

# **ECON 672**

## **Week 8: Difference-in-Differences**

**Samuel Rowe, PhD 1/19/2023**

# Overview

- The Takeaway
- Background of Diff-in-Diff
- Estimator
- DD Example: Minimum Wage
- Sample Averages vs Regressions
- Inference
- Testing Assumptions
- Diff-in-Diff and Triple Difference Examples

# Overview

## Difference-in-Differences

- The difference-in-differences estimator (or diff-in-diff) is a very common estimator used in academic research
  - Not utilized enough in federal evaluation
- It can handle observed and unobserved heterogeneity/confounders
  - Fixed Effects estimator only handles time-invariant confounders
  - Diff-in-diff can handle both time-invariant and time-varying confounders, as long as the parallel assumption holds

# The Takeaway

## Difference-in-Differences

- Strengths
  - An identification strategy that can handle both time-invariant and time-varying confounders both observed and unobserved
  - It's a powerful, yet simple identification strategy
  - We can indirectly test the main assumption, but not quite as robust as RDD
- Weaknesses
  - The diff-in-diff estimator only estimates the ATT and not the ATE
  - We have a dichotomous issue with our bias-variance tradeoff
    - When we have more than 1 treatment group, we may have biased coefficient if adoption time varies
    - When we have only 1 treatment group, we may have inconsistent and biased errors and higher rejection rate of a true null (Type I Errors)

# The Takeaway

## Difference-in-Differences

- Assumptions
  - Parallel Trends assumption is our main assumption
    - The trends of the control group are the true counterfactual trends of the treatment group
  - Stable Unit Treatment Value Assumption
    - There is no spillover from treatment onto control group
    - This is a bit more concern with triple difference-in-difference-in-differences due to contamination within-treatment group

# The Takeaway

## Difference-in-Differences

- Testable Assumptions
  - Parallel trends assumption is not directly testable, since we can never observe the counterfactual
  - We have three types of indirect tests of the parallel trends assumption
    - Pre-treatment balance - plotting or descriptive statistics
    - Event Studies - pre-treatment leads and lags
    - Placebo - Falsification tests including Triple DDD

# Thorny Issue

## Difference-in-Differences

- Just restating the weakness
  - If we only have 1 or 2 treatment groups, our standard errors will likely be inconsistent and biased
    - This will result in rejecting the true null hypothesis
  - However, our estimate will be biased if we have multiple treatment groups that vary in time adoption
    - If treatment is adopted at the same time, then this is less of a concern

# **Background Difference-in-Differences**

# Background

## Difference-in-Differences

- John Snow (1855) famous cholera study
  - He utilized a rudimentary difference-in-differences study design to study London's cholera outbreaks
  - Snow had a theory that cholera was not an inanimate object but a living organism that entered the body through food and water
  - Snow's results help show that cholera is transmitted through water and not air
- Most good Diff-in-Diff come from “natural experiments”
  - There is a naturally occurring variation in treatment that affects only some units over time

# Background

## Difference-in-Differences

- In 1800s London, there were several water companies that supplied water to different parts of the city
  - They typically took water from the Thames, which was polluted from human waste including bodily waste
- In 1849, the Lambeth company moved their intake pipes upriver along the Thames before human waste polluted the water
  - Snow utilizes this change to observe if cholera changes for customers of Lambeth
  - If cholera is water transmitted then Lambeth should have lower death rates of cholera

# Background

## Difference-in-Differences

- Snow uses Southwark and Vauxhall Waterworks Company as the counterfactual for Lambeth
  - Snow documents the similarities between Southwark/Vauxhall and Lambeth, so Southwark/Vauxhall is a true counterfactual of Lambeth
- Snow meticulously collected data on water consumption company going door to door, and he matches households with city level cholera death
- The only difference should be what water the households drank between treatment (Lambeth) and control (Southwark/Vauxhall)

# Background

## Difference-in-Differences

- In 1849, there were 135 cases of cholera out of 10,000 households for Southwark/Vauxhall and 85 cases out of 10,000 households for Lambeth
- In 1854, there were 147 cases per 10,000 households for Southwark/Vauxhall and 19 cases per 10,000 households for Lambeth
- We can calculate an ATT of 78 fewer deaths per household
- This resulted in the reject of cholera as an airborne disease and saved lives

Company name	1849	1854
Southwark and Vauxhall	135	147
Lambeth	85	19

# Estimator

## Difference-in-Differences

- Our simple case Difference-in-Difference estimator will be a 2-by-2 Diff-in-Diff
  - We will get into more complex Diff-in-Diff estimators next week
  - We are usually using panel data not cross-sectional data
  - You can use cross-sectional data but you run the risk of compositional changes
- The basis
  - Difference before and after in the treatment group
  - Difference before and after in the control group
  - Difference the differences to get an estimate of ATT

# Estimator

## Difference-in-Differences

- We need repeated observations over time for both the treatment and control
  - We use the first difference to get rid of **observed and unobserved time-invariant confounders** (like Fixed Effects)
  - We use the second difference to control for **observed and unobserved time-varying confounders** (with the parallel trends assumption)
- We eliminate selection bias with the pre-post comparisons
  - As long as the parallel trends assumption holds

# Estimator

## Difference-in-Differences

- The basis of 2x2 DD
  - We have a mean outcome before and a mean outcome after the treatment for the treatment group  $k$
  - We have a mean outcome before and a mean outcome after the treatment for the control group  $U$
  - $\hat{\delta}_{kU}^{2x2} = (\bar{y}_k^{post(k)} - \bar{y}_k^{pre(k)}) - (\bar{y}_U^{post(U)} - \bar{y}_U^{pre(U)})$
  - $\hat{\delta}_{kU}^{2x2}$  is the estimator for the ATT

# Estimator

## Difference-in-Differences

- What is being estimated?

- $\hat{\delta}_{kU}^{2x2} = (E[Y_k | Post] - E[Y_k | Pre]) - (E[Y_U | Post] - E[Y_U | Pre])$

- Use potential outcome and add a zero

$$\hat{\delta}_{kU}^{2x2} = (E[Y_k^1 | Post] - E[Y_k^0 | Pre]) - (E[Y_U^0 | Post] - E[Y_U^0 | Pre]) + 0$$

Switching Equation

$$E[Y_k^0 | Post] - E[Y_k^0 | Post]$$

equals 0

# Estimator

## Difference-in-Differences

- $\hat{\delta}_{kU}^{2x2} = (E[Y_k^1 | Post] - E[Y_k^0 | Pre]) - (E[Y_U^0 | Post] - E[Y_U^0 | Pre]) + E[Y_k^0 | Post] - E[Y_k^0 | Post]$
- We can rearrange to get ATT and Parallel Trends Assumption
- $$\hat{\delta}_{kU}^{2x2} = \underbrace{E[Y_k^1 | Post] - E[Y_k^0 | Post]}_{\text{ATT}} + \underbrace{[E[Y_k^0 | Post] - E[Y_k^0 | Pre]] - [E[Y_U^0 | Post] - E[Y_U^0 | Pre]]}_{\text{Non-Parallel Trends Bias in 2x2 Case}}$$

# Estimator

## Difference-in-Differences

- Let's decompose the second term

- $\hat{\delta}_{kU}^{2x2} = ATT + NonParallel\ Trends\ Bias$

$$= ATT + \underbrace{[E[Y_k^0 | Post] - E[Y_k^0 | Pre]] - [E[Y_U^0 | Post] - E[Y_U^0 | Pre]]}_{\text{Non-Parallel Trends Bias in 2x2 Case}}$$

$$= ATT + \underbrace{[E[Y_k^0 | Post] - E[Y_k^0 | Pre]]}_{\text{Counterfactual}} - \underbrace{[E[Y_U^0 | Post] - E[Y_U^0 | Pre]]}_{\text{Control Trend}}$$

$\overbrace{\hspace{10em}}$  Treatment Trends

# Estimator

## Difference-in-Differences

- If trends are not the same between treatment and control
  - $\hat{\delta}_{kU}^{2x2} = ATT + NonParallel\ Trends\ Bias$
- If trends are the same between treatment and control
  - $\hat{\delta}_{kU}^{2x2} = ATT + 0$  and  $\hat{\delta}_{kU}^{2x2} = \delta_{ATT}$
- We **cannot directly test the parallel trends assumption**, since we cannot observe the counterfactual,  $E[Y_k^0 | Post]$ 
  - We assume that the trend in the control is the same as the treatment and we extrapolate  $E[Y_k^0 | Post]$  from this trend

# **Minimum Wage and Diff-in-Diff**

## **Example**

# Minimum Wage and Diff-in-Diff Example

- The classic minimum wage study by Card and Krueger (1994)
  - The authors utilize a difference-in-difference design to study the impact of minimum wage on employment
  - The authors found an ATT effect of a 2.76 increase in full-time employment
- It challenged what economists think about minimum wage and negative impact on employment
  - If the labor market was perfectly competitive, then employment would have declined
  - There was a lot of backlash due to economists having their priors challenged, but it provided a solid research design to counter the status quo model

# Minimum Wage and Diff-in-Diff Example

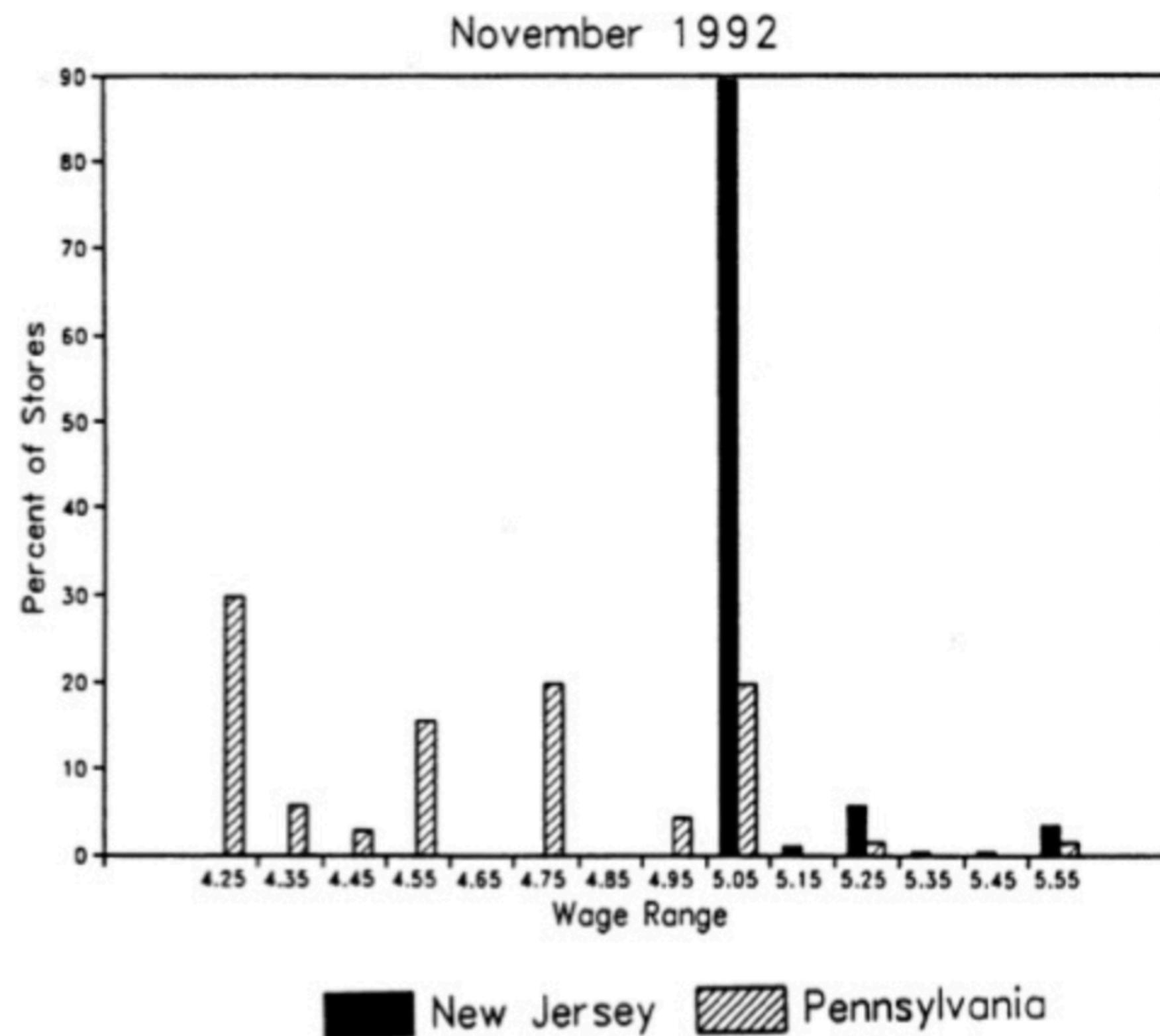
- Research Question: what is the impact of minimum wage on employment?
  - Theory suggests that in a perfectly competitive labor market, minimum wage will create a deadweight loss and decrease employment
  - In a monopsony, minimum wages can increase employment
  - The impact of minimum wage may depend on the context of local labor markets

