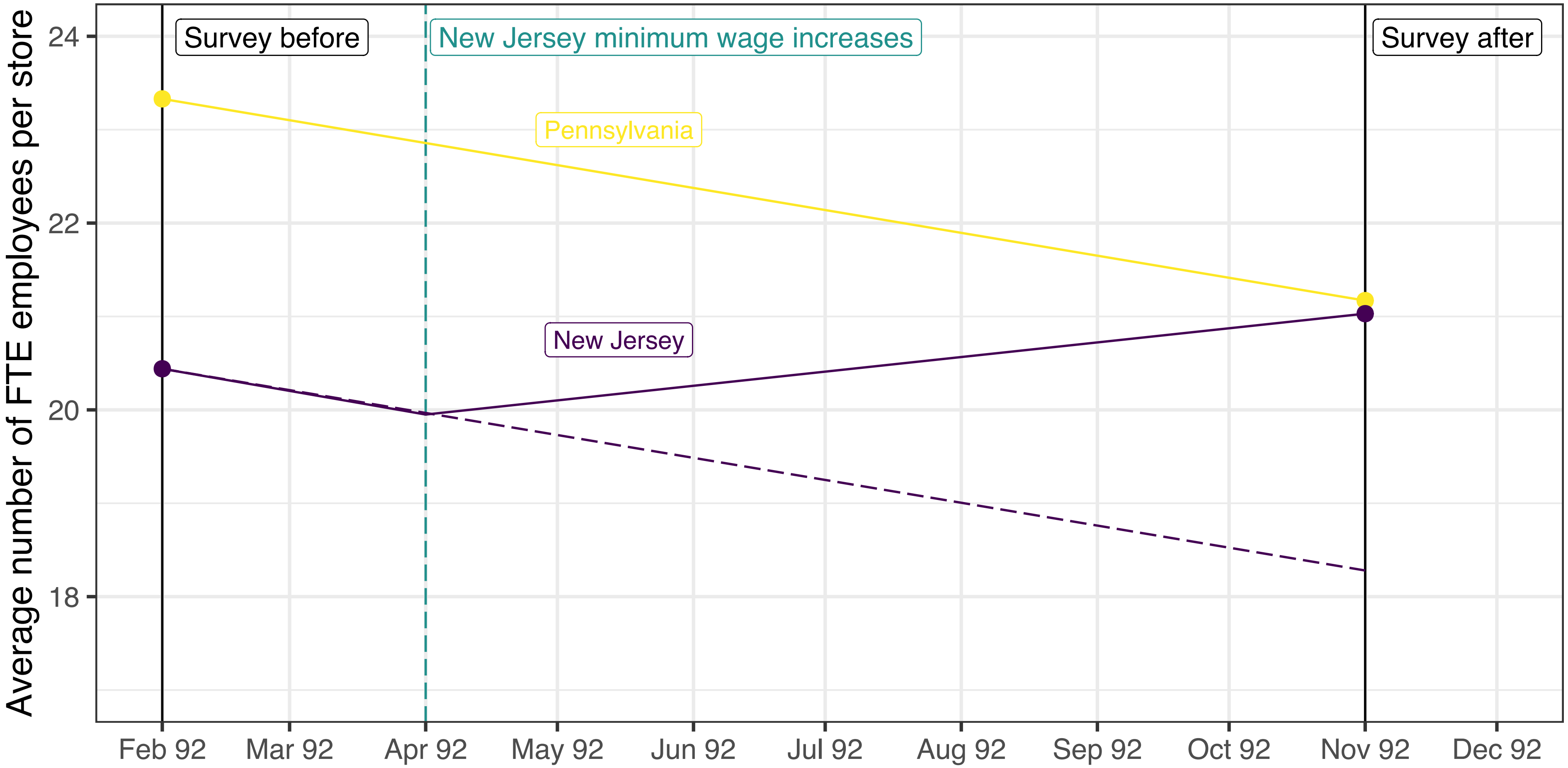
DiD Graphically



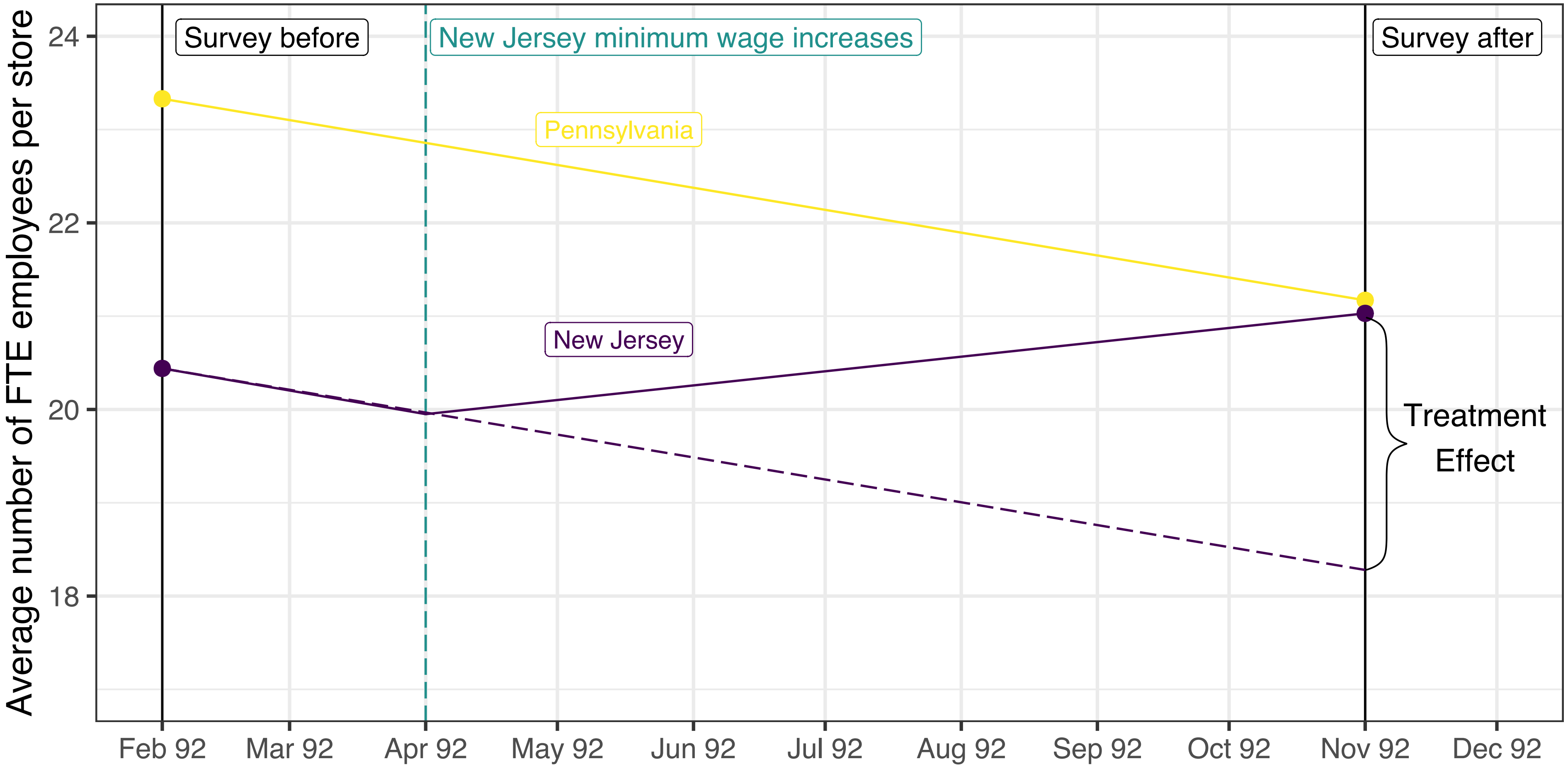
DiD Graphically



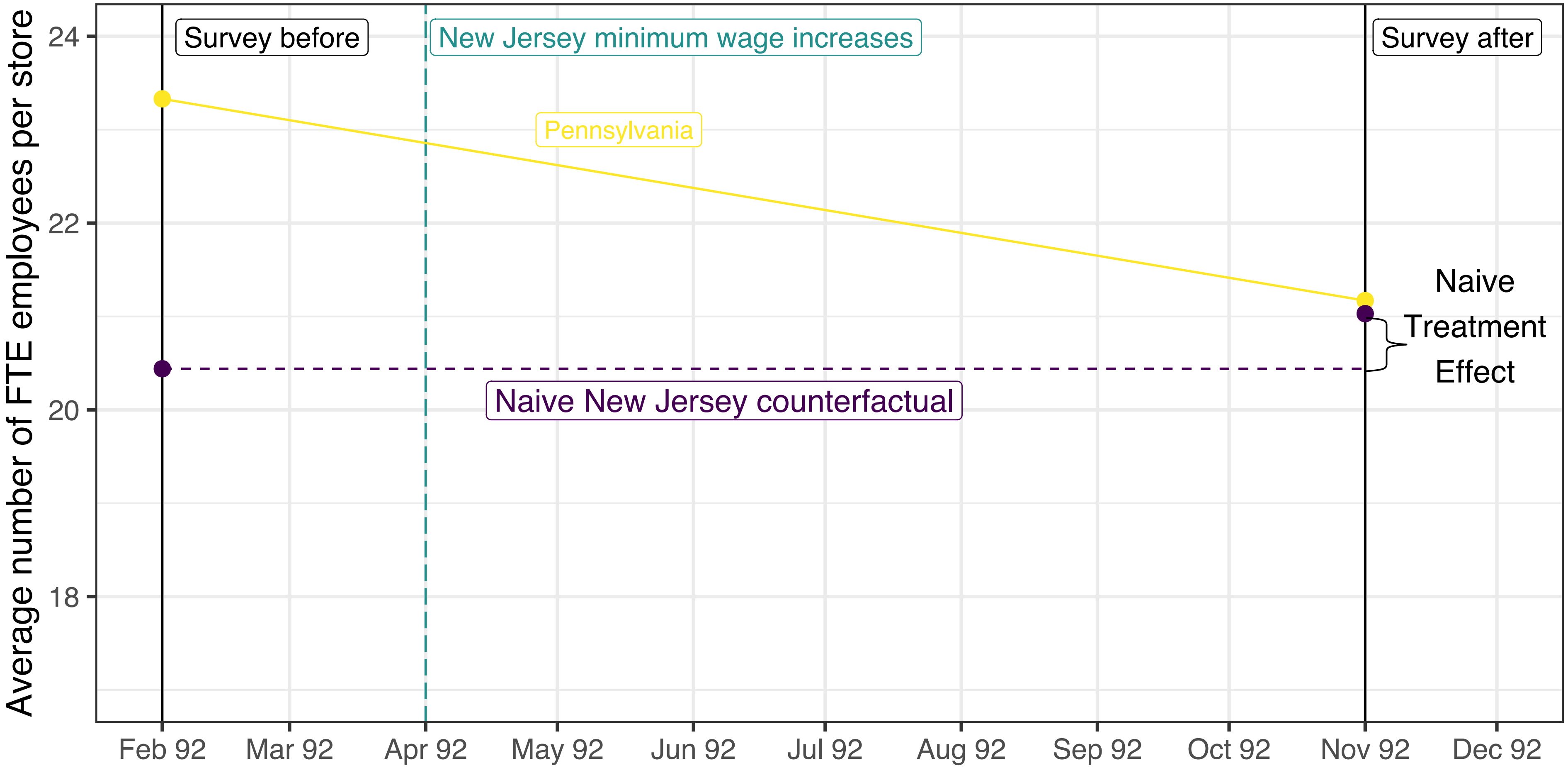
DiD Graphically



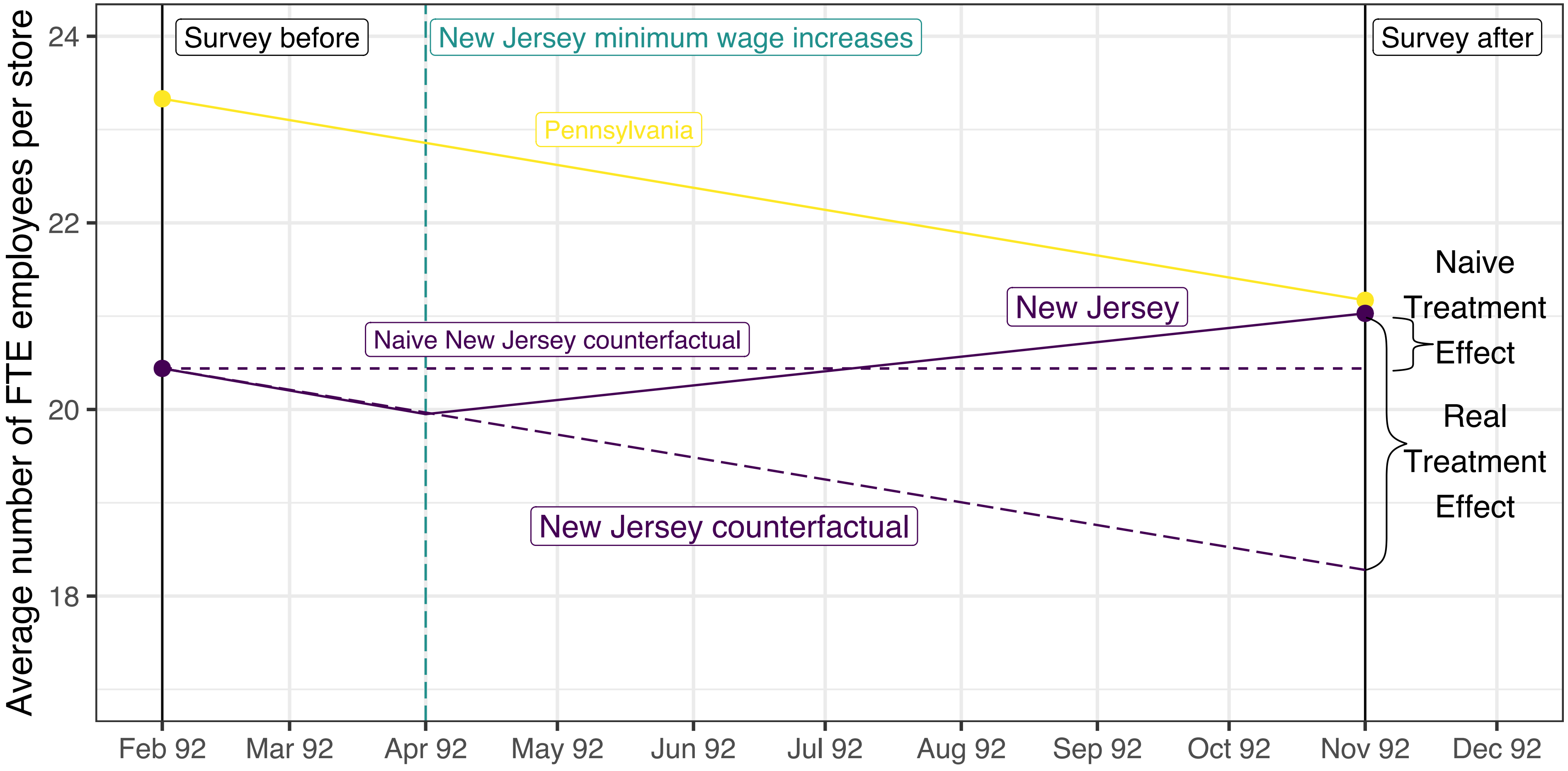
DiD Graphically



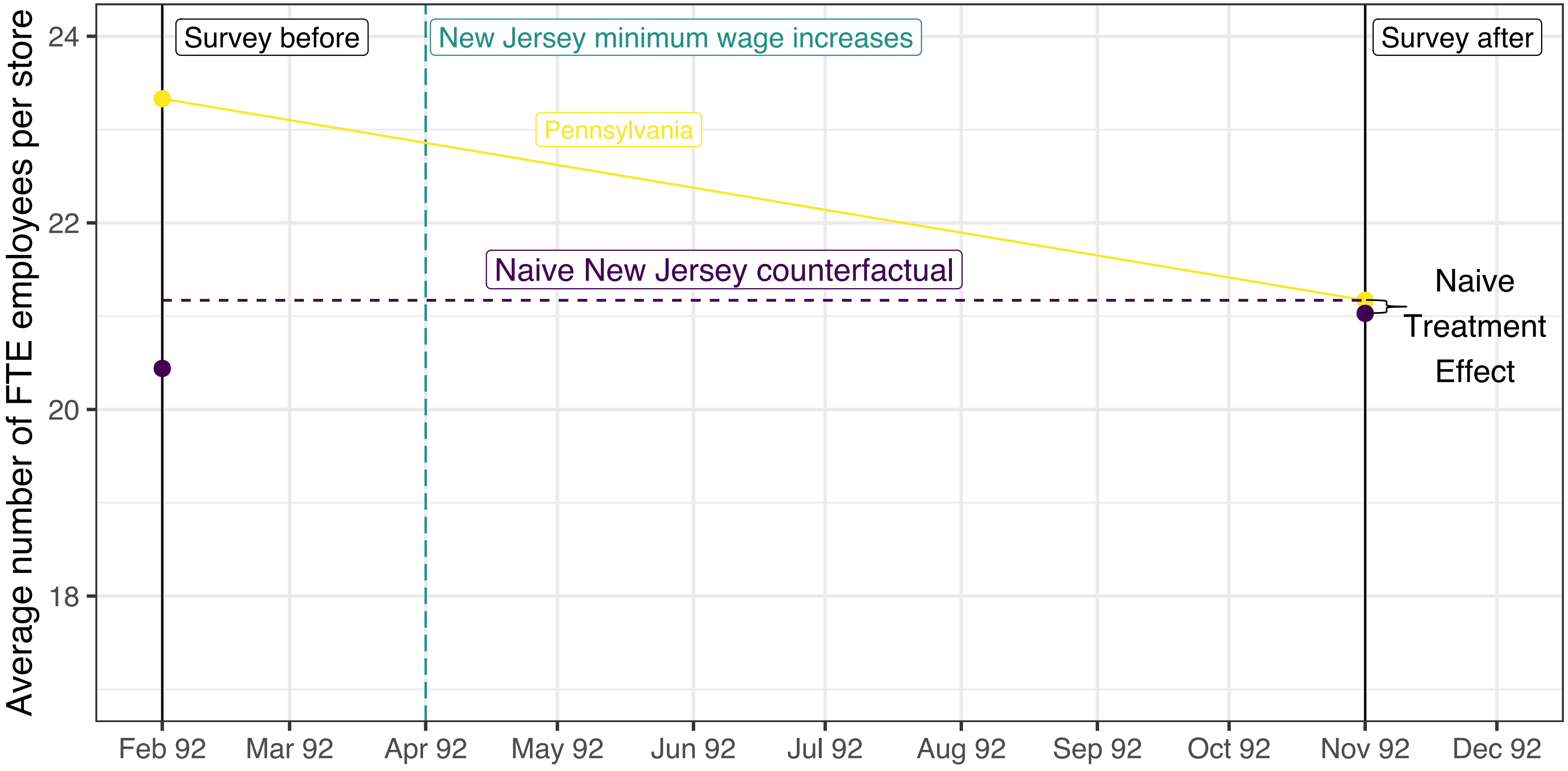
What if we had done a naive after/before comparison?



