How to Make Causal Inferences with Time-Series Cross-Sectional Data

Matthew Blackwell
University of Rochester

Adam Glynn
Harvard University

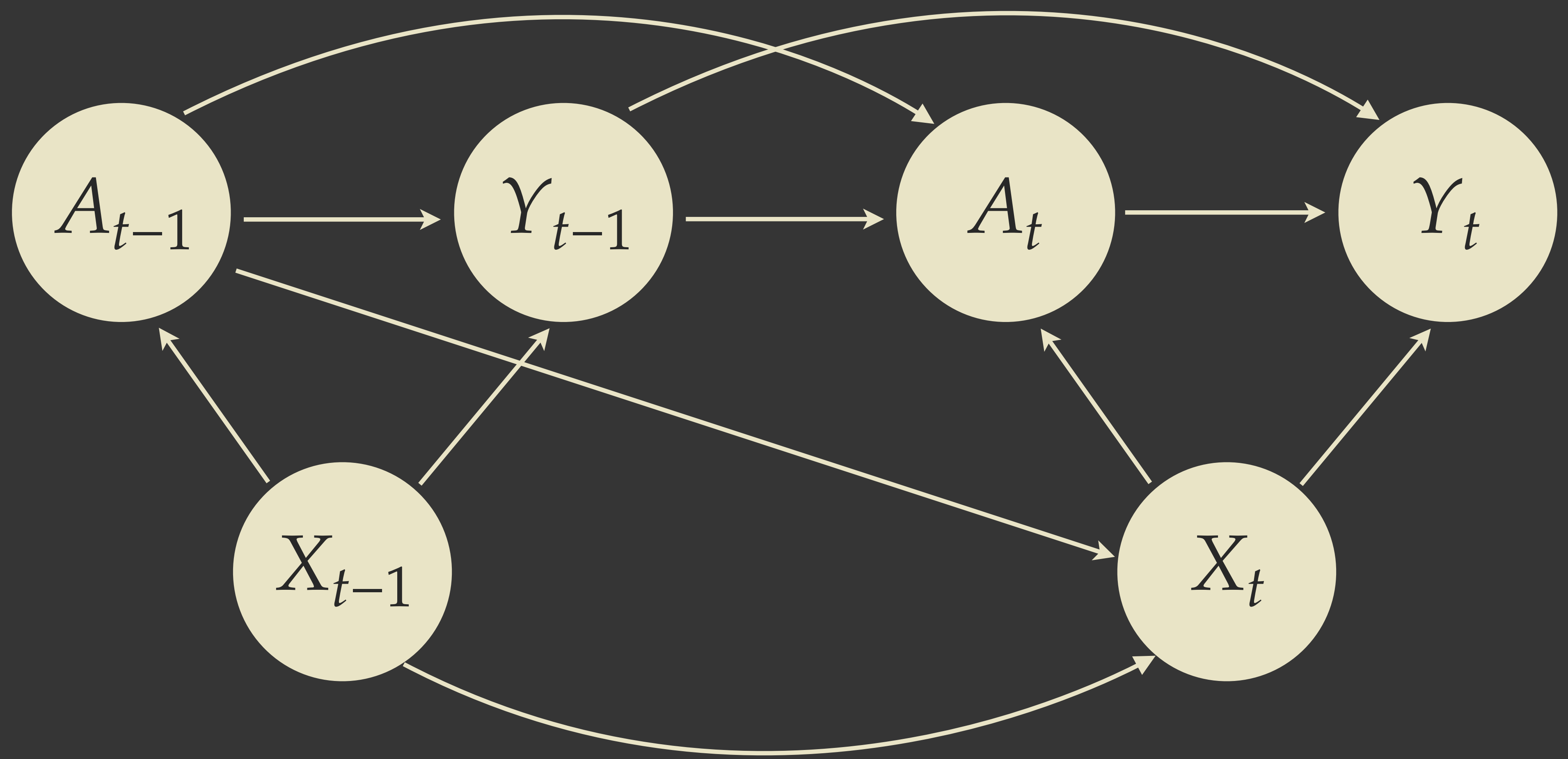
How to Make Causal Inferences with Time-Series Cross-Sectional Data

How to Make Causal Inferences with Time-Series Cross-Sectional Data

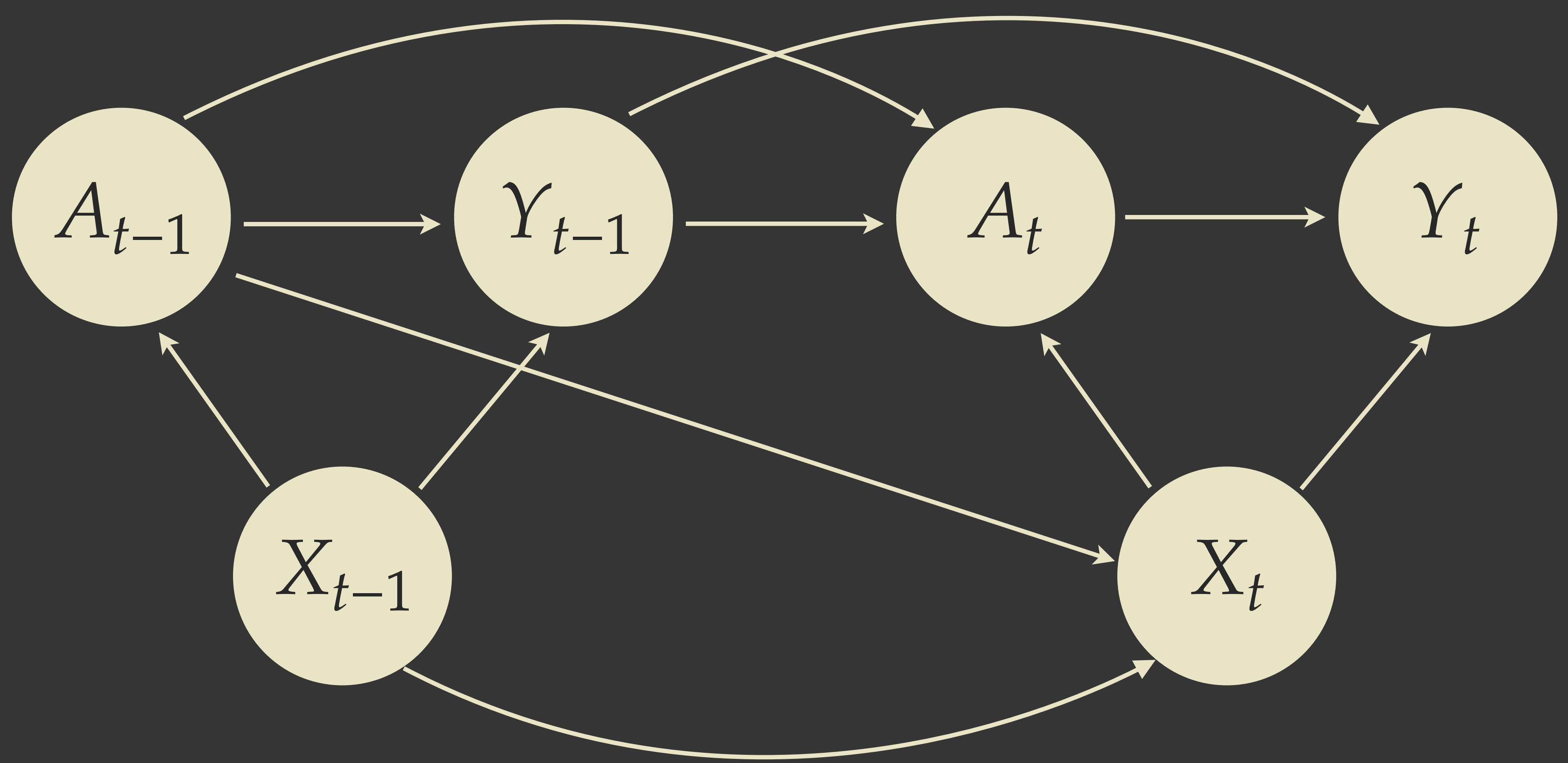
Very Carefully.

How to Make Causal Inferences with Time-Series Cross-Sectional Data

Using weights.

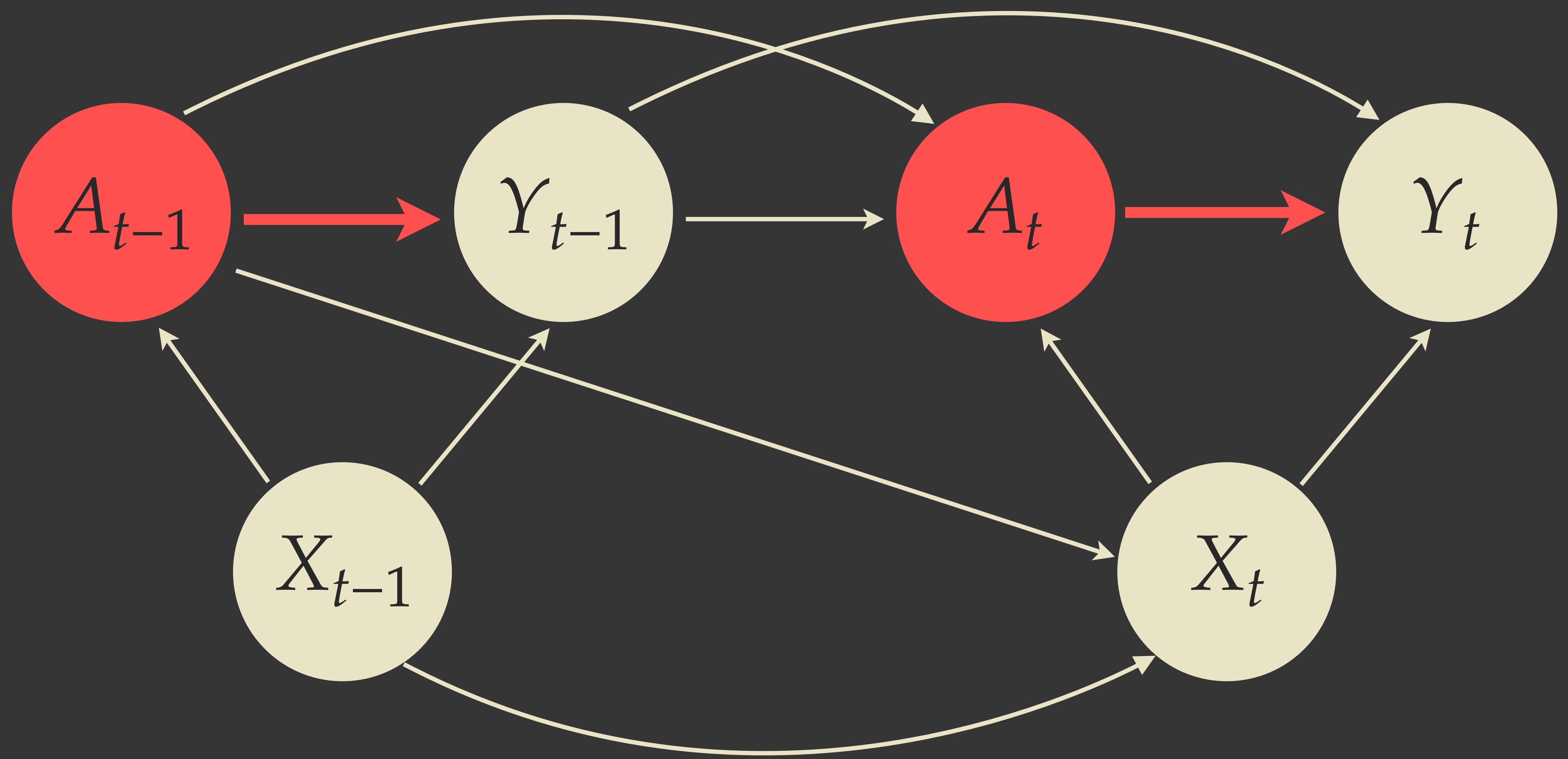

A_{t-1}
Y_{t-1}
A_t
Y_t
X_{t-1}
X_t

What is the effect of A on Y?



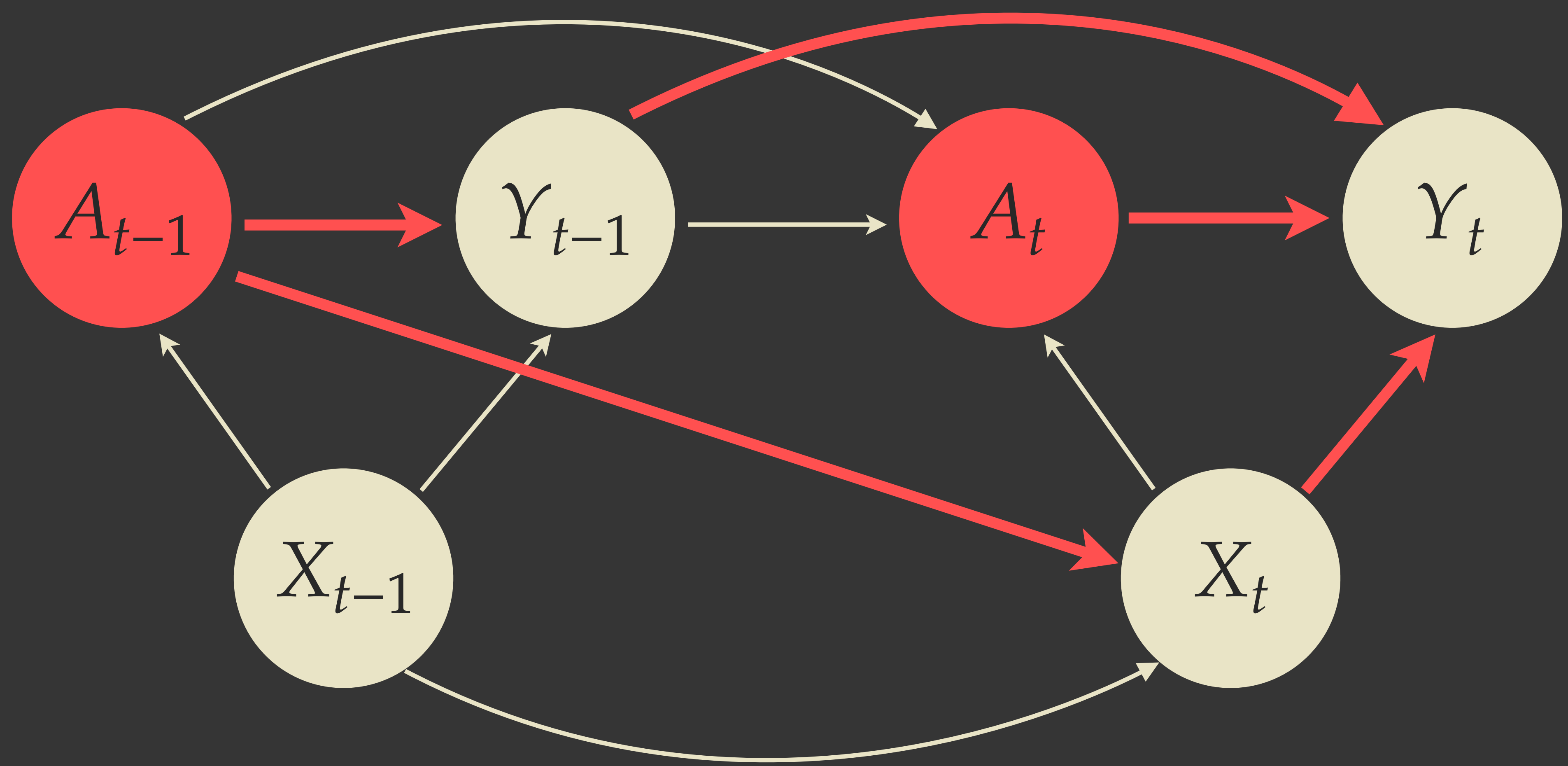
What is the effect of A on Y?

contemporaneous



What is the effect of A on Y?

treatment history



Shouldn't we have more notation?

Shouldn't we have more notation?

$$\underline{A}_t = (A_1, \dots, A_t)$$

Treatment history

Shouldn't we have more notation?

$$\underline{A}_t = (A_1, \dots, A_t)$$

Treatment history

$$\underline{a}_t = (a_1, \dots, a_t)$$

Specific instance of a treatment history

Shouldn't we have more notation?

$$\underline{A}_t = (A_1, \dots, A_t)$$

Treatment history

$$\underline{a}_t = (a_1, \dots, a_t)$$

Specific instance of a treatment history

$$Y_t(\underline{a}_t)$$

Potential outcomes

The effect of history

The effect of history

Average Treatment History Effect

$$\tau(\underline{a}_t, \underline{a}'_t) = E[Y_t(\underline{a}_t) - Y_t(\underline{a}'_t)]$$

The effect of history

Average Treatment History Effect

ATHE

$$\tau(\underline{a}_t, \underline{a}'_t) = E[Y_t(\underline{a}_t) - Y_t(\underline{a}'_t)]$$

The effect of history

Average Treatment History Effect

ATHE

$$\tau(\underline{a}_t, \underline{a}'_t) = E[Y_t(\underline{a}_t) - Y_t(\underline{a}'_t)]$$

The effect of history

Average Treatment History Effect

ATHE

$$\tau(\underline{a}_t, \underline{a}'_t) = E[Y_t(\underline{a}_t) - Y_t(\underline{a}'_t)]$$

1 1 1 1 1 1 1

VS

0 0 0 0 0 0 0

The effect of history

The effect of history

Blip Effect $\tau_b(\underline{a}_{t-1}) = E[Y_t(\underline{a}_{t-1}, 1) - Y_t(\underline{a}_{t-1}, 0)]$

The effect of history

Blip Effect $\tau_b(\underline{a}_{t-1}) = E[Y_t(\underline{a}_{t-1}, 1) - Y_t(\underline{a}_{t-1}, 0)]$