# Minimum Wage and Diff-in-Diff Example

- Data Methodology
  - There is a lack of randomized assignment for minimum wage, so Card and Krueger (1994) analyzes the labor market in New Jersey near the border of Pennsylvania, and utilize a design fairly close to Snow (1855)
  - They compare New Jersey with Pennsylvania before and after New Jersey increases its state-level minimum wage from \$4.25 to \$5.05
  - They surveyed over 400 fast-food establishments in both states near the border in February 1992 before the policy and in November 1992 after the policy

# Minimum Wage and Diff-in-Diff Example

- The policy appeared to have been binding and establishments complied with the new state-level minimum wage in New Jersey (kind of like a 1st Stage)



# Minimum Wage and Diff-in-Diff Example

- What are we estimating?
  - The average treatment effect on the treated (NJ) for minimum wage on employment

$$\hat{\delta}_{NJ,PA}^{2x2} = \underbrace{E[Y_{NJ}^1 | Post] - E[Y_{NJ}^0 | Post]}_{\text{ATT}} + \underbrace{[E[Y_{NJ}^0 | Post] - E[Y_{NJ}^0 | Pre]] - [E[Y_{PA}^0 | Post] - E[Y_{PA}^0 | Pre]]}_{\text{Non-Parallel Trends Bias in 2x2 Case}}$$

# Minimum Wage and Diff-in-Diff Example

- Results
  - Our estimated ATT is 2.76 full-time employment

Table 1

		Before	After	1st Difference
NJ		20.44 (0.51)	21.03 (0.52)	0.59 (0.54)
PA		23.3 (0.94)	21.147 (1.35)	-2.16 (1.25)
				2nd Difference
				2.76 (1.36)

# Sample Average vs Regressions

- You can use sample average pre-post for difference-in-differences
- You can also use regressions as well
  - You will get 3 of out of the 4 dummies in our regression (we need to exclude one for the intercept)
  - You can use covariates to increase precision around your DD estimate
  - It's a good idea to use specification tests for your Diff-in-Diff estimator with and without covariates to see how sensitive your estimate is

# Sample Average vs Regressions

## Regression

- Difference-in-Difference 2x2 case Regression (Using NJ-PA Example)
  - $Y_{its} = \alpha + \gamma NJ_s + \lambda D_t + \delta(NJ \times D)_{st} + \varepsilon_{its}$
  - Where  $i$  is individual in state  $s$  in time period  $t$
  - Where  $NJ_s = 1$  if the observation lives in NJ
  - Where  $D_t = 1$  if the observation is in the POST period or November 1992

# Sample Average vs Regressions

## Decomposition of the Regression

- Difference-in-Difference 2x2 case Regression (Using NJ-PA Example)

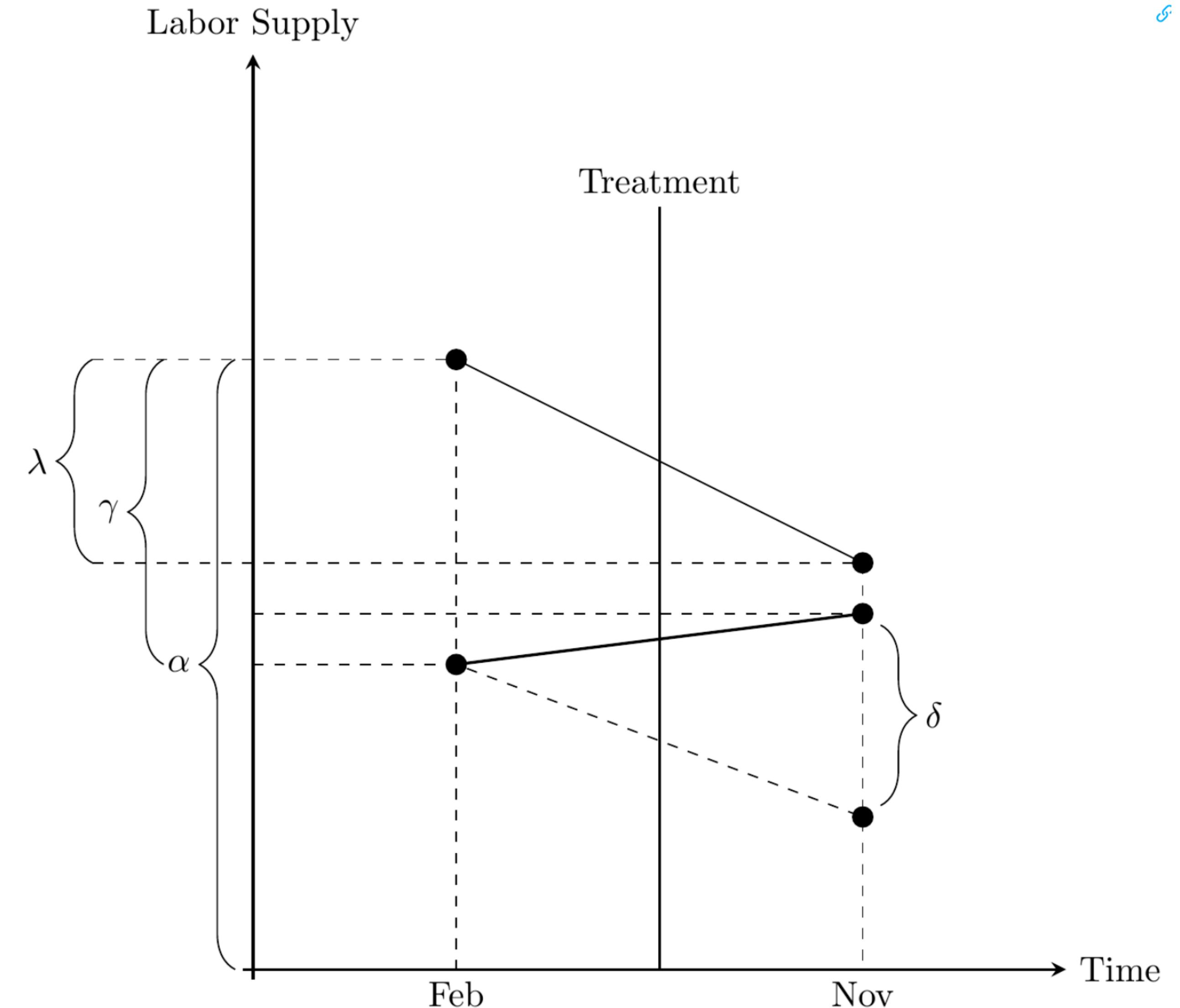
$$Y_{its} = \alpha + \gamma NJ_s + \lambda D_t + \delta(NJ \times D)_{st} + \varepsilon_{its}$$

- PA Pre:  $\alpha$
- PA Post:  $\alpha + \lambda$
- NJ Pre:  $\alpha + \gamma$
- NJ Post:  $\alpha + \gamma + \lambda + \delta$

DD	Post	Pre	Difference
NJ	$\alpha + \gamma + \lambda + \delta$	$\alpha + \gamma$	$\lambda + \delta$
PA	$\alpha + \lambda$	$\alpha$	$\lambda$
Second Difference			$\delta$