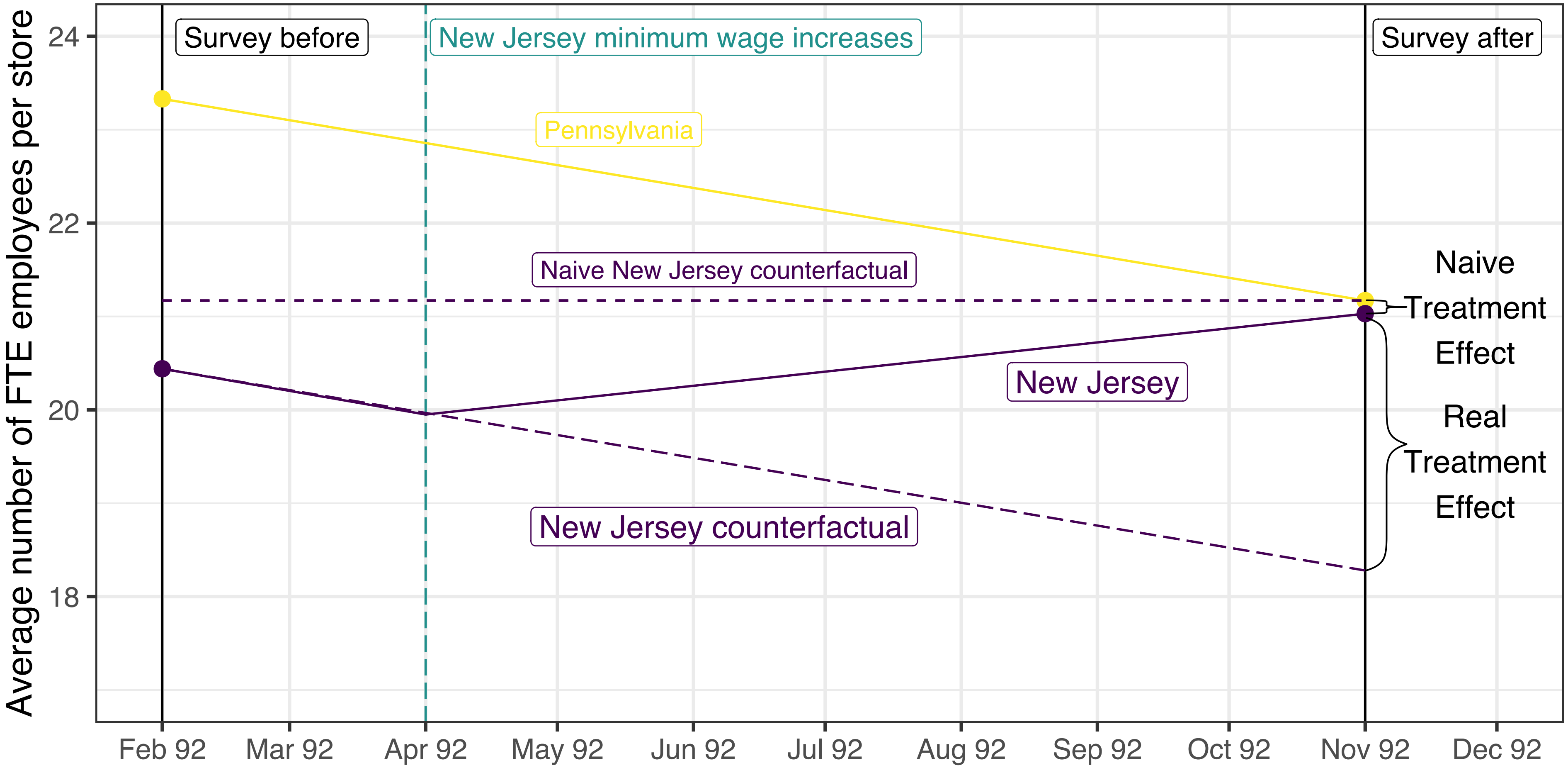
What if we had done a naive after/before comparison?



What if we had done a naive after NJ/PA comparison?



What if we had done a naive after NJ/PA comparison?



Estimation

DiD in Regression Form

- In practice, DiD is usually estimated on more than 2 periods (4 observations)
- There are more data points before and after the policy change

DiD in Regression Form

- In practice, DiD is usually estimated on more than 2 periods (4 observations)
- There are more data points before and after the policy change

3 ingredients:

DiD in Regression Form

- In practice, DiD is usually estimated on more than 2 periods (4 observations)
- There are more data points before and after the policy change

3 ingredients:

1. **Treatment dummy variable**: $TREAT_s$ where the s subscript reminds us that the treatment is at the state level

DiD in Regression Form

- In practice, DiD is usually estimated on more than 2 periods (4 observations)
- There are more data points before and after the policy change

3 ingredients:

1. **Treatment dummy variable**: $TREAT_s$ where the s subscript reminds us that the treatment is at the state level
2. **Post-treatment periods dummy variables**: $POST_t$ where the t subscript reminds us that this variable varies over time

DiD in Regression Form

- In practice, DiD is usually estimated on more than 2 periods (4 observations)
- There are more data points before and after the policy change

3 ingredients:

1. **Treatment dummy variable**: $TREAT_s$ where the s subscript reminds us that the treatment is at the state level
2. **Post-treatment periods dummy variables**: $POST_t$ where the t subscript reminds us that this variable varies over time
3. **Interaction term between the two**: $TREAT_s \times POST_t$ 👉 the ***coefficient on this term is the DiD causal effect***!

DiD in Regression Form

Treatment dummy variable

$$TREAT_s = \begin{cases} 0 & \text{if } s = \text{Pennsylvania} \\ 1 & \text{if } s = \text{New Jersey} \end{cases}$$

DiD in Regression Form

Treatment dummy variable

$$TREAT_s = \begin{cases} 0 & \text{if } s = \text{Pennsylvania} \\ 1 & \text{if } s = \text{New Jersey} \end{cases}$$

Post-treatment periods dummy variable

$$POST_t = \begin{cases} 0 & \text{if } t < \text{April 1, 1992} \\ 1 & \text{if } t \geq \text{April 1, 1992} \end{cases}$$

DiD in Regression Form

Treatment dummy variable

$$TREAT_s = \begin{cases} 0 & \text{if } s = \text{Pennsylvania} \\ 1 & \text{if } s = \text{New Jersey} \end{cases}$$

Post-treatment periods dummy variable

$$POST_t = \begin{cases} 0 & \text{if } t < \text{April 1, 1992} \\ 1 & \text{if } t \geq \text{April 1, 1992} \end{cases}$$

Which observations correspond to $TREAT_s \times POST_t = 1$?

DiD in Regression Form

Treatment dummy variable

$$TREAT_s = \begin{cases} 0 & \text{if } s = \text{Pennsylvania} \\ 1 & \text{if } s = \text{New Jersey} \end{cases}$$

Post-treatment periods dummy variable

$$POST_t = \begin{cases} 0 & \text{if } t < \text{April 1, 1992} \\ 1 & \text{if } t \geq \text{April 1, 1992} \end{cases}$$

Which observations correspond to $TREAT_s \times POST_t = 1$?

- Let's put all these ingredients together:

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

- δ: causal effect of the minimum wage increase on employment

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

We have the following:

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

We have the following:

$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 0) = \alpha$

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

We have the following:

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 0) = \alpha$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 1) = \alpha + \gamma$$

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

We have the following:

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 0) = \alpha$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 1) = \alpha + \gamma$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 0) = \alpha + \beta$$

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

We have the following:

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 0) = \alpha$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 1) = \alpha + \gamma$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 0) = \alpha + \beta$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 1) = \alpha + \beta + \gamma + \delta$$

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

We have the following:

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 0) = \alpha$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 1) = \alpha + \gamma$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 0) = \alpha + \beta$$

$$\mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 1) = \alpha + \beta + \gamma + \delta$$

$$[\mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 1) - \mathbb{E}(EMP_{st} \mid TREAT_s = 1, POST_t = 0)] -$$
$$[\mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 1) - \mathbb{E}(EMP_{st} \mid TREAT_s = 0, POST_t = 0)] = \delta$$

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

In table form:

	Pre mean	**Post mean**	**Δ(post - pre)**
Pennsylvania (PA)	α	$\alpha + \gamma$	γ
New Jersey (NJ)	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$	$\gamma + \delta$
Δ(NJ - PA)	β	$\beta + \delta$	δ

Understanding the Regression

$$EMP_{st} = \alpha + \beta TREAT_s + \gamma POST_t + \delta(TREAT_s \times POST_t) + \varepsilon_{st}$$

In table form:

	Pre mean	**Post mean**	**Δ(post - pre)**
Pennsylvania (PA)	α	$\alpha + \gamma$	γ
New Jersey (NJ)	$\alpha + \beta$	$\alpha + \beta + \gamma + \delta$	$\gamma + \delta$
Δ(NJ - PA)	β	$\beta + \delta$	δ

This table generalizes to other settings by substituting *Pennsylvania* with *Control* and *New Jersey* with *Treatment*

Task 2 (10 minutes)

1. Create a full-time equivalent (FTE) employees variable called `empfte` equal to `empft` + 0.5*`emppt` + `nmgrs`. `empft` and `emppt` correspond respectively to the number of full-time and part-time employees. `nmgrs` corresponds to the number of managers. This is how Card and Krueger compute their full-time equivalent (FTE) employment variable (p.775 of the paper).

2. Create a dummy variable, `treat`, equal to `FALSE` if `state` is Pennsylvania and `TRUE` if New Jersey.

3. Create a dummy variable, `post`, equal to `FALSE` if `observation` is February 1992 and `TRUE` otherwise.

4. Estimate the following regression model. Do you obtain the same results as in slide 9 (≈ 2.76)?

$$empfte_{st} = \alpha + \beta treat_s + \gamma post_t + \delta(treat_s \times post_t) + \varepsilon_{st}$$

Identifying Assumptions

DiD Crucial Assumption: Parallel Trends

> **Common or parallel trends assumption**: absent any minimum wage increase, Pennsylvania's fast-food employment trend would have been what we should have expected to see in New Jersey.

DiD Crucial Assumption: Parallel Trends

> **Common or parallel trends assumption**: absent any minimum wage increase, Pennsylvania's fast-food employment trend would have been what we should have expected to see in New Jersey.

- This assumption states that Pennsylvania's fast-food employment trend between February and November 1992 provides a reliable counterfactual employment trend New Jersey's fast-food industry *would have experienced* had New Jersey not increased its minimum wage.

DiD Crucial Assumption: Parallel Trends

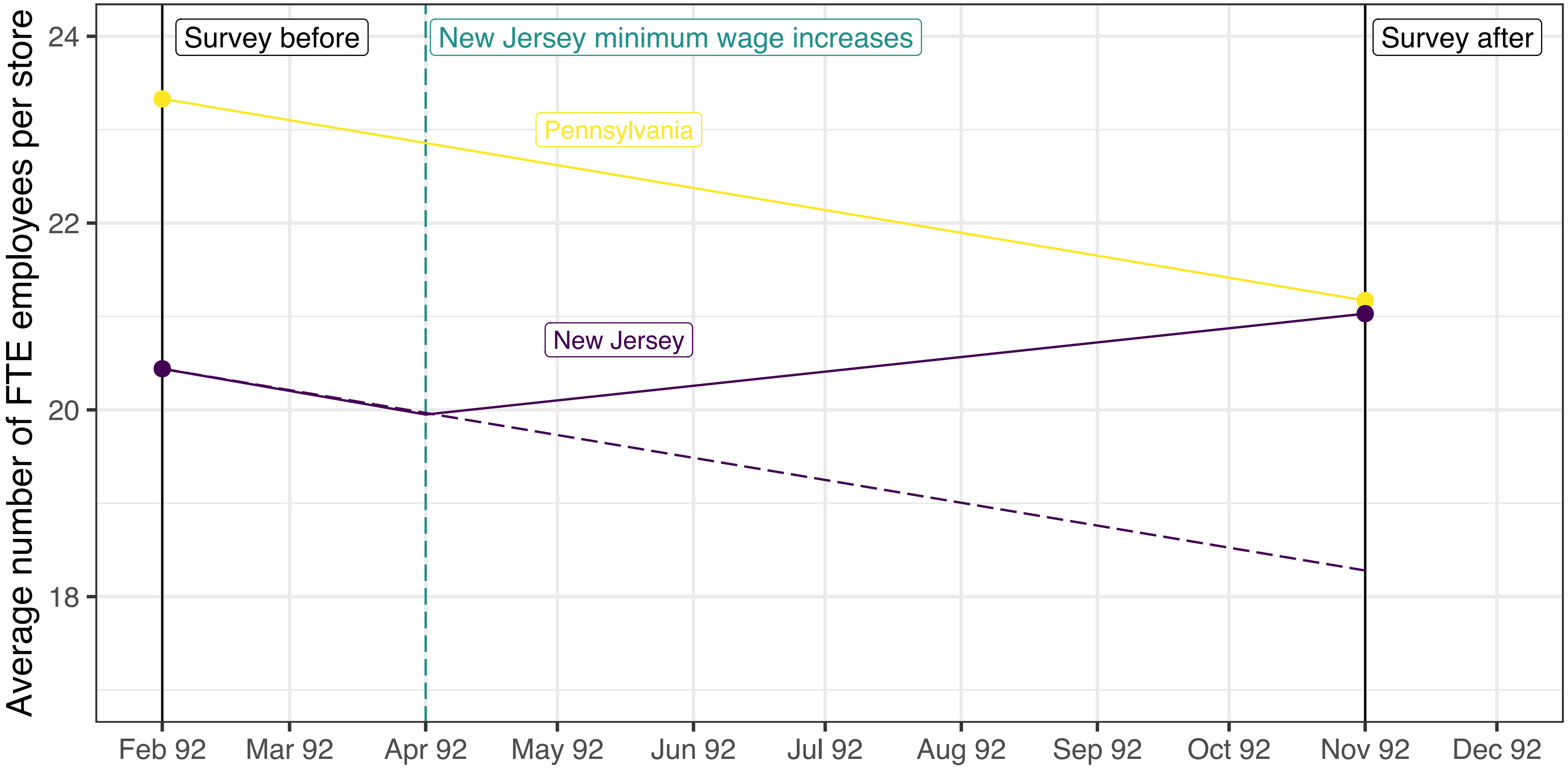
> **Common or parallel trends assumption**: absent any minimum wage increase, Pennsylvania's fast-food employment trend would have been what we should have expected to see in New Jersey.

- This assumption states that Pennsylvania's fast-food employment trend between February and November 1992 provides a reliable counterfactual employment trend New Jersey's fast-food industry *would have experienced* had New Jersey not increased its minimum wage.

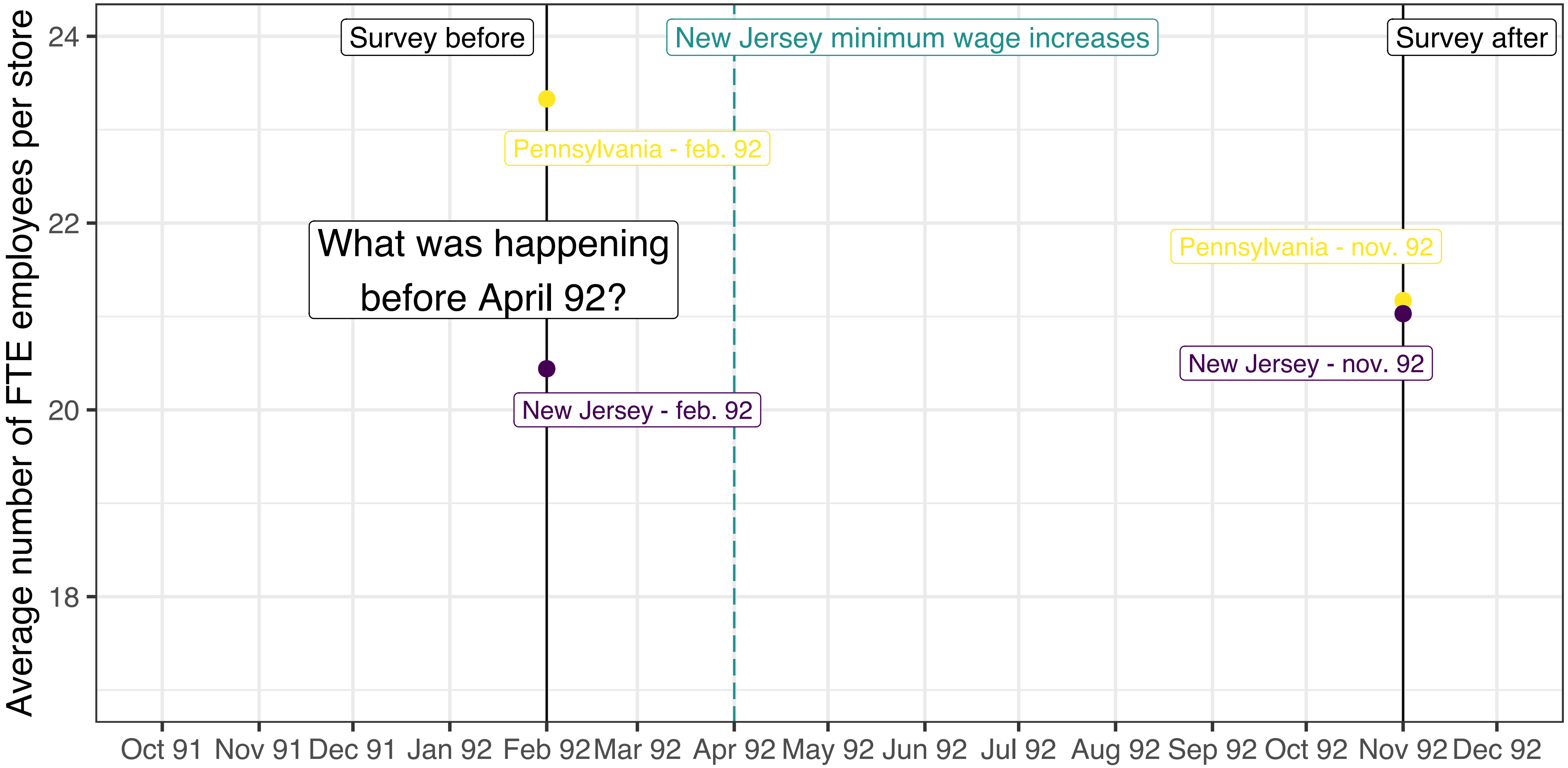
- Impossible to completely validate or invalidate this assumption.