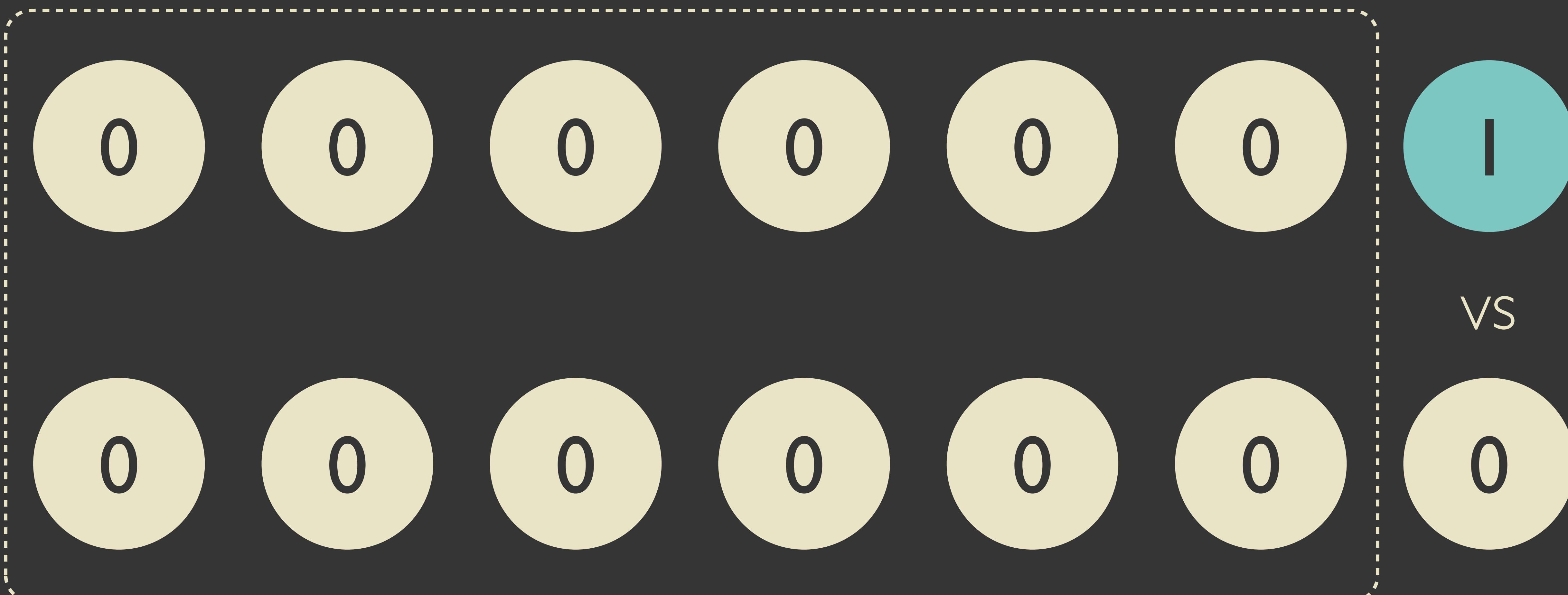
0 0 0 0 0 0 1

vs

0 0 0 0 0 0 0

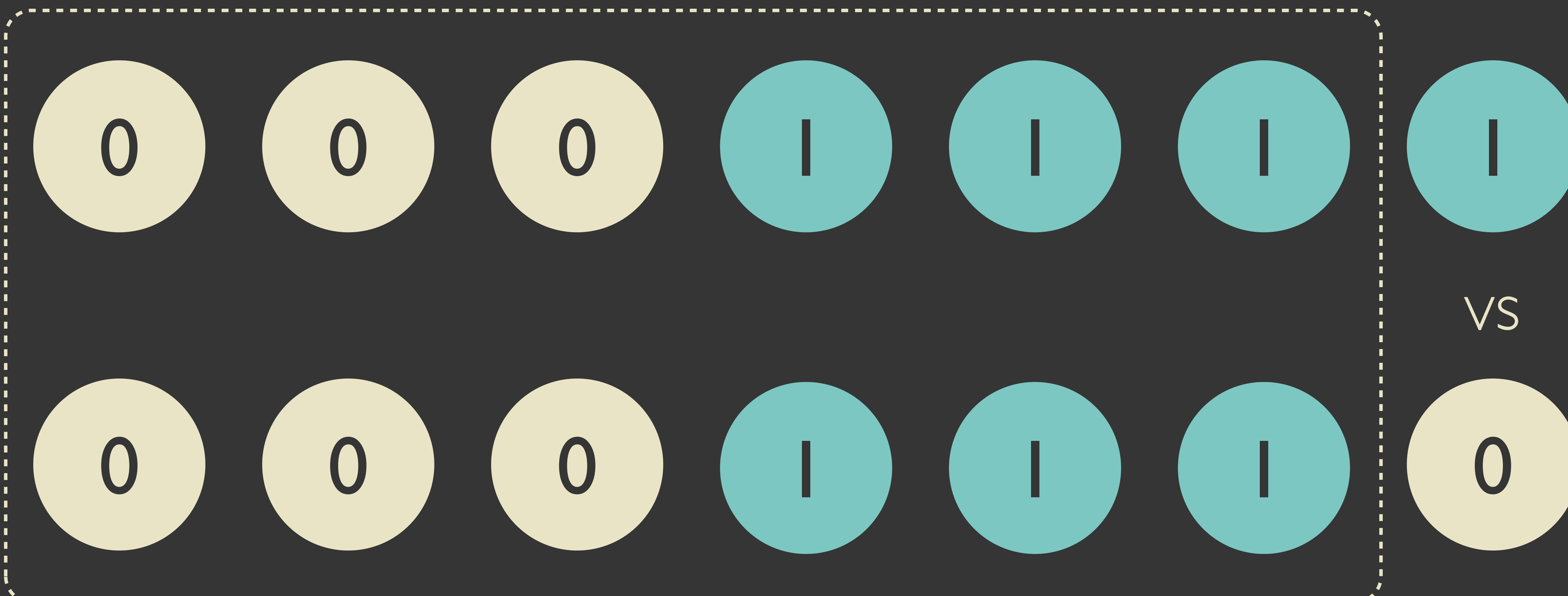
The effect of history

Blip Effect $\tau_b(\underline{a}_{t-1}) = E[Y_t(\underline{a}_{t-1}, 1) - Y_t(\underline{a}_{t-1}, 0)]$



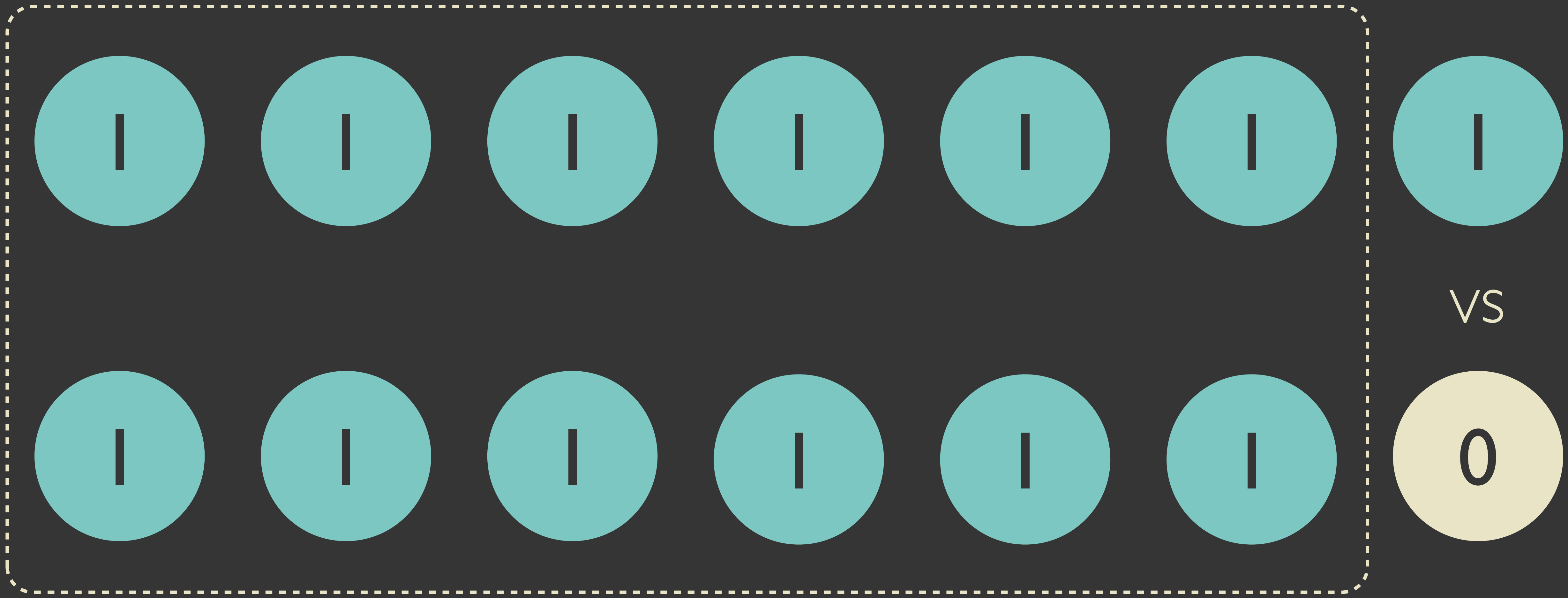
The effect of history

Blip Effect $\tau_b(\underline{a}_{t-1}) = E[Y_t(\underline{a}_{t-1}, 1) - Y_t(\underline{a}_{t-1}, 0)]$



The effect of history

Blip Effect $$\tau_b(\underline{a}_{t-1}) = E[Y_t(\underline{a}_{t-1}, 1) - Y_t(\underline{a}_{t-1}, 0)]$$



The effect of history

The effect of history

Contemporaneous
Effect of Treatment

$$\tau_t = E[\tau_b(\underline{a}_{t-1})]$$

The effect of history

Contemporaneous
Effect of Treatment

CET

$$\tau_t = E[\tau_b(\underline{a}_{t-1})]$$

The effect of history

Contemporaneous
Effect of Treatment

CET

$$\tau_t = E[\tau_b(\underline{a}_{t-1})]$$

1

vs

0

The effect of history

Contemporaneous
Effect of Treatment

CET

$$\tau_t = E[\tau_b(\underline{a}_{t-1})]$$

Marginalize over the past

1

vs

0

TSCS data under sequential ignorability

$$Y_t(\underline{a}_t) \perp\!\!\!\perp A_t | \underline{X}_t, \underline{Y}_{t-1}, \underline{A}_{i,t-1} = \underline{a}_{t-1}$$

Treatment is unrelated
to the potential outcomes

...conditional on the
covariate history.

How conditioning leads you astray

How conditioning leads you astray

...for some questions.

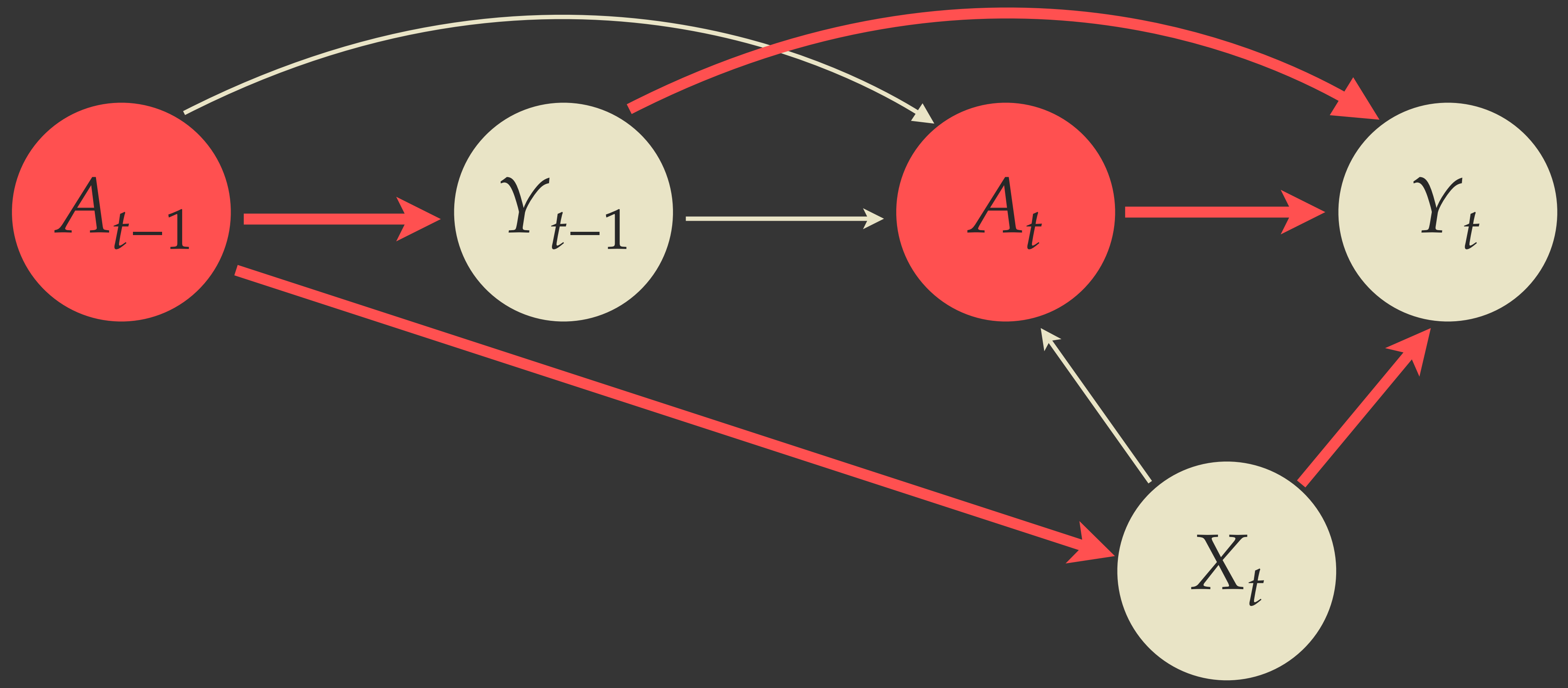
How conditioning leads you astray

...for some questions.

$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

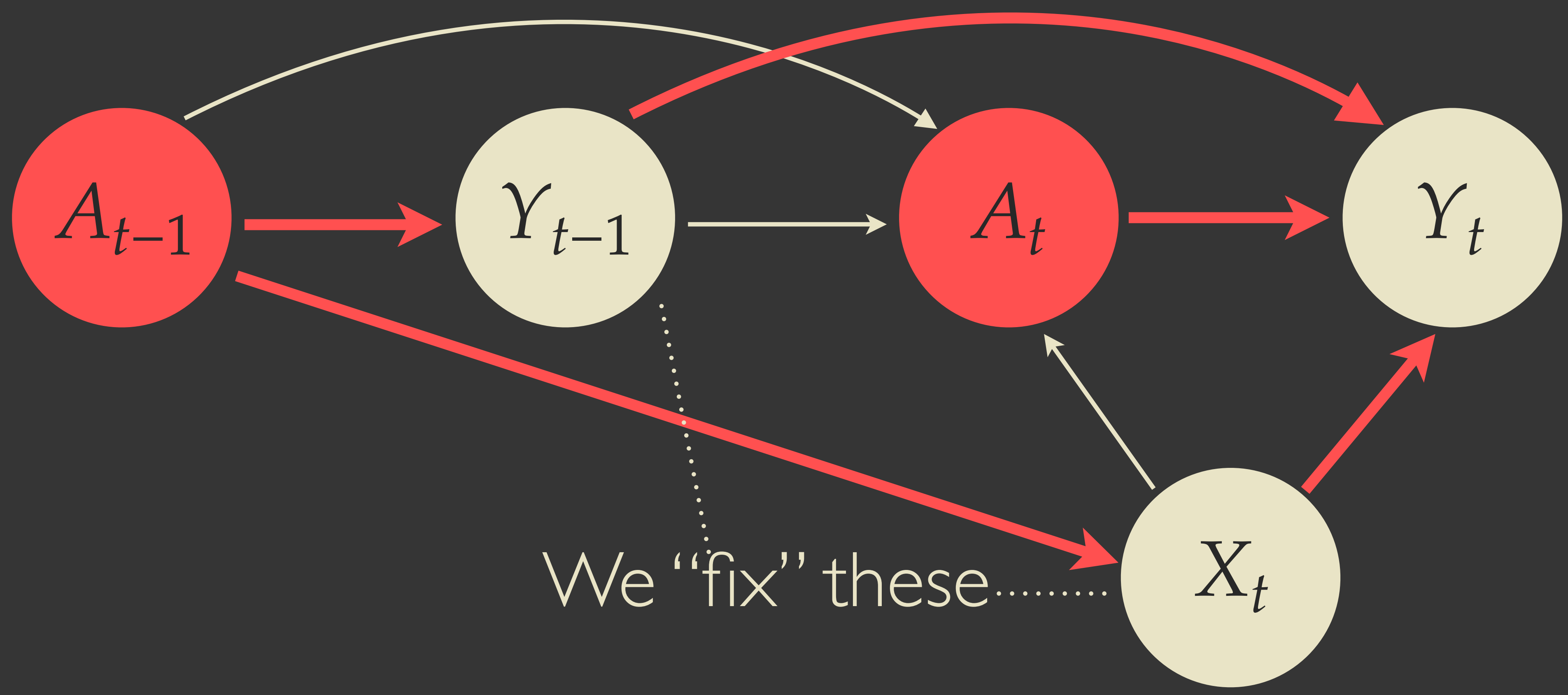
...for some questions.



$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

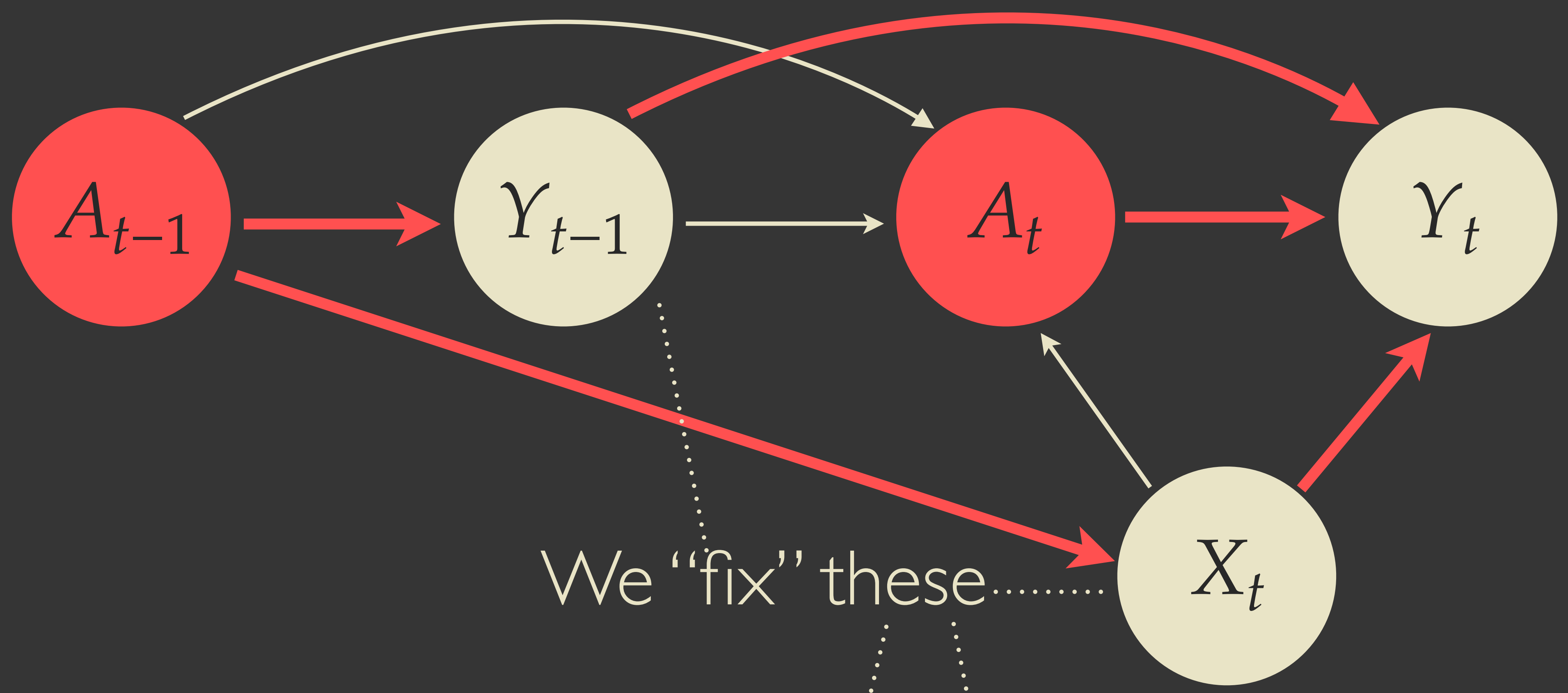
...for some questions.



