Causation and difference-in-differences

Raising the minimum wage

What happens if you raise the minimum wage?

Economic theory says there should be fewer jobs

New Jersey in 1992

$4.25 → $5.05

Before vs. after

Average fast food jobs in NJ

Before: 20.44

After: 21.03

Δ: 0.59

Is this the causal effect?

Treatment vs. control

Average fast food jobs in states

PA_{after}: 21.17

NJ_{after}: 21.03

Δ: −0.14

Is this the causal effect?

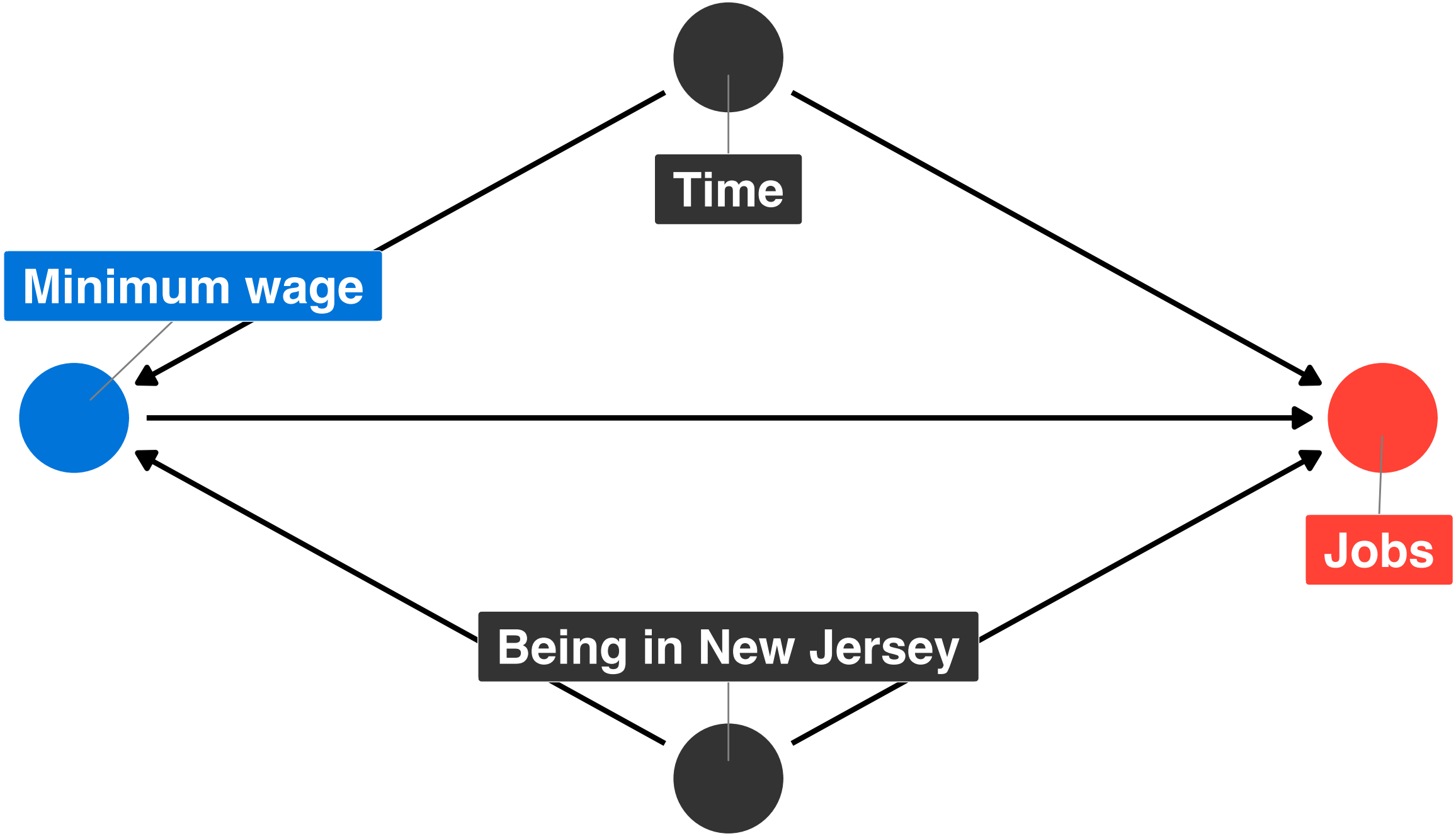
Problems

Comparing only before/after

Impossible to know if growth happened because of treatment or just naturally

Comparing only treatment/control

Impossible to know if any changes happened because of natural growth


Time
Minimum wage
Jobs
Being in New Jersey

	Pre mean	Post mean
Treatment	A (not yet treated)	B (treated)
Control	C (never treated)	D (never treated)

	Pre mean	Post mean	Δ (post–pre)
Treatment	A (not yet treated)	B (treated)	B–A
Control	C (never treated)	D (never treated)	D–C

Growth!