- *Intuitive check:* compare trends before policy change (and after policy change if no expected medium-term effects)

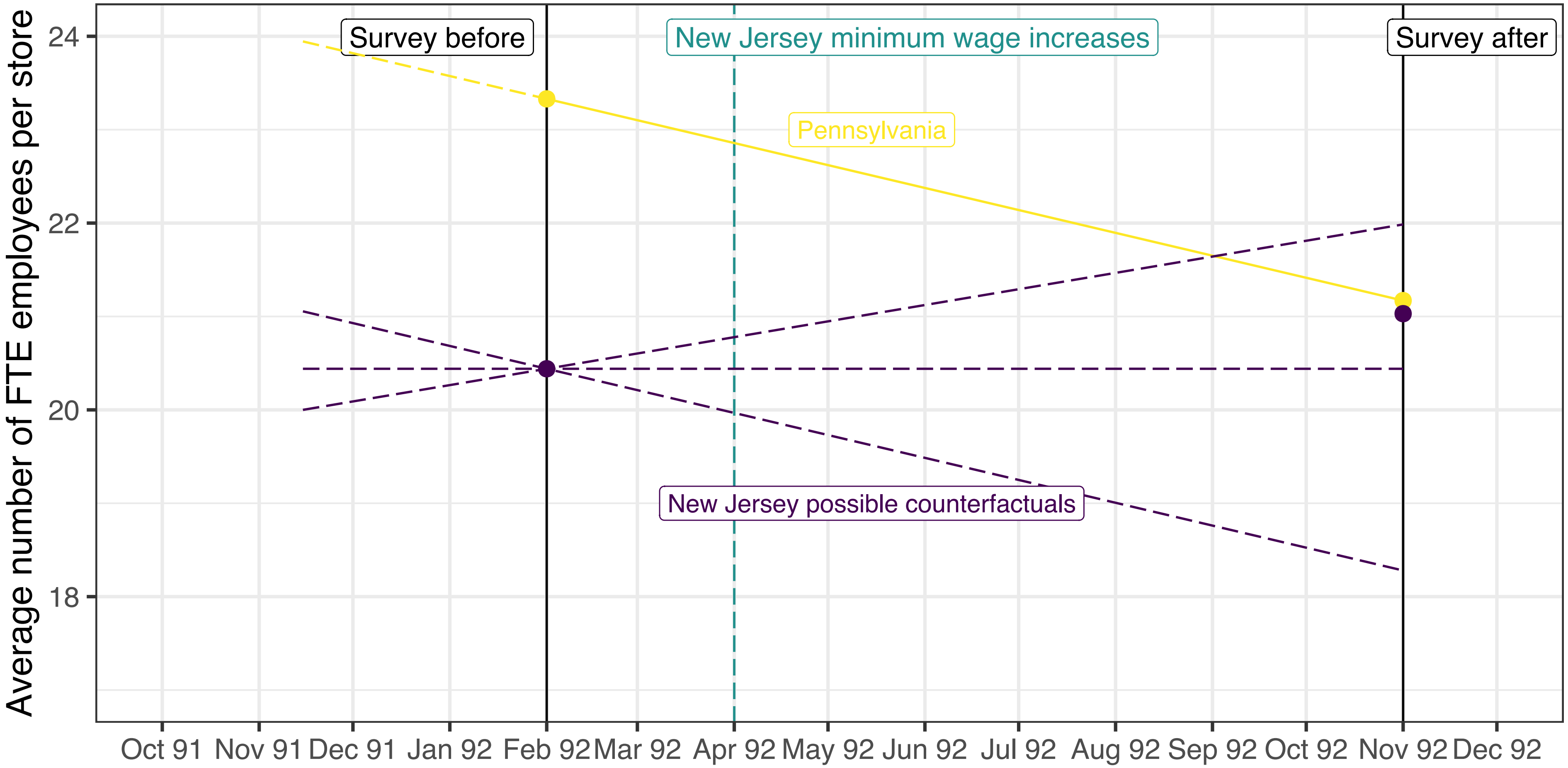
Parallel Trends: Graphically



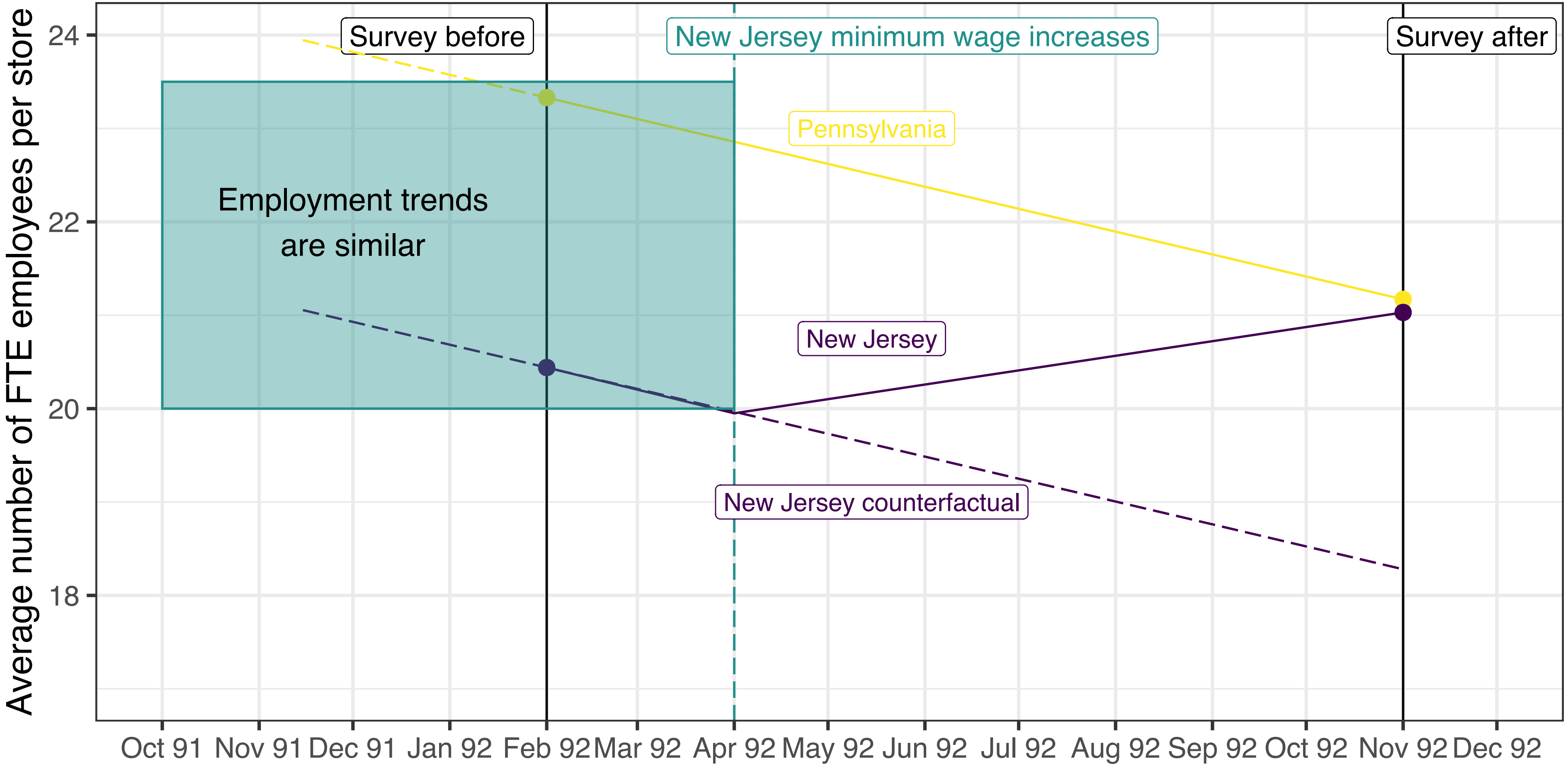
Checking the parallel trends assumption



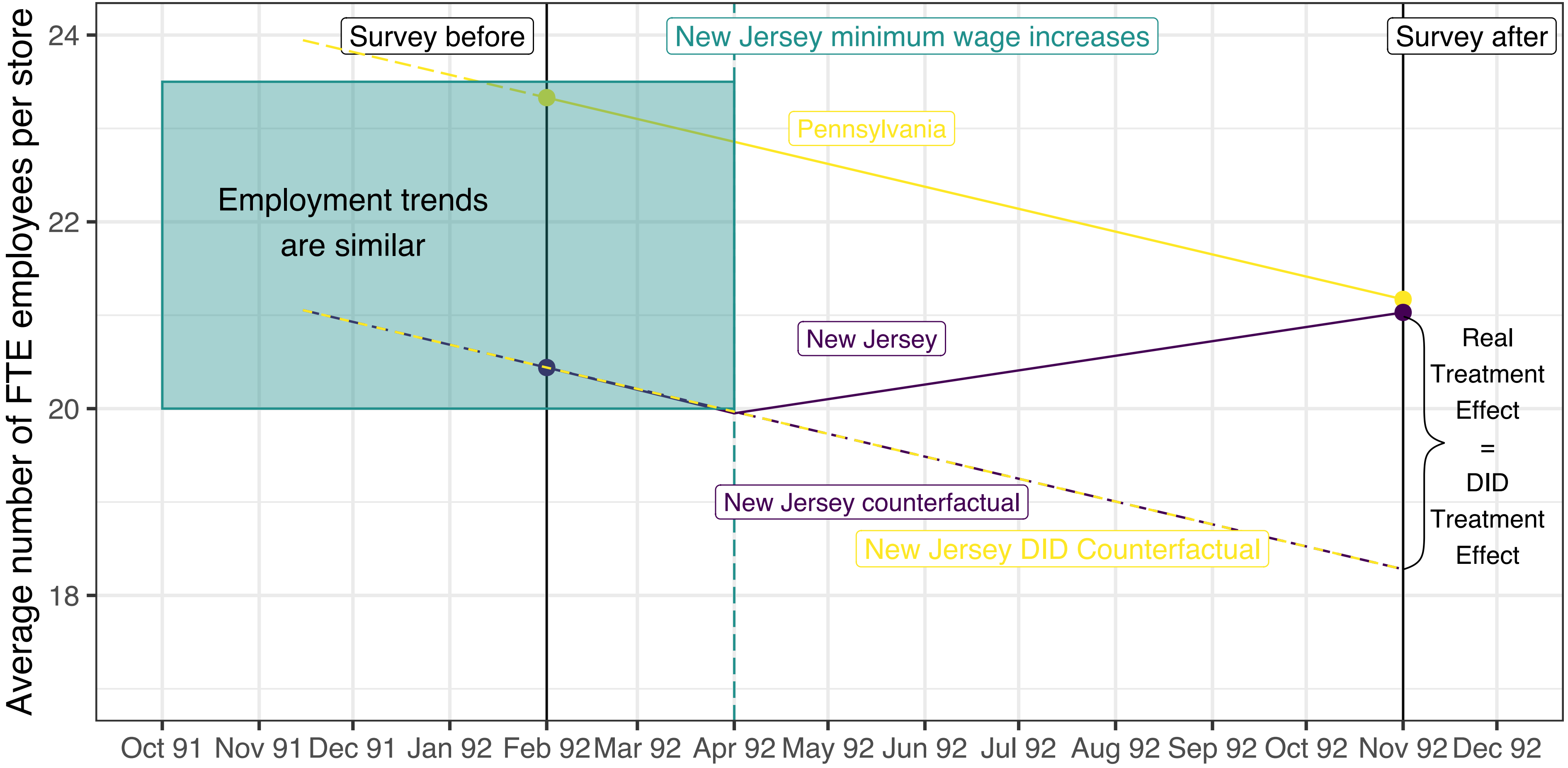
Checking the parallel trends assumption



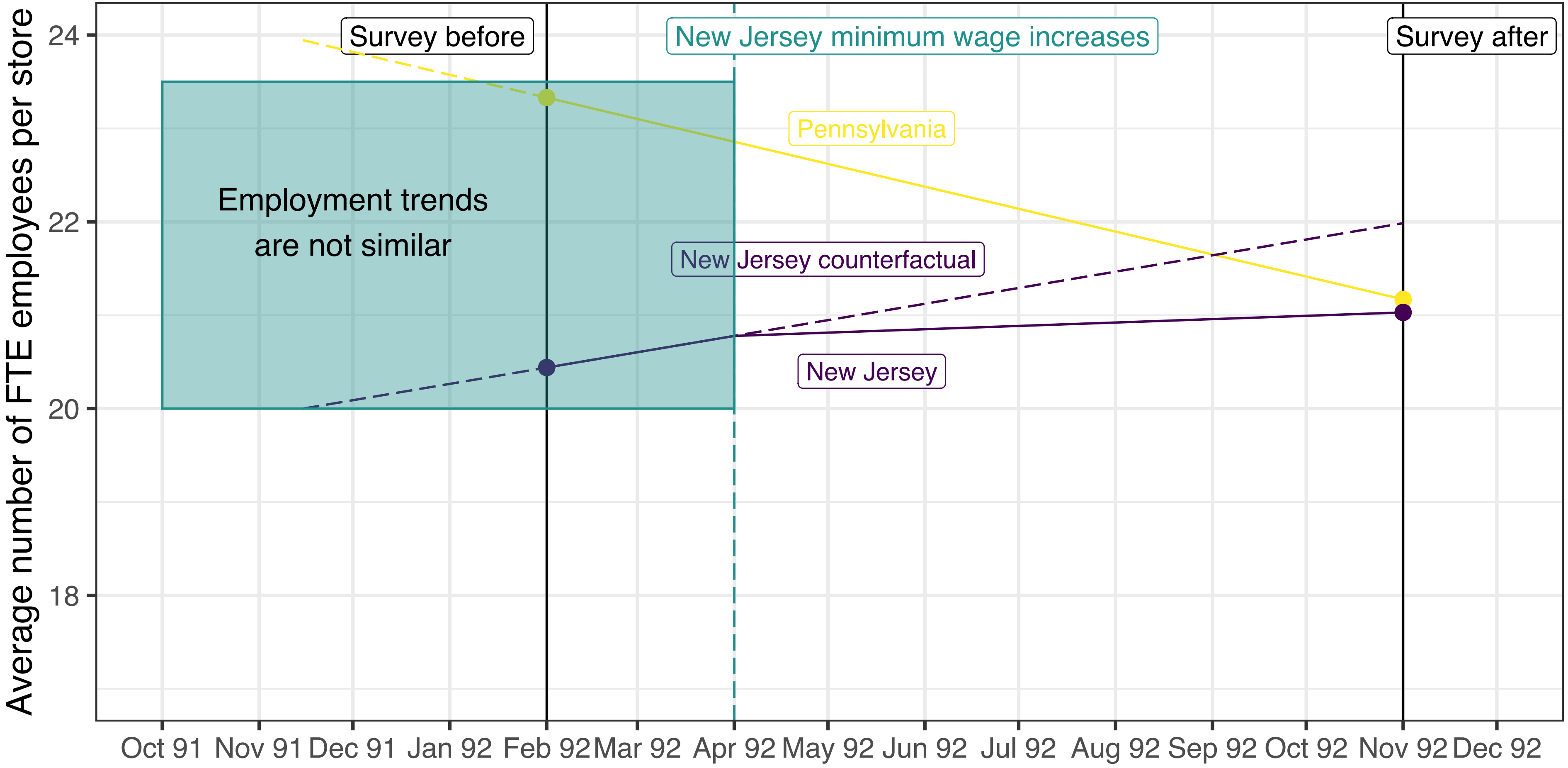
Parallel trends assumption → Verified ✅



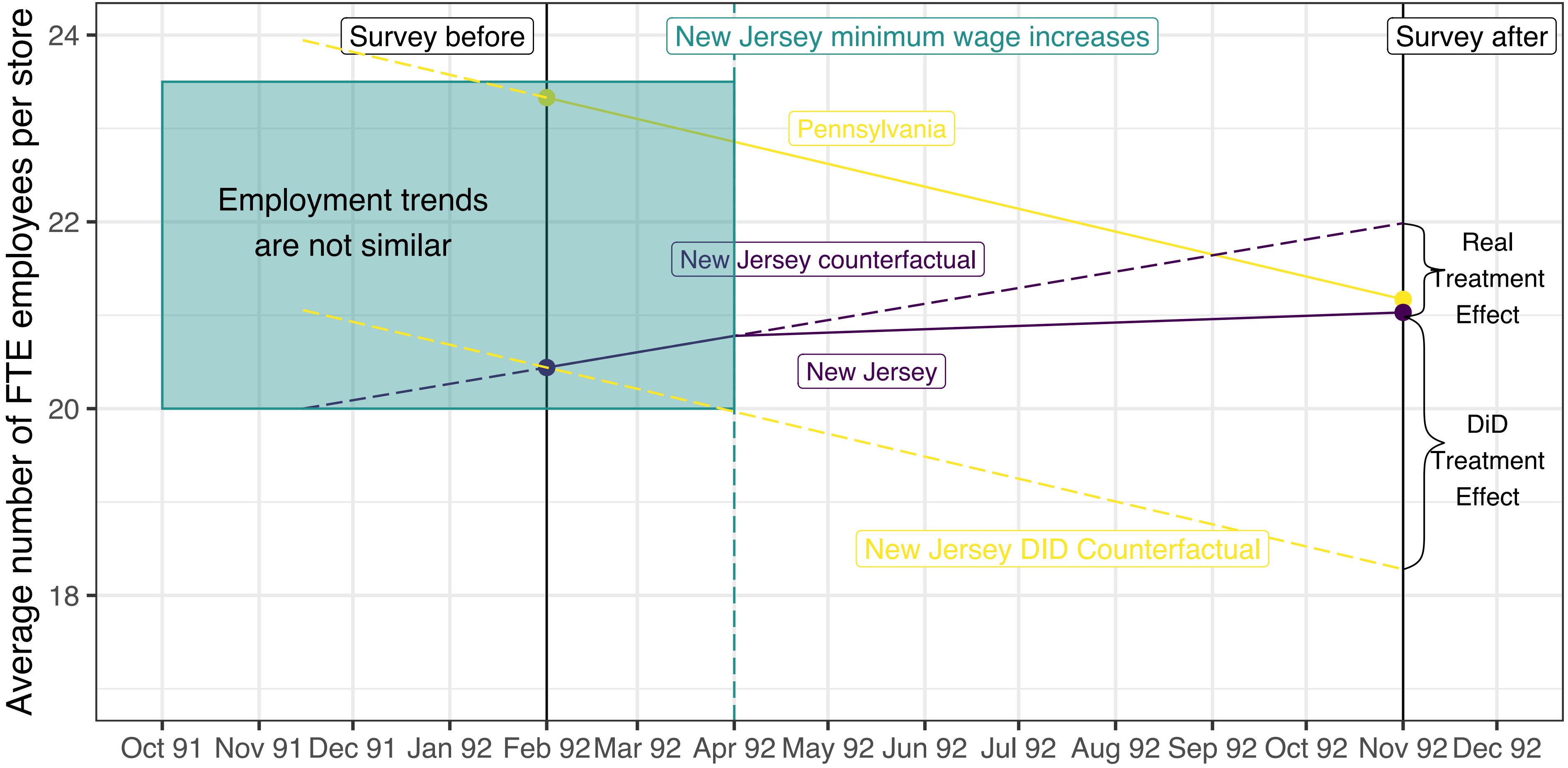
Parallel trends assumption → Verified ✅



Parallel trends assumption → Not verified ❌



Parallel trends assumption → Not verified ❌



Regression Discontinuity Design

Regression Discontinuity Design (RDD)

- Very common research design in applied research because it provides credible causal estimates.