$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

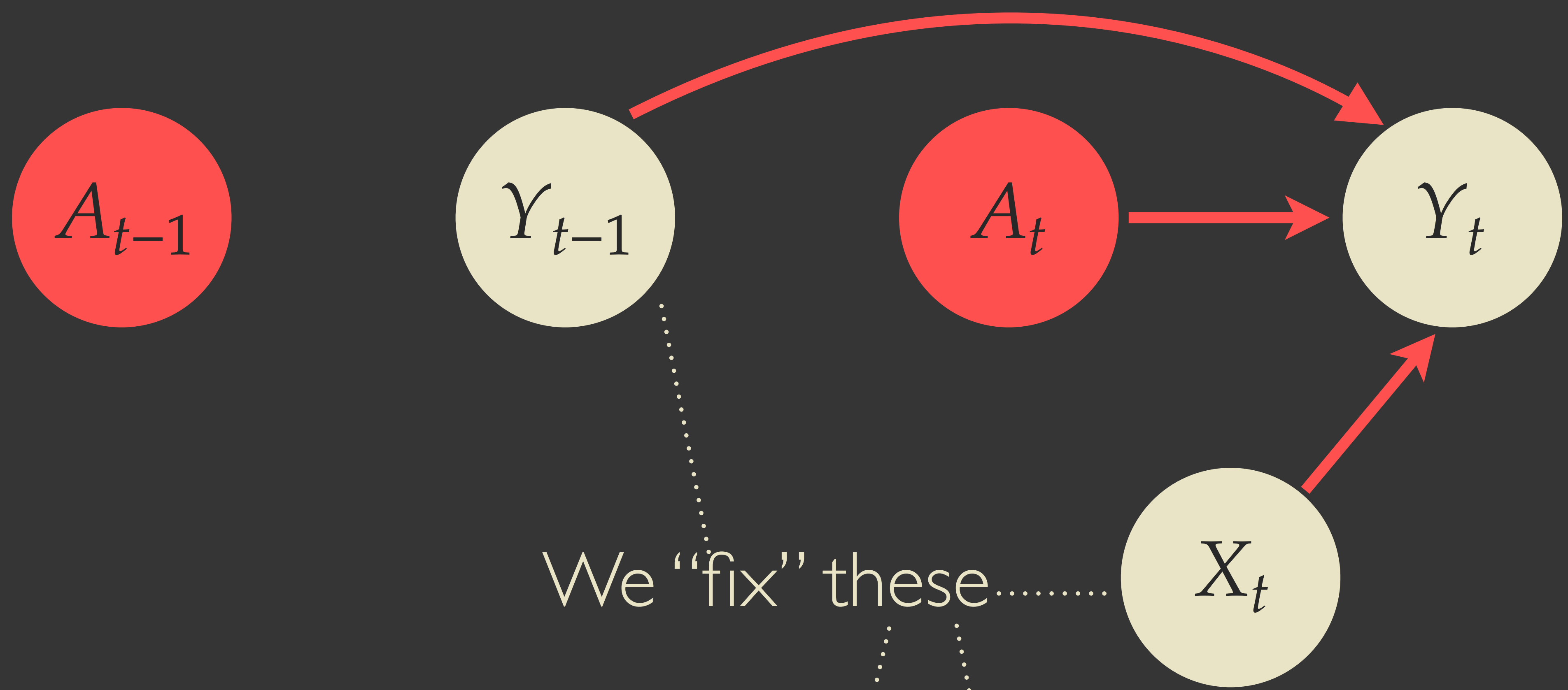
...for some questions.



$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

...for some questions.

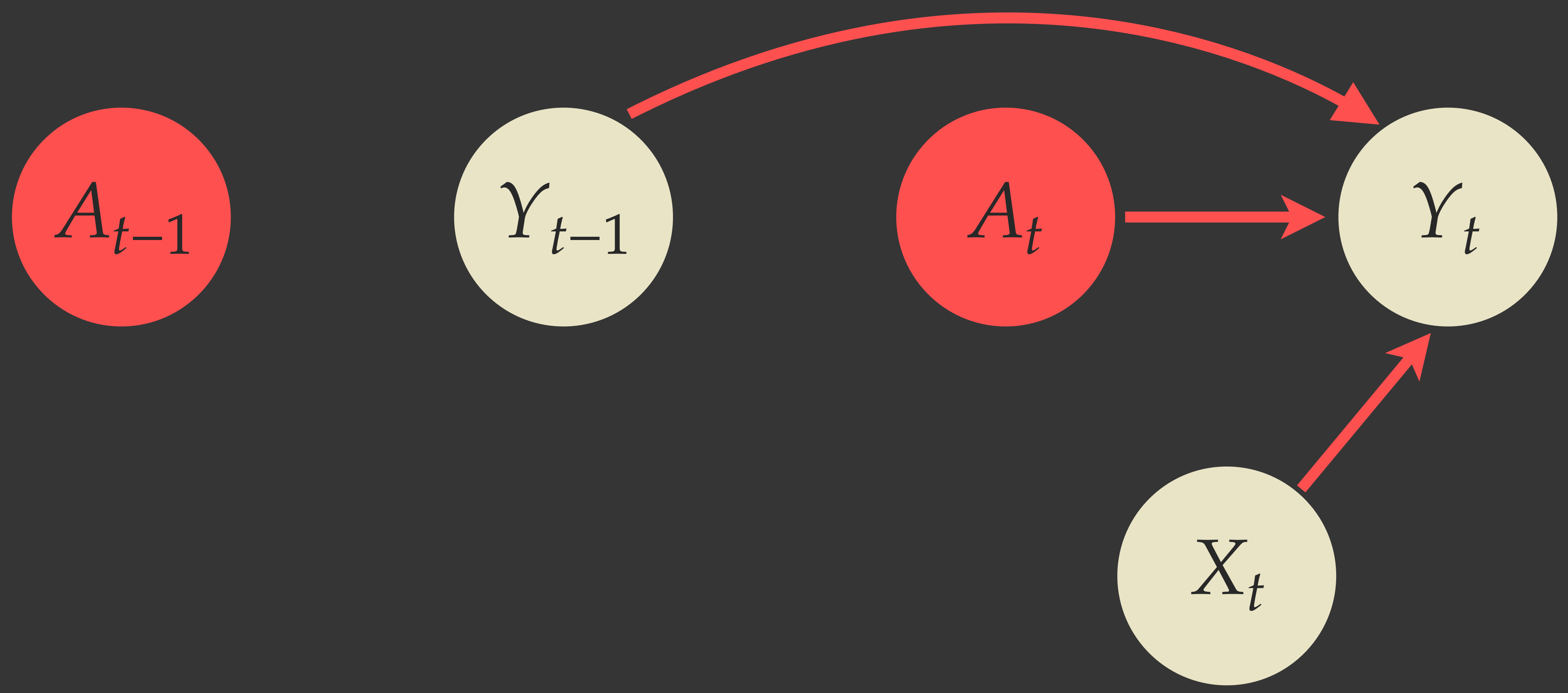


We “fix” these

$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

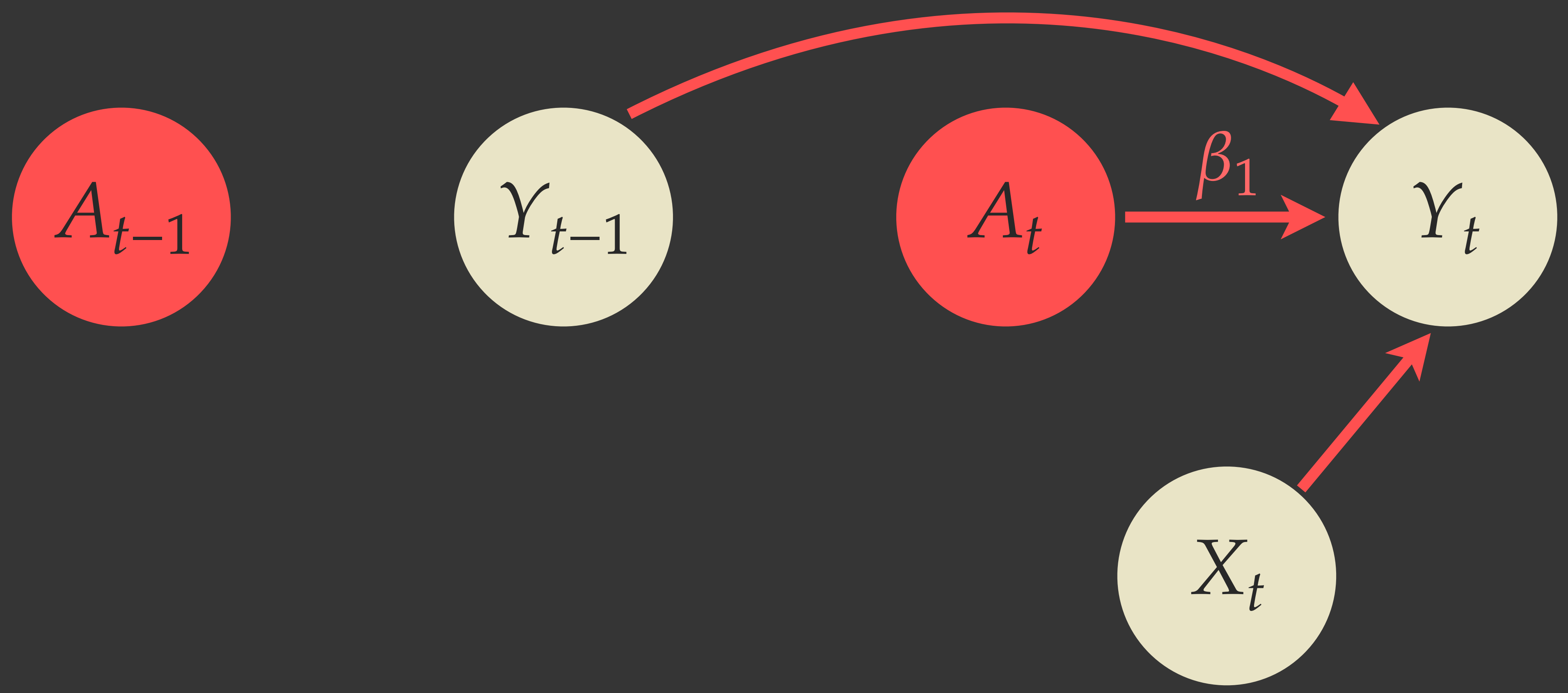
...for some questions.



$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

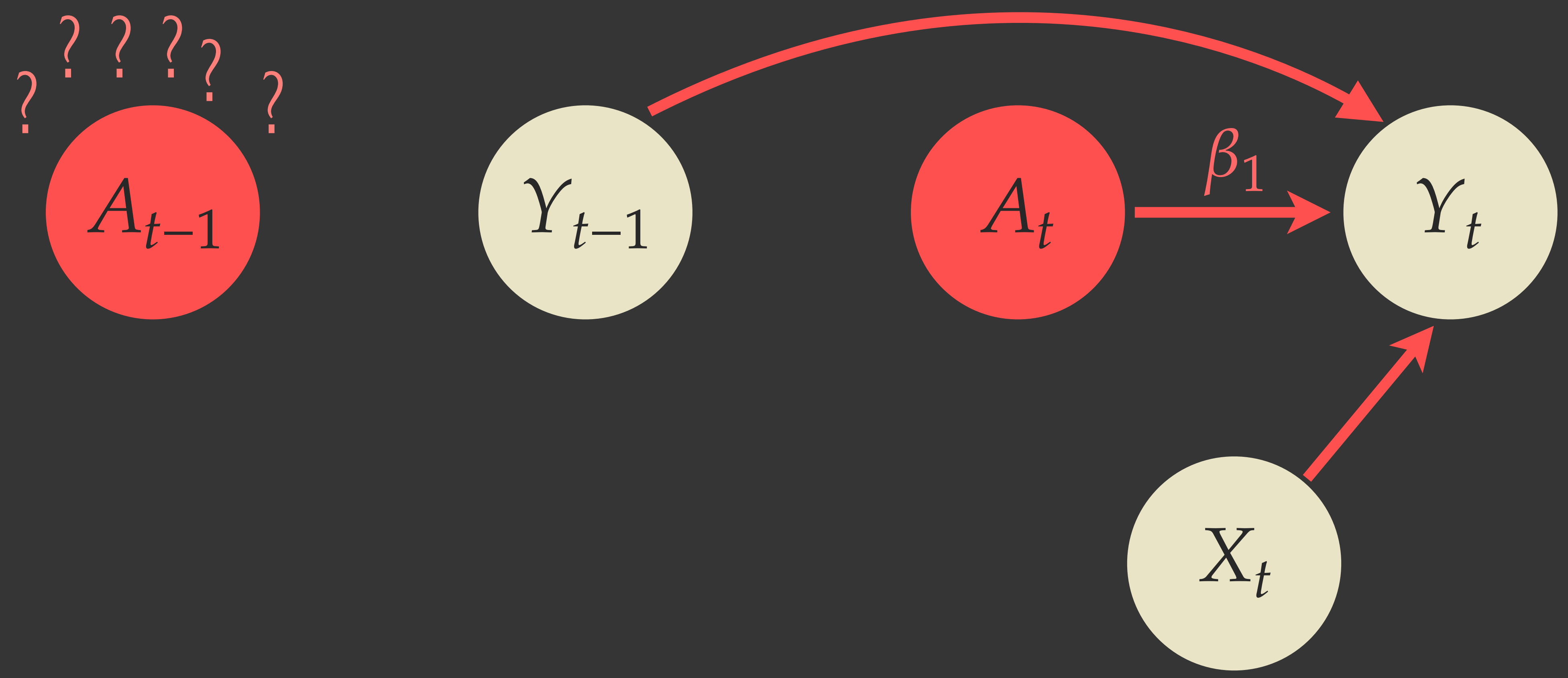
...for some questions.



$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

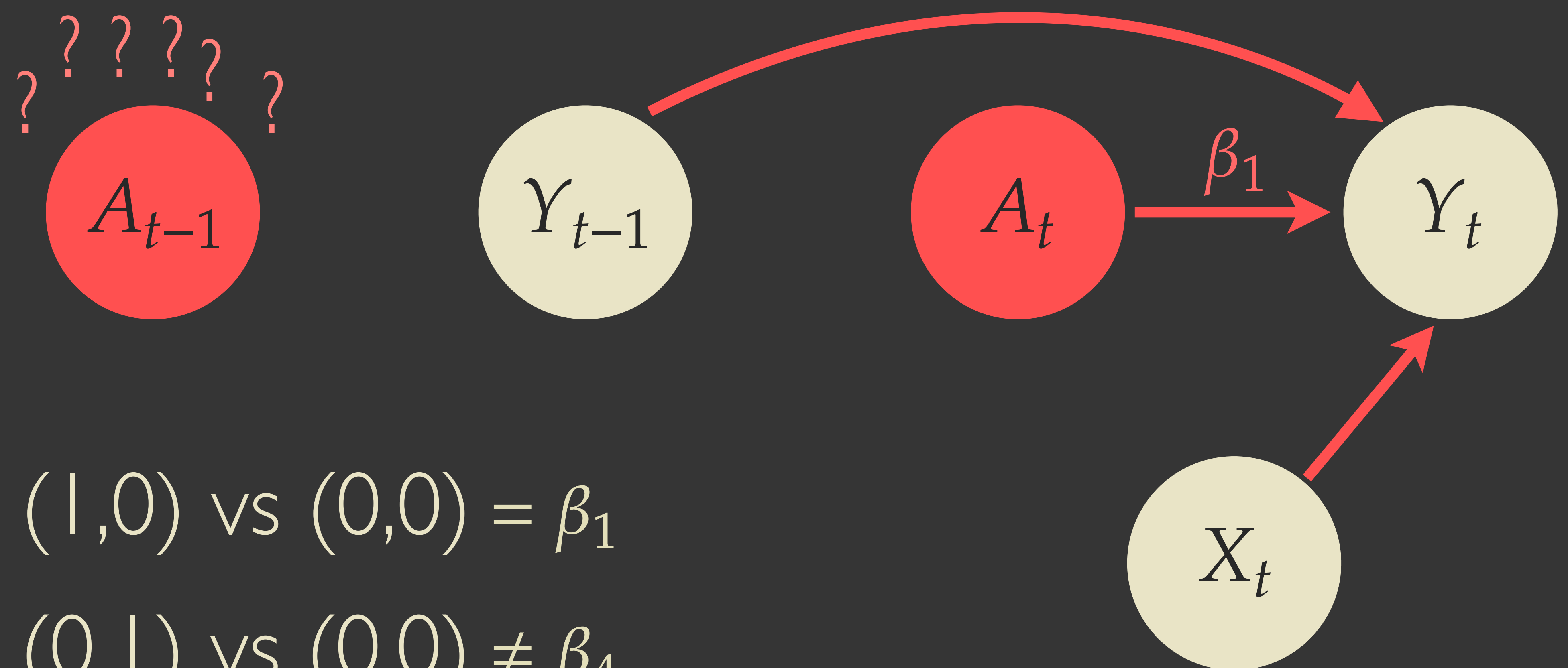
...for some questions.



$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How conditioning leads you astray

...for some questions.