	Pre mean	Post mean
Treatment	A (not yet treated)	B (treated)
Control	C (never treated)	D (never treated)
Δ (trtmt–ctrl)	A–C	B–D

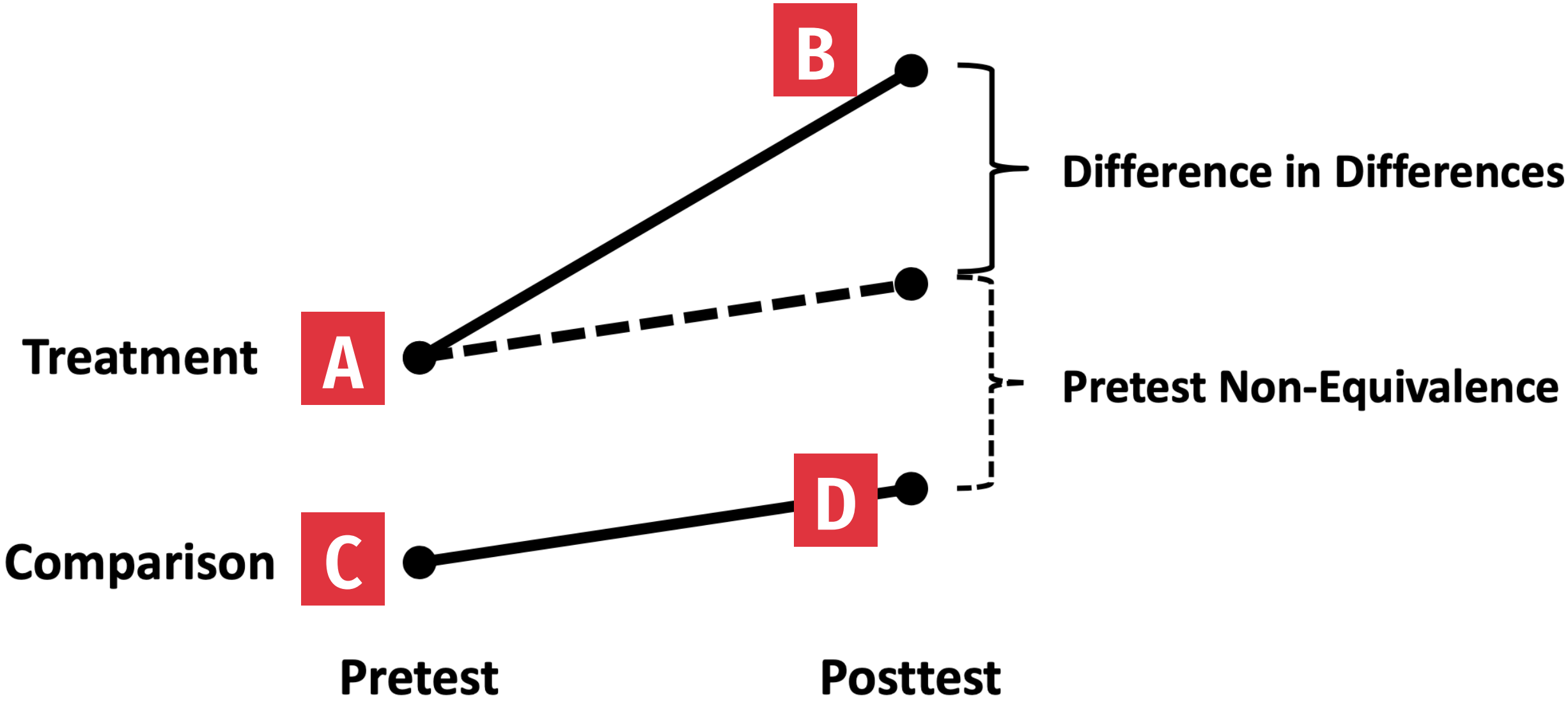
Within-group effects

	Pre mean	Post mean	Δ (post–pre)
Treatment	A (not yet treated)	B (treated)	B–A
Control	C (never treated)	D (never treated)	D–C
Δ (trtmt–ctrl)	A–C	B–D	(B–A) – (D–C)

Growth of treatment – growth of control (DiD!)

$$\begin{aligned}\text{DD} =& (\bar{x}_{\text{treatment, post}} - \bar{x}_{\text{treatment, pre}}) \\ & - (\bar{x}_{\text{control, post}} - \bar{x}_{\text{control, pre}})\end{aligned}$$

	Pre mean	Post mean	Δ (post−pre)
NJ	A **20.44**	B **21.03**	B−A **0.59**
PA	C **23.33**	D **21.17**	D−C **−2.16**
Δ (trtmt−ctrl)	A−C **−2.89**	B−D **−0.14**	**(0.59) − (−2.16) =** **2.76**