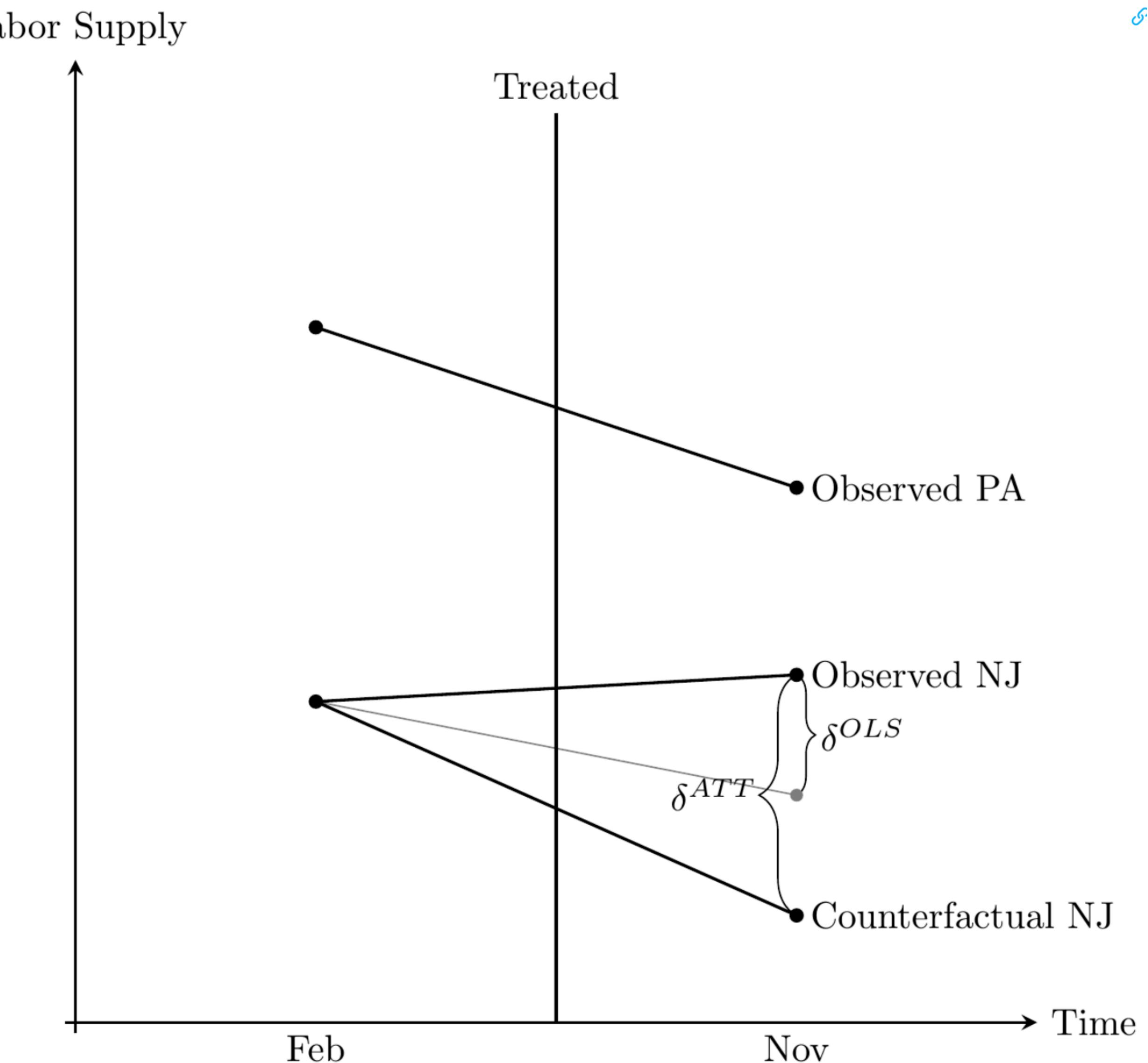
# Sample Average vs Regressions

- We use the trend in the control to get the counterfactual for the treatment group for  $E[Y_k^0 | Post]$
- The parallel trends assumption is key here



# Sample Average vs Regressions

- Note: OLS will still estimate the DD estimator even if the parallel trends assumption does not hold
- If Pennsylvania is not a true counterfactual trend of New Jersey, then the DD estimator will be biased
- $\hat{\delta}_{kU}^{2x2} = ATT + Bias$
- The Parallel Trends assumption is just a restatement of the strict exogeneity assumption



# **Inference**

## **Difference-in-Differences**

# Inference

## Difference-in-Differences

- 2 Issues with Diff-in-Diff Inference
- 1) Errors are correlated within groups, typically at the state level
  - Bertrand, Duflo, and Mullainathan (2004) note that conventional standard errors will underestimate the standard deviation of the estimator and more likely to commit a Type I error (reject a true null)
- 2) Conley and Taber (2011) mention the problem of small clusters
  - When we cluster at the state level and 1 or a few states adopt the treatment, then standard errors will be inconsistent and downward biased (Type 1 Error)

# Inference

## Difference-in-Differences

- 1) Bertrand, et al. (2004) propose three methods to deal with serial correlations within large clusters
  - Block bootstrapping standard errors
    - bootstrapping at with replacement at the cluster level (usually state)
    - Aggregating data into one pre and one post period
    - Clustering standard errors at the treatment level (usually state)
      - Can be implemented easily with Stata
      - Small clusters are still a problem discussed next

# Inference

## Difference-in-Differences

- 2) What to do when we have small clusters?
  - Cunningham (2021) suggests using randomized inference
  - Conley and Taber (2011) suggest 2-stage estimates of standard errors
  - Cameron and Miller (2015) suggest using synthetic control method
- Santa'Anna Updates
  - Ferman and Pinto (2019) suggest using bootstrap-based inference
  - Hagemann (2020) suggest using rearrangement permutation-based method
  - Canay, Santos, and Shaikh (2021) suggest using cluster wild bootstrap

# **Testing Assumptions**

## **Difference-in-Difference 2x2 case**

# Testing Assumptions

## Difference-in-Difference 2x2 case

- There are four indirectly tests for the parallel trends assumption
  - 1) Plotting pre-treatment trends
  - 2) Pre-treatment trends - Event Studies
  - 3) Placebo or falsification tests
  - 4) Triple Difference-in-Difference-in-Differences
- Remember if the parallel trends assumption fails
  - $\hat{\delta}_{kU}^{2x2} = ATT + NonParallel\ Trend\ Bias$
  - We never observe the counterfactual  $E[Y_k^0 | Post]$

# Testing Assumptions

## Problems with testing pre-treatment trends

- Let's discuss the problem of testing pre-treatment trends
  - If the parallel trends assumption fails, our estimator is biased
  - We can never directly test the parallel trends assumption
- Justification for using pre-treatment trends
  - If they were similar trends before the treatment, then they should be similar in the absence of the treatment group adopting the policy/program/treatment
  - However, just because treatment and control were similar in the past does not mean they will be the same other than treatment in the post period
  - This is a reason to use placebo tests and falsification tests

# Testing Assumptions

## Problems with testing pre-treatment trends

- Another problem with testing pre-treatment trends
  - Is there self-selection with the treatment?
  - You cannot be vigilant enough with being cautious of your control groups, especially for states
    - $(Y^1, Y^0) \perp D$  needs to hold
    - E.g.: Texas and California will self-select into different types of treatment/policies/programs
  - Pre-treatment trends will not find this
  - If you have poor control group, the error term will be endogenous with the treatment

# Testing Assumptions

## Plotting pre-treatment trends

- Show the data
  - You can provide pre-treatment data for treatment and control
  - You can inspect the summary statistics of the pre-treatment trends
- Plot the data
  - You can simply plot the pre-treatment trends and visually inspect
  - This usually requires lots of graphs in the appendix but it is transparent
  - If there are a lot of treatment groups, this test gets harder to implement

# Testing Assumptions

## Pre-treatment trends - Event Studies

- We can check pre-treatment and post-treatment differences between treatment and control with Event Studies for each time period
  - Add pre-treatment leads and post-treatment lags for each time period
  - Pre-treatment leads should be statistically insignificant between treatment and control since the treatment group has not received the treatment
- Pro:
  - You plot coefficients of the regression instead of raw data
- Con
  - Does not work if one of your control groups adopts a similar post in post-treatment

# Testing Assumptions

## Pre-treatment trends - Event Studies

- We can check pre-treatment and post-treatment differences between treatment and control with Event Studies for each time period
  - Add pre-treatment leads and post-treatment lags for each time period
  - Pre-treatment leads should be statistically insignificant between treatment and control since the treatment group has not received the treatment
- Pro:
  - You plot coefficients of the regression instead of raw data
- Con
  - Does not work if one of your control groups adopts a similar post in post-treatment

# Event Studies

## Miller, Altekrouse, Johnson, and Wherry (2019)

- Miller, Altekrouse, Johnson, and Wherry (2019) test the impact of the Medicaid expansion under the Affordable Care Act (ACA)
  - Were there any impacts on mortality outcomes?
  - We saw earlier that there were some doubts about Medicaid's expansion on mortality from Finkelstein, et al. (2012) and Baicker, et al. (2013) using instrumental variable estimators for the Oregon Medicaid Lottery
  - These authors revisit the question with a larger set of data and more states
  - They link a large-scale federal survey with admin data on death
- The authors find that Medicaid expansion reduces annual mortality rates by 0.13 percentage points (9.3% reduction in the sample mean)

# Event Studies

**Miller, Altekrose, Johnson, and Wherry (2019)**

- Most contemporary DD studies utilize pre-treatment leads instead of plotting the raw data
- Including both pre-treatment leads and post-treatment allows readers to check
  - Post-treatment dynamics were different between treatment and control
  - Pre-treatment dynamics were similar between treatment and control
- Event Study Model

$$Y_{its} = \gamma_s + \lambda_t + \sum_{\tau=-q}^{-1} \gamma_\tau D_{s\tau} + \sum_{\tau=0}^m \delta_\tau D_{s\tau} + x_{its} + \varepsilon_{its}$$

# Event Studies

**Miller, Altekroose, Johnson, and Wherry (2019)**

$$Y_{its} = \gamma_s + \lambda_t + \underbrace{\sum_{\tau=-q}^{-1} \gamma_\tau D_{s\tau}}_{\text{pre-treatment leads}} + \underbrace{\sum_{\tau=0}^m \delta_\tau D_{s\tau}}_{\text{post-treatment lags}} + x_{its} + \varepsilon_{its}$$

- Treatment occurs when  $\tau = 0$ 
  - There are  $q$  leads or (pre-treatment) anticipatory effects
  - There are  $m$  lags or post-treatment effects
  - We should expect no impacts of pre-treatment leads or no anticipatory effects

# Event Studies

Miller, Altekroose, Johnson, and Wherry (2019)

- The first stage would be that the policy affects Medicaid eligibility
- ACA Medicaid expansions shows a clear impact on eligibility
- We see no difference between treatment and control in the pre-treatment leads, but differences in the post-treatment lags

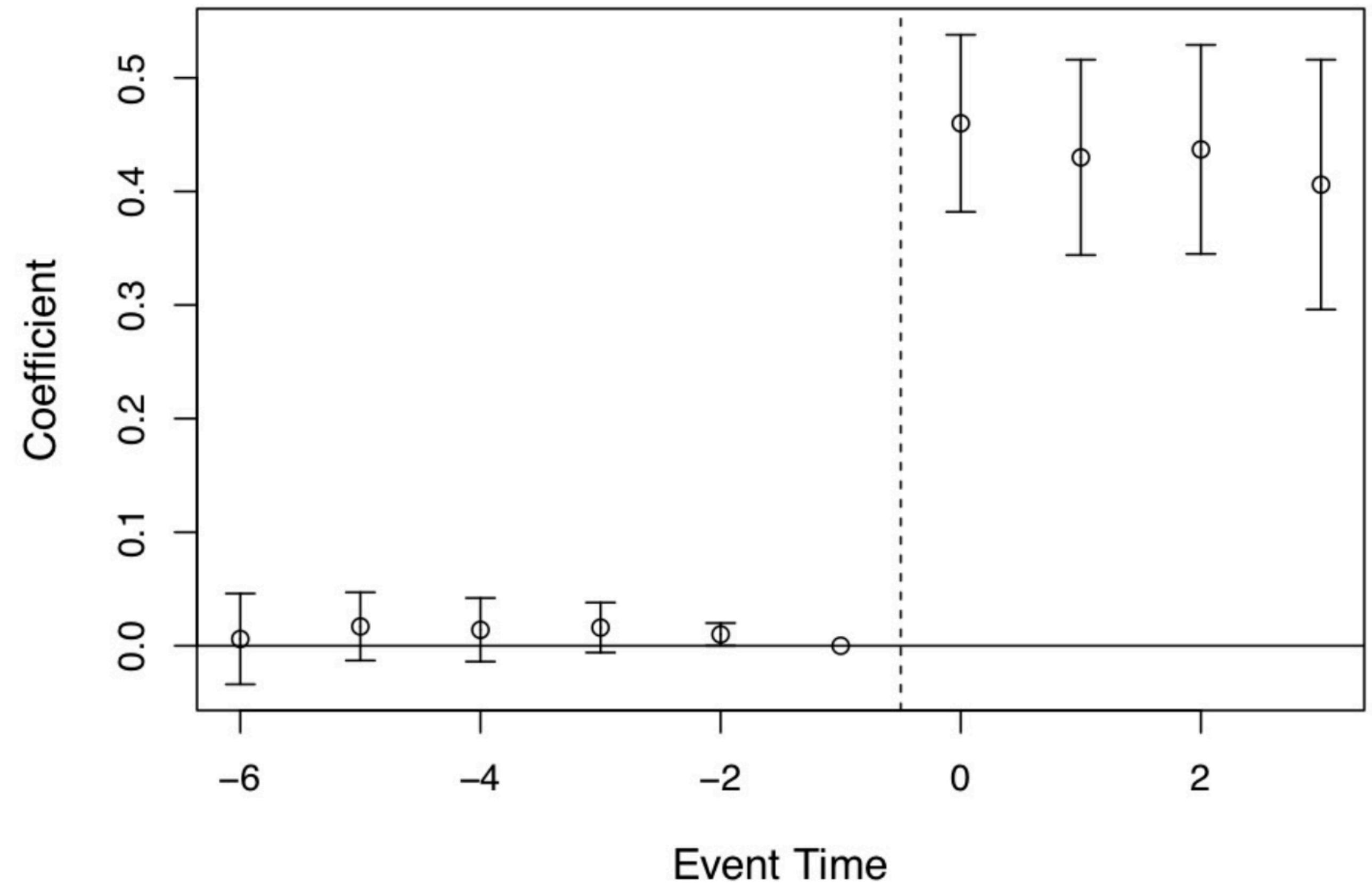


Figure 9.4: Estimates of Medicaid expansion's effects on eligibility using leads and lags in an event study model. Reprint from Miller et al. (2019).