CET: (1,0) vs (0,0) $= \beta_1$

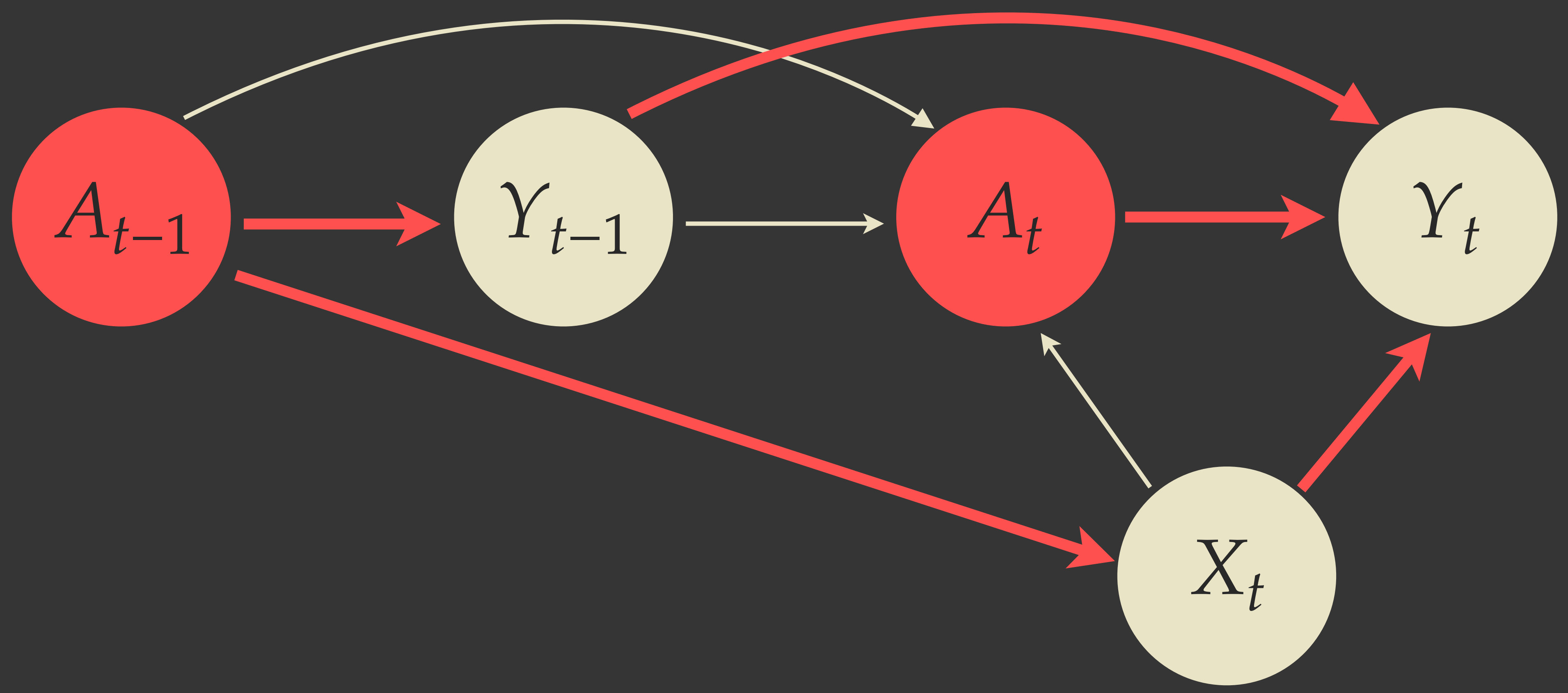
ATHE: (0,1) vs (0,0) $\neq \beta_4$

ATHE: (1,1) vs (0,0) $\neq \beta_1 + \beta_4$

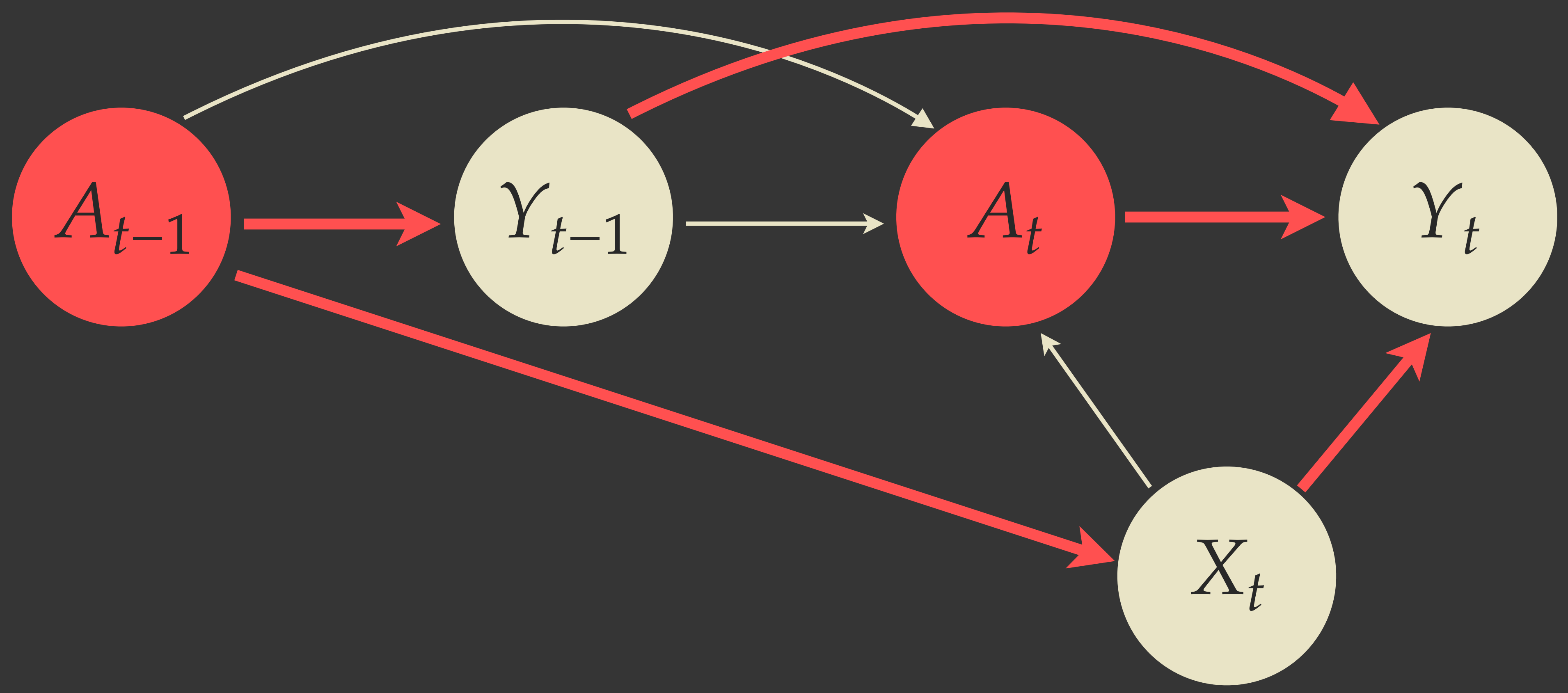
$$Y_t = \beta_0 + \beta_1 A_t + \beta_2 X_t + \beta_3 Y_{t-1} + \beta_4 A_{t-1}$$

How weighting can help

How weighting can help

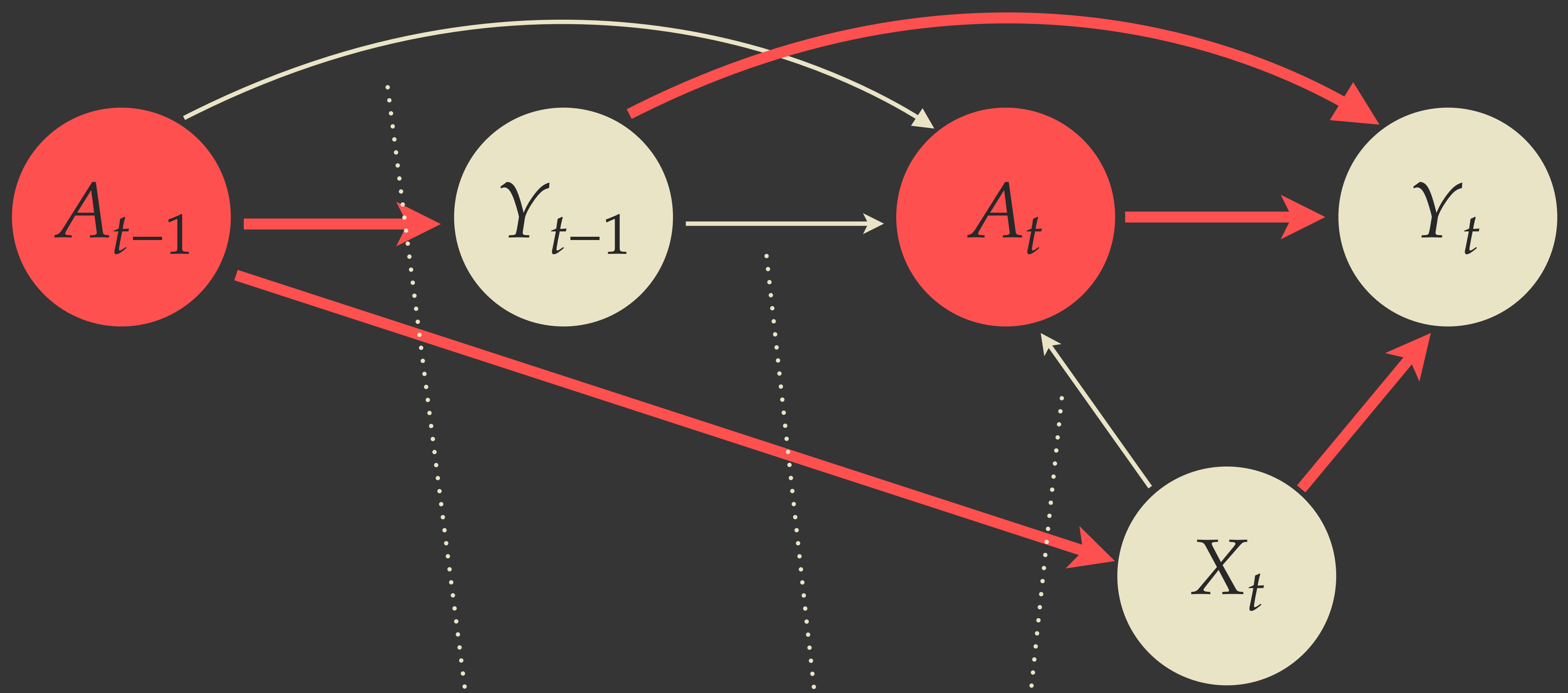


How weighting can help



$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is}|A_{t-1}, X_t, Y_{t-1}]}.$$

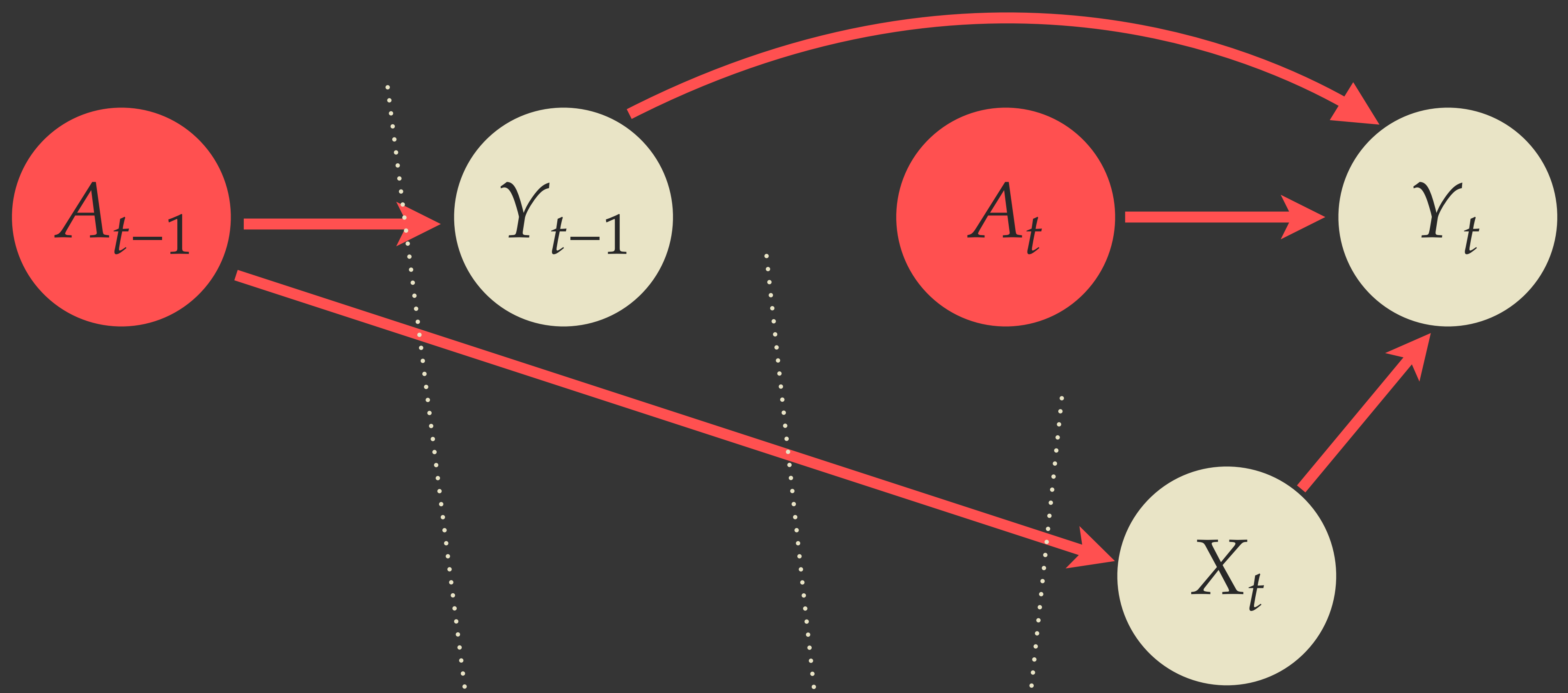
How weighting can help



We weight to create balance

$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is} | A_{t-1}, X_t, Y_{t-1}]}.$$

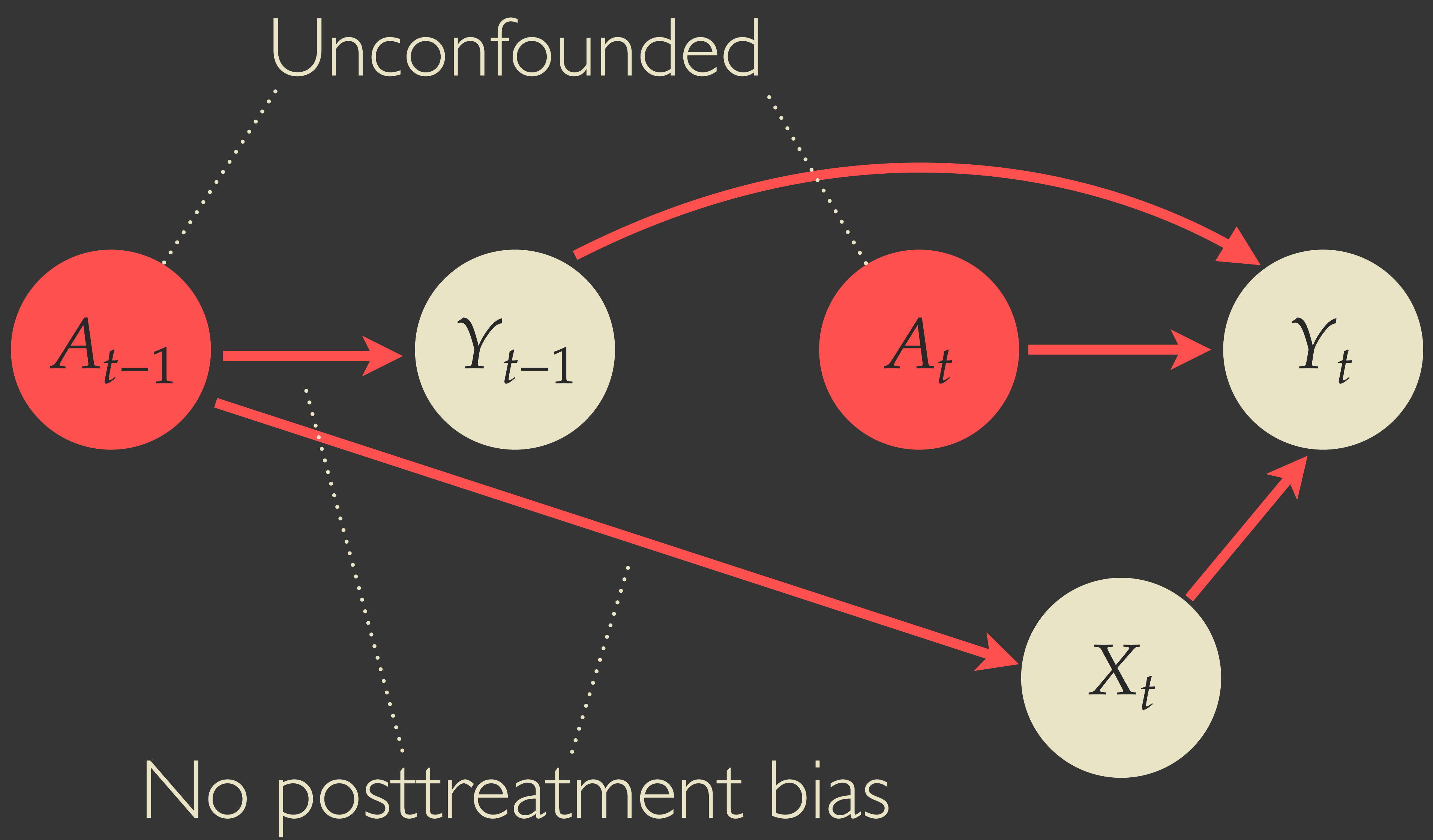
How weighting can help



We weight to create balance

$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is} | A_{t-1}, X_t, Y_{t-1}]}.$$

How weighting can help



$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is} | A_{t-1}, X_t, Y_{t-1}]}.$$

How weighting can help

How weighting can help

$$
\begin{aligned}
E[Y_t(a_t, a_{t-1})] &= E_W[Y_t | A_t = a_t, A_{t-1} = a_{t-1}] \\
&= \beta_0 + \beta_1 a_t + \beta_2 a_{t-1}
\end{aligned}
$$

How weighting can help

WLS

$$E[Y_t(a_t, a_{t-1})] = E_W[Y_t | A_t = a_t, A_{t-1} = a_{t-1}]$$
$$= \beta_0 + \beta_1 a_t + \beta_2 a_{t-1}$$

How weighting can help

WLS

$$E[Y_t(a_t, a_{t-1})] = E_W[Y_t | A_t = a_t, A_{t-1} = a_{t-1}]$$
$$= \beta_0 + \beta_1 a_t + \beta_2 a_{t-1}$$