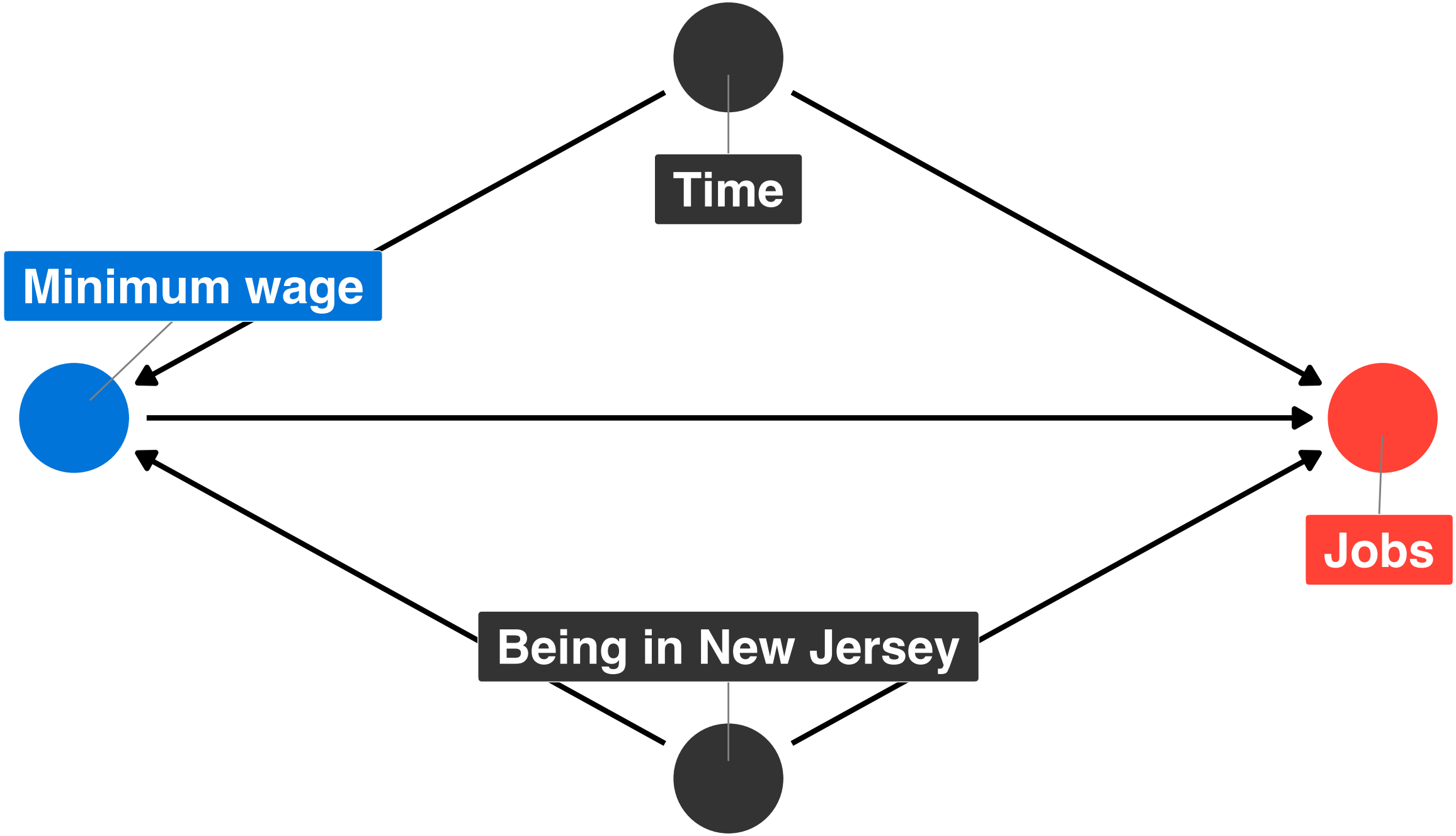

B
Difference in Differences
Treatment
A
Pretest Non-Equivalence
D
Comparison
C
Pretest
Posttest

	Pre mean	Post mean	Δ (post−pre)
Math story	A	B	B−A
Normal story	C	D	D−C
Δ (trtmt−ctrl)	A−C	B−D	(B−A) − (D−C)

Finding all the group means is tedious though!

What if there are other backdoors to worry about?

Regression to the rescue!


Time
Minimum wage
Jobs
Being in New Jersey

$$Y_{it} = \alpha + \beta\ \text{Group}_i + \gamma\ \text{Time}_t + \delta\ (\text{Group}_i \times \text{Time}_t) + \epsilon_{it}$$

```
model <- lm(outcome ~ group + time + group * time)
```

Group = 1/TRUE if treatment

Time = 1/TRUE if after

$$Y_{it} = \alpha + \beta\ \text{Group}_i + \gamma\ \text{Time}_t + \delta\ (\text{Group}_i \times \text{Time}_t) + \epsilon_{it}$$

```
model <- lm(outcome ~ group + time + group * time)
```

α = Mean of control, pre-treatment

β = Increase in outcome across groups

γ = Increase in outcome across time

δ = Difference in differences!