# Event Studies

Miller, Altekroose, Johnson, and Wherry (2019)

- Another first stage test is the ACA Medicaid expansion's impact on coverage
- There are no significant differences in the pre-treatment leads, but we see statistically significant differences in the post-treatment lags

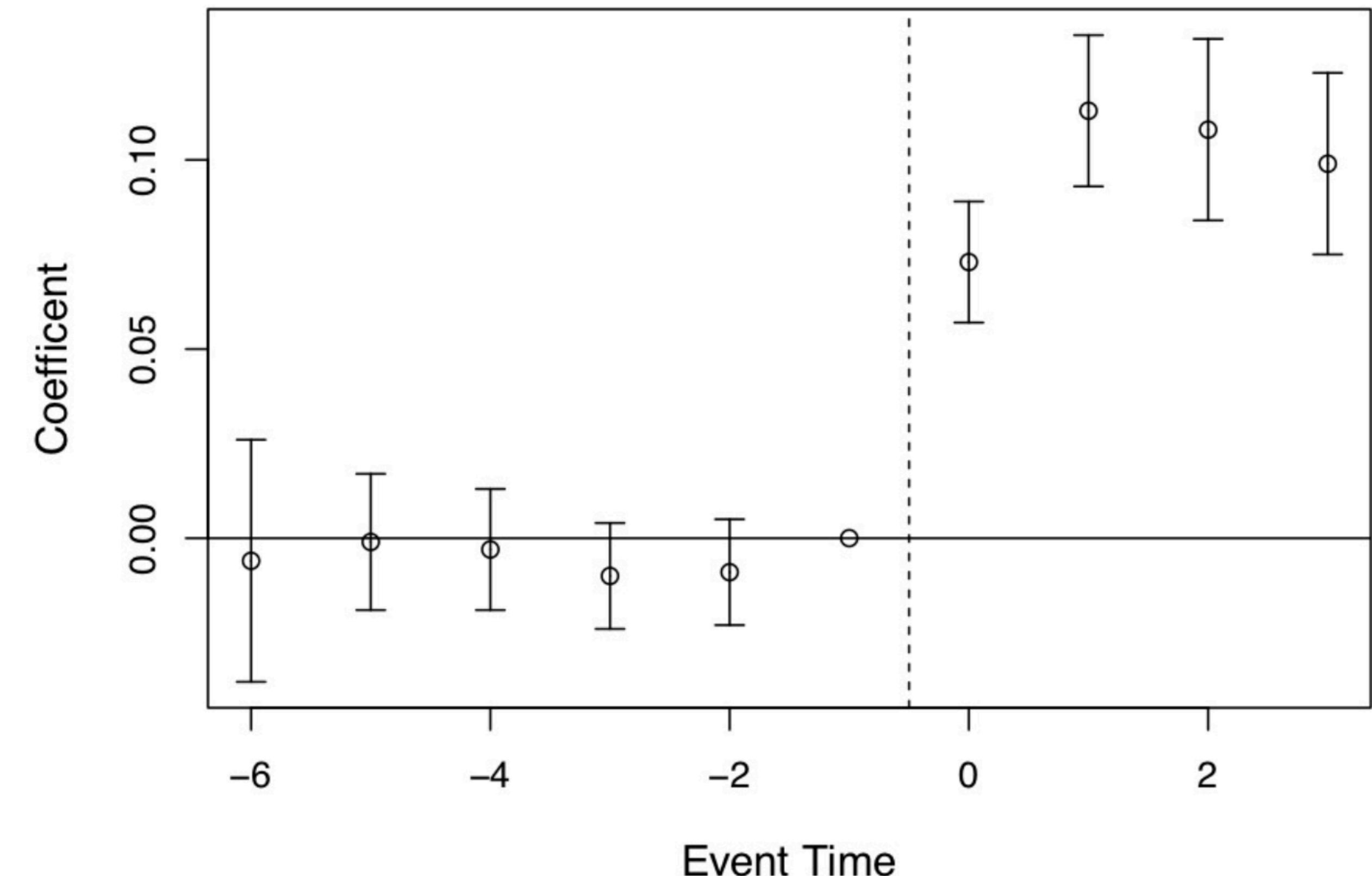


Figure 9.5: Estimates of Medicaid expansion's effects on **coverage** using leads and lags in an event study model. Reprint from Miller et al. (2019).

# Event Studies

Miller, Altekroose, Johnson, and Wherry (2019)

- The main result for the authors assess the the impact of ACA's Medicaid expansion on annual mortality rate
- If the parallel trends assumption holds, then ACA's Medicaid expansion lead to a 0.13 percentage point decline in annual mortality rate, which translates to a 9.3% reduction

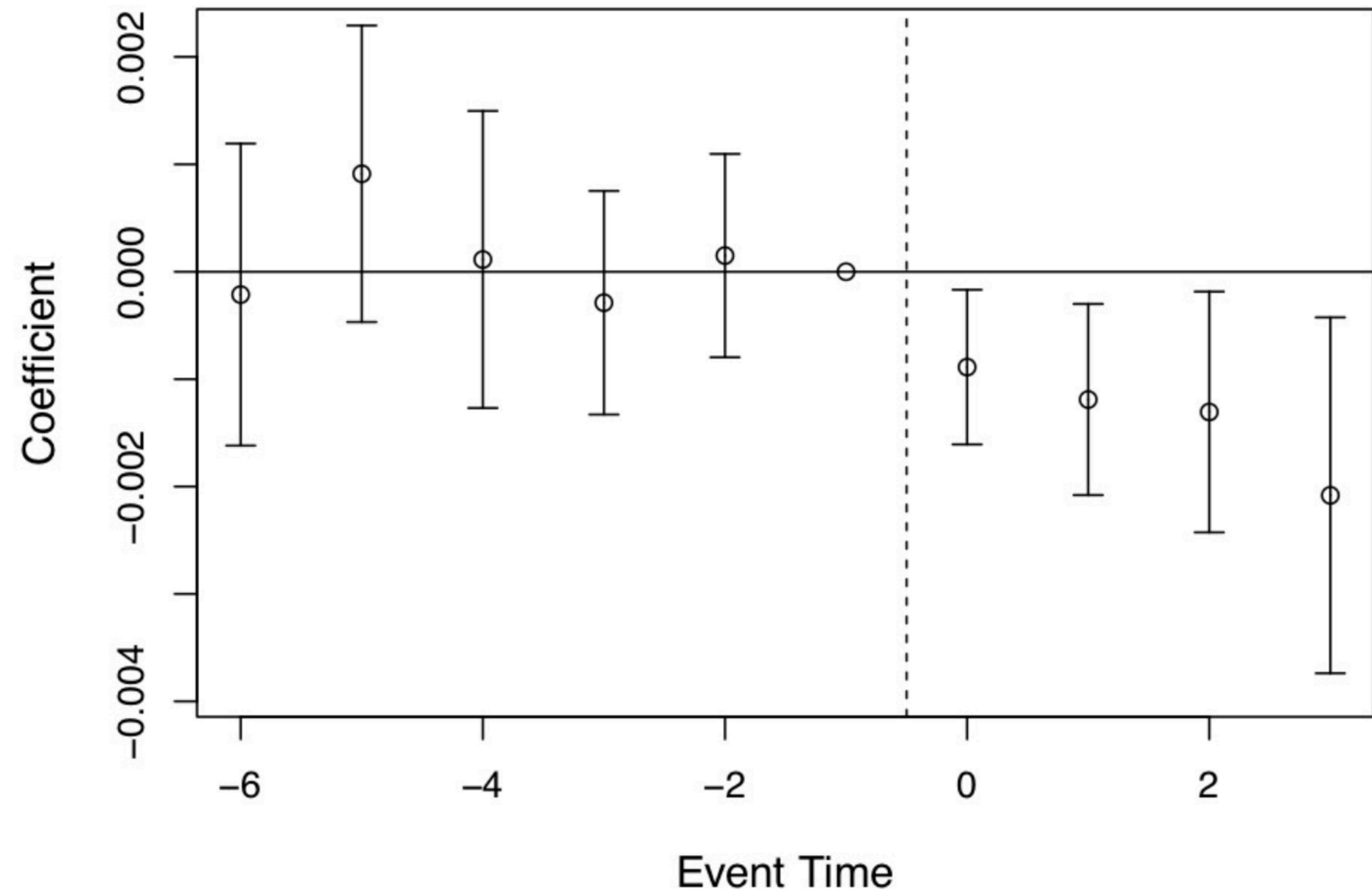


Figure 9.7: Miller et al. (2019) estimates of Medicaid expansion's effects on on annual mortality using leads and lags in an event study model

# Event Studies

Miller, Altekroose, Johnson, and Wherry (2019)

- Note: Just because the pre-treatment leads are not statistically significant, does not mean that the control group is a true counterfactual of the treatment group
- However, it is compelling that there is “bite” in the 1st stage, such that Medicaid expansion increases eligibility & coverage while reducing uninsured