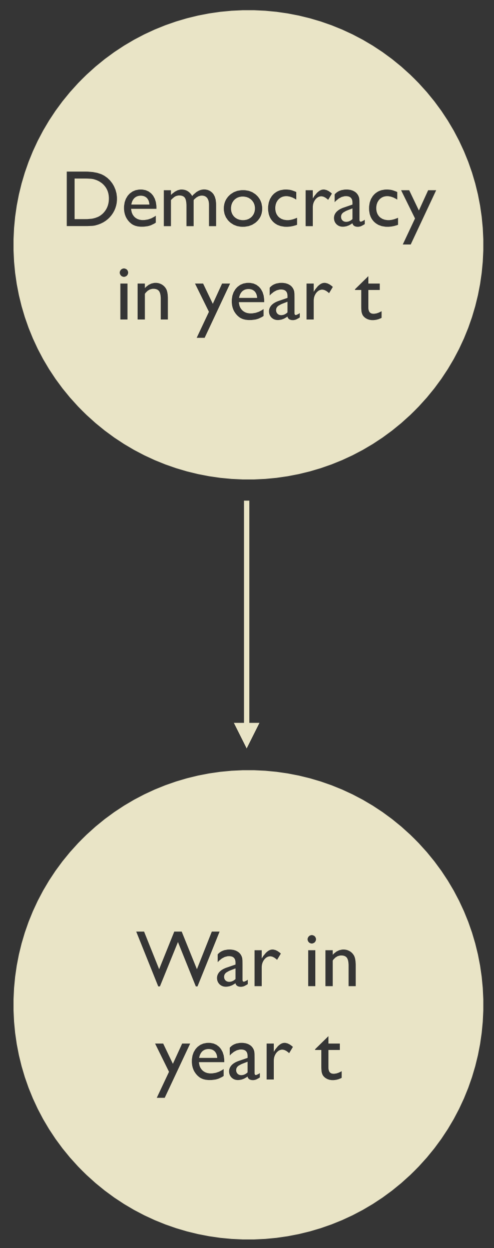
CET: (1,0) vs (0,0) $= \beta_1$

ATHE: (0,1) vs (0,0) $= \beta_2$

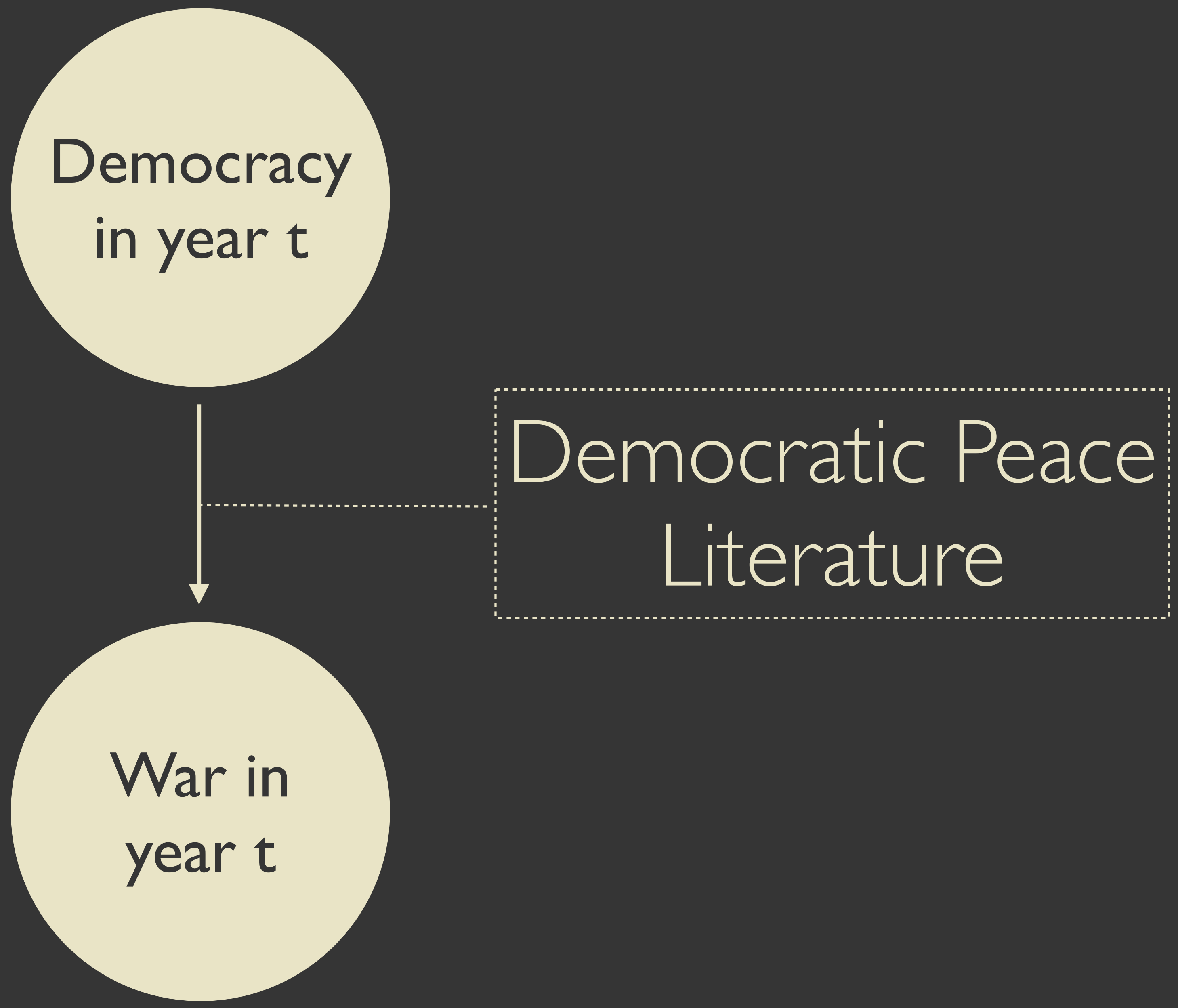
ATHE: (1,1) vs (0,0) $= \beta_1 + \beta_2$

The Long Arm of the Democratic Peace?

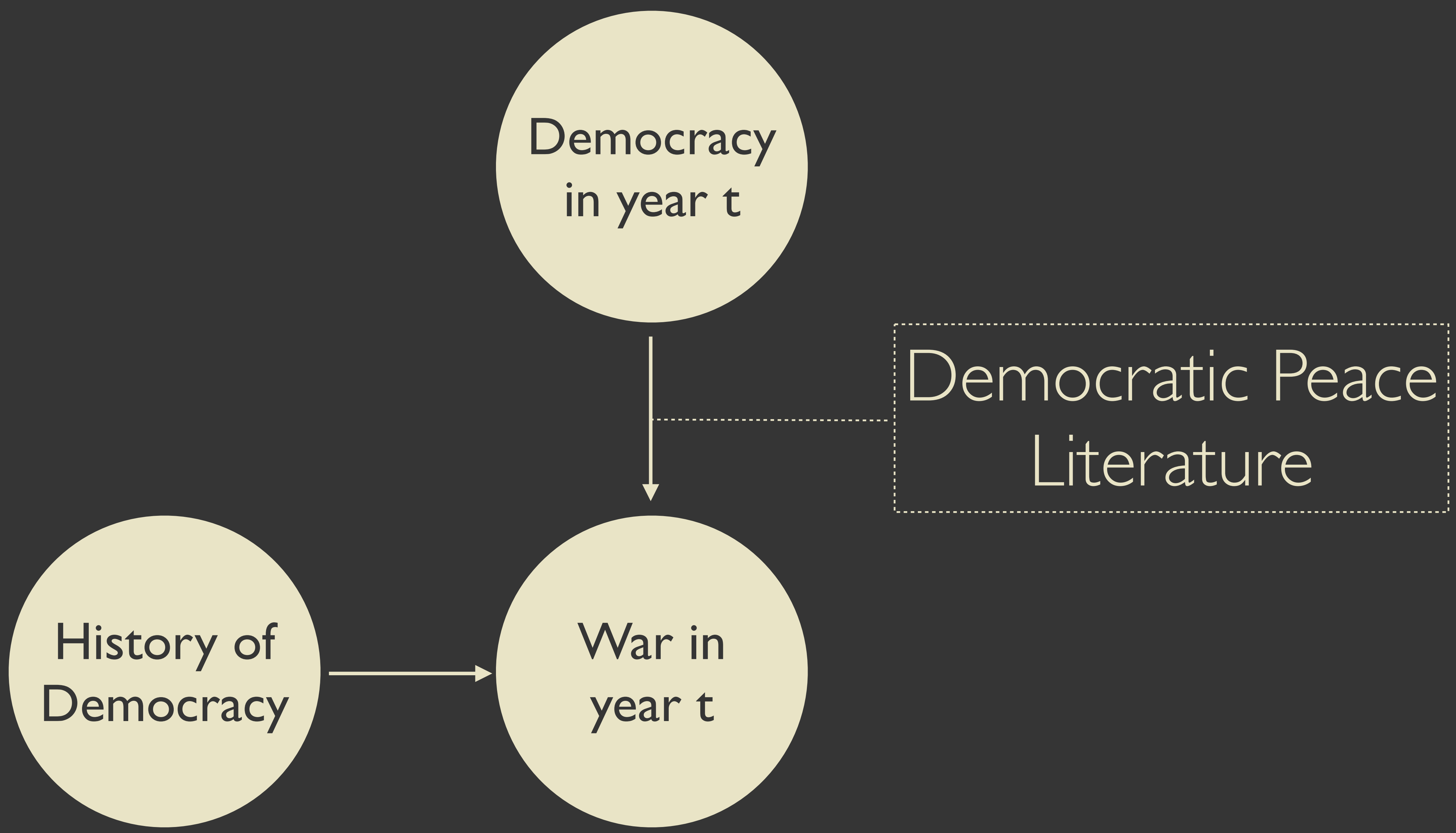
The Long Arm of the Democratic Peace?



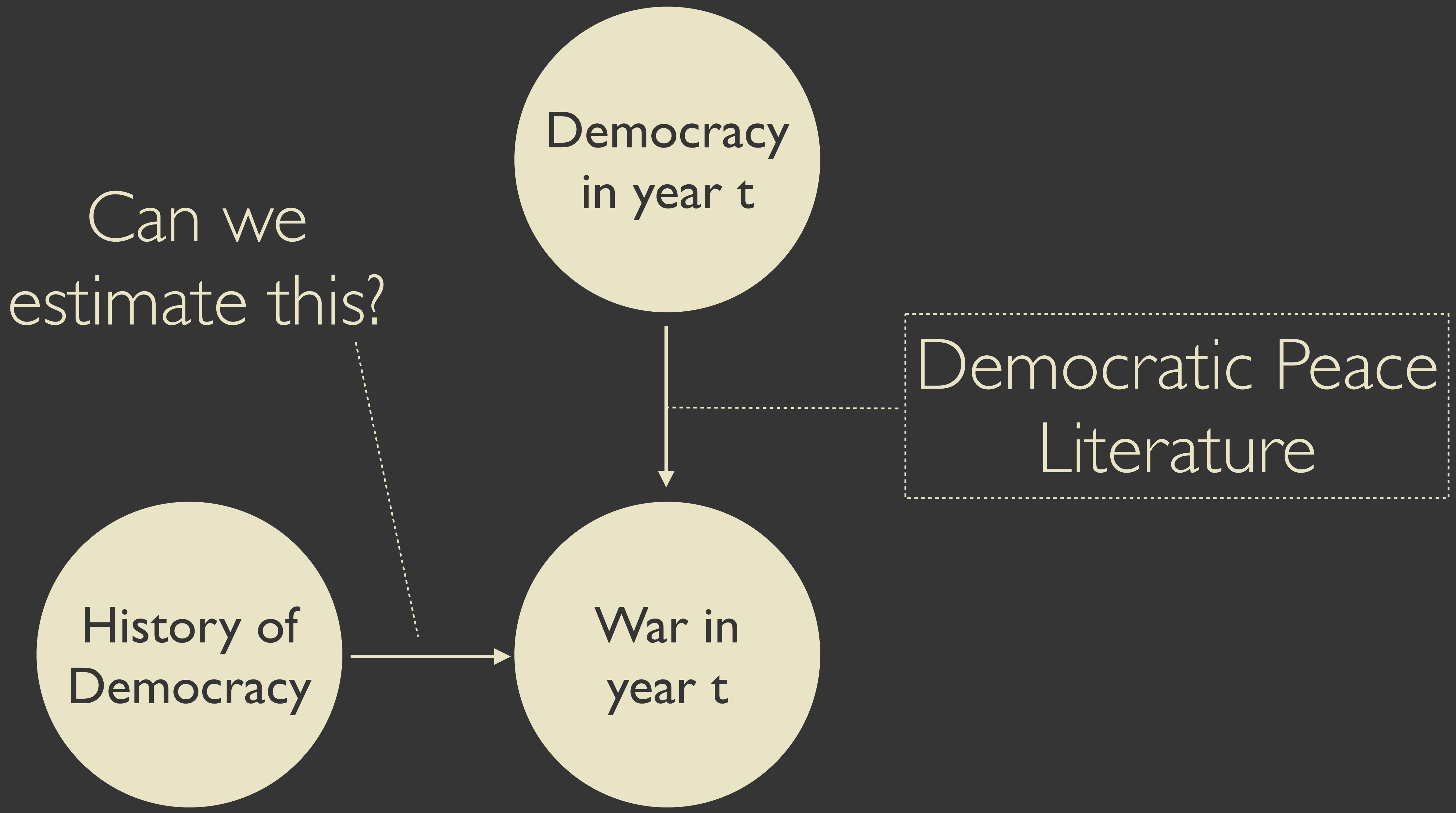
The Long Arm of the Democratic Peace?



The Long Arm of the Democratic Peace?



The Long Arm of the Democratic Peace?



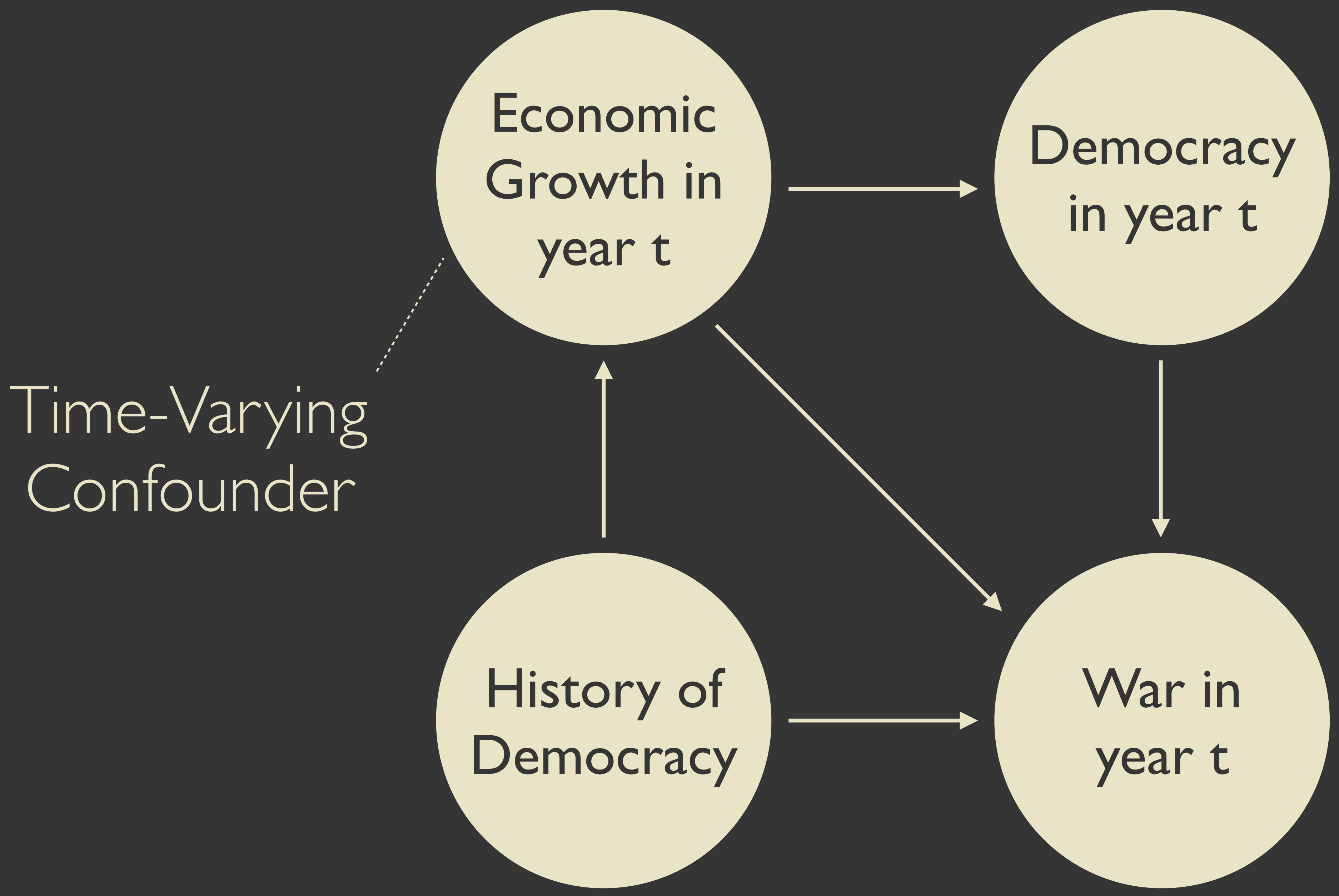
Revisiting Beck, Katz, and Tucker (1998)

	Dependent variable: Dispute
	BKT Model
	(1)
Democracy Blip	-0.651^{***}
	(0.160)
Cumulative Democracy	
Growth	-3.837^{***}
Observations	20,448
Note:	*p<0.1; **p<0.05; ***p<0.01

Revisiting Beck, Katz, and Tucker (1998)

	Dependent variable: Dispute	
	BKT Model	Misspecified Cumulative Model
	(1)	(2)
Democracy Blip	-0.651^{***}	
	(0.160)	
Cumulative Democracy		-0.010
		(0.012)
Growth	-3.837^{***}	-4.360^{***}
Observations	20,448	20,448
Note:		*p<0.1; **p<0.05; ***p<0.01

Misspecification of an ATHE



Revisiting Beck, Katz, and Tucker (1998)

	Dependent variable: Dispute		
	BKT Model	Misspecified Cumulative Model	IPTW + MSM
	(1)	(2)	(3)
Democracy Blip	-0.651^{***}		
	(0.160)		
Cumulative Democracy		-0.010	-0.045^{***}
		(0.012)	(0.013)
Growth	-3.837^{***}	-4.360^{***}	
Observations	20,448	20,448	20,448
Note:		*p<0.1; **p<0.05; ***p<0.01	

TSCS data under unmeasured confounding

TSCS data under unmeasured confounding

$$Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{it}, \underline{A}_{i,t-1} = \underline{a}_{t-1}, U$$

TSCS data under unmeasured confounding

$$Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{it}, \underline{A}_{i,t-1} = \underline{a}_{t-1}, U$$

Treatment is unrelated
to the potential outcomes

TSCS data under unmeasured confounding

$$Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{it}, \underline{A}_{i,t-1} = \underline{a}_{t-1}, U$$

Treatment is unrelated
to the potential outcomes

...conditional on the
covariate history

TSCS data under unmeasured confounding

...and a time-fixed
unmeasured confounder.