Regression Discontinuity Design (RDD)

- Very common research design in applied research because it provides credible causal estimates.
- Starting point: subjects are ***not*** randomly allocated to treatment ⚠️

Regression Discontinuity Design (RDD)

- Very common research design in applied research because it provides credible causal estimates.
- Starting point: subjects are ***not*** randomly allocated to treatment ⚠️
- RDD can be applied when we have specific information about the rules determining treatment.

Regression Discontinuity Design (RDD)

- Very common research design in applied research because it provides credible causal estimates.
- Starting point: subjects are ***not*** randomly allocated to treatment ⚠️
- RDD can be applied when we have specific information about the rules determining treatment.
- **RDD** exploits this precise information about allocation to treatment!

Discontinuities are Everywhere

There are many arbitrary rules in life that determine assignment to some treatment:

Discontinuities are Everywhere

There are many arbitrary rules in life that determine assignment to some treatment:

- In North Carolina, you used to have to have reached the age of five by October 16 in the relevant year to be eligible to enter kindergarten (Cook and Kang, 2016);

Discontinuities are Everywhere

There are many arbitrary rules in life that determine assignment to some treatment:

- In North Carolina, you used to have to have reached the age of five by October 16 in the relevant year to be eligible to enter kindergarten (Cook and Kang, 2016);
- In the US, a new born baby weighing less than 1,500 grams is considered to be of "very low birth weight" and receive additional treatment (Almond et al., 2010);

Discontinuities are Everywhere

There are many arbitrary rules in life that determine assignment to some treatment:

- In North Carolina, you used to have to have reached the age of five by October 16 in the relevant year to be eligible to enter kindergarten (Cook and Kang, 2016);
- In the US, a new born baby weighing less than 1,500 grams is considered to be of "very low birth weight" and receive additional treatment (Almond et al., 2010);
- Flagship state universities use a certain SAT cutoff level to select their students (Hoekstra, 2009);

Discontinuities are Everywhere

There are many arbitrary rules in life that determine assignment to some treatment:

- In North Carolina, you used to have to have reached the age of five by October 16 in the relevant year to be eligible to enter kindergarten (Cook and Kang, 2016);
- In the US, a new born baby weighing less than 1,500 grams is considered to be of "very low birth weight" and receive additional treatment (Almond et al., 2010);
- Flagship state universities use a certain SAT cutoff level to select their students (Hoekstra, 2009);
- In Italy, there are quotas of residence permits for illegal immigrants that are allocated on a first-come first-served basis until quota is exhausted (Pinotti, 2017);

Discontinuities are Everywhere

There are many arbitrary rules in life that determine assignment to some treatment:

- In North Carolina, you used to have to have reached the age of five by October 16 in the relevant year to be eligible to enter kindergarten (Cook and Kang, 2016);
- In the US, a new born baby weighing less than 1,500 grams is considered to be of "very low birth weight" and receive additional treatment (Almond et al., 2010);
- Flagship state universities use a certain SAT cutoff level to select their students (Hoekstra, 2009);
- In Italy, there are quotas of residence permits for illegal immigrants that are allocated on a first-come first-served basis until quota is exhausted (Pinotti, 2017);

We will focus our analysis on the following discontinuity:

- In the US, the legal drinking age is 21 years old (Carpenter and Dobkin, 2009).

An Example: Alcohol Consumption and Mortality

An Example: Alcohol Consumption and Mortality

- Imagine you are interested in assessing the **causal** impact of alcohol consumption by young adults on mortality.

An Example: Alcohol Consumption and Mortality

- Imagine you are interested in assessing the **causal** impact of alcohol consumption by young adults on mortality.
- Why is this not that straightforward? Why can't you just regress alcohol consumption on dying age and cause of death?

An Example: Alcohol Consumption and Mortality

- Imagine you are interested in assessing the **causal** impact of alcohol consumption by young adults on mortality.
- Why is this not that straightforward? Why can't you just regress alcohol consumption on dying age and cause of death?
  - Because there may be unobserved selection into alcohol consumption that may also be a determinant of mortality.

An Example: Alcohol Consumption and Mortality

- Imagine you are interested in assessing the **causal** impact of alcohol consumption by young adults on mortality.
- Why is this not that straightforward? Why can't you just regress alcohol consumption on dying age and cause of death?
  - Because there may be unobserved selection into alcohol consumption that may also be a determinant of mortality.
- In the US, alcohol consumption is prohibited before the age of 21.

An Example: Alcohol Consumption and Mortality

- Imagine you are interested in assessing the **causal** impact of alcohol consumption by young adults on mortality.
- Why is this not that straightforward? Why can't you just regress alcohol consumption on dying age and cause of death?
  - Because there may be unobserved selection into alcohol consumption that may also be a determinant of mortality.
- In the US, alcohol consumption is prohibited before the age of 21.
- Debate on whether the minimum legal drinking age (MLDA) should be lowered to 18, as was the case in the Vietnam-era.

Key Terms and Intuition

> ***Running variable:*** variable that determines assignment to treatment.

Key Terms and Intuition

> ***Running variable:*** variable that determines assignment to treatment.

$\rightarrow a$ = age

Key Terms and Intuition

> ***Running variable:*** variable that determines assignment to treatment.

$\rightarrow a$ = age

> ***Cutoff level:*** level of the ***running variable*** above (or below) which individuals are treated (or not).

Key Terms and Intuition

> ***Running variable:*** variable that determines assignment to treatment.

$\rightarrow a$ = age

> ***Cutoff level:*** level of the ***running variable*** above (or below) which individuals are treated (or not).

$\rightarrow c = 21$ year old birthday

Key Terms and Intuition

> ***Running variable:*** variable that determines assignment to treatment.

$\rightarrow a$ = age

> ***Cutoff level:*** level of the ***running variable*** above (or below) which individuals are treated (or not).

$\rightarrow c = 21$ year old birthday

Causal intuition:

- How different are individuals *just before* and *just after* their 21st birthday, other than legal access to alcohol?

Key Terms and Intuition

> *Running variable:* variable that determines assignment to treatment.

$\rightarrow a$ = age

> *Cutoff level:* level of the *running variable* above (or below) which individuals are treated (or not).

$\rightarrow c = 21$ year old birthday

Causal intuition:

- How different are individuals *just before* and *just after* their 21st birthday, other than legal access to alcohol?
- Around the threshold, allocation to treatment is ***as good as random***.

Key Terms and Intuition

> *Running variable:* variable that determines assignment to treatment.

$\rightarrow a$ = age

> *Cutoff level:* level of the *running variable* above (or below) which individuals are treated (or not).

$\rightarrow c = 21$ year old birthday

Causal intuition:

- How different are individuals *just before* and *just after* their 21st birthday, other than legal access to alcohol?
- Around the threshold, allocation to treatment is *as good as random*.
- 👉 *Regression discontinuity design* exploits this allocation to treatment!

Carpenter and Dobkin's data

- Let's take a closer at the data used in the paper

```r
# install package containing data
devtools::install_github("jrnold/masteringmetrics",
                         subdir = "masteringmetrics")

# load data
data("mlda", package = "masteringmetrics")
```

Carpenter and Dobkin's data

- Let's take a closer at the data used in the paper

```
# install package containing data
devtools::install_github("jrnold/masteringmetrics",
                         subdir = "masteringmetrics")

# load data
data("mlda", package = "masteringmetrics")
```

```
## # A tibble: 6 x 7
##   agecell   all internal external alcohol homicide suicide
##     <dbl> <dbl>    <dbl>    <dbl>   <dbl>    <dbl>   <dbl>
## 1    19.1  92.8     16.6     76.2   0.639     16.3    11.2
## 2    19.2  95.1     18.3     76.8   0.677     16.9    12.2
## 3    19.2  92.1     18.9     73.2   0.866     15.2    11.7
## 4    19.3  88.4     16.1     72.3   0.867     16.7    11.3
## 5    19.4  88.7     17.4     71.3   1.02      14.9    11.0
## 6    19.5  90.2     17.9     72.3   1.17      15.6    12.2
```

Carpenter and Dobkin's data

- Let's take a closer at the data used in the paper

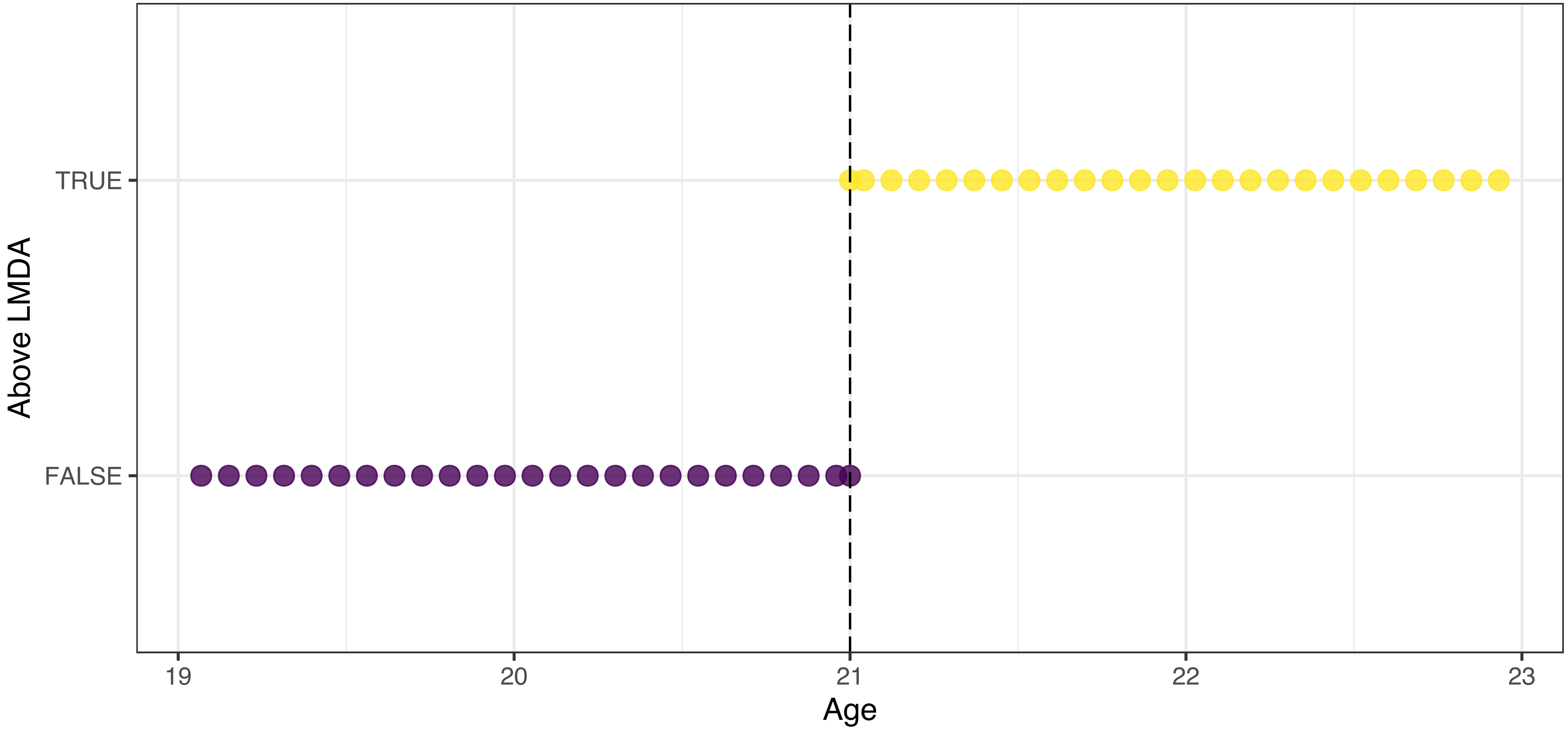
```r
# install package containing data
devtools::install_github("jrnold/masteringmetrics",
                         subdir = "masteringmetrics")

# load data
data("mlda", package = "masteringmetrics")
```

```
## # A tibble: 6 x 7
##   agecell   all internal external alcohol homicide suicide
##     <dbl> <dbl>    <dbl>    <dbl>   <dbl>    <dbl>   <dbl>
## 1    19.1  92.8     16.6     76.2   0.639     16.3    11.2
## 2    19.2  95.1     18.3     76.8   0.677     16.9    12.2
## 3    19.2  92.1     18.9     73.2   0.866     15.2    11.7
## 4    19.3  88.4     16.1     72.3   0.867     16.7    11.3
## 5    19.4  88.7     17.4     71.3   1.02      14.9    11.0
## 6    19.5  90.2     17.9     72.3   1.17      15.6    12.2
```

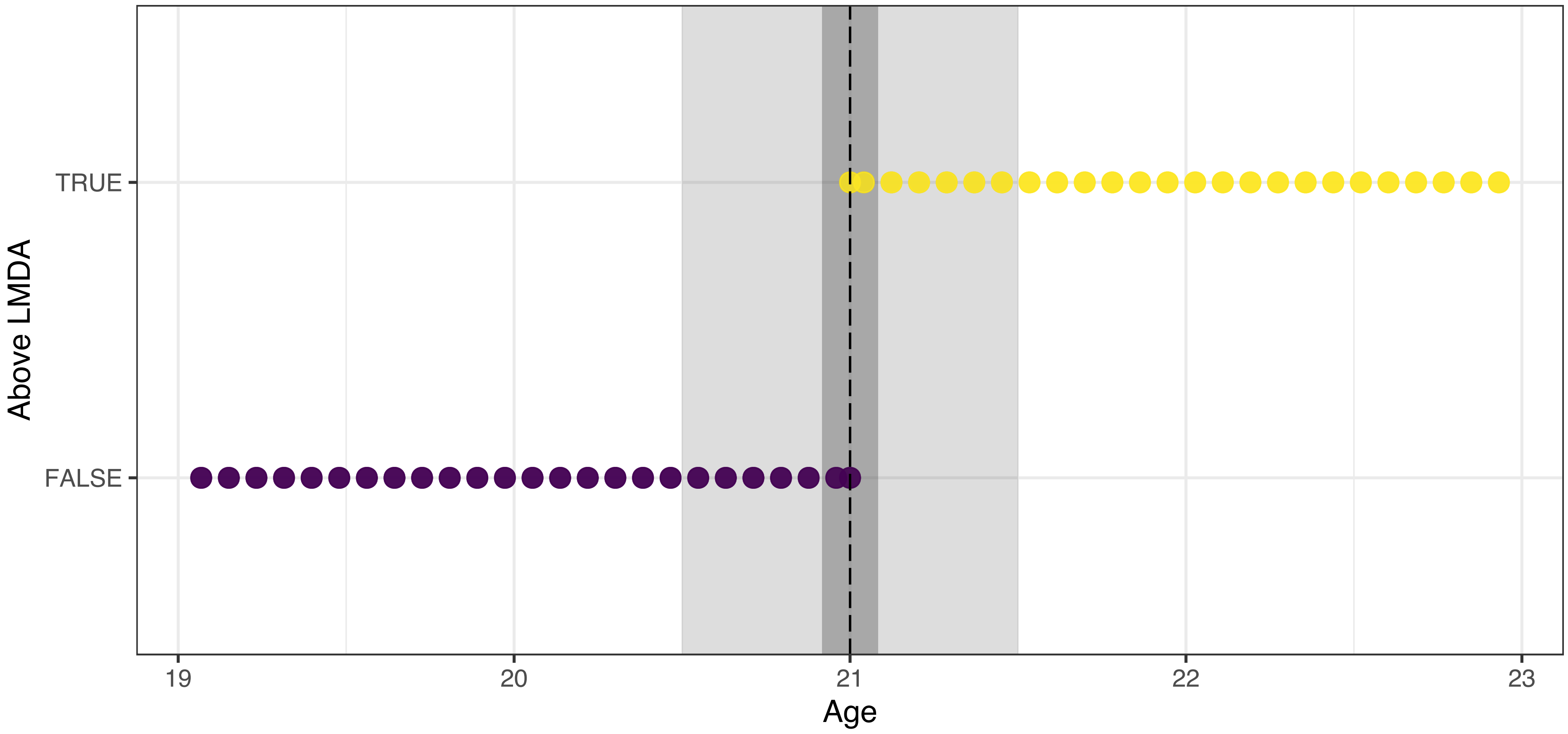
- This dataset contains aggregate death rates (and their causes) for different age groups (`agecell`) between 19 and 23 years old.