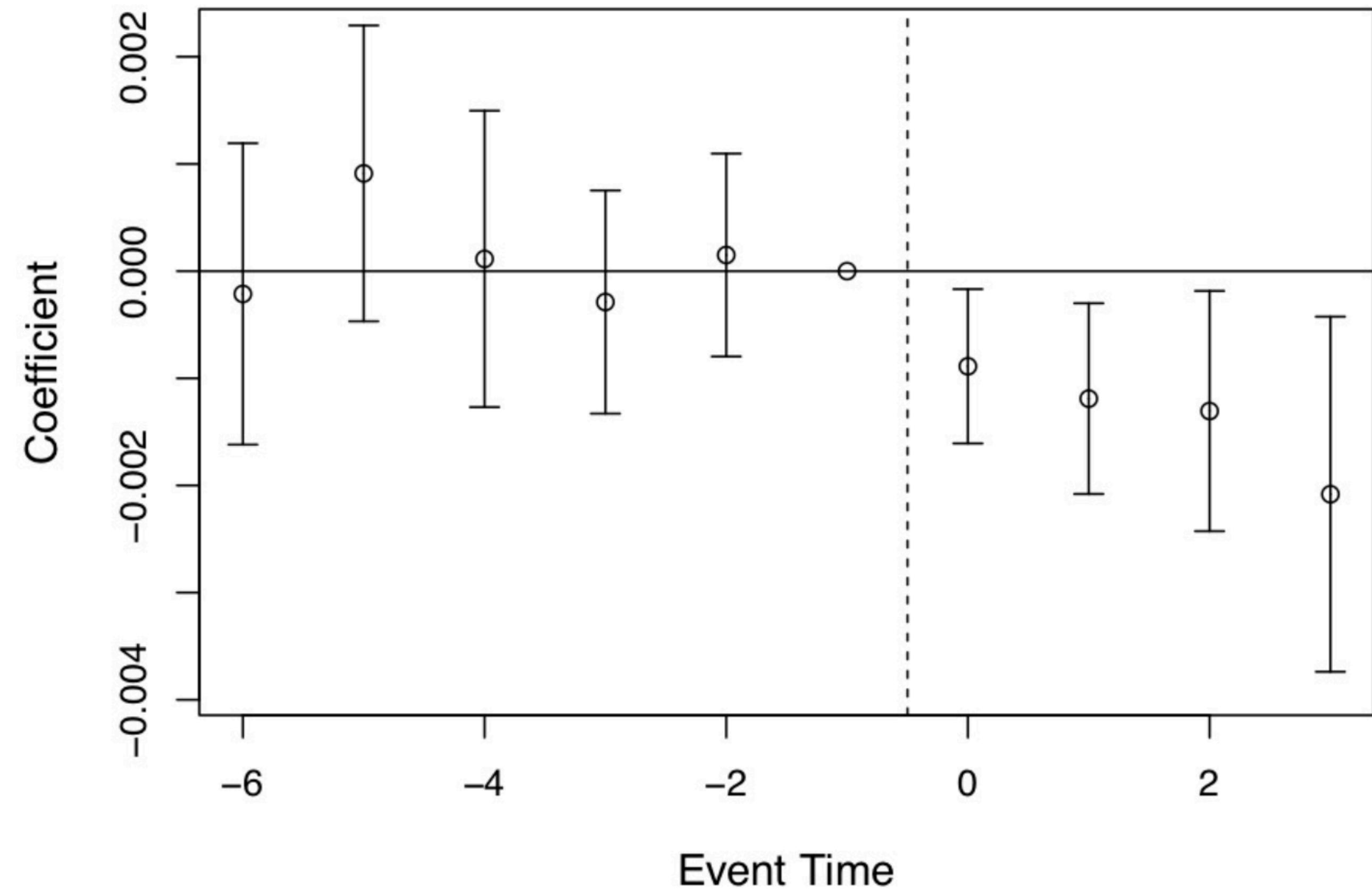


Figure 9.7: Miller et al. (2019) estimates of Medicaid expansion's effects on annual mortality using leads and lags in an event study model

# Testing Assumptions

## Placebo and Falsification Tests

- Placebo tests can be used as a robustness check
  - These are essential in contemporary DD analysis for indirect tests of the parallel trends assumption
  - Our Stata example will provide a nice example of how a preliminary result can fail robustness checks
- Placebo tests strengthen our main DD analysis
  - Being unable to reject the null hypothesis of the placebo boost your main result
  - You may have picked up a spurious correlation in your main result
    - Such as cyclical factors or other time-varying unobserved confounders and testing falsification and placebo tests help prevent this

# Testing Assumptions

## Placebo and Falsification Tests

- There are three types of Placebo tests we will discuss
  - 1) False time adoption of treatment
  - 2) Placebo test for control group that did not receive the treatment
  - 3) Triple Difference-in-Difference-in-Differences
    - Sometimes researchers use Triple DDD as a main analysis

# Testing Assumptions

## Placebo and Falsification Tests

- False Time Adoption (Falsification)
  - We can conduct an diff-in-diff design for the treatment and control groups, but in a time period before the policy, program, or treatment is adopted
  - Let's a program is implemented in 2015, we would want to test the impact of a false adoption time in 2010 (5 years before adoption)
  - We should not expect any significant differences between treatment and control
  - If we do find differences, then we need to reconsider our original DD design

# Testing Assumptions

## Placebo and Falsification Tests

- Another test if a placebo test for a group that did not receive treatment
  - For example, supervisory and managerial workers should not be directly affected by a unionization program treatment
  - Another example are high wage workers that should not be directly affected by a minimum wage policy. In Card and Krueger (1994), high wage workers should not be affected by the NJ minimum wage increase
- You can also try applying a treatment placebo to a group that was in control
  - You know that a particular state did not adopt a program or policy, then you can test that state as the treatment state in a placebo test

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- Similar to the last placebo test, we will test a triple difference-in-difference-in-differences or Triple DDD
- The basis
  - We will apply our difference-in-difference to 2 subgroups within the treatment group
    - A subgroup within treatment that **will be affected** by the program
    - A subgroup within treatment that **will not be affected** by the program
  - The subgroup that should not be impacted by the program should be unable to reject the null hypothesis

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- DDD Estimator
- $$Y_{ijt} = \alpha + \psi X_{ijt} + \beta_1 \tau_t + \beta_2 \delta_j + \beta_3 D_i + \beta_4 (\delta \times \tau)_{jt} + \beta_5 (\tau \times D)_{ti} + \beta_6 (\delta \times D)_{ji} + \beta_7 (\delta \times \tau \times D)_{ijt} + \varepsilon_{ijt}$$
  - All interactions are included in a DDD design
  - Group dummy  $\delta_j$
  - Post-treatment dummy  $\tau_t$
  - Treatment group dummy  $D_i$
  - $X_{ijt}$  is a set of covariates of interest

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- $Y_{ijt} = \alpha + \psi X_{ijt} + \beta_1 \tau_t + \beta_2 \delta_j + \beta_3 D_i + \beta_4 (\delta \times \tau)_{jt} + \beta_5 (\tau \times D)_{ti} + \beta_6 (\delta \times D)_{ji} + \beta_7 (\delta \times \tau \times D)_{ijt} + \varepsilon_{ijt}$
- Our parameter of interest is  $\beta_7$  to compare to our original DD
  - If our original DD estimate is robust, the  $\beta_7$  should be similar to our original estimate
- Our other coefficient of interest is  $\beta_5$ 
  - This is the DD estimator for the placebo subgroup (or the subgroup in the treatment group that does not get the treatment)
  - This should be statistically insignificant

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- Let's go back to our Card and Krueger (1994) example
  - We had our treatment state NJ ( $D_i$ ) and our control state PA
  - We had our pre-period February 1992 and our post-period November 1992 ( $\tau_t$ )
  - Now look at low-wage workers and high-wage workers ( $\delta_j$ )
  - High-wage workers should not be directly impacted by the increase in minimum wage laws

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- Before the treatment, we have 4 fixed groups
  - $NJ_{low}$  low-wage workers in New Jersey (treatment)
  - $NJ_{high}$  high-wage workers in New Jersey (treatment)
  - $PA_{low}$  low-wage workers in Pennsylvania (control)
  - $PA_{high}$  high-wage workers in Pennsylvania (control)

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- After the treatment, we have 4 changes for New Jersey (treatment) low-wage workers
  - $T$  national (secular) trends that cause employment to change
  - $NJ_t$  is New Jersey specific time shocks to employment
  - $l_t$  is generic low-wage worker trends that cause employment to change
  - $D$  is the impact of the treatment (minimum wage)

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- After the treatment, we have 3 changes for Pennsylvania (control) low-wage workers
  - $T$  national (secular) trends that cause employment to change
  - $PA_t$  is Pennsylvania specific time shocks to employment
  - $l_t$  is generic low-wage worker trends that cause employment to change

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- After the treatment, we have 3 changes to New Jersey (treatment) high-wage workers (placebo)
  - $T$  national (secular) trends that cause employment to change
  - $NJ_t$  is New Jersey specific time shocks to employment
  - $h_t$  is generic high-wage worker trends that cause employment to change
- After treatment, we have 3 change to Pennsylvania (control) high-wage workers
  - $T$  national (secular) trends that cause employment to change
  - $PA_t$  is Pennsylvania specific time shocks to employment
  - $h_t$  is generic high-wage worker trends that cause employment to change

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

States	Group	Before	After	First Difference	Second Difference	Third Difference
NJ	Low	$NJ_{low}$	$NJ_{low} + T + NJ_t + l_t + D$	$T + NJ_t + l_t + D$	$(l_t - h_t) + D$	$D$
	High	$NJ_{high}$	$NJ_{high} + T + NJ_t + h_t$	$T + NJ_t + h_t$		
PA	Low	$PA_{low}$	$PA_{low} + T + PA_t + l_t$	$T + PA_t + l_t$	$(l_t - h_t)$	
	High	$PA_{high}$	$PA_{high} + T + PA_t + h_t$	$T + PA_t + h_t$		

# Testing Assumptions

## Triple Difference-in-Difference-in-Differences

- Our first difference
  - Takes care of fixed effects
  - We still have confounders of such as  $T, NJ_t, PA_t, h_t, l_t$
- Our second difference
  - Takes care of trends  $T, NJ_t, PA_t$
  - We still have confounders  $l_t - h_t$
- Our third difference
  - Takes care of  $l_t - h_t$
  - Identifies the causal effects if the parallel trends assumption holds

# **Difference-in-Difference Examples**

## **Diff-in-Diff and Triple Difference Examples**

# Difference-in-Difference-in-Differences

## State-Mandated Maternity Benefits

- Gruber (1994) introduced the Triple Difference design
  - The author studied state-mandated maternity benefits on hourly wages
  - Gruber uses a Triple Difference to check to robustness of the Diff-in-Diff
    - When we do a Triple difference design, the variation should come from the treatment group (and focus subgroup) and not the control group (or placebo group)
    - Gruber's treatment subgroup is women in child-bearing age, while the placebo subgroup is men 20-40 and women over 40

# Difference-in-Difference-in-Differences

## State-Mandated Maternity Benefits

- Gruber (1994) results show that women in child-bearing age had a loss of hourly wages of 6%

Table 9.3: DDD Estimates of the Impact of State Mandates on Hourly Wages

Location/year	Pre-law	Post-law	Difference
A. Treatment: Married women, 20-40yo			
Experimental states	1.547	1.513	-0.034
	(0.012)	(0.012)	(0.017)
Control states	1.369	1.397	0.028
	(0.010)	(0.010)	(0.014)
Difference	0.178	0.116	
	(0.016)	(0.015)	
Difference-in-difference	-0.062		
	(0.022)		

# Difference-in-Difference-in-Differences

## State-Mandated Maternity Benefits

- Gruber (1994) Triple Difference check shows that Diff-in-Diff for women 40 and older and men 20-40 had a statistically insignificant impact on hourly wages from mandated maternity benefits
- However, women 20-40 had a reduction in hourly wages of 5.3%