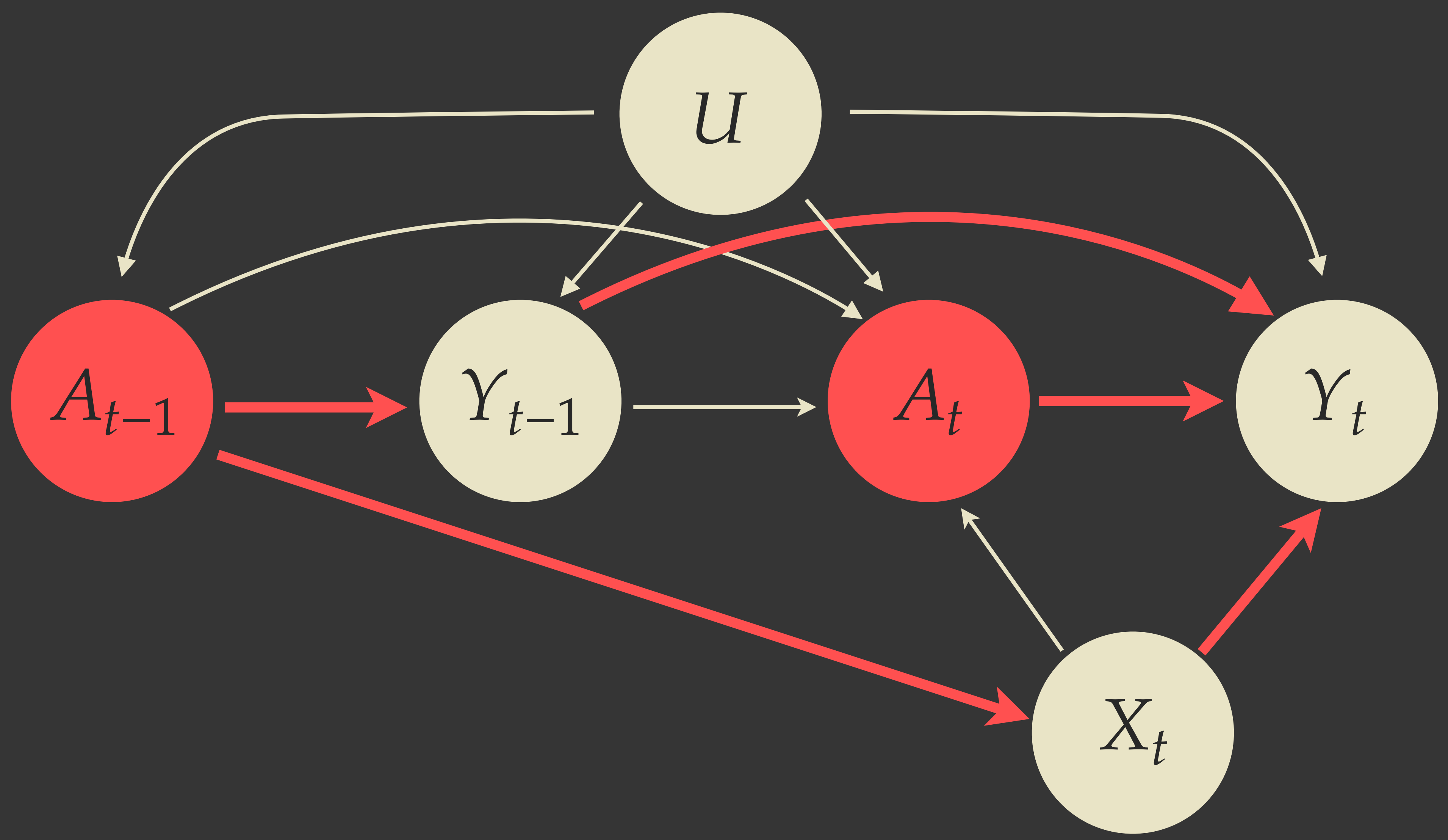
$$Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{it}, \underline{A}_{i,t-1} = \underline{a}_{t-1}, U$$

Treatment is unrelated
to the potential outcomes

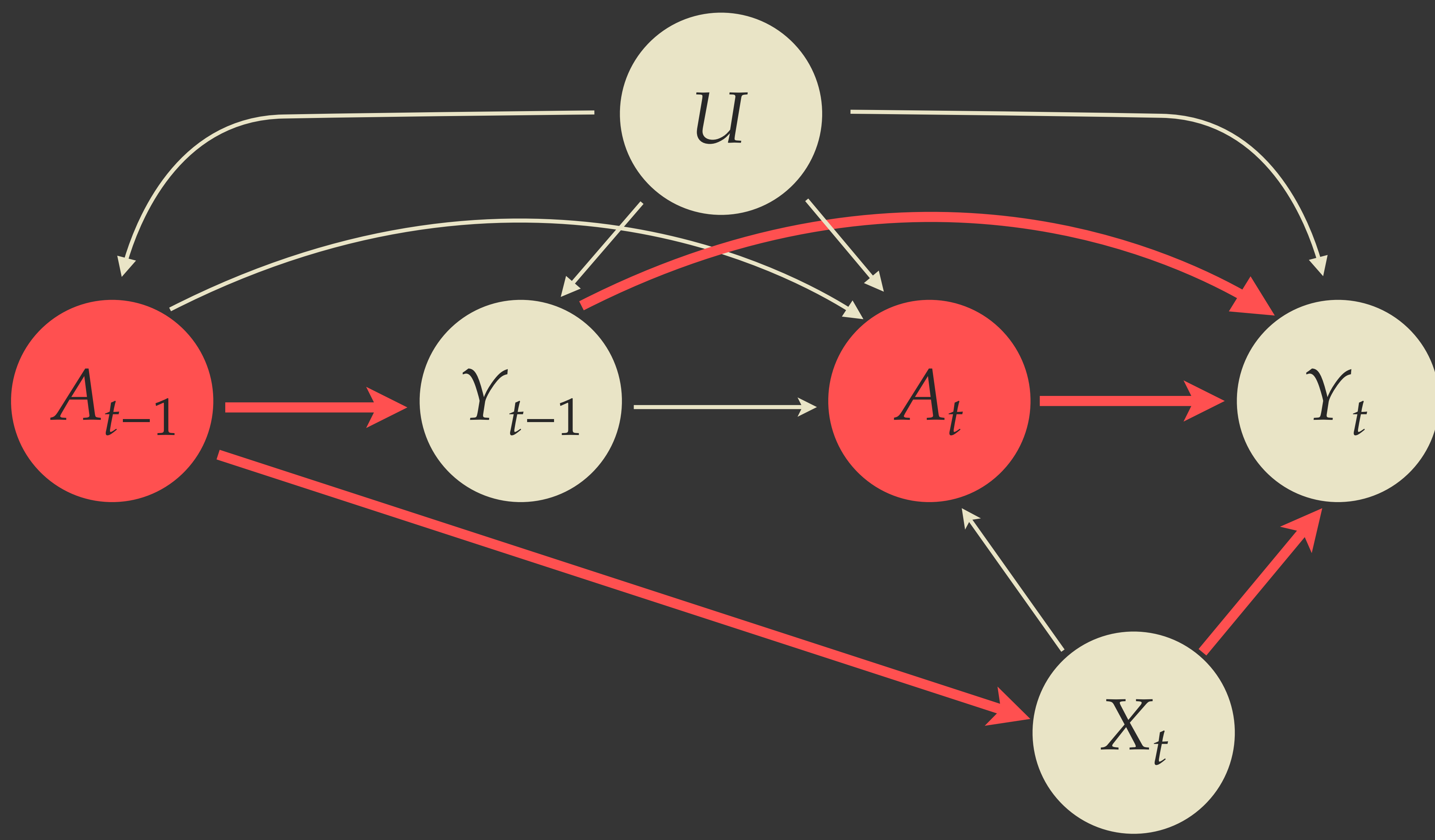
...conditional on the
covariate history

How unit-specific weighting can help

How unit-specific weighting can help

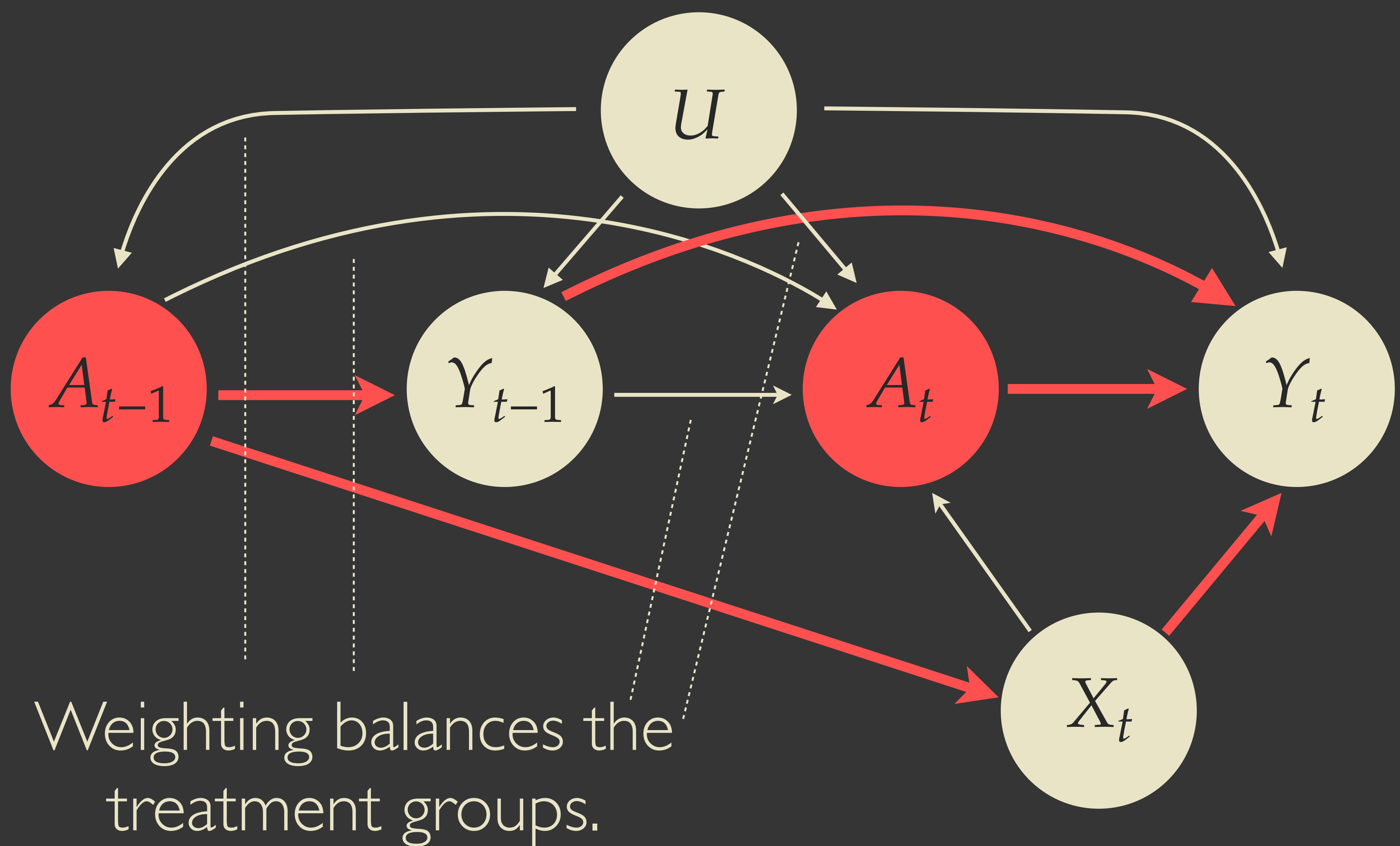


How unit-specific weighting can help



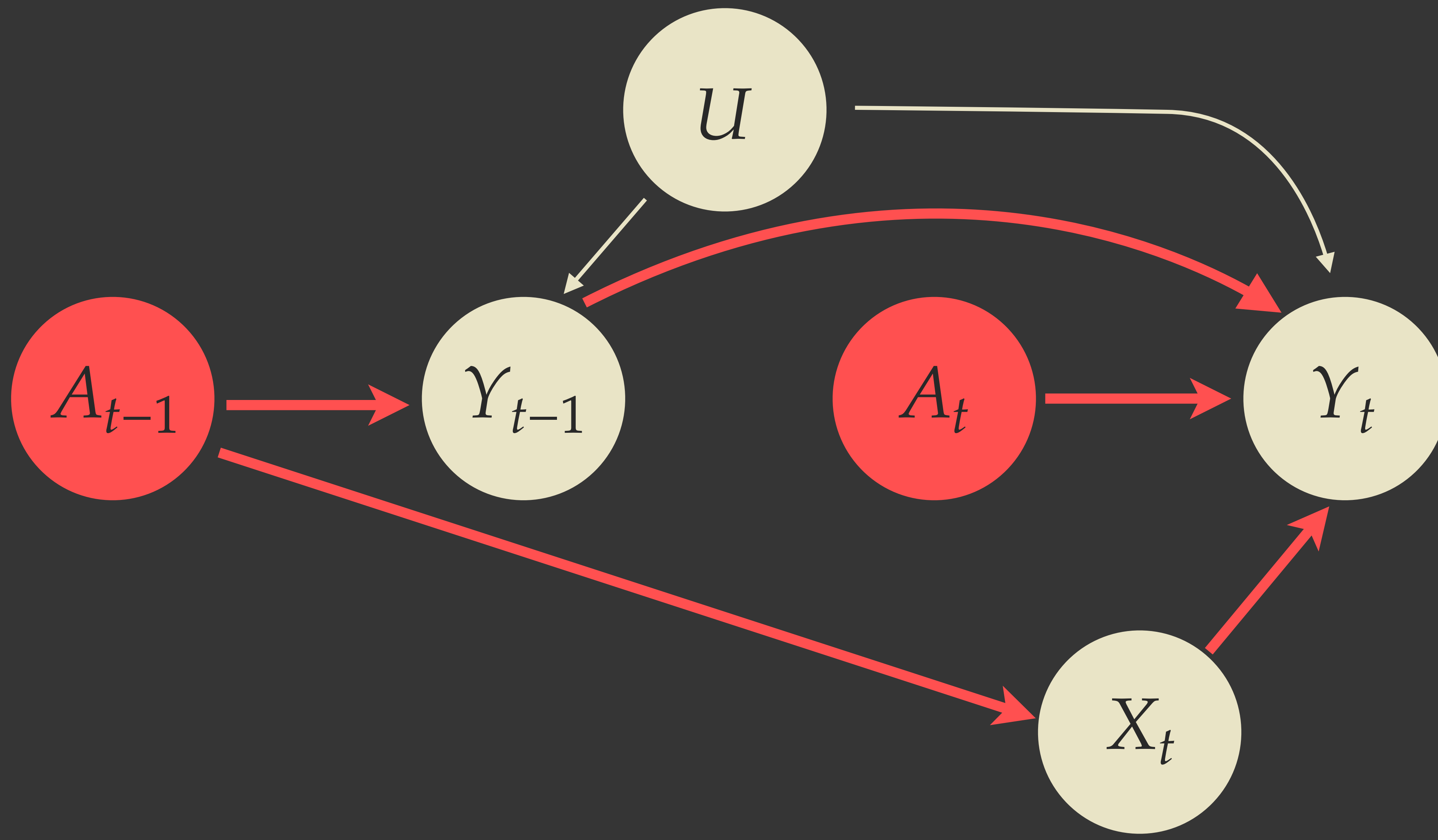
$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is} | A_{t-1}, X_t, Y_{t-1}, U]}$$

How unit-specific weighting can help



$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is} | A_{t-1}, X_t, Y_{t-1}, U]}$$

How unit-specific weighting can help



$$W_{it} = \prod_{s=1}^{t} \frac{1}{\Pr[A_{is} | A_{t-1}, X_t, Y_{t-1}, U]}$$

A weighting approach to fixed effects

A weighting approach to fixed effects

1. Estimate unit-specific probability of treatment over time and construct weights.

A weighting approach to fixed effects

1. Estimate unit-specific probability of treatment over time and construct weights.
2. Estimate a pooled outcome model with unit-specific weights

k-order sequential ignorability

k-order sequential ignorability

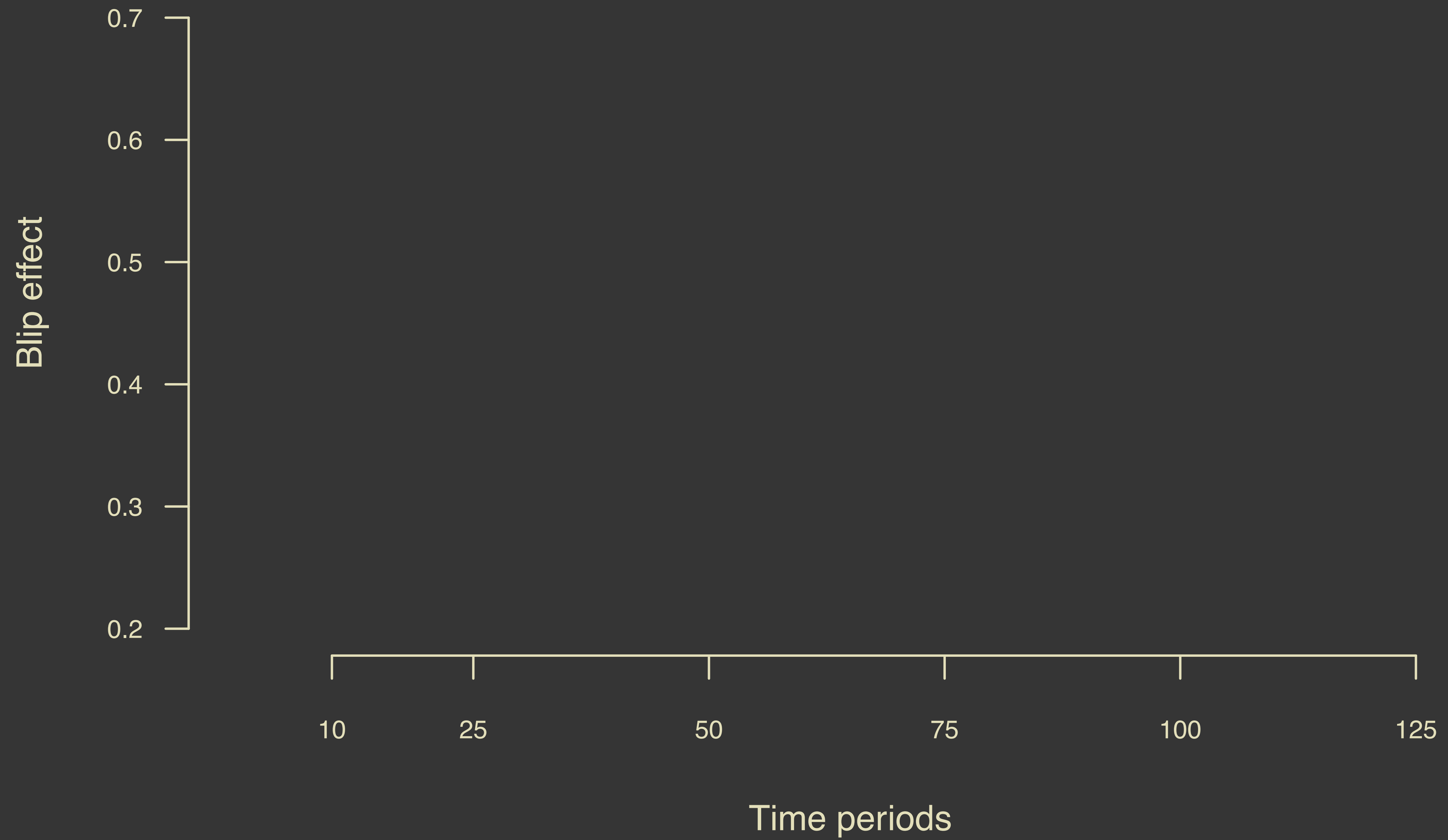
$$Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{i,t:t-k}, \underline{A}_{i,t-1:t-k} = \underline{a}_{t-1:t-k}, U$$