Sharp Discontinuity at Cutoff

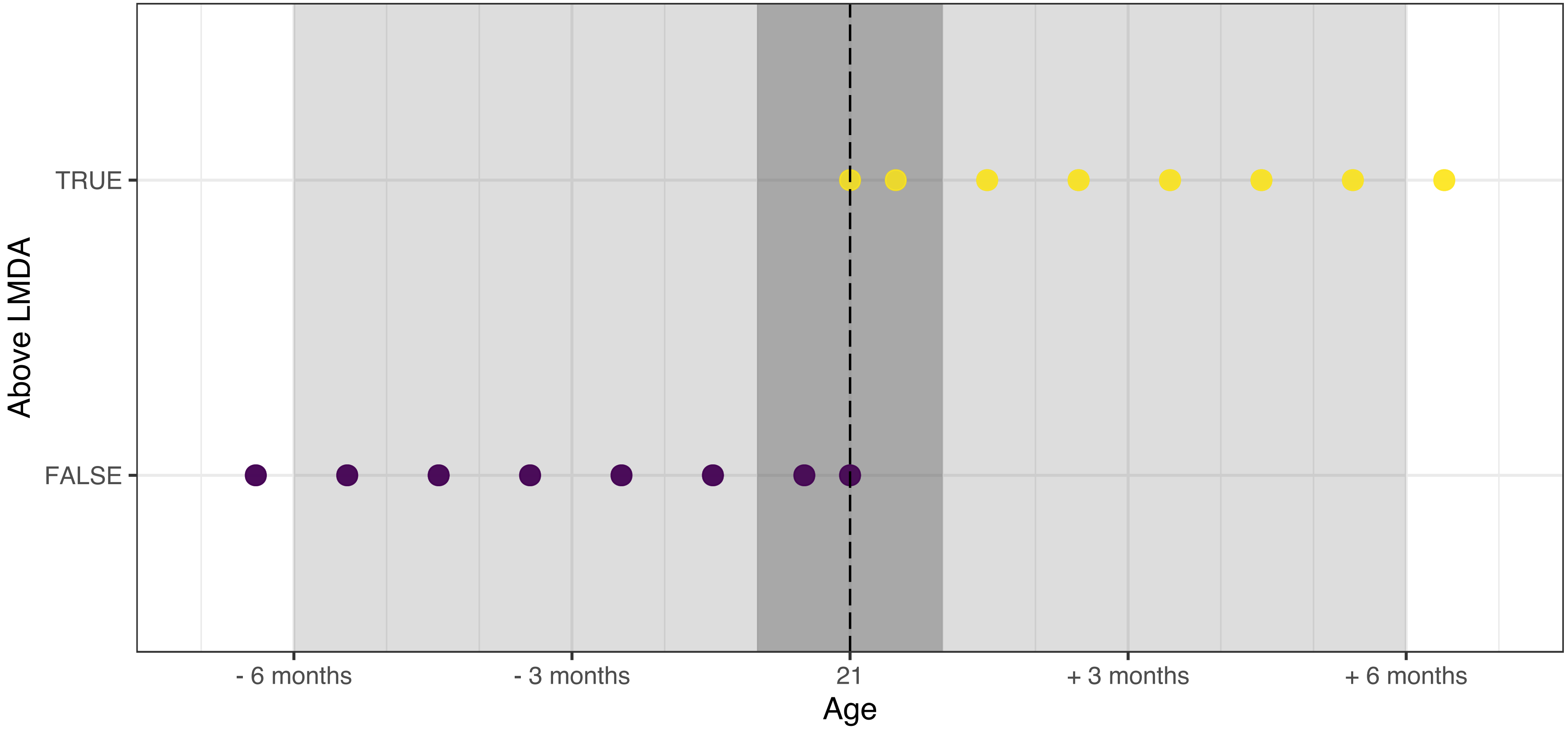


At the threshold, the probability of being treated jumps from 0 to 1.

Sharp Discontinuity at Cutoff



Sharp Discontinuity at Cutoff



RDD Framework

- ***Treatment variable***: D_a

RDD Framework

- ***Treatment variable***: D_a
    - $D_a = 1$ if individual is over 21 years old, $D_a = 0$ if not.

RDD Framework

- ***Treatment variable***: D_a
  - $D_a = 1$ if individual is over 21 years old, $D_a = 0$ if not.
  - D_a is a function of the individual's age, a, which is the ***running variable***.

RDD Framework

- ***Treatment variable***: D_a
  - D_a = 1 if individual is over 21 years old, D_a = 0 if not.
  - D_a is a function of the individual's age, a, which is the ***running variable***.
- The ***cutoff*** age, 21, separates those who can drink legally and those who can't:

$$D_a = \begin{cases} 1 & \text{if } a \geq 21 \\ 0 & \text{if } a < 21 \end{cases}$$

Key features of RD designs

1. Treatment status is a **deterministic** function of $a \rightarrow$ we know the assignment rule

RDD Framework

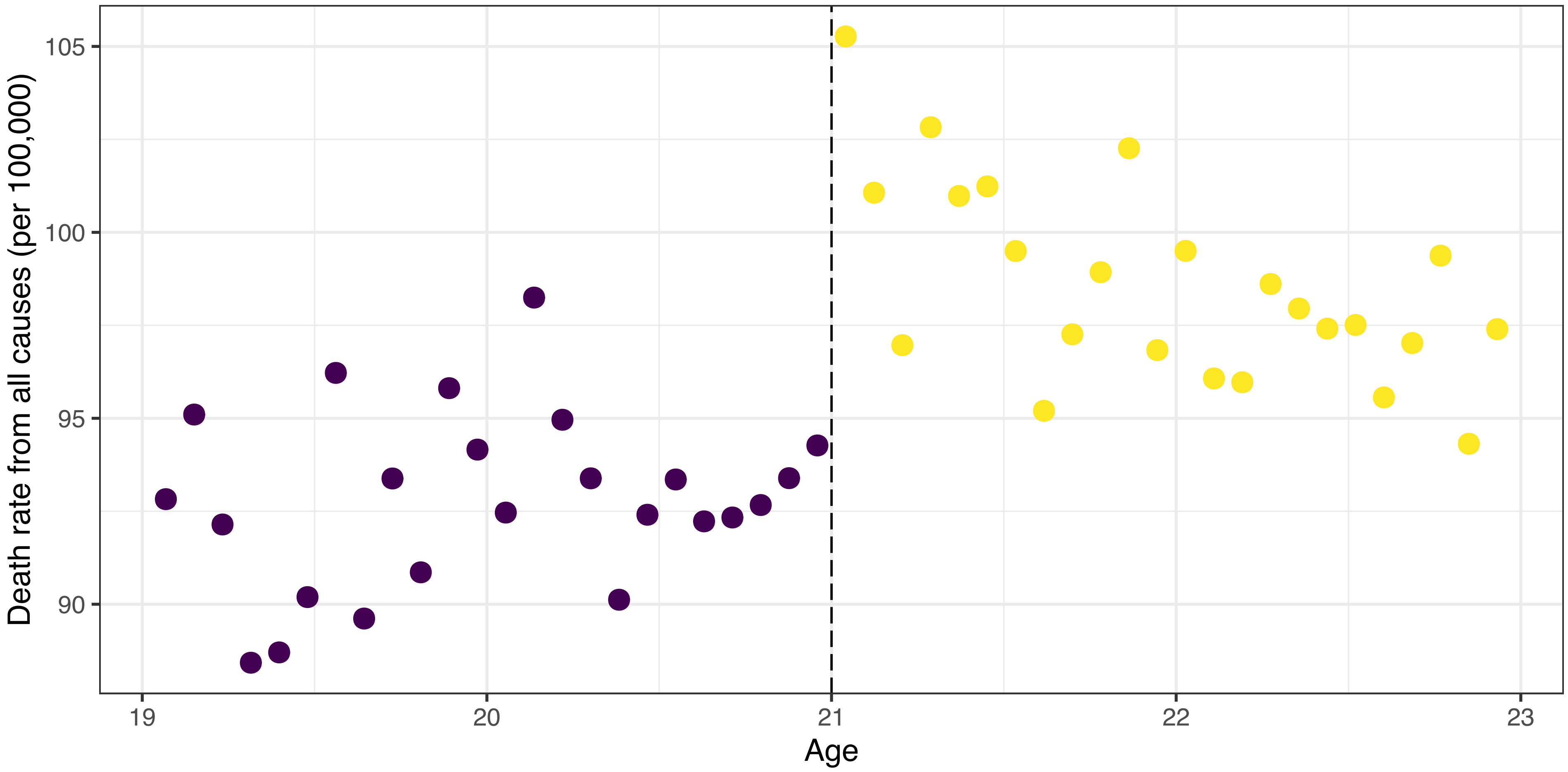
- ***Treatment variable***: D_a
    - D_a = 1 if individual is over 21 years old, D_a = 0 if not.
    - D_a is a function of the individual's age, a, which is the ***running variable***.
- The ***cutoff*** age, 21, separates those who can drink legally and those who can't:

$$D_a = \begin{cases} 1 & \text{if } a \geq 21 \\ 0 & \text{if } a < 21 \end{cases}$$

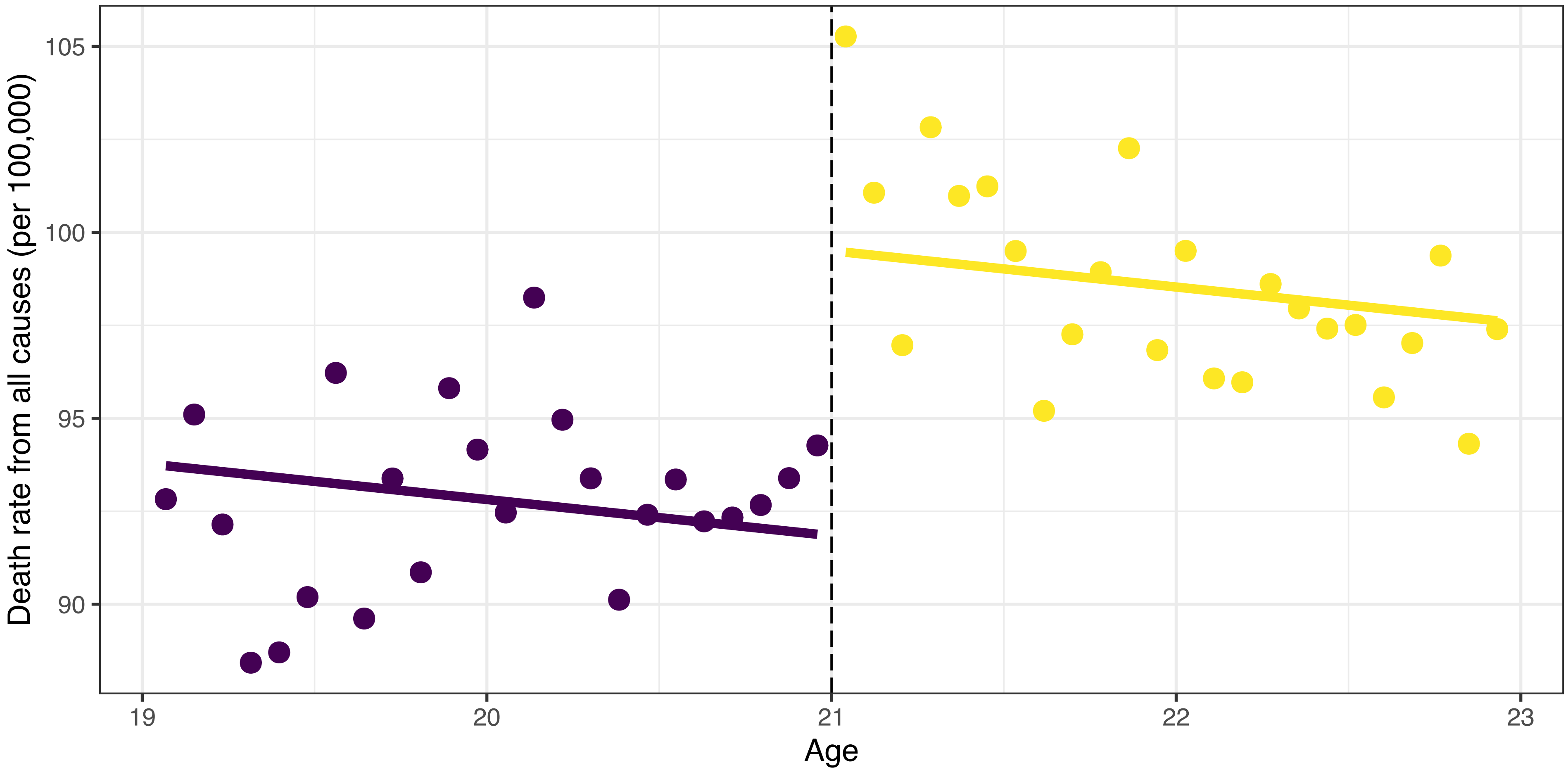
Key features of RD designs

1. Treatment status is a **deterministic** function of $a \rightarrow$ we know the assignment rule
2. Treatment status is a **discontinuous** function of $a \rightarrow$ there is some cutoff level

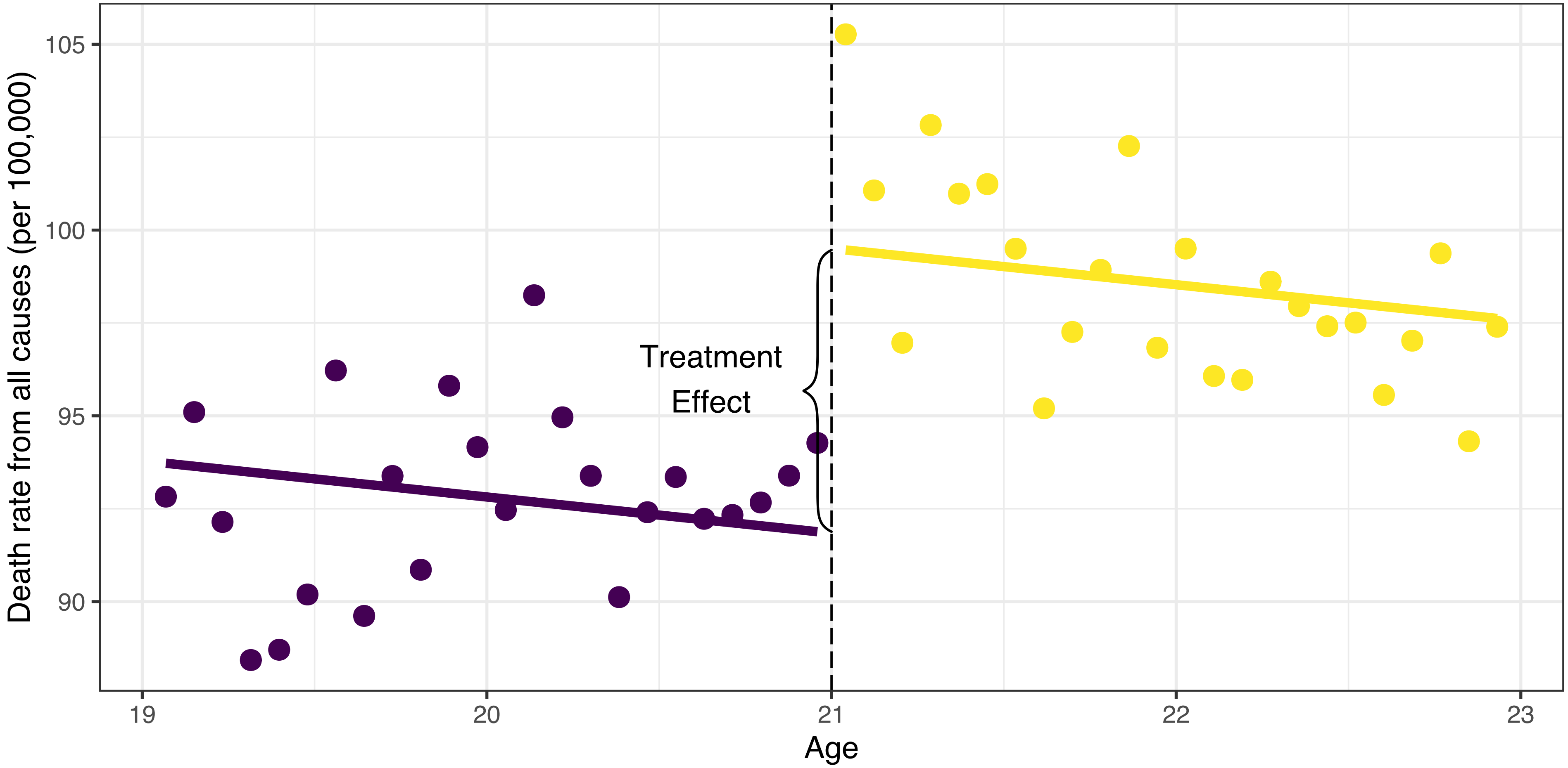
Graphical Results: All Death Rates



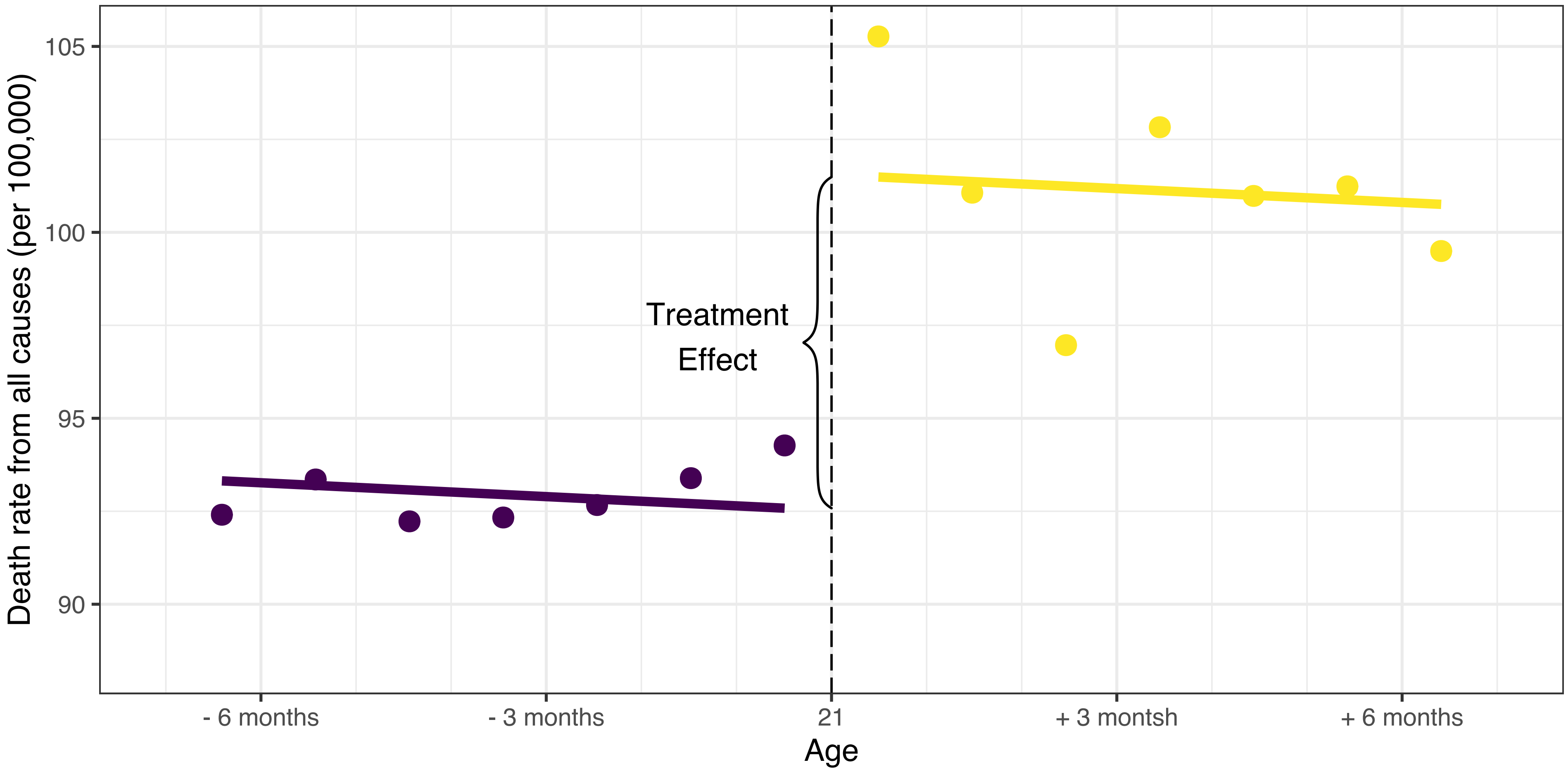
Graphical Results: All Death Rates



Graphical Results: All Death Rates



Graphical Results: All Death Rates



RDD as Local Average Treatment Effect (LATE)

- The RD estimator is a **local average treatment effect (LATE)**.