B. Control: Over 40 and Single Males 20-40

	Experimental states	1.759	1.748	-0.011
	(0.007)	(0.007)	(0.010)	
	Control states	1.630	1.627	-0.003
	(0.007)	(0.007)	(0.010)	
Difference		1.09	1.21	
		(0.010)	(0.010)	
Difference-in-difference		-0.008		
		(0.014)		
DDD		-0.054		
		(0.026)		

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Stata Example
- Cunningham and Cornwell (2013) assess the impact of abortion legalization on long-term incidences of gonorrhea
  - This research builds off of the reproductive health literature and the abortion legalization hypothesis
  - Gruber, Levine, and Staiger (1999) look at characteristics of the marginal child that was aborted
  - Donohue and Levitt (2001) test abortion legalization and crime

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Donohue and Levitt (2001) test the abortion legalization hypothesis
  - Abortion legalization hypothesis says that the availability of abortion reduces children born in high-risk of committing a crime
  - They find that abortion legalization reduced crime
  - Levitt (2004) finds that abortion legalization led to about a 10% decline in crime between 1991 and 2001

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Joyce (2004,2009) and Foot and Goetz (2008) dispute the results
  - Joyce (2009) challenges long-term changes from legalized abortion through a Triple DDD
  - Joyce argues that if there was extreme negative selection for crime, then it should show up everywhere not just in crime
  - “If abortion lowers homicide rates by 20-30%, then it is likely to have affected an entire spectrum of outcomes associated with well-being: infant health, child development, schooling, earnings, and marital status. Similarly, the policy implications are broader than abortion. Other interventions that affect fertility control and that lead to fewer unwanted births—contraception or sexual abstinence—have huge potential payoffs. In short, a causal relationship between legalized abortion and crime has such significant ramifications for social policy and at the same time is so controversial, that further assessment of the identifying assumptions and their robustness to alternative strategies is warranted. (p.112)”

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham and Cornwell (2013) takes up Joyce (2009) challenge by testing legalized abortion with long-run incidences of gonorrhea
  - Single-parent households have a higher risk of sexual activity and unprotected sex Levine, et al. (1999)
  - Charles and Stephen (2006) find that child exposed to legalized abortion were less likely to use illegal substances, which is linked with sexual activity
- Cunningham and Cornwell (2013) use a difference-in-difference design as opposed to Donohue and Levitt (2001) that use lagged ratio values at the state level

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham makes a good point about theories
  - Good theories provide very specific falsifiable hypotheses
  - More specific the hypothesis, the more compelling the theory if evidence supports the theory
- The theory of abortion legalization makes very specific hypotheses
  - If we find evidence of the hypothesis, we have to take the theory seriously
  - What are usually yet testable predictions from the abortion legalization theory?

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham and Cornwell (2013) argue that the abortion legalization theory predicts the shape of the effects with staggered adoption
  - 5 states legalized abortion and then all states are exposed to legalized abortion
  - There is a three-year lag between the 5 states and Roe v Wade decision
  - The shape should be nonlinear parabolic treatment effect
    - There should be increasingly negative effect of abortion legalization on the outcome between the treatment states and control states
    - The treatment effects should recede after the control states are forced to legalize abortion

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- The authors use age cohorts due to data limitations
  - For 15-19 year olds born between 1971-1973, there should be increasingly negative treatment effects for this cohort from 1986-1987
  - A plateau treatment effect between 1988-1991
  - And a weakening treatment effects between 1991-1992

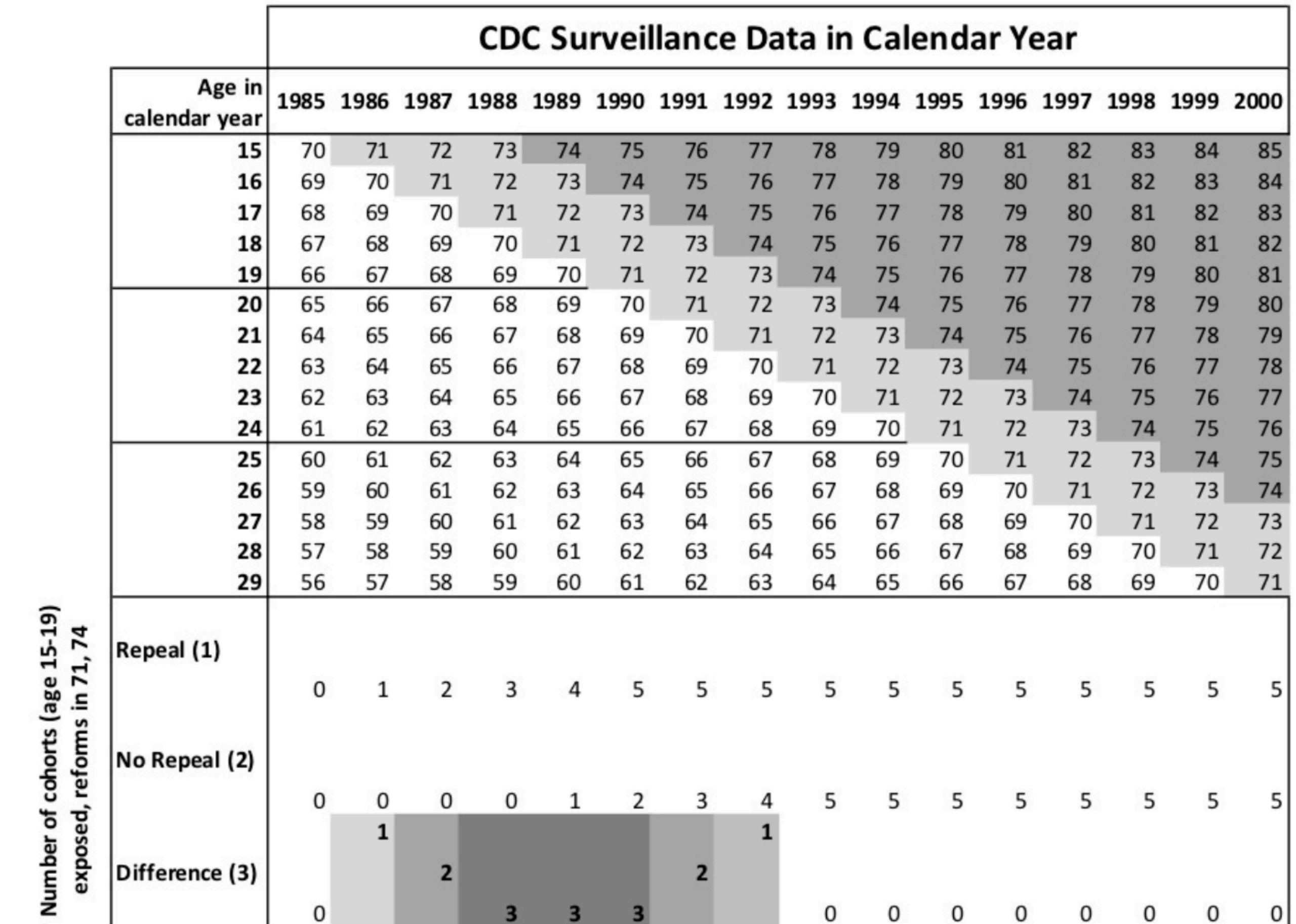


Figure 9.9: Theoretical predictions of abortion legalization on age profiles of gonorrhea incidence. Reprinted from Cunningham and Cornwell (2013).

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- There should be two testable hypotheses with a staggered rollout
  - Are there treatment effects during the transition period?
  - Are there no treatment effects after the transition period?
- Plotting the raw data can be seen where treatment states have declining rates of gonorrhea between 1986 and 1992

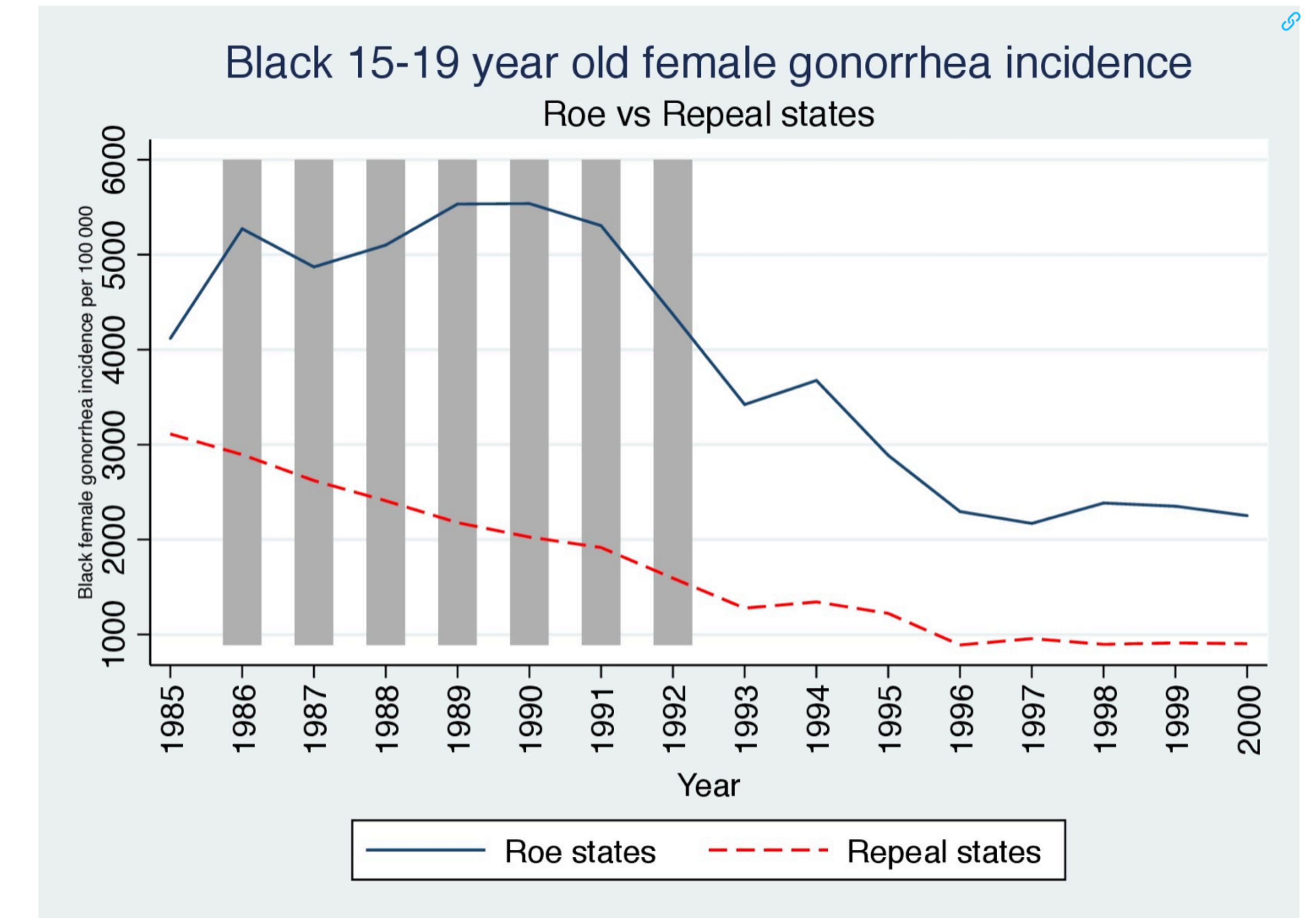


Figure 9.10: Differences in black female gonorrhea incidence between repeal and Roe cohorts expressed as coefficient plots. Reprinted from Cunningham and Cornwell (2013).