$$Y_{it} = \alpha + \beta\ \text{Group}_i + \gamma\ \text{Time}_t + \delta\ (\text{Group}_i \times \text{Time}_t) + \epsilon_{it}$$

	Pre mean	Post mean	Δ (post−pre)
Control	α	α + γ	γ
Treatment	α + β	α + β + γ + δ	γ + δ
Δ (trtmt−ctrl)	β	β + δ	δ

Our turn

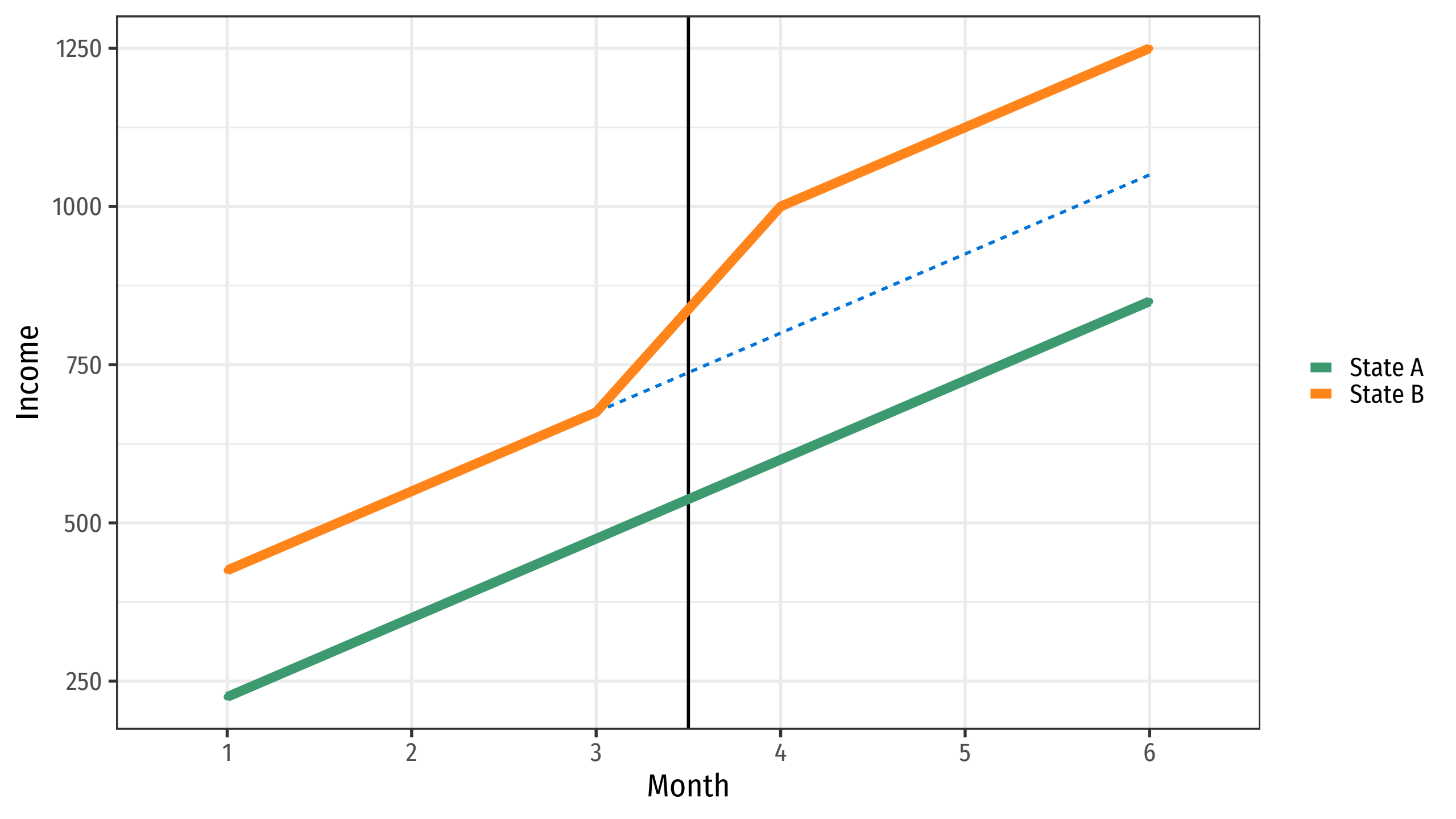
Let’s calculate diff-in-diff estimates with R!