k-order sequential ignorability

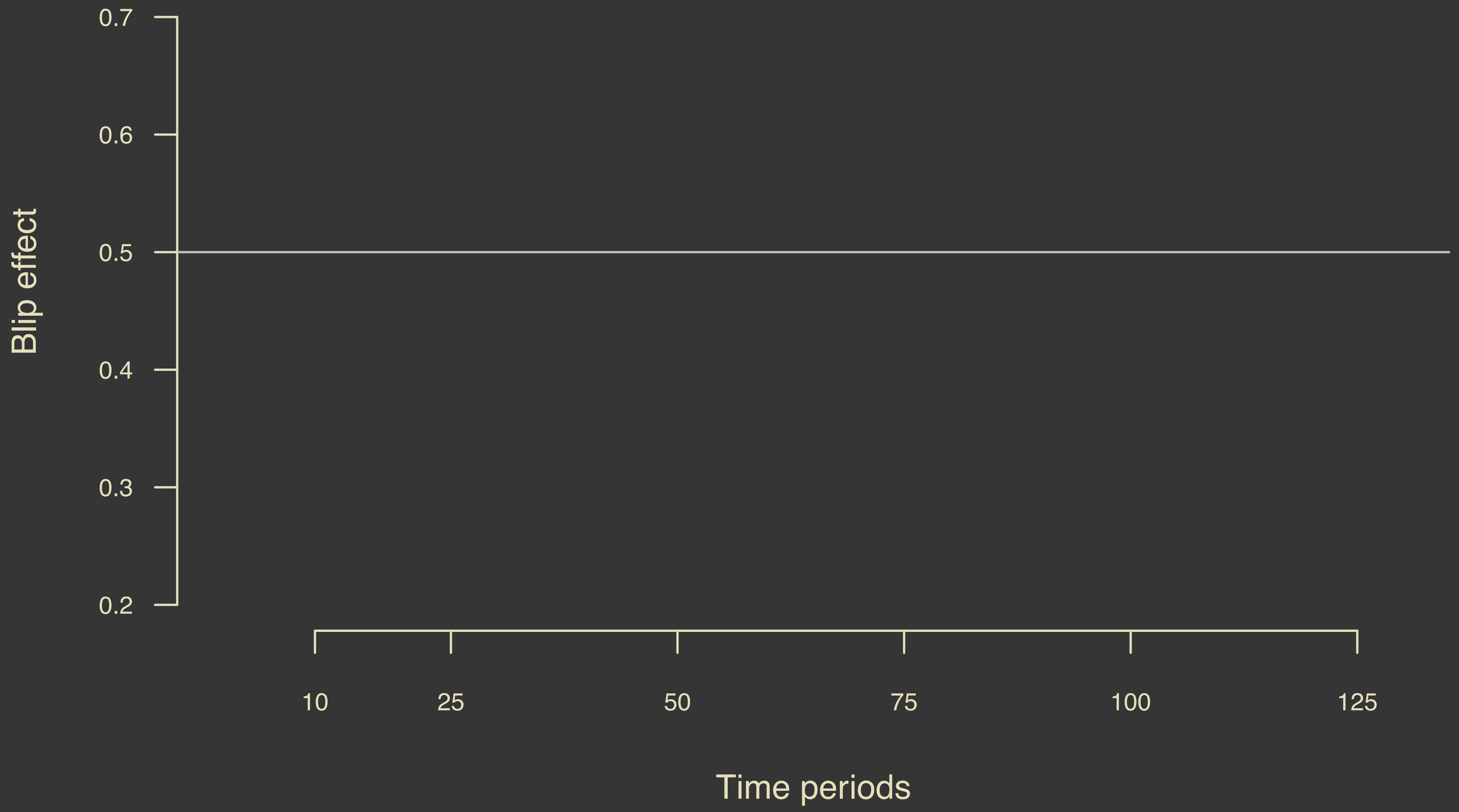
$$Y_{it}(\underline{a}_t) \perp\!\!\!\perp A_{it} | \underline{X}_{i,t:t-k}, \underline{A}_{i,t-1:t-k} = \underline{a}_{t-1:t-k}, U$$

Only the last k periods matter.

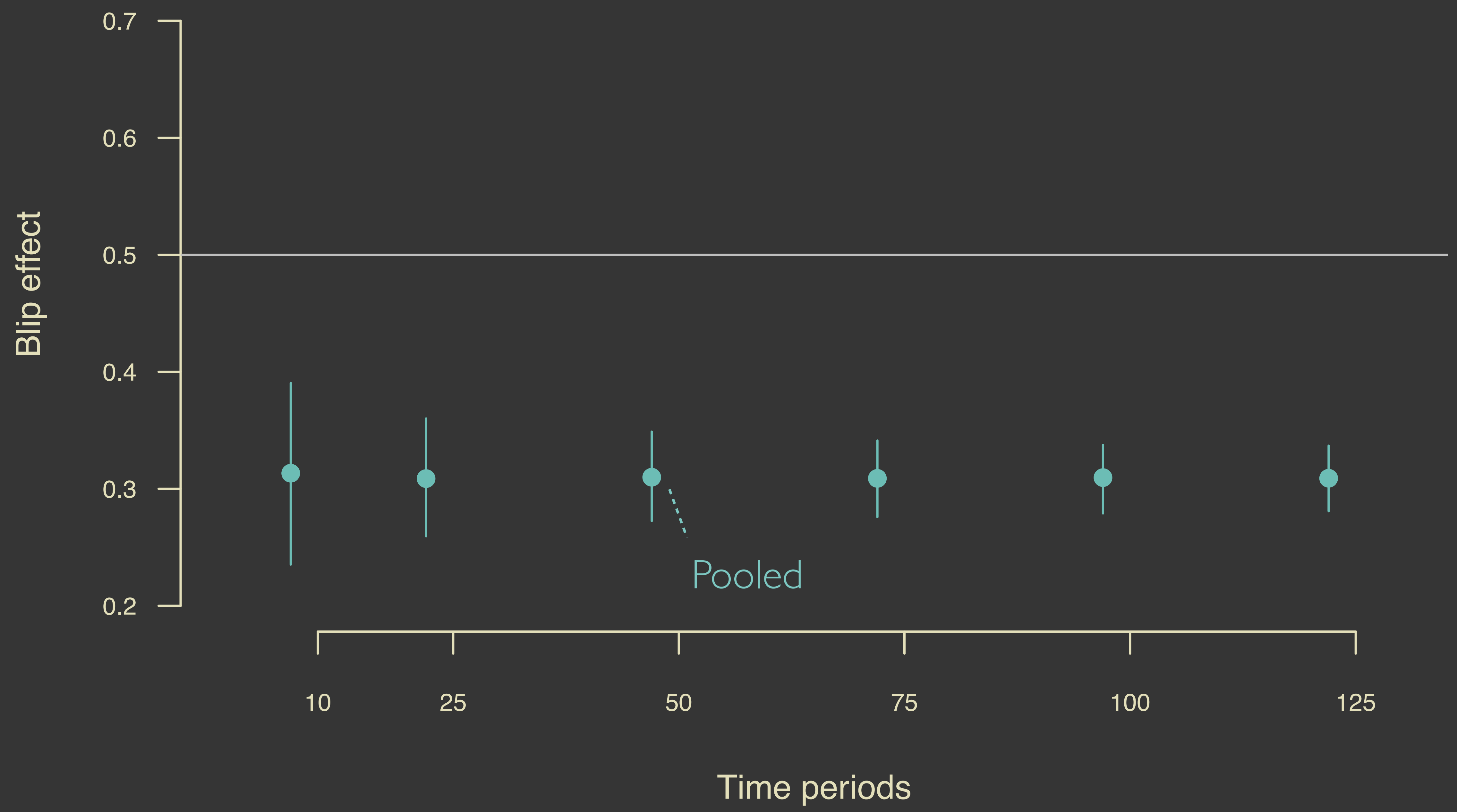
Blip effect: (1,0) vs (0,0)



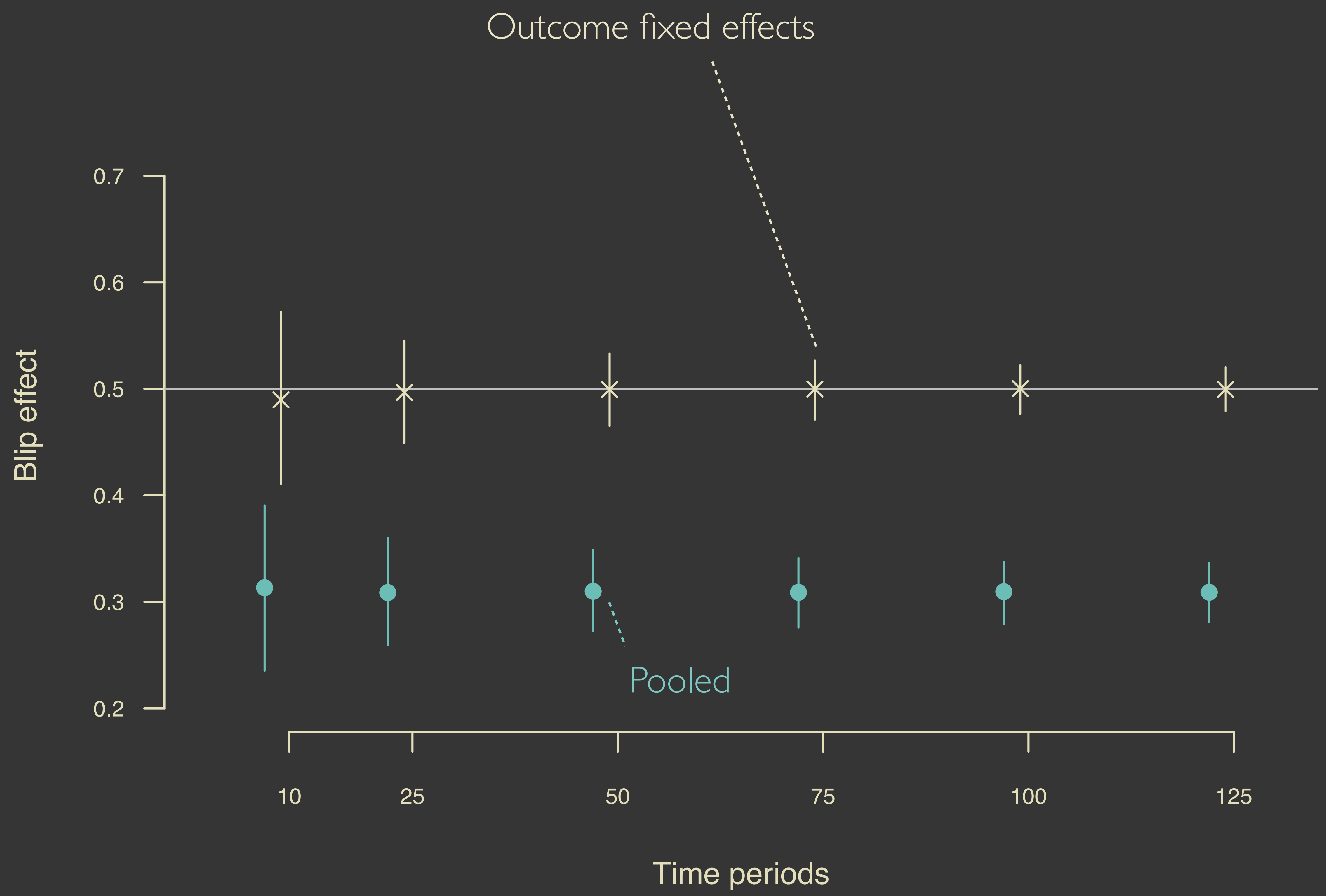
Blip effect: (1,0) vs (0,0)



Blip effect: (1,0) vs (0,0)

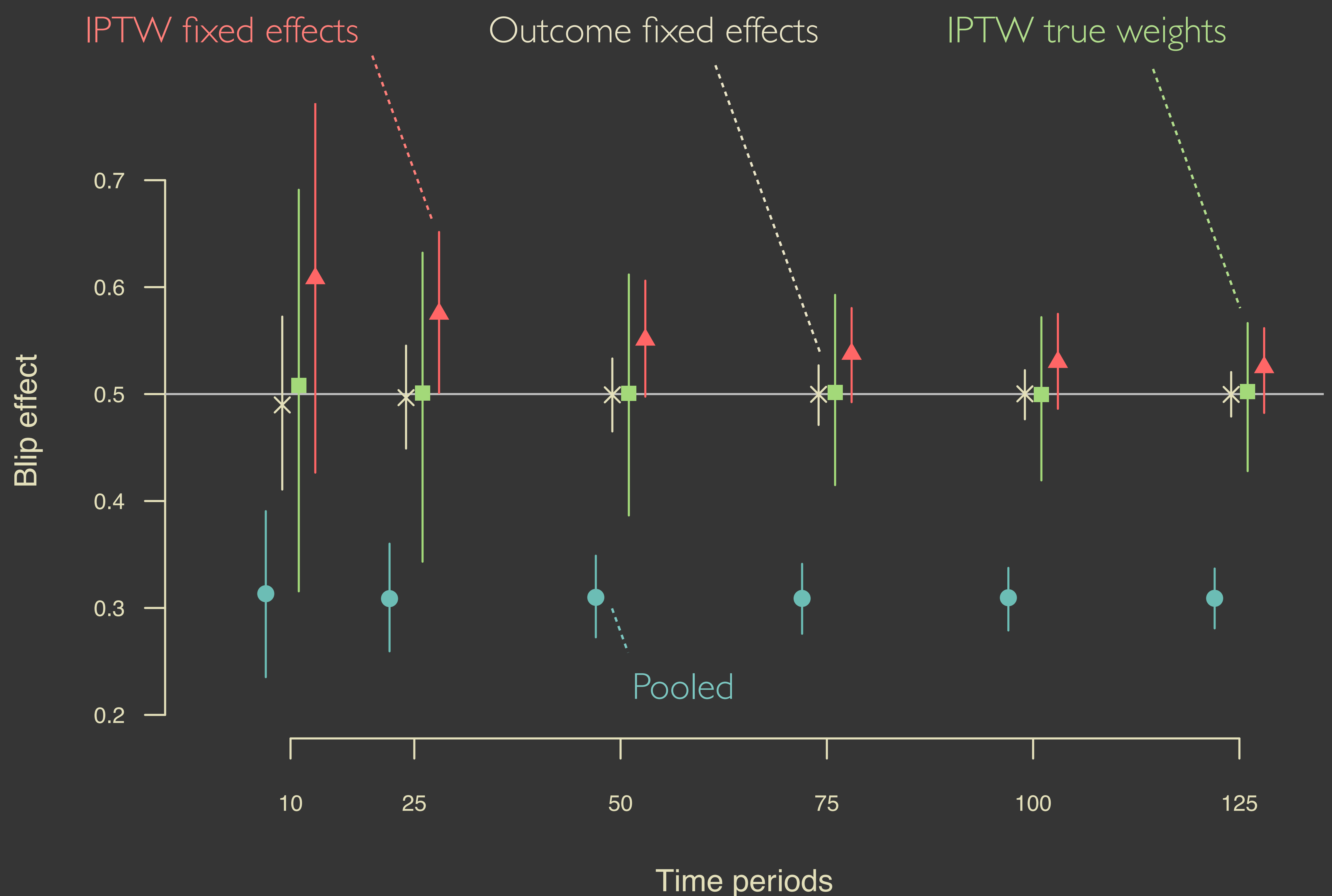


Blip effect: (1,0) vs (0,0)

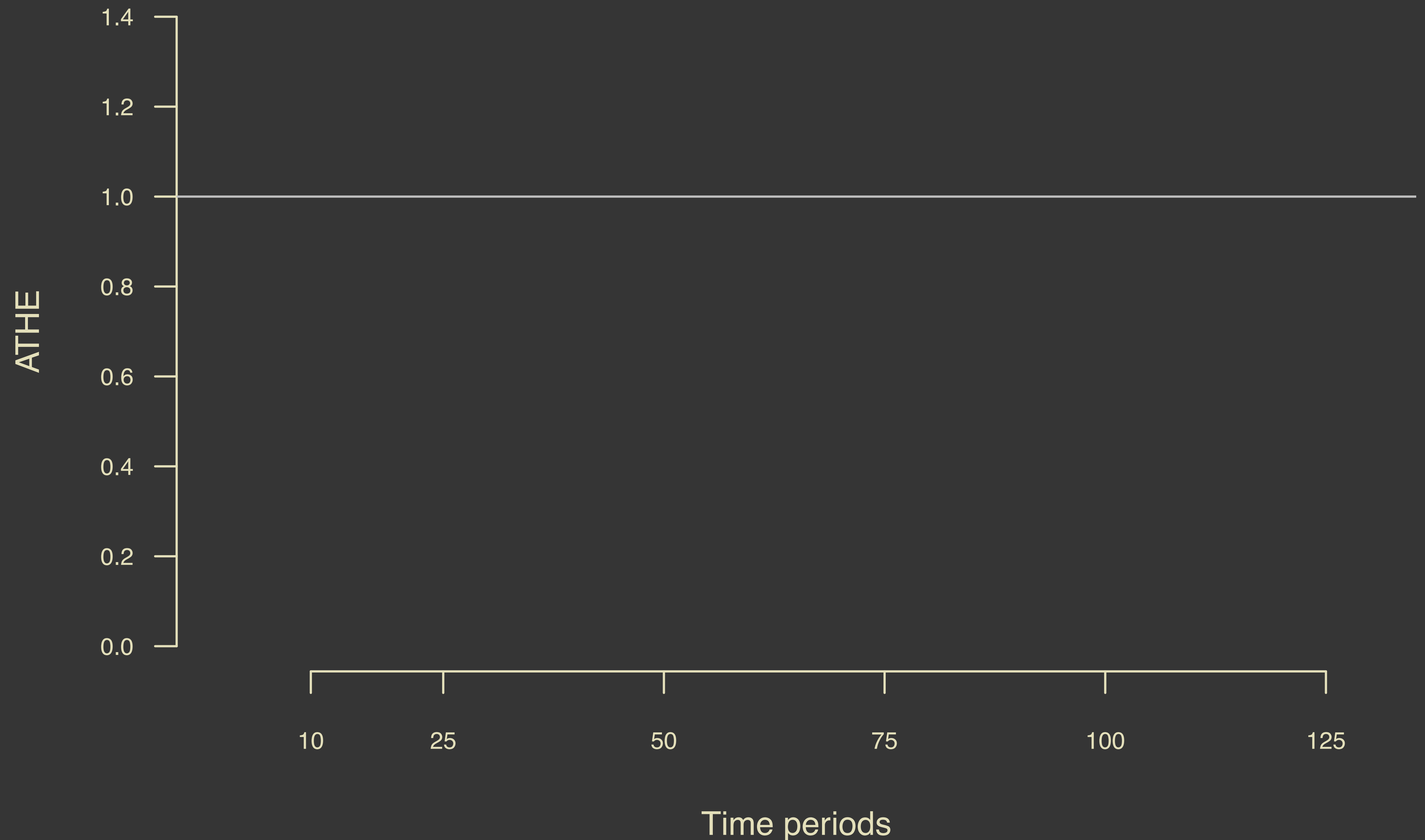


Blip effect: (1,0) vs (0,0)