# Difference-in-Difference Example

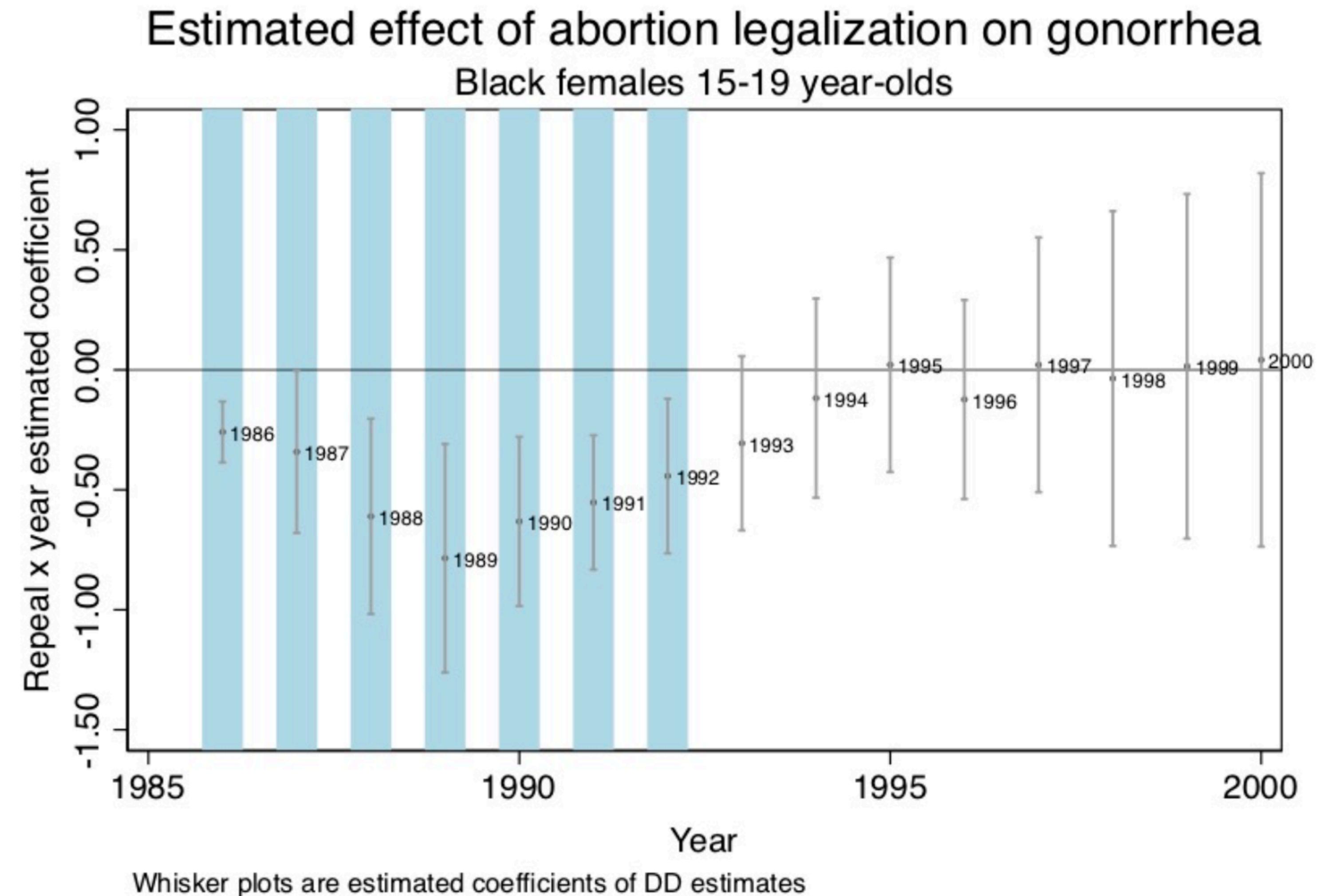
## Abortion Legalization and Incidence of Gonorrhea

- Main model
  - $Y_{st} = \beta_1 Repeal_s + \beta_2 DT_t + \beta_3 (Repeal \times DT)_{st} + \psi X_{st} + \alpha DS_s + \varepsilon_{st}$
  - Where
  - $Y_{st}$  is the outcome of interest - new gonorrhea cases for 15-19 year olds per 100,000
  - $Repeal_s$  equal 1 if the state legalized abortion before Roe v Wade
  - $DT_t$  is a year dummy
  - $DS_s$  is a state-specific dummy
  - $X_{st}$  are a set of covariates

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham and Cornwell (2013) provide the DD estimates by year
- A parabolic shape is seen between 1986 and 1992
- The effect is no longer significant after 1992
- This does give evidence for the abortion legalization hypothesis



# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham does placebo tests on his own work using a Triple DDD to rule out alternative hypotheses
- Triple DDD
  - He chooses a group close to 15-19 year olds for trends, but not close enough for SUTVA spillovers
    - 25-29 year olds were the placebo group within treatment states
  - $$Y_{ast} = \beta_1 Repeal_s + \beta_2 DT_t + \beta_3 (Repeal \times DT)_{st} + \delta_1 DA_a + \delta_2 (Repeal \times DA)_{sa} + \delta_3 (DA \times DT)_{at} + \delta_4 (DA \times Repeal \times DT)_{ast} + \alpha_1 DS_s + \alpha_2 (DS \times DA)_{sa} + \psi X_{ast} + \gamma_1 t + \gamma_2 (DS \times t)_{st} + \gamma_3 (DA \times t)_{at} + \gamma_4 (DA \times DS \times t)_{ast} + \varepsilon$$
  - Our treatment of interest is  $\delta_4$

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- The prediction begins to break down with the DDD placebo tests
- There are statistically significant impacts between 1986 and 1989, but no difference between treatment and control groups after 1989
- There is also no parabolic effect
- Key: There might be some unobservable confounders in treatment states that are not in the Roe v Wade states

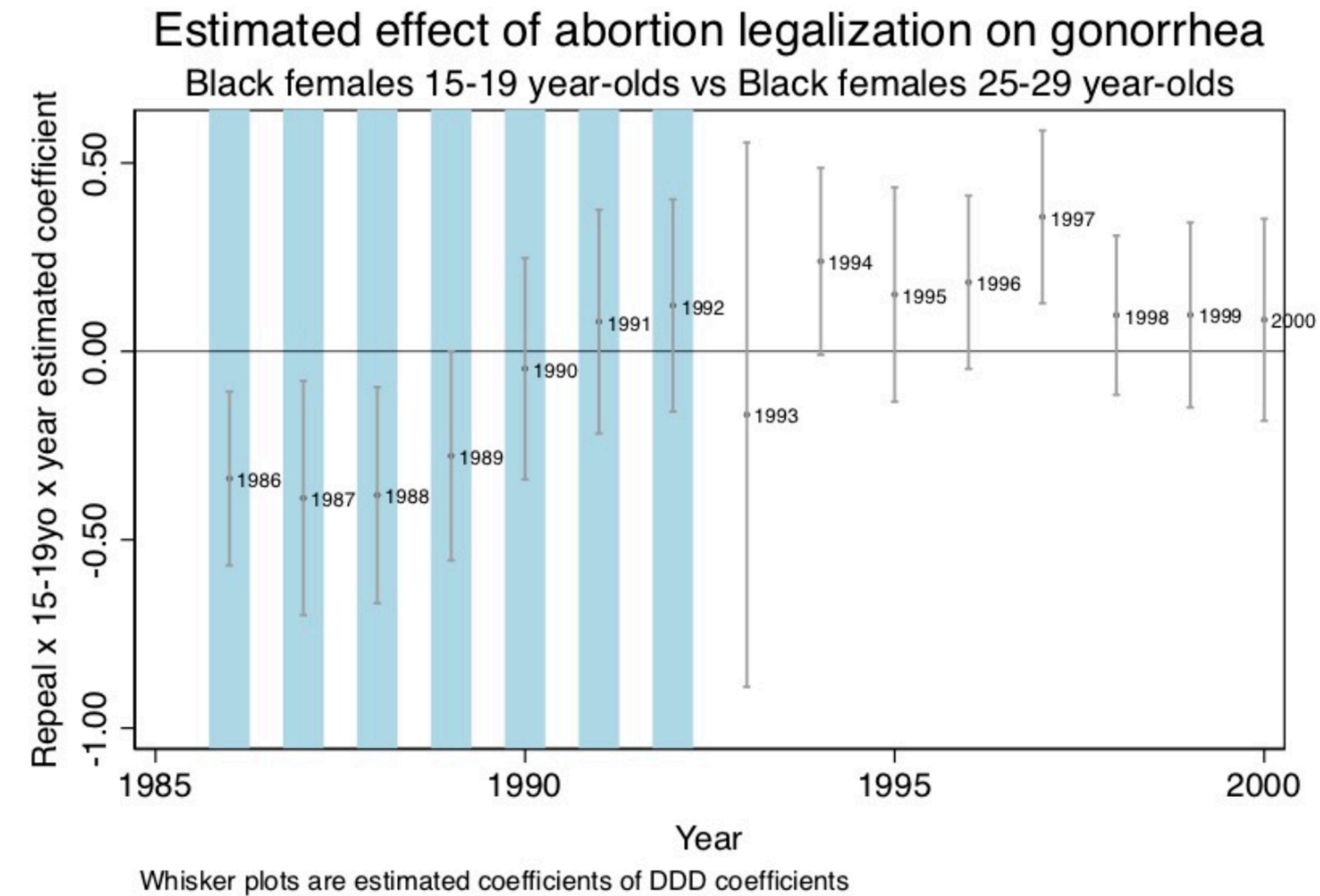


Figure 9.12: DDD Estimates of Abortion Legalization on 15-19yo Black Female Log Gonorrhea

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham then tests the DD estimator for 20-24 year olds who would have been exposed to legal abortion between 1991-1993 in treatment states
- The treatment show grow between 1991 to 1993
- The treatment show plateau between 1993 and 1995
- The treatment show converge to zero between 1995 and 1997