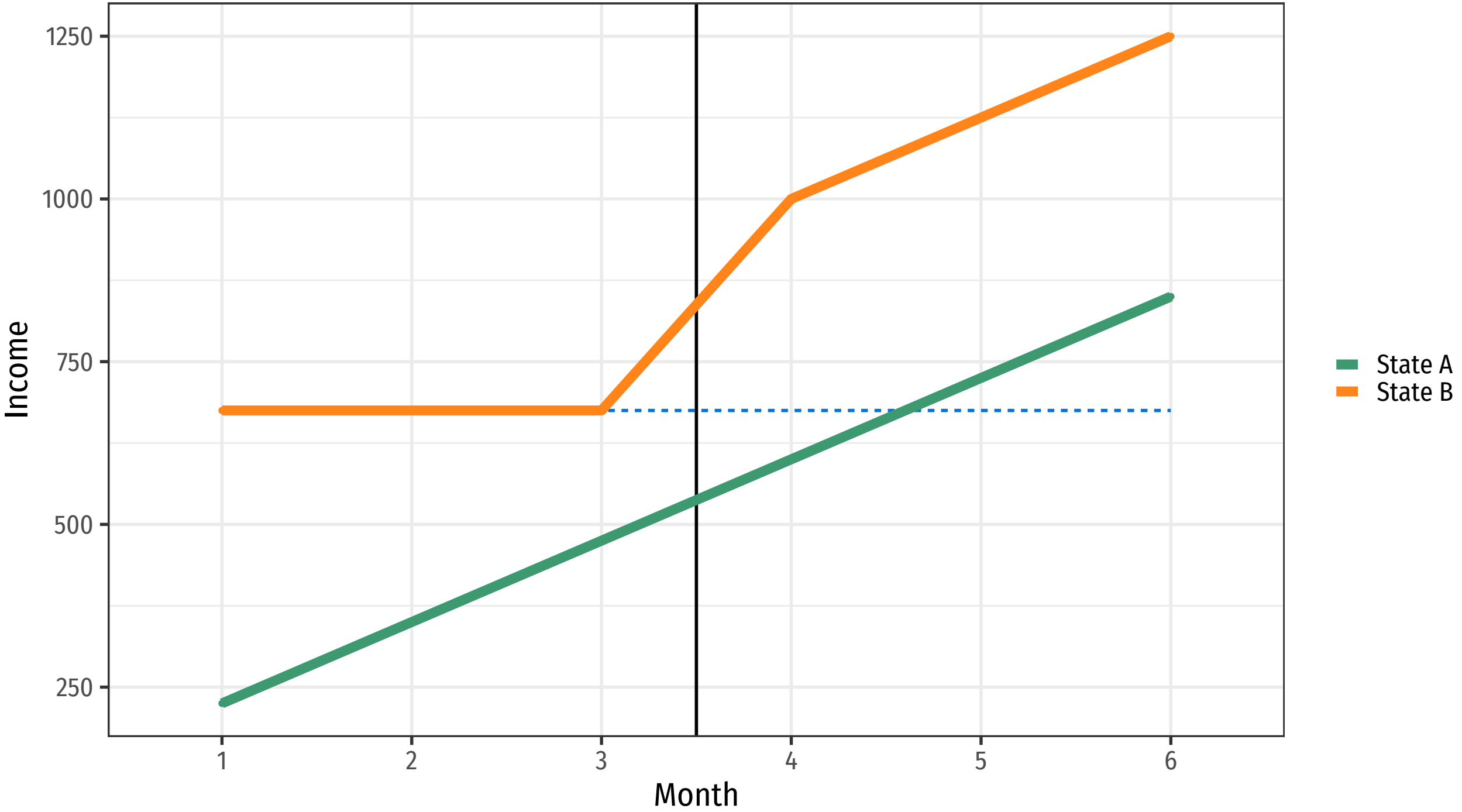
DiD assumptions

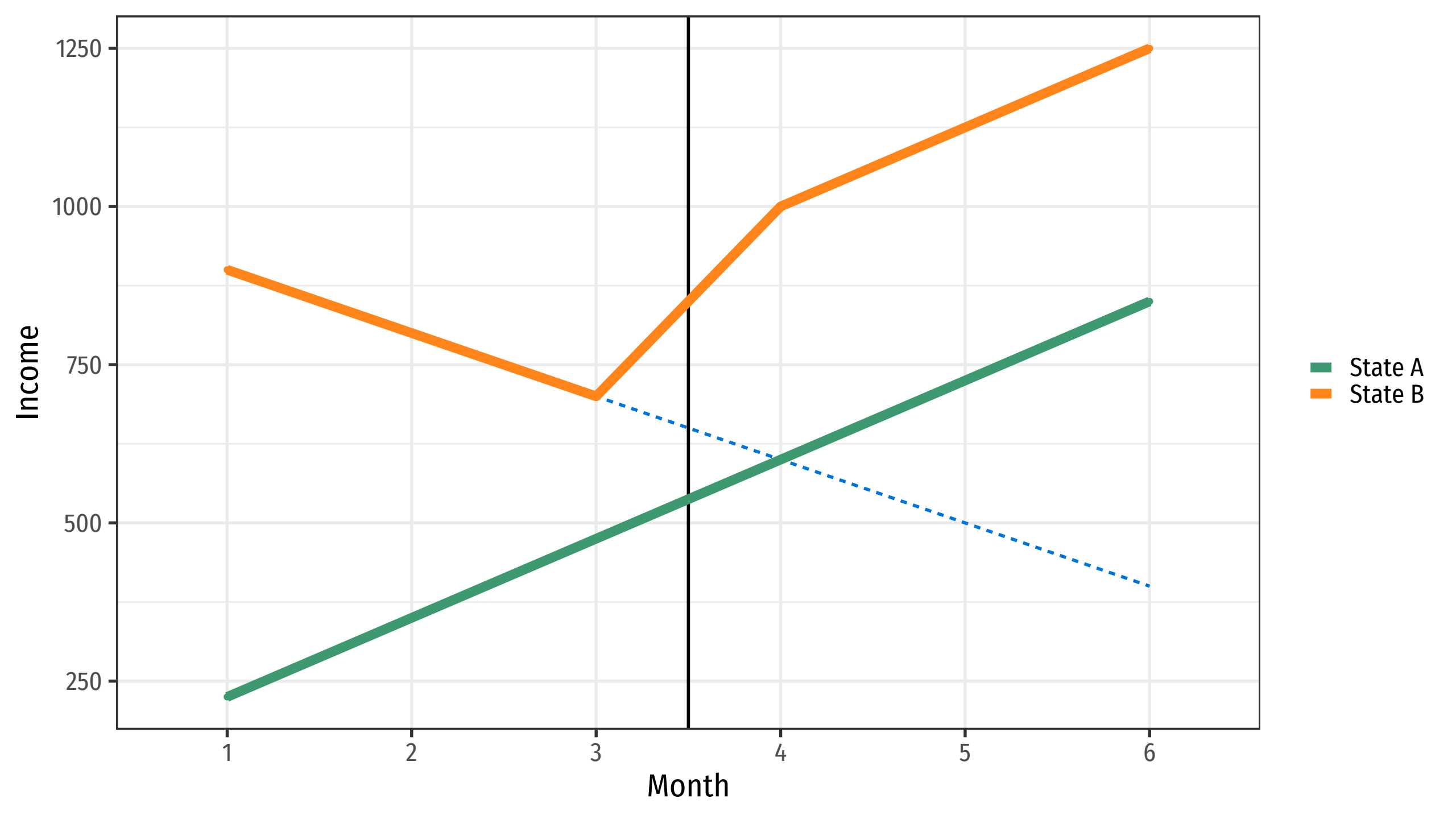
Assumptions

Parallel trends

Treatment and control might have different values at first, but we assume treatment group would have changed like control in absence of treatment