RDD as Local Average Treatment Effect (LATE)

- The RD estimator is a **local average treatment effect (LATE)**.
- It only tells you the impact of treatment D on outcome Y ***around*** the cutoff value of the running variable.

RDD as Local Average Treatment Effect (LATE)

- The RD estimator is a **local average treatment effect (LATE)**.
- It only tells you the impact of treatment D on outcome Y ***around*** the cutoff value of the running variable.
- Limited ***external validity*** $\rightarrow$ you cannot extrapolate to the entire population.

RDD as Local Average Treatment Effect (LATE)

- The RD estimator is a **local average treatment effect (LATE)**.
- It only tells you the impact of treatment D on outcome Y ***around*** the cutoff value of the running variable.
- Limited ***external validity*** $\rightarrow$ you cannot extrapolate to the entire population.
- Using the 21 year old alcohol restriction age in the RD context will only tell you the effect of this restriction on death rates but not the general effect of alcohol consumption.

RDD as Local Average Treatment Effect (LATE)

- The RD estimator is a **local average treatment effect (LATE)**.
- It only tells you the impact of treatment D on outcome Y ***around*** the cutoff value of the running variable.
- Limited ***external validity*** $\rightarrow$ you cannot extrapolate to the entire population.
- Using the 21 year old alcohol restriction age in the RD context will only tell you the effect of this restriction on death rates but not the general effect of alcohol consumption.
- One may easily argue that all results from quantitative empirical analyses have a local nature.

Estimation

Estimation

- *Objective:* measure ***gap*** between the two lines at the cutoff.

Estimation

- *Objective:* measure ***gap*** between the two lines at the cutoff.
- In its simplest form, we can write the following regression model:

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_i,$$

where $DEATHRATE_a$ is the death rate at age a, D_a is the treatment dummy, and a is age (defined in months relative to 21st birthday).

Estimation

- *Objective:* measure ***gap*** between the two lines at the cutoff.
- In its simplest form, we can write the following regression model:

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_i,$$

where $DEATHRATE_a$ is the death rate at age a, D_a is the treatment dummy, and a is age (defined in months relative to 21st birthday).

$\rightarrow \delta$ captures the **jump in death rate** between individuals above and below 21 years old.

Estimation

- *Objective:* measure ***gap*** between the two lines at the cutoff.
- In its simplest form, we can write the following regression model:

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_i,$$

where $DEATHRATE_a$ is the death rate at age a, D_a is the treatment dummy, and a is age (defined in months relative to 21st birthday).

$\rightarrow \delta$ captures the **jump in death rate** between individuals above and below 21 years old.

- The RDD estimator exploits a discontinuity at $a = 21$ in the conditional expectation function:

$$\underbrace{\lim_{c \to 21^+} \mathbb{E}[DEATHRATE_a | a = c]}_{\alpha+\delta} - \underbrace{\lim_{c \to 21^-} \mathbb{E}[DEATHRATE_a | a = c]}_{\alpha} = \delta$$

Task 2 (5 minutes)

1. Estimate the following model on `all` death causes.

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_i,$$

   Does the RDD coefficient correspond to the graphical illustration?

2. How do you interpret each coefficient?

3. What is the causal effect of legal access to alcohol on death rates?

Estimation #1: Simple Linear Model

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_a,$$

```r
mlda <- mlda %>%
  mutate(over21 = (agecell >= 21),
         agecell_21 = agecell - 21)
rdd <- lm(all ~ agecell_21 + over21, mlda)

library(broom)
tidy(rdd)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   91.8       0.805    114.   4.59e-57
## 2 agecell_21    -0.975     0.632     -1.54 1.30e- 1
## 3 over21TRUE     7.66      1.44       5.32 3.15e- 6
```

Estimation #1: Simple Linear Model

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_a,$$

```r
mlda <- mlda %>%
  mutate(over21 = (agecell >= 21),
         agecell_21 = agecell - 21)
rdd <- lm(all ~ agecell_21 + over21, mlda)

library(broom)
tidy(rdd)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   91.8       0.805    114.   4.59e-57
## 2 agecell_21    -0.975     0.632     -1.54 1.30e- 1
## 3 over21TRUE     7.66      1.44       5.32 3.15e- 6
```

Interpretation:

Estimation #1: Simple Linear Model

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_a,$$

```r
mlda <- mlda %>%
  mutate(over21 = (agecell >= 21),
         agecell_21 = agecell - 21)
rdd <- lm(all ~ agecell_21 + over21, mlda)

library(broom)
tidy(rdd)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   91.8       0.805    114.   4.59e-57
## 2 agecell_21    -0.975     0.632     -1.54 1.30e- 1
## 3 over21TRUE     7.66      1.44       5.32 3.15e- 6
```

Interpretation:

On average, the MLDA increases death rates from all causes by 7.66 percentage points.

Estimation #1: Simple Linear Model

$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \varepsilon_a,$$

```
mlda <- mlda %>%
  mutate(over21 = (agecell >= 21),
         agecell_21 = agecell - 21)
rdd <- lm(all ~ agecell_21 + over21, mlda)

library(broom)
tidy(rdd)
```

```
## # A tibble: 3 x 5
##   term        estimate std.error statistic  p.value
##   <chr>          <dbl>     <dbl>     <dbl>    <dbl>
## 1 (Intercept)   91.8       0.805    114.   4.59e-57
## 2 agecell_21    -0.975     0.632     -1.54 1.30e- 1
## 3 over21TRUE     7.66      1.44       5.32 3.15e- 6
```

Interpretation:

On average, the MLDA increases death rates from all causes by 7.66 percentage points.

This is a big effect considering the average death rate for individuals between 19 and 22 is:

```
mean(mlda$all, na.rm = TRUE)
```

```
## [1] 95.67272
```

Estimation Issues

- The ***functional form*** used to approximate the lines really matters!

Estimation Issues

- The ***functional form*** used to approximate the lines really matters!

  $\longrightarrow$ an insufficiently flexible specification runs the risk of mistaking nonlinearity for treatment effect;

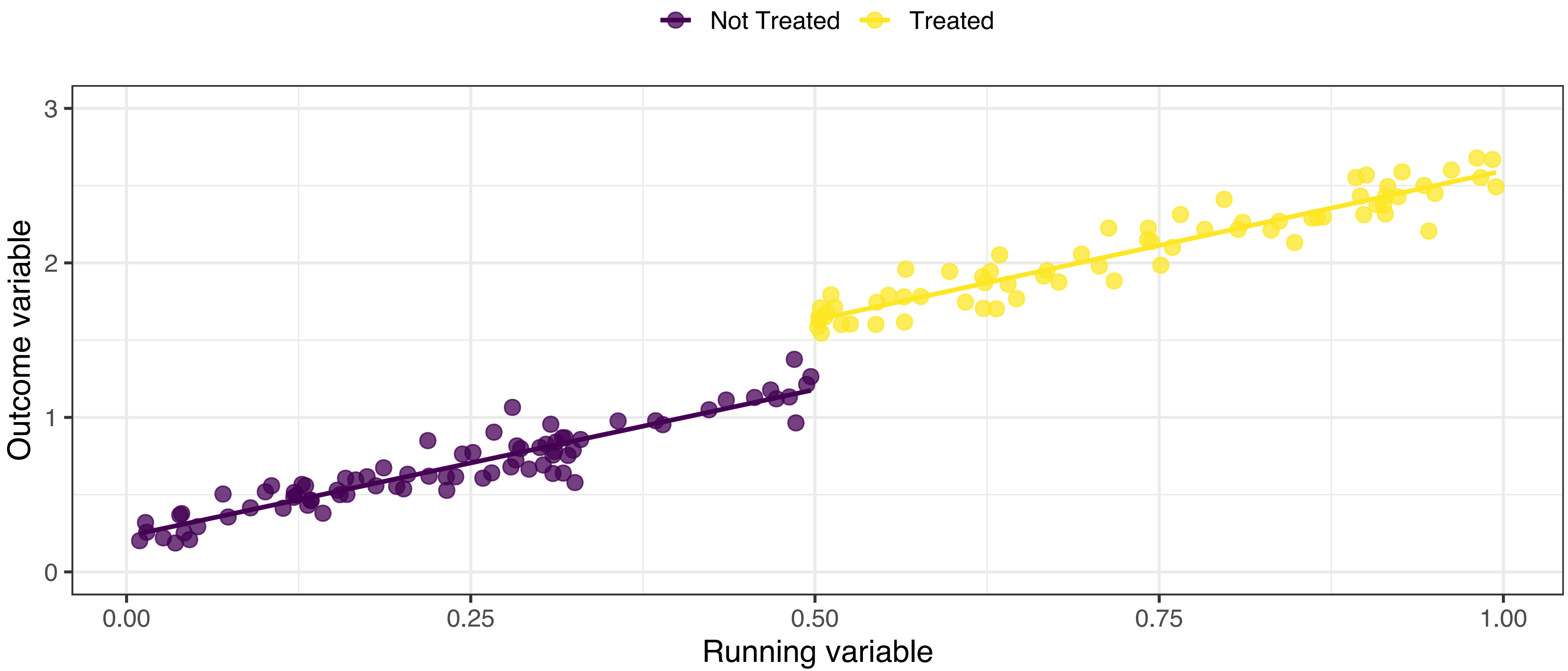
Estimation Issues

- The *functional form* used to approximate the lines really matters!

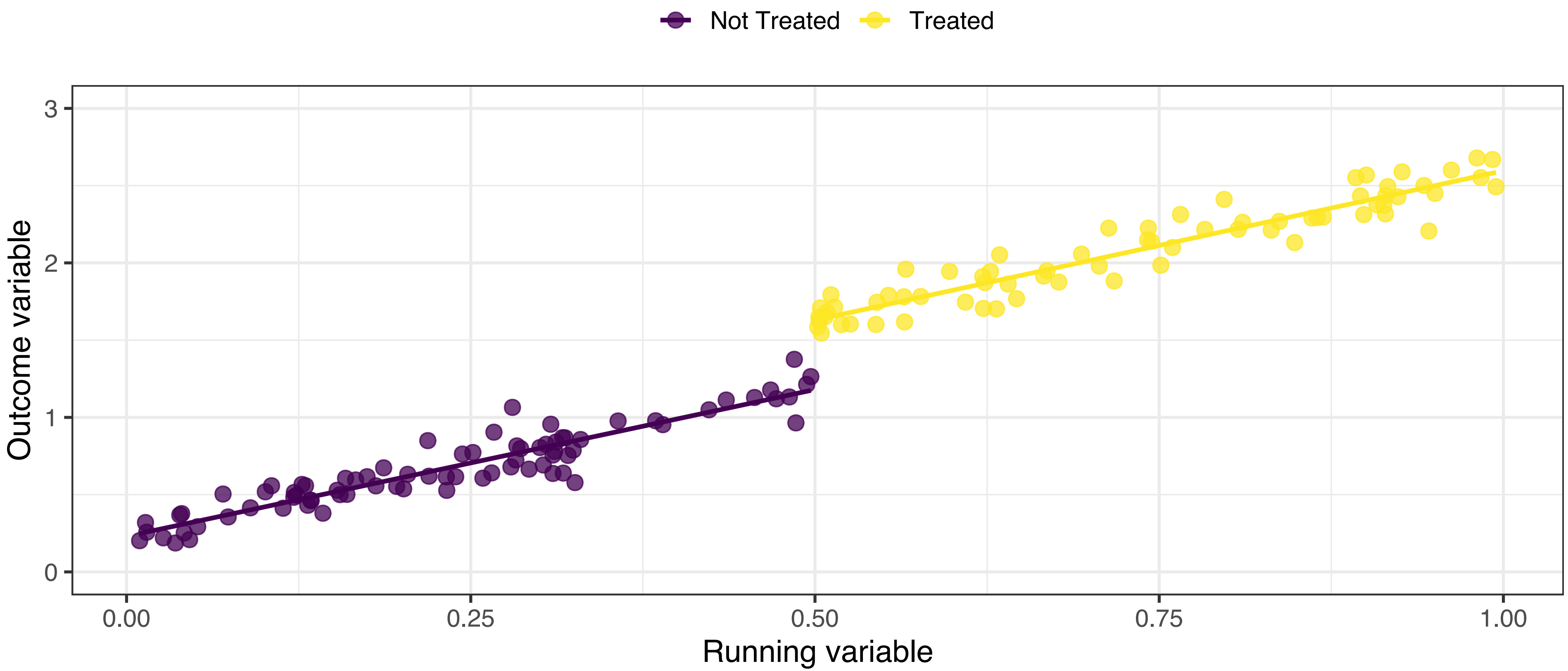
  $\rightarrow$ an insufficiently flexible specification runs the risk of mistaking nonlinearity for treatment effect;

  $\rightarrow$ an overly flexible specification reduces precision and runs the risk of overfitting.

Simulations - Linear Relationship and Clear Discontinuity

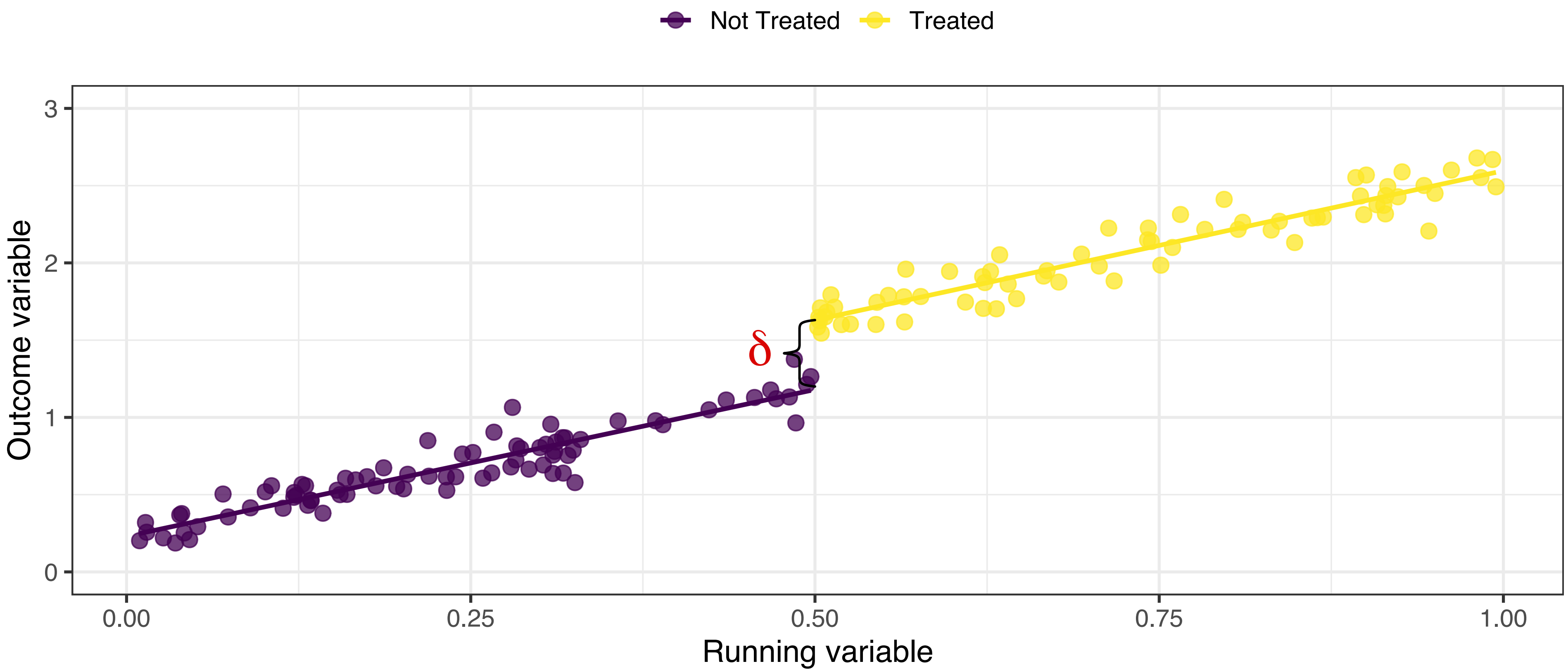


Simulations - Linear Relationship and Clear Discontinuity



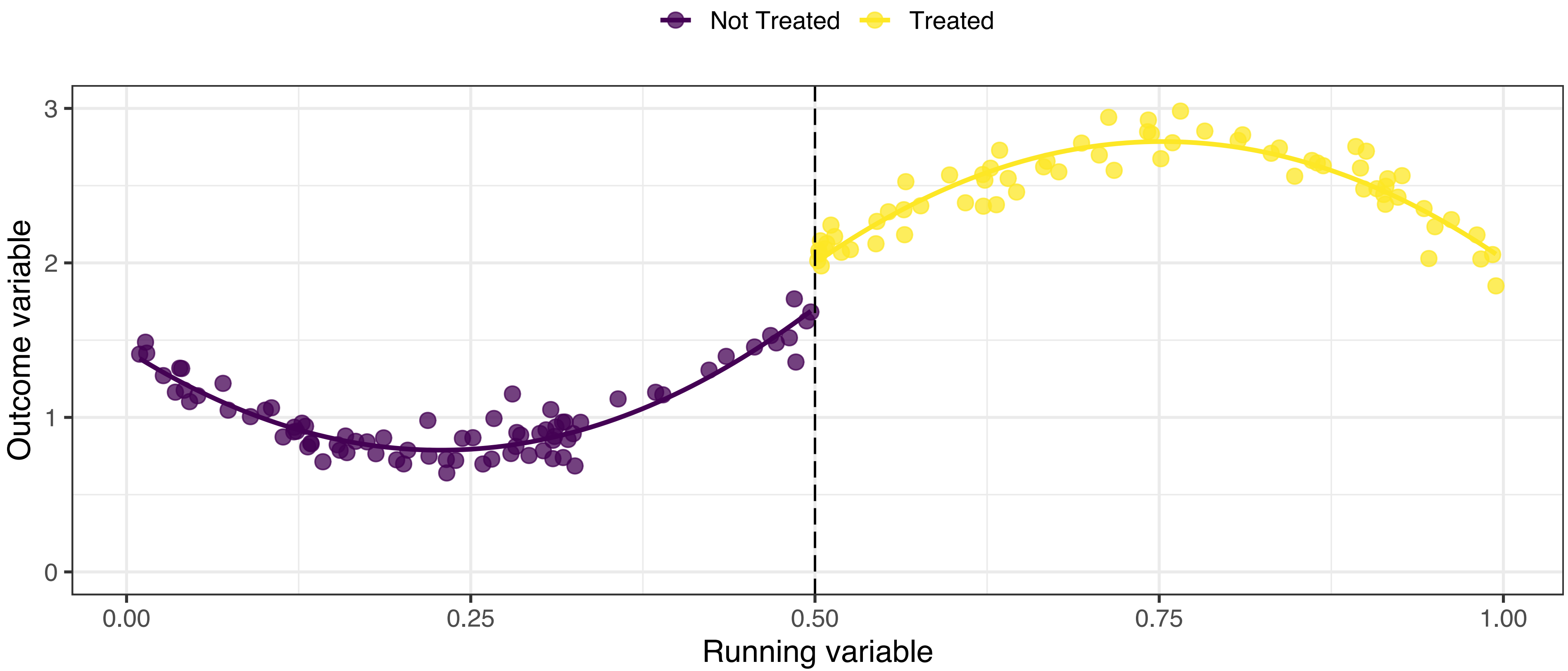
$$outcome_i = \alpha + \delta treatment_i + \beta running_i + e_i,$$