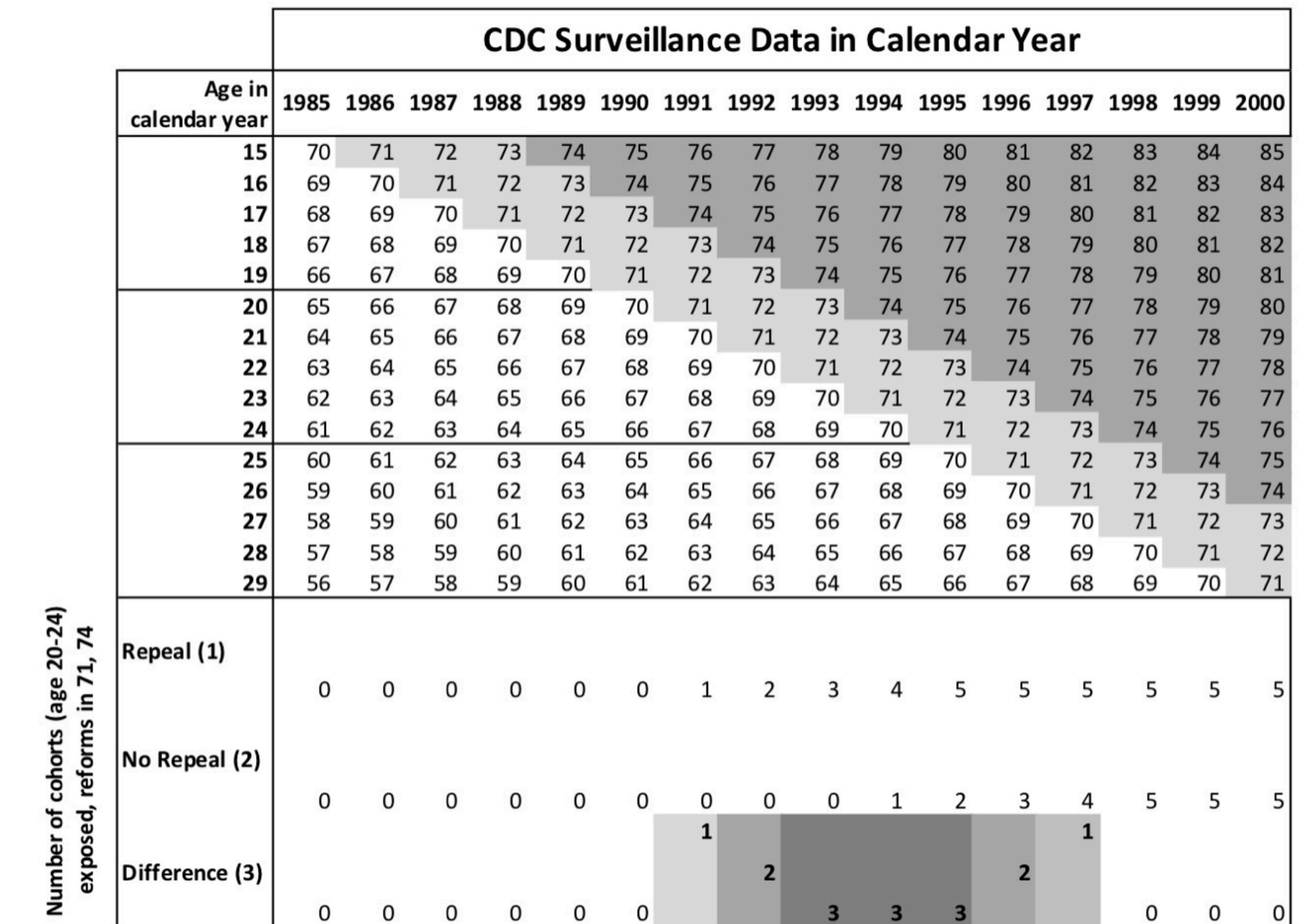


Figure 9.13: Theoretical predictions of abortion legalization on age profiles of gonorrhea incidence for 20-24 year olds

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- The prediction breaks down here as well for the 20-24 year olds
- There is a parabolic effect but it occurs before the predicted time period
- The impacts should show up in 1991, but they begin to appear in 1989
- The impact become statistically insignificant in 1993 instead of 1998

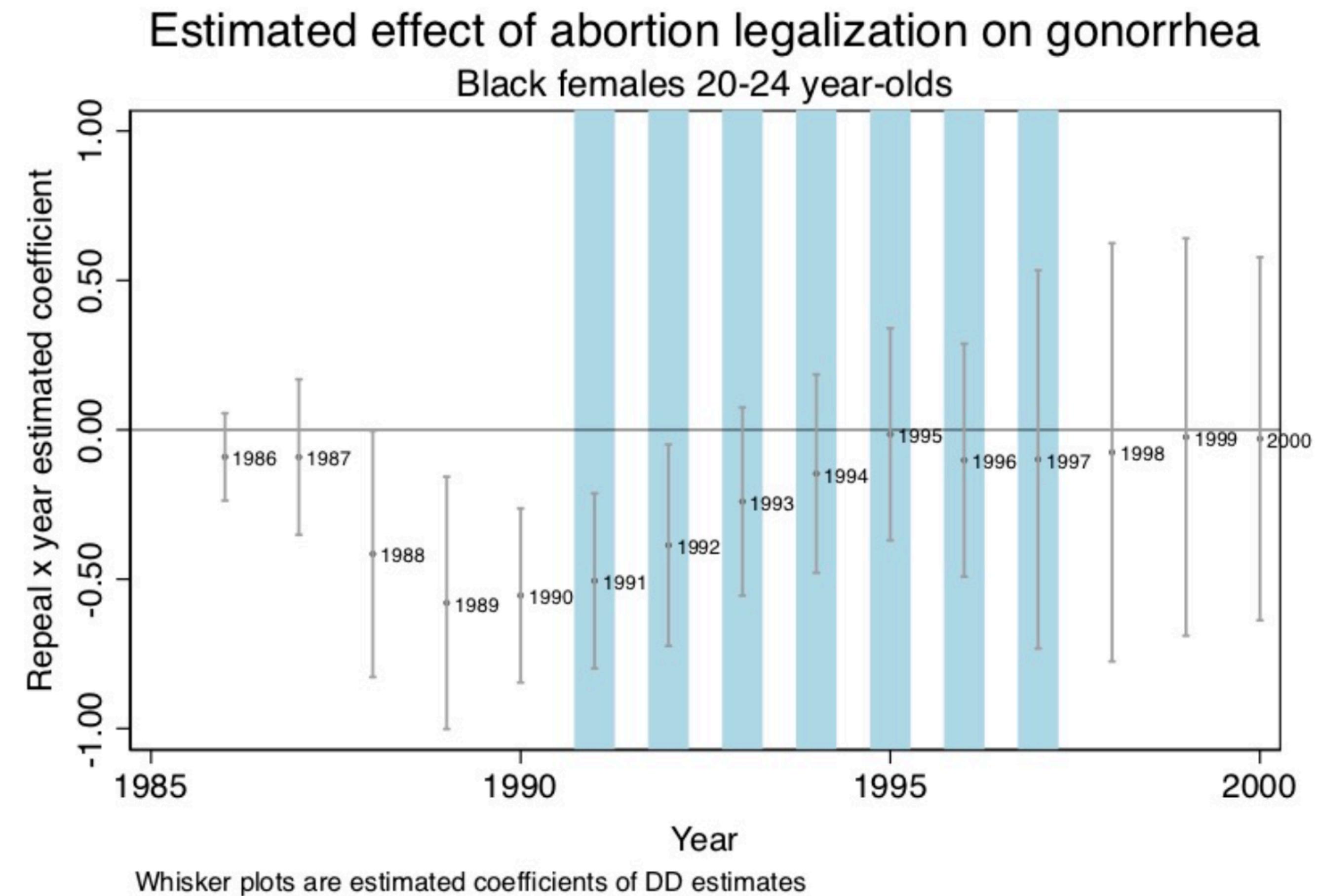


Figure 9.14: Coefficients and standard errors from DD regression equation for the 20-24 year olds

# Difference-in-Difference Example

## Abortion Legalization and Incidence of Gonorrhea

- Cunningham tests a placebo Triple DDD comparing 20-24 year olds with 25-29 year olds
- The prediction breaks down even more
- The original design provided support the abortion legalization hypothesis, but additional tests show that there is little evidence

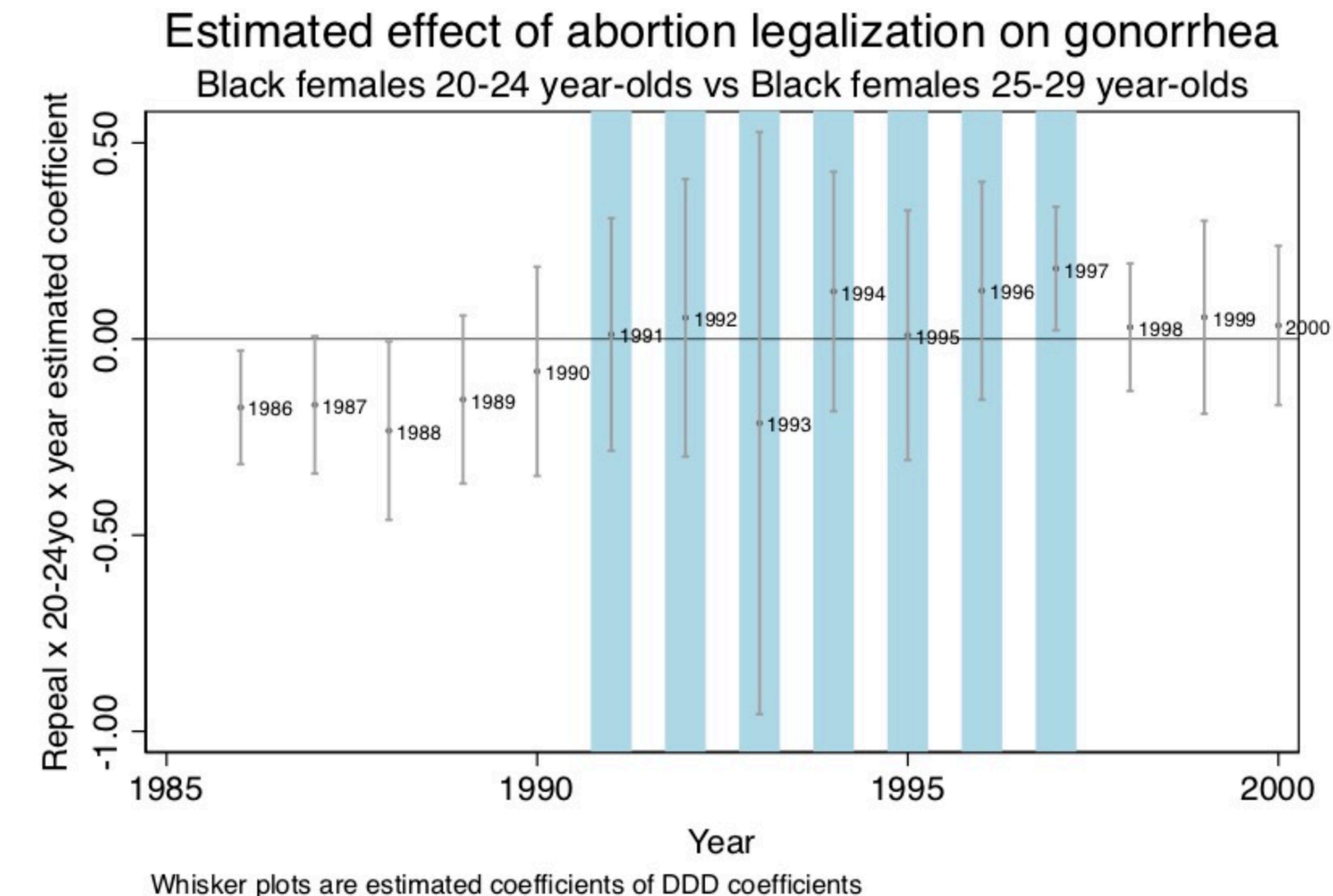


Figure 9.15: Coefficients and standard errors from DDD regression equation for the 20-24 year olds vs 25-29 year olds