1250
1000
750
500
250
Income
1
2
3
4
5
6
Month
State A
State B

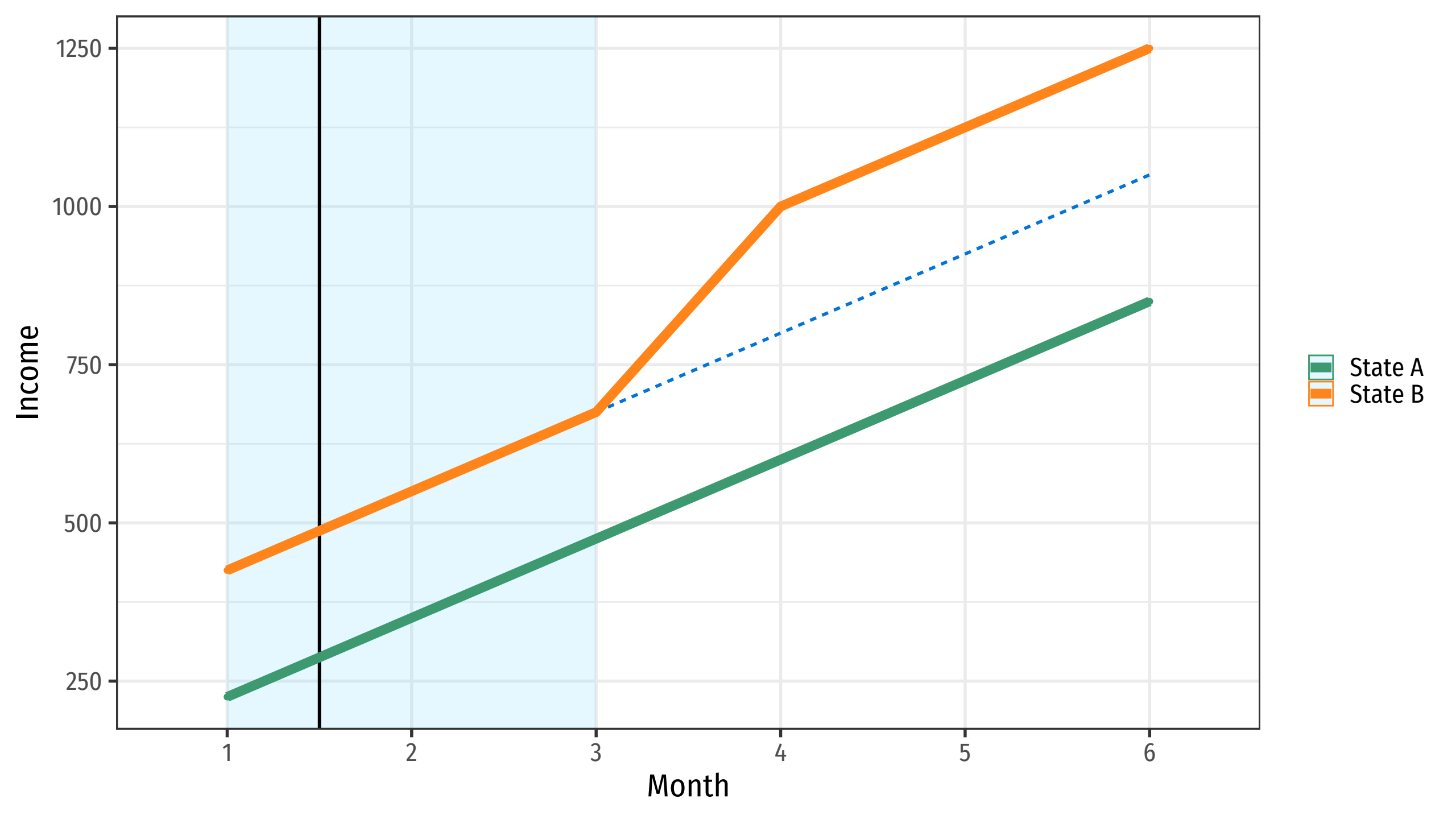

1250
1000
750
500
250
Income
1
2
3
4
5
6
Month
State A
State B

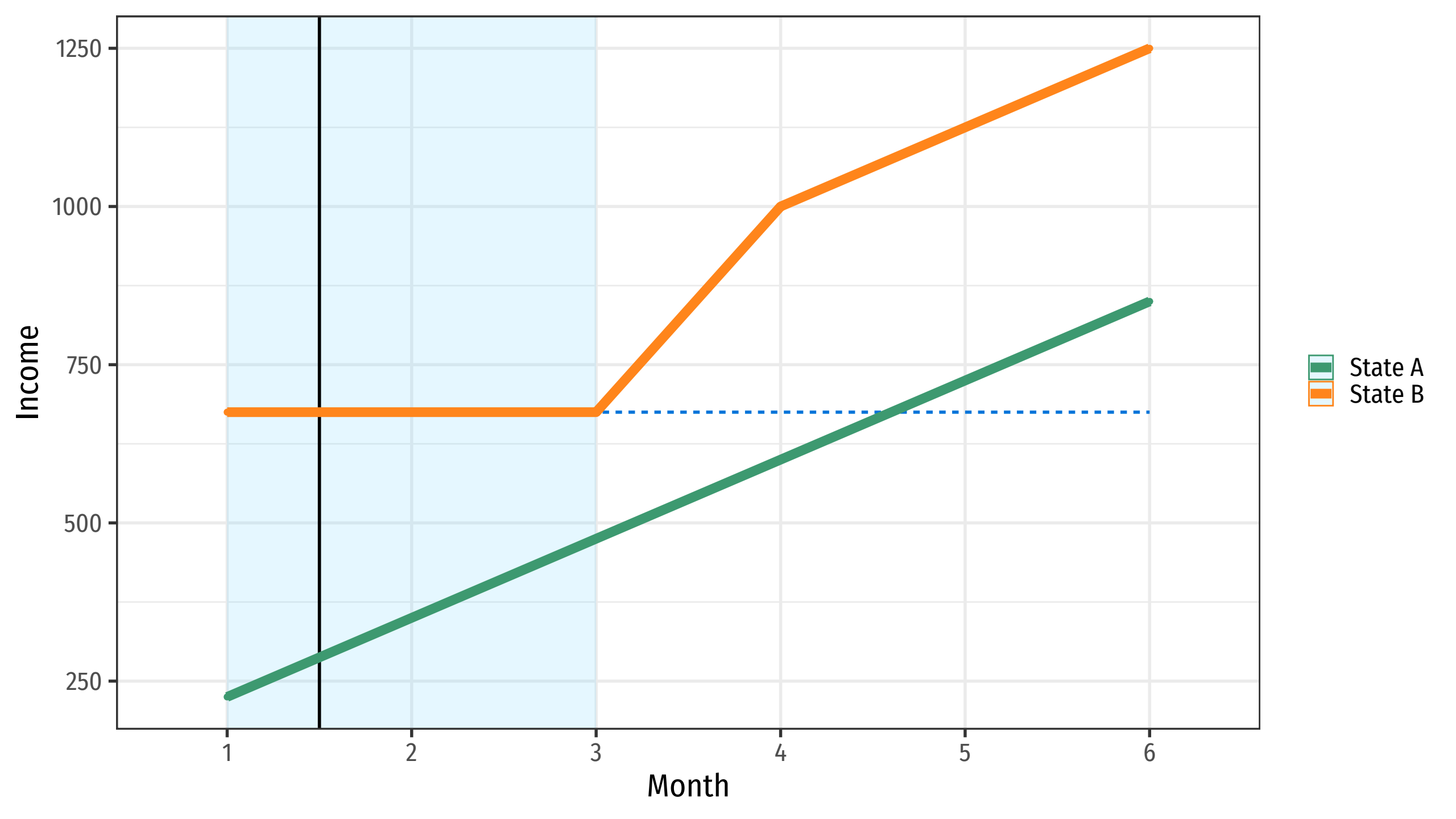

1250
1000
750
500
250
Income
1
2
3
4
5
6
Month
State A
State B

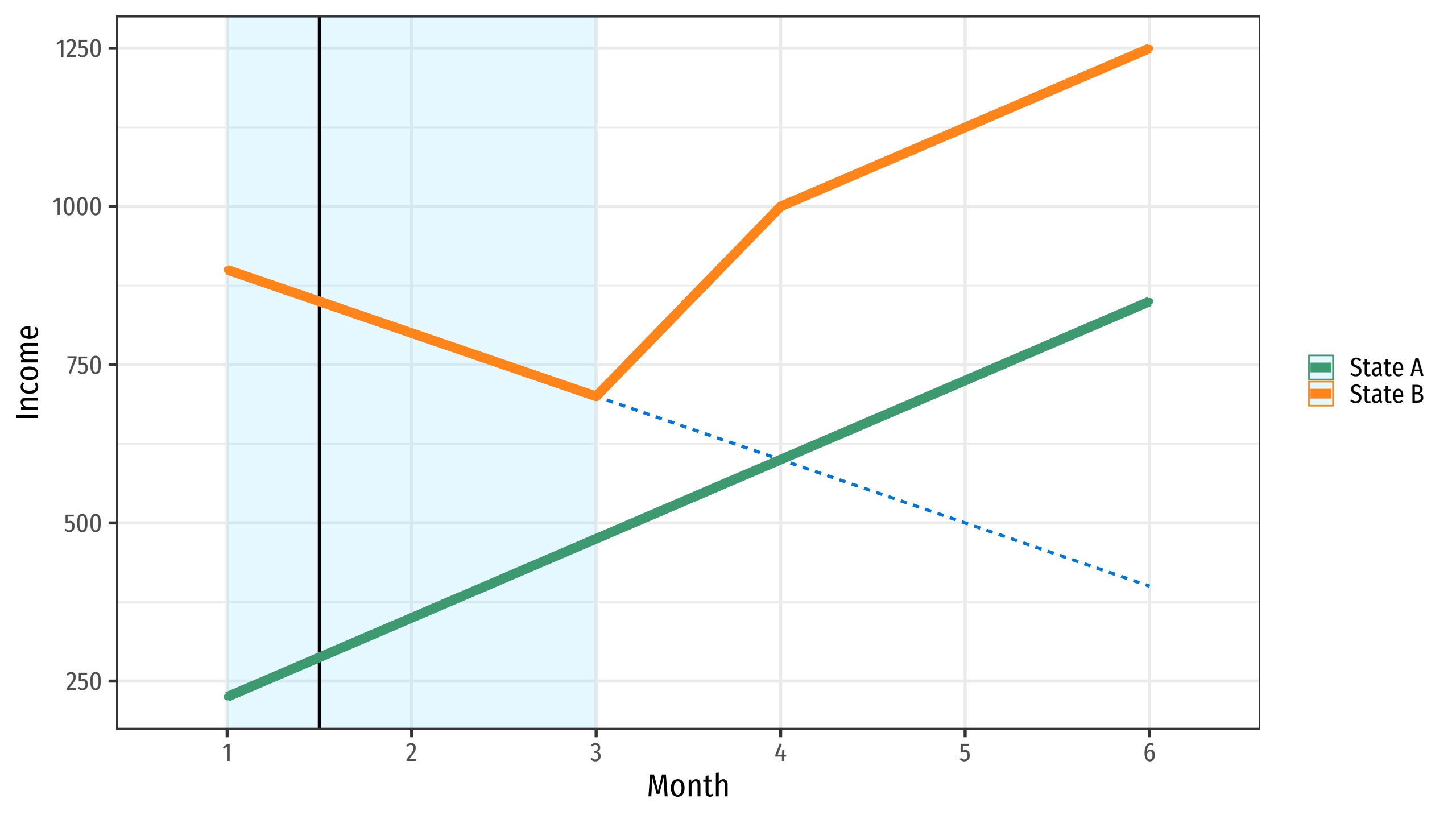
Assumptions

Parallel trends

Check by pretending the treatment happened earlier. If there's an effect, there's an underlying trend.


1250
1000
750
500
250
Income
1
2
3
4
5
6
Month
State A
State B


1250
1000
750
500
250
Income
1
2
3
4
5
6
Month
State A
State B


1250
1000
750
500
250
Income
1
2
3
4
5
6
Month
State A
State B