Simulations - Linear Relationship and Clear Discontinuity

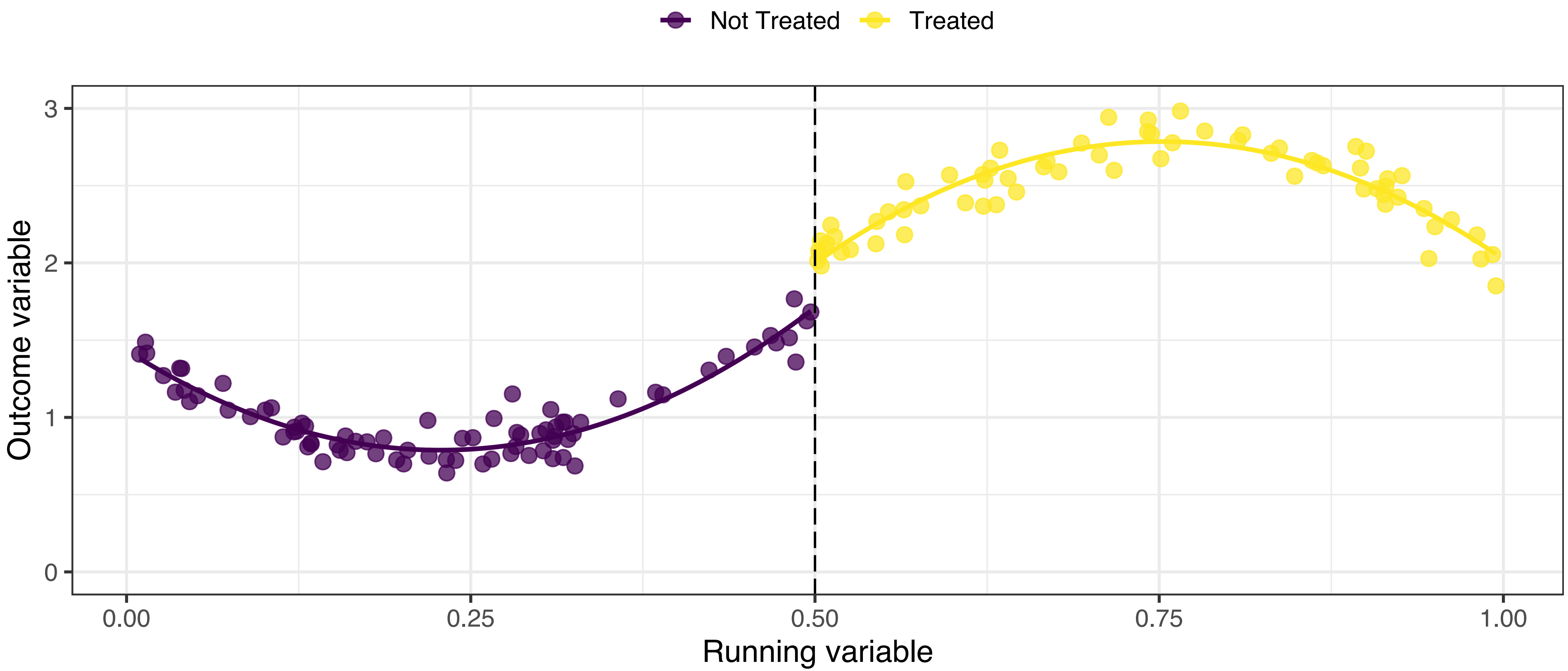


$$outcome_i = \alpha + \delta treatment_i + \beta running_i + e_i,$$

Simulations - Quadratic Relationship and Clear Discontinuity

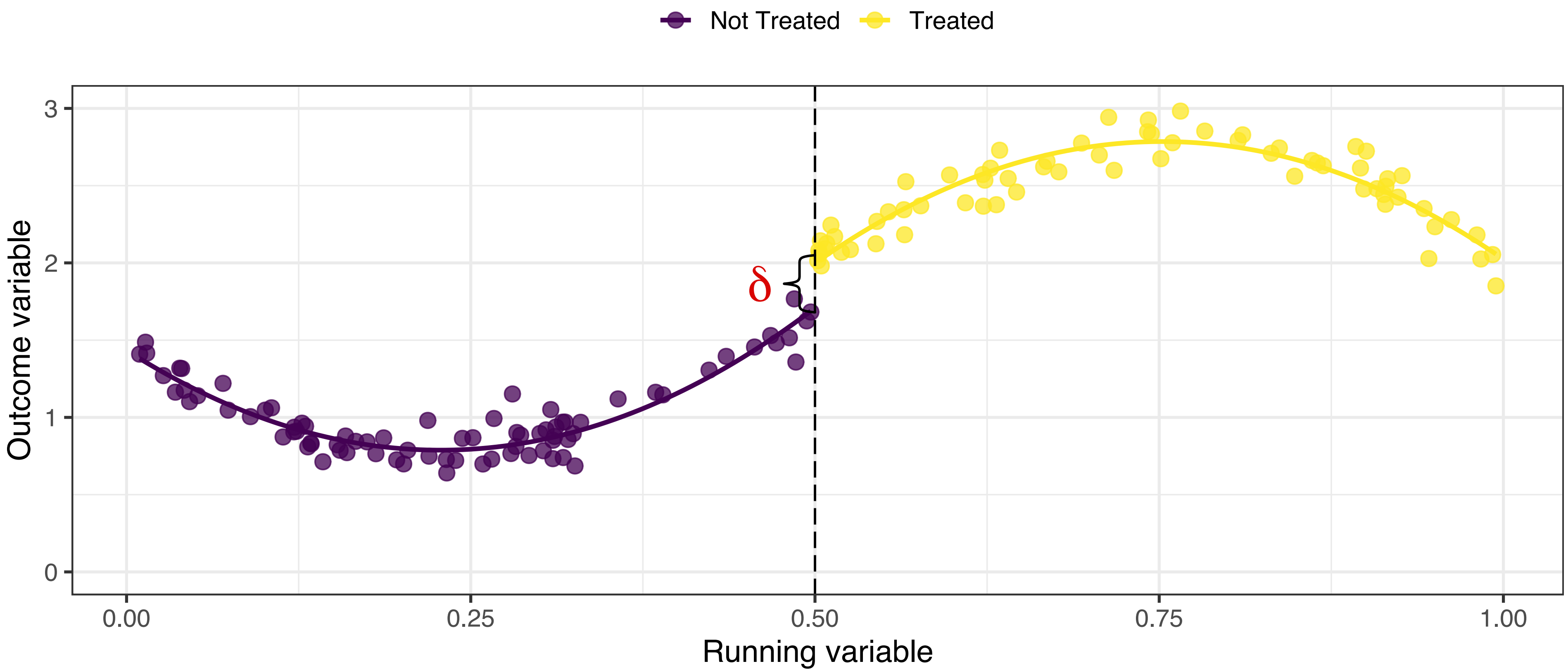


Simulations - Quadratic Relationship and Clear Discontinuity



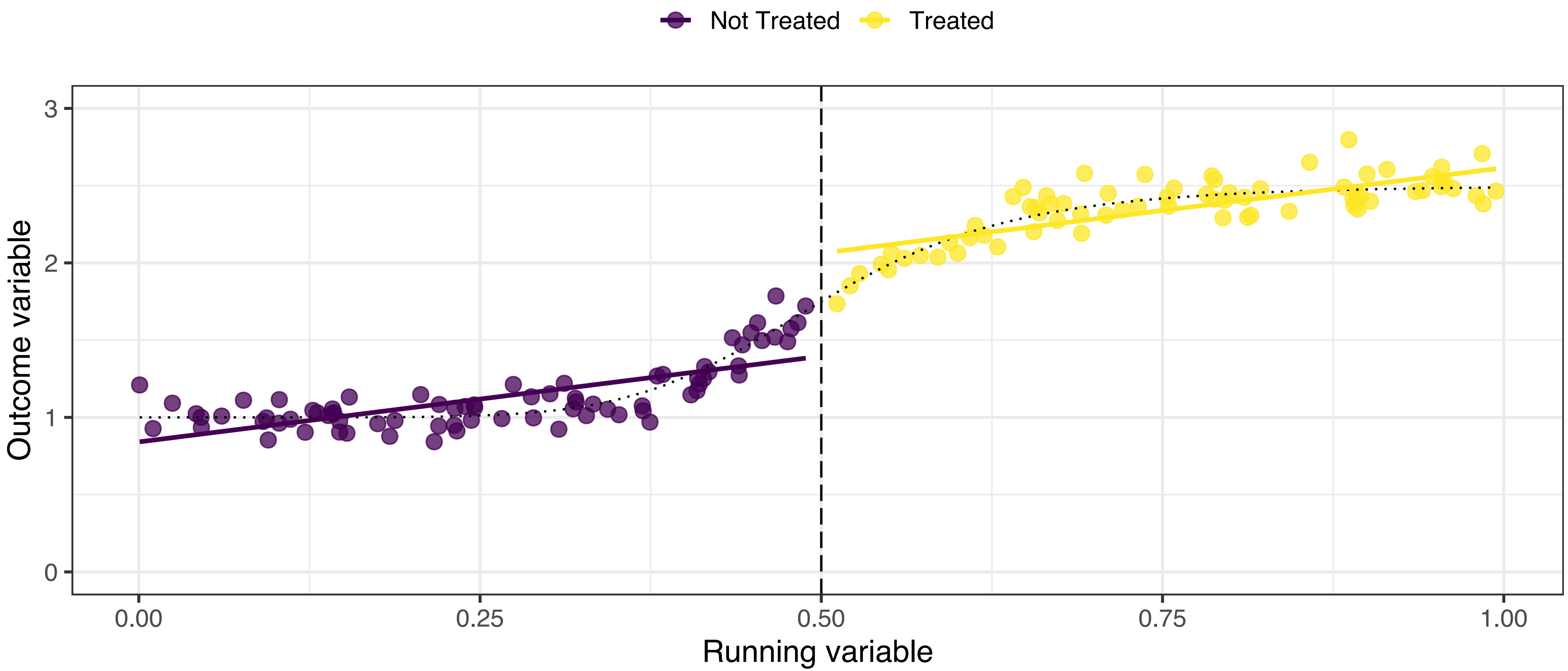
$$outcome_i = \alpha + \delta treatment_i + \beta_1 running_i + \beta_2 running_i^2 + e_i,$$

Simulations - Quadratic Relationship and Clear Discontinuity

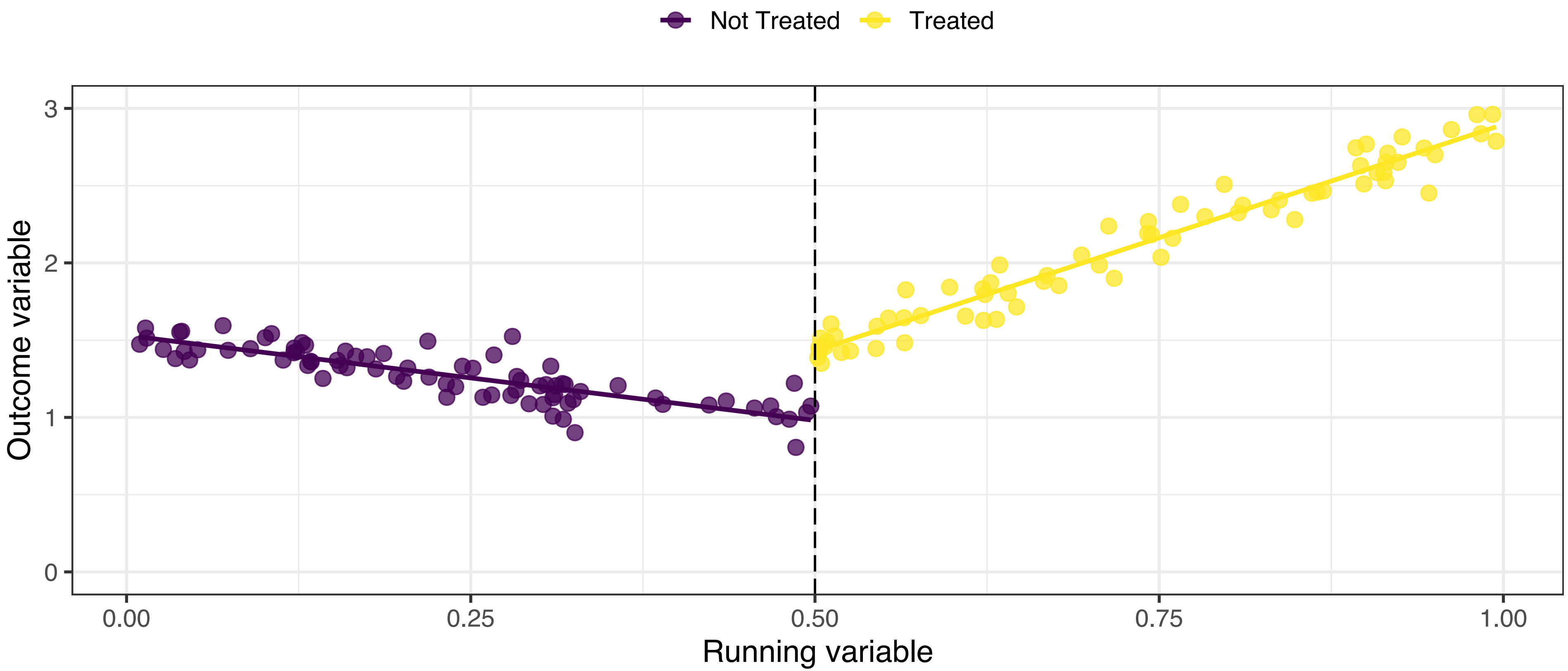


$$outcome_i = \alpha + \delta treatment_i + \beta_1 running_i + \beta_2 running_i^2 + e_i,$$

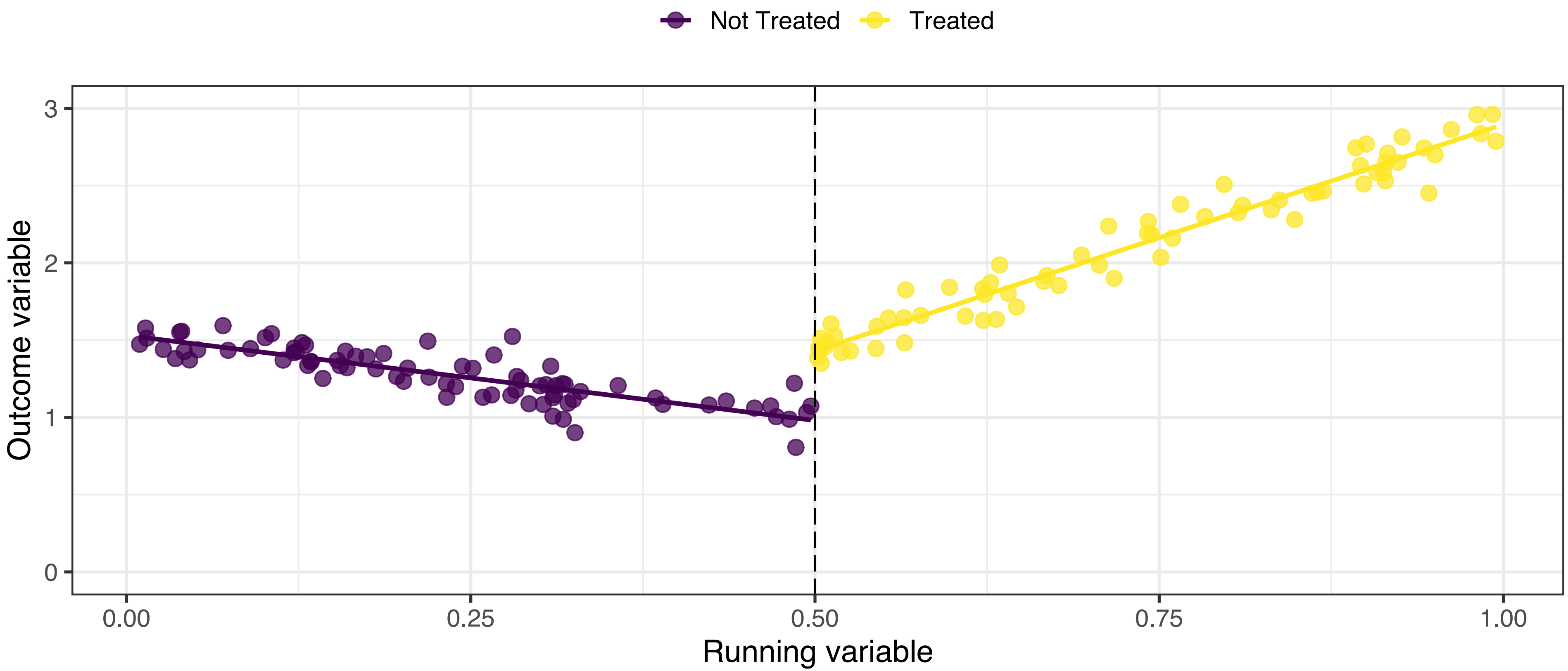
Simulations - Linear Relationship but NO Discontinuity



Simulations - Different (Linear) Slopes

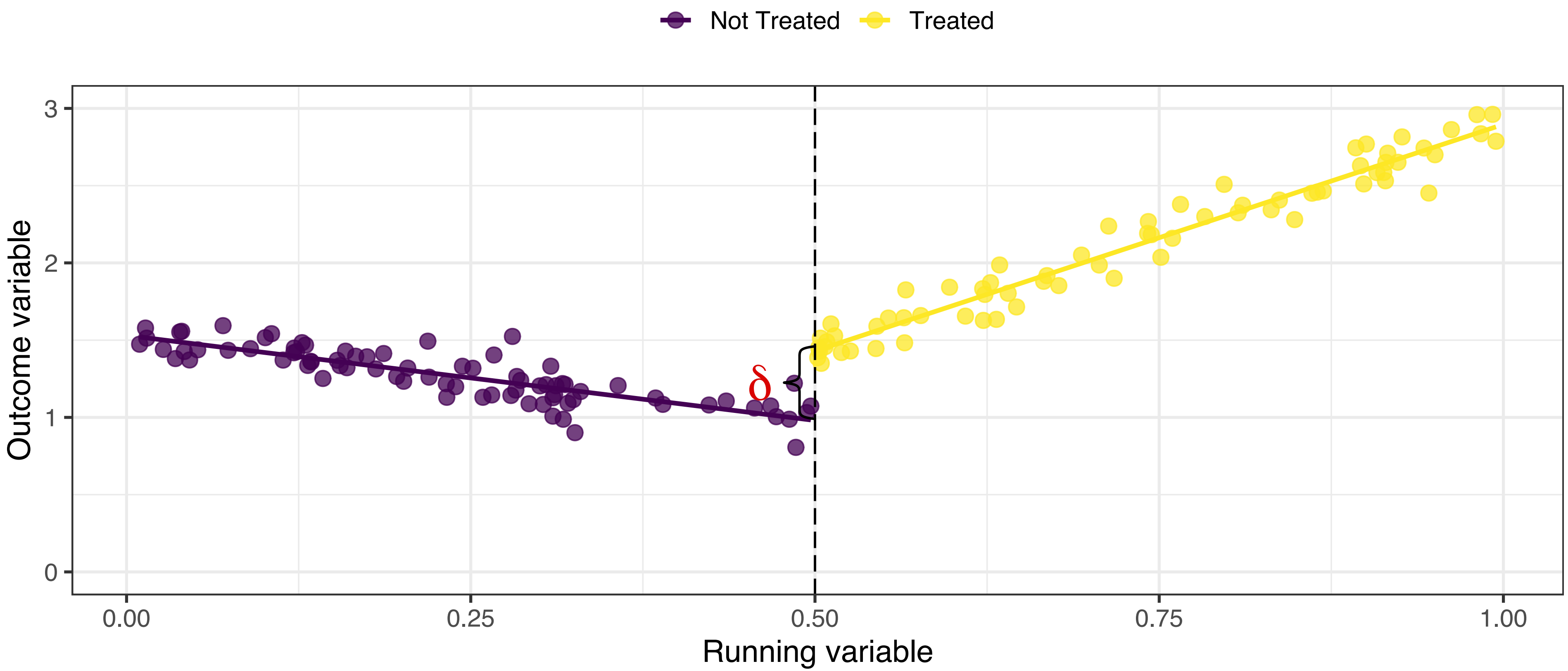


Simulations - Different (Linear) Slopes



$$outcome_i = \alpha + \delta treatment_i + \beta(running_i - cutoff) +$$
$$\color{red}{\gamma treatment_i * (running_i - cutoff)} + e_i,$$

Simulations - Different (Linear) Slopes



$$outcome_i = \alpha + \delta treatment_i + \beta(running_i - cutoff) + \gamma treatment_i * (running_i - cutoff) + e_i,$$

How to Choose Appropriate Functional Form?

- Essential to **visualise** the data!

How to Choose Appropriate Functional Form?

- Essential to **visualise** the data!
- Coefficients across models shouldn't vary too much.

How to Choose Appropriate Functional Form?

- Essential to **visualise** the data!
- Coefficients across models shouldn't vary too much.
- Should we expect the relationship between the outcome variable and the running variable to be nonlinear? Should we expect it to differ around the cutoff?

How to Choose Appropriate Functional Form?

- Essential to **visualise** the data!
- Coefficients across models shouldn't vary too much.
- Should we expect the relationship between the outcome variable and the running variable to be nonlinear? Should we expect it to differ around the cutoff?
- Gelman and Imbens (2019), "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs":
  "We recommend researchers [...] use estimators based on local linear or quadratic polynomials or other smooth functions."

Task 3 (10 minutes)

1. Estimate the following *quadratic* model on `all` death causes. Does the RDD coefficient differ from the linear model?

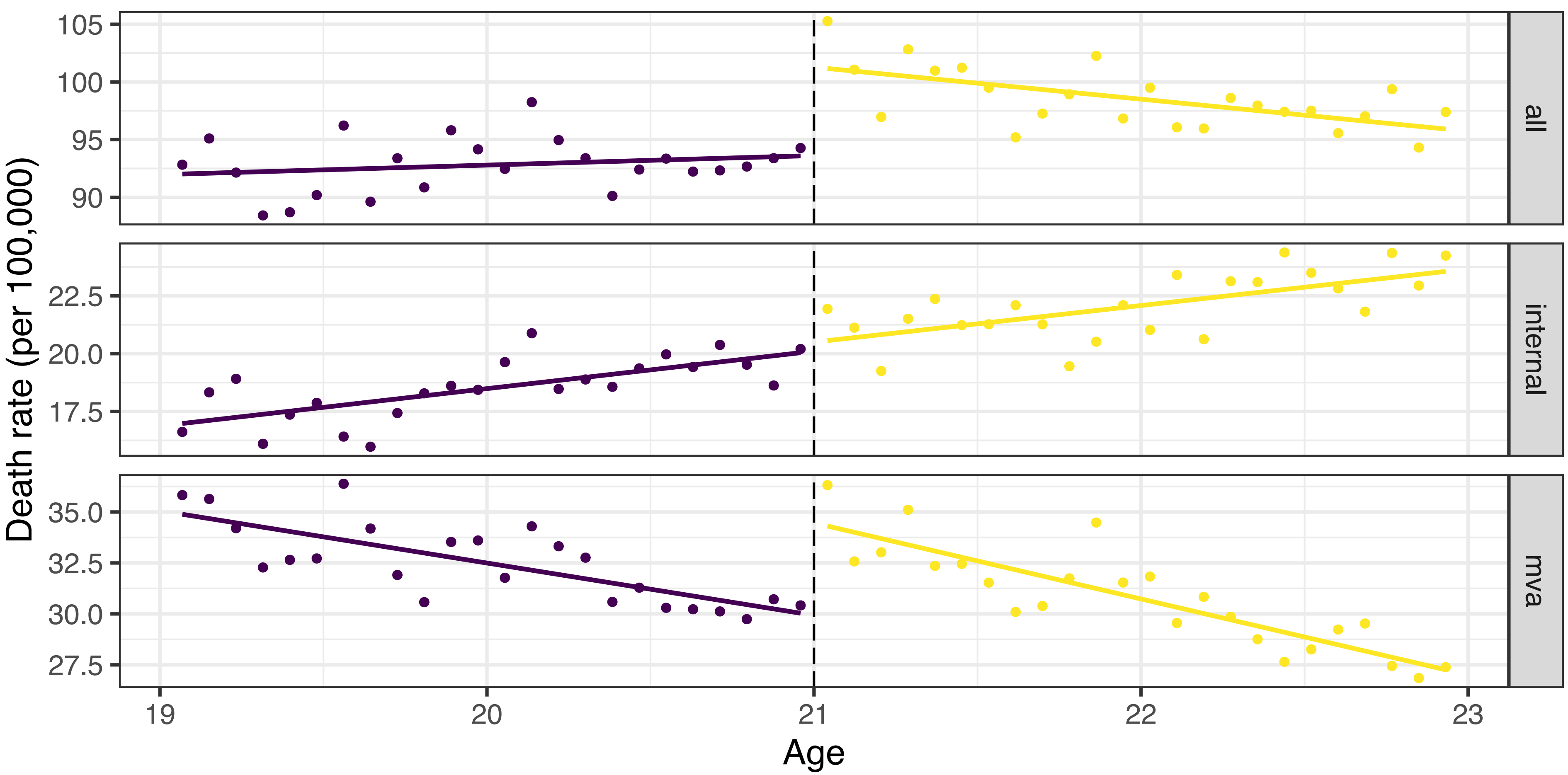
$$DEATHRATE_a = \alpha + \delta D_a + \beta a + \beta a^2 + \varepsilon_a,$$

2. Recall that the regression model allowing for different slopes on each side of the cutoff is:

$$DEATHRATE_a = \alpha + \delta D_a + \beta(a - 21) + \gamma D_a * (a - 21) + \varepsilon_a,$$

   - Why do we need to substract the 21 from `a`? (Hint: compute $\mathbb{E}(DEATHRATE_a | a = 21)$)
   - Should we expect the relationship between death rates and age to change at 21?
   - Estimate this model. How different is the RDD coefficient from the other models you have estimated?

Graphical Representation of the Regression Results



Nonparametric Estimation

- Give more weight to observations close to the cutoff level

Nonparametric Estimation

- Give more weight to observations close to the cutoff level

2 settings:

- How much more weight?

Nonparametric Estimation

- Give more weight to observations close to the cutoff level

2 settings:

- How much more weight?

  $\rightarrow$ depends on the chosen ***kernel***.

Nonparametric Estimation

- Give more weight to observations close to the cutoff level

2 settings:

- How much more weight?

  $\rightarrow$ depends on the chosen ***kernel***.

- How far away from the cutoff do observations need to be to be discarded?

Nonparametric Estimation

- Give more weight to observations close to the cutoff level

2 settings:

- How much more weight?

  $\longrightarrow$ depends on the chosen ***kernel***.

- How far away from the cutoff do observations need to be to be discarded?

  $\longrightarrow$ depends on the chosen ***bandwidth***.

Nonparametric Estimation

- Give more weight to observations close to the cutoff level

2 settings:

- How much more weight?

  $\rightarrow$ depends on the chosen ***kernel***.

- How far away from the cutoff do observations need to be to be discarded?

  $\rightarrow$ depends on the chosen ***bandwidth***.

Luckily there's an `R` package that chooses these settings optimally based on fancy algorythms: `rdrobust`.

Identifying Assumptions

RDD Assumptions

> *Key assumption*: ***Potential outcomes are smooth at the threshold.***

RDD Assumptions

> *Key assumption*: ***Potential outcomes are smooth at the threshold.***

$\longrightarrow$ assignment variable cannot be manipulated!

RDD Assumptions

> *Key assumption*: ***Potential outcomes are smooth at the threshold.***

$\rightarrow$ assignment variable cannot be manipulated!

- The population just below must not be discretely different from the population just above the cutoff.

RDD Assumptions

> *Key assumption*: ***Potential outcomes are smooth at the threshold.***

$\rightarrow$ assignment variable cannot be manipulated!

- The population just below must not be discretely different from the population just above the cutoff.
- Assumption is violated if people can manipulate the running variable because they know the cutoff value.

RDD Assumptions

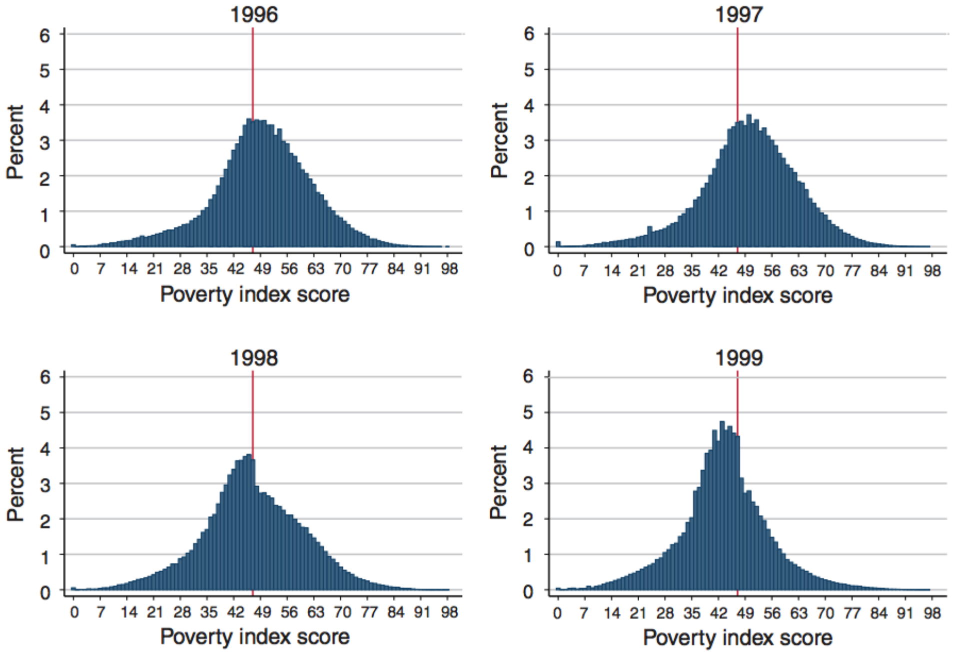
> *Key assumption*: ***Potential outcomes are smooth at the threshold.***

$\rightarrow$ assignment variable cannot be manipulated!

- The population just below must not be discretely different from the population just above the cutoff.
- Assumption is violated if people can manipulate the running variable because they know the cutoff value.
  - Knowing the cutoff value in itself does not violate the assumption, only ability to manipulate running variable does.

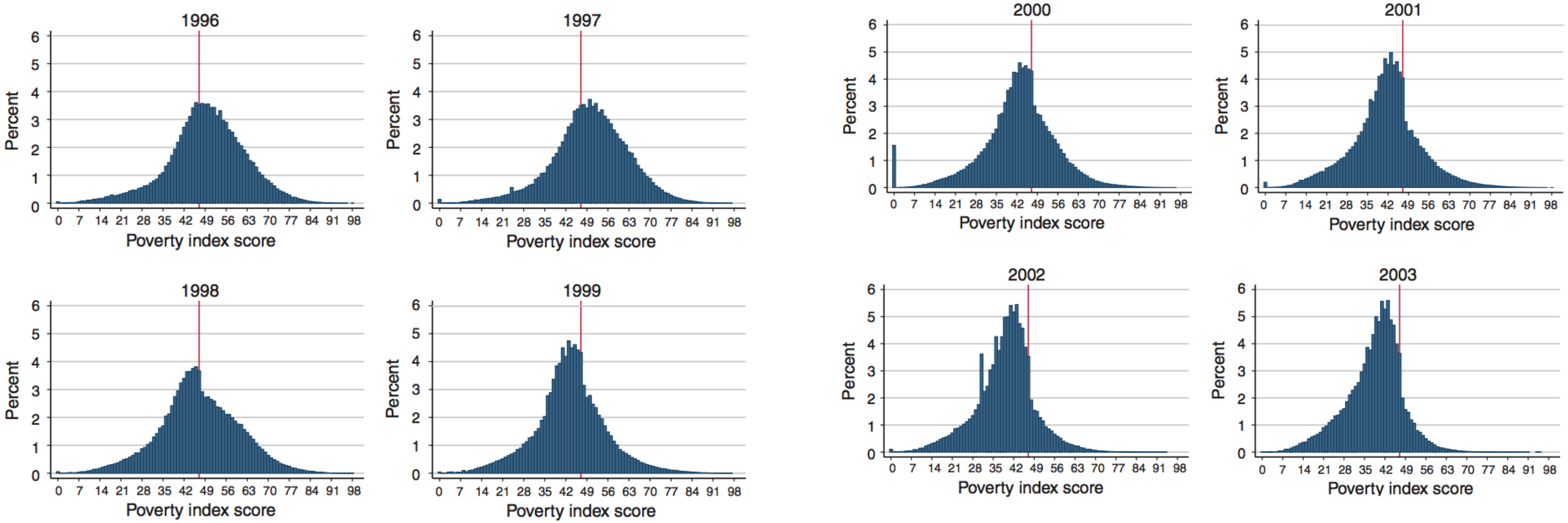
Example of Manipulation: Camacho and Conover (2011)

What happens when threshold for eligibility to social assistance programs becomes known?



Example of Manipulation: Camacho and Conover (2011)

What happens when threshold for eligibility to social assistance programs becomes known?



Noncompliance

What if the running variable does not *fully* determine assignment to treatment?

$\rightarrow$ ***Fuzzy RDD***

- Even if all observations that satisfy the treatment condition are not treated, there is still a jump in the probability of being treated.
- For you, just know that problem of imperfect determination of allocation to treatment can still be solved

5 Steps for Conducting RDD in Practice[1]

Step #1: ***Is assignment to treatment rule-based?***

[1] Taken from Andrew Heiss' wonderful course on RDD.

5 Steps for Conducting RDD in Practice[1]

Step #1: ***Is assignment to treatment rule-based?***

Step #2: ***Is design sharp or fuzzy?***

[1] Taken from Andrew Heiss' wonderful course on RDD.

5 Steps for Conducting RDD in Practice[1]

Step #1: *Is assignment to treatment rule-based?*

Step #2: *Is design sharp or fuzzy?*

Step #3: *Is there a discontinuity in running variable at cutoff?*

[1] Taken from Andrew Heiss' wonderful course on RDD.

5 Steps for Conducting RDD in Practice[1]

Step #1: *Is assignment to treatment rule-based?*

Step #2: *Is design sharp or fuzzy?*

Step #3: *Is there a discontinuity in running variable at cutoff?*

Step #4: *Is there a discontinuity in outcome variable at cutoff in running variable?*

[1] Taken from Andrew Heiss' wonderful course on RDD.

5 Steps for Conducting RDD in Practice[1]

Step #1: *Is assignment to treatment rule-based?*

Step #2: *Is design sharp or fuzzy?*

Step #3: *Is there a discontinuity in running variable at cutoff?*

Step #4: *Is there a discontinuity in outcome variable at cutoff in running variable?*

Step #5: *How big is the gap?*

[1] Taken from Andrew Heiss' wonderful course on RDD.

SciencesPo
DEPARTMENT OF ECONOMICS

END

	florian.oswald@sciencespo.fr
	Slides
	Book
	@ScPoEcon
	@ScPoEcon