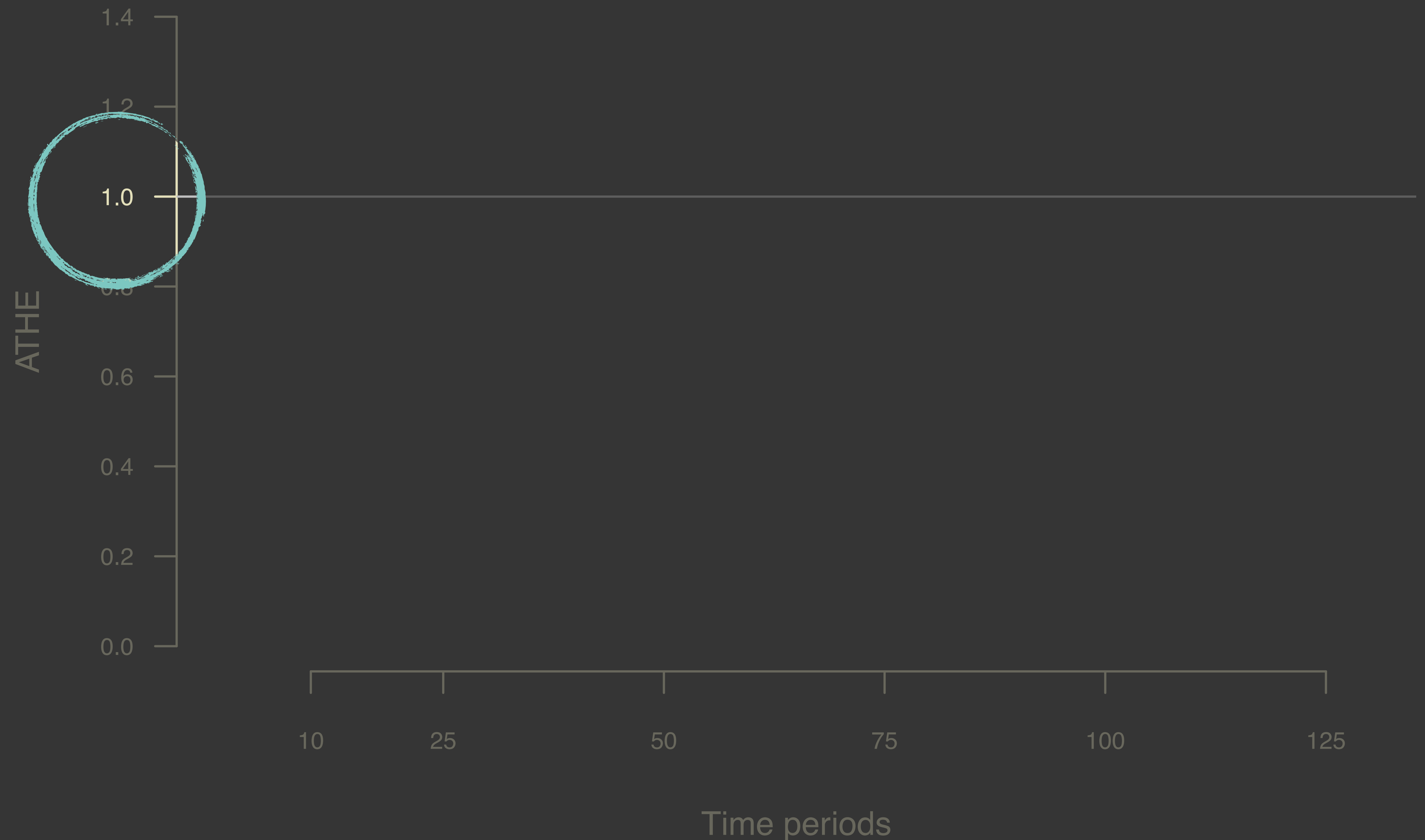
Blip effect: (1,0) vs (0,0)



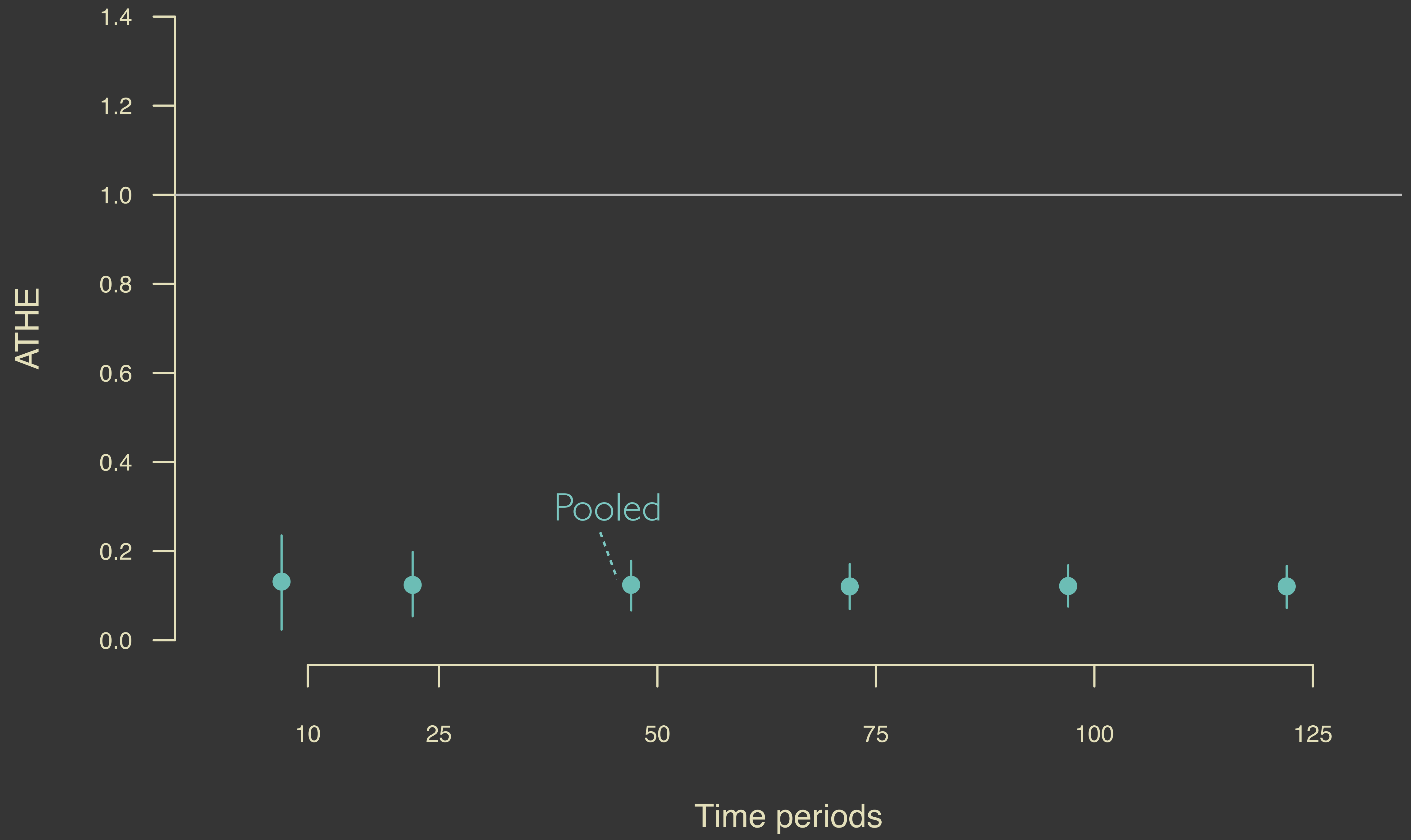
Treatment History Effect: (1,1) vs (0,0)



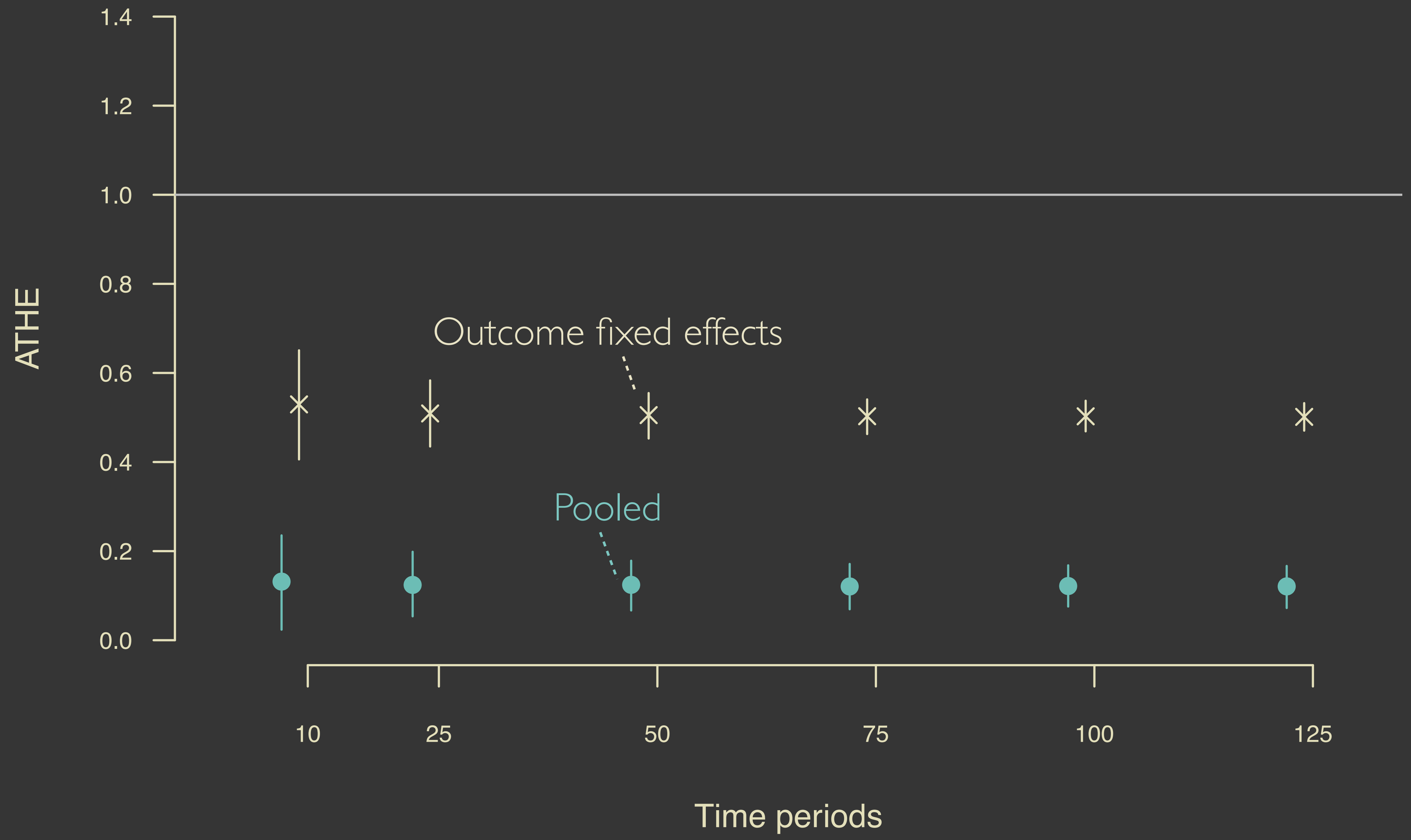
Treatment History Effect: (1,1) vs (0,0)



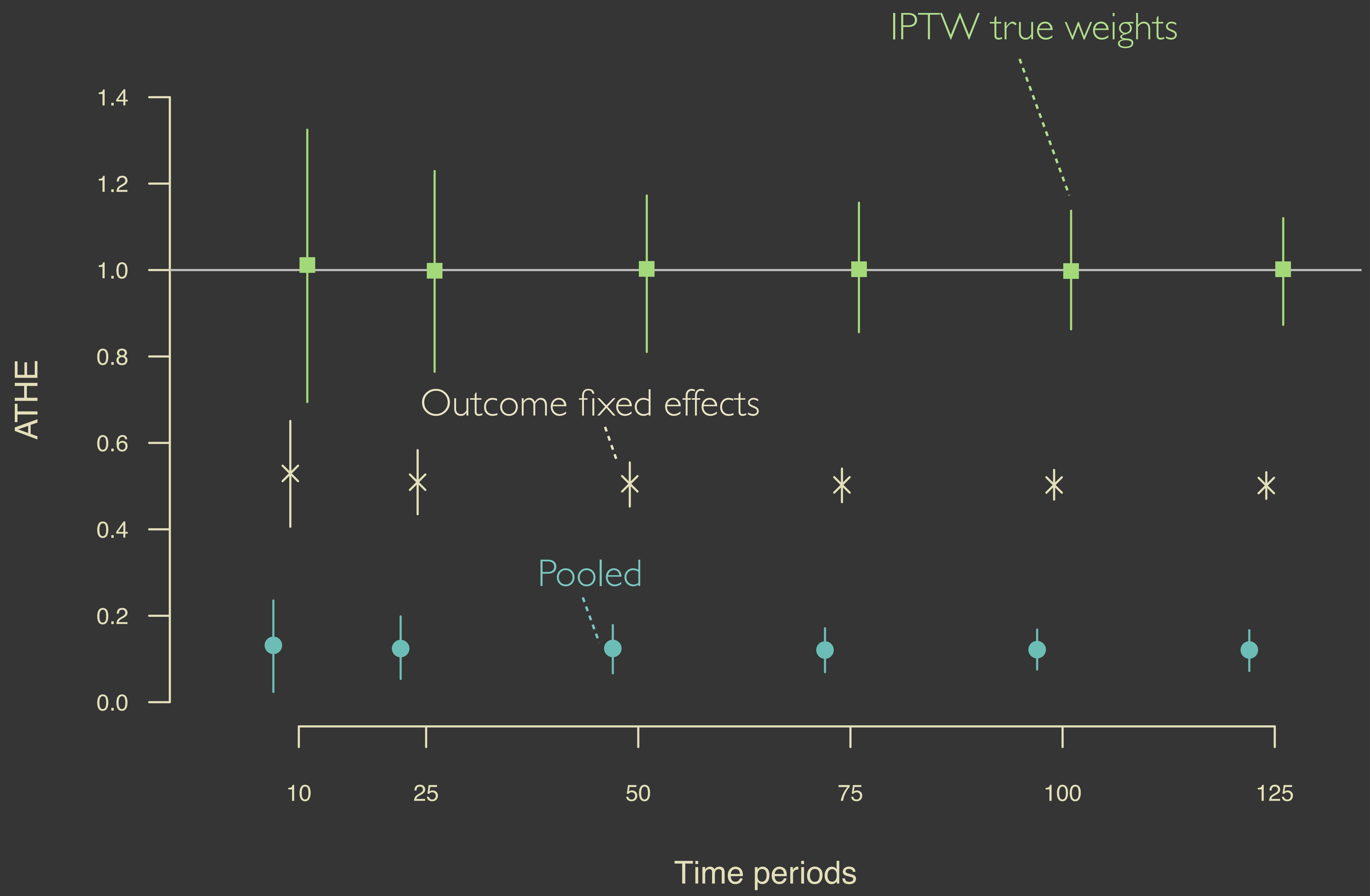
Treatment History Effect: (1,1) vs (0,0)



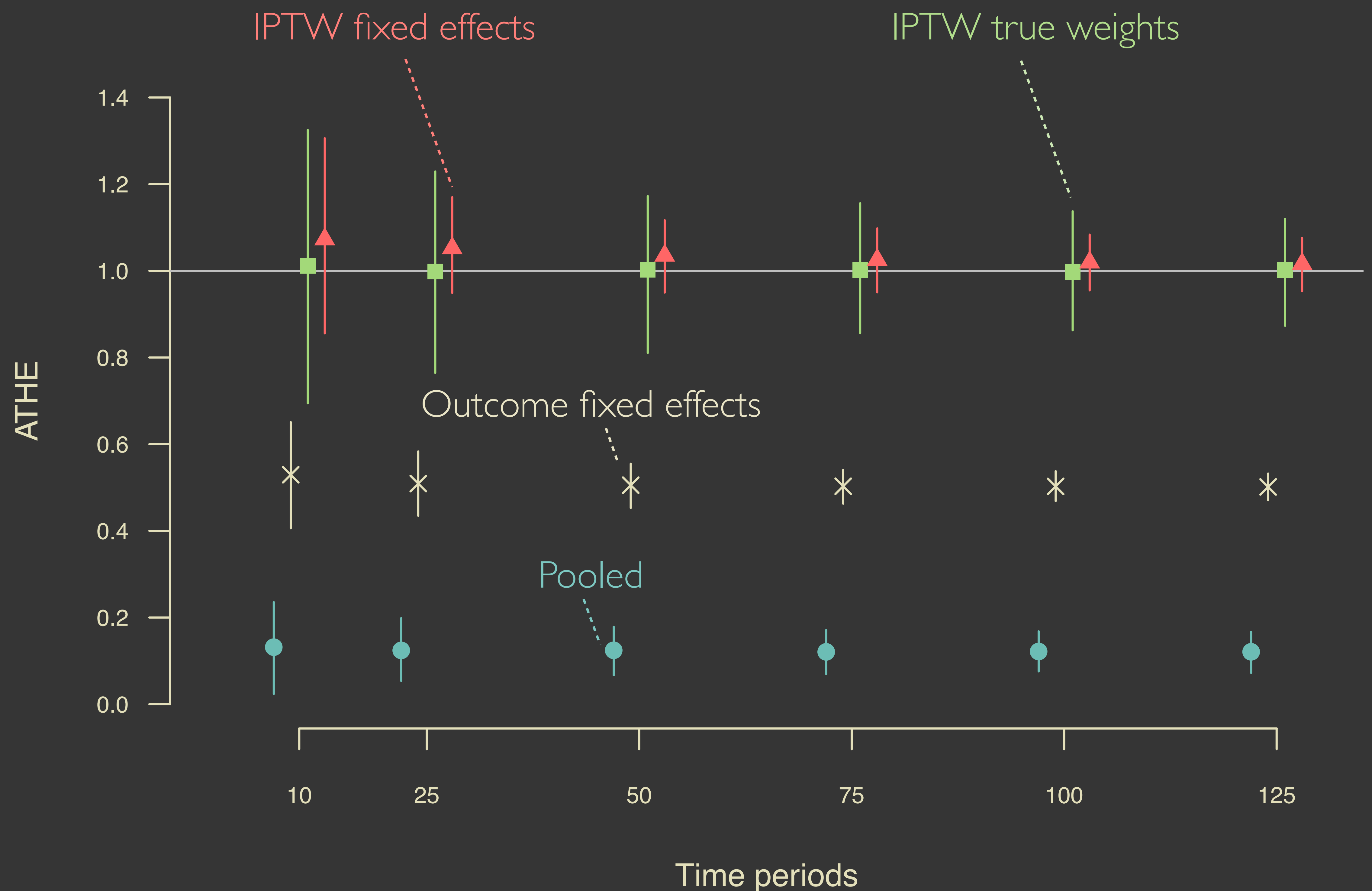
Treatment History Effect: (1,1) vs (0,0)



Treatment History Effect: (1,1) vs (0,0)



Treatment History Effect: (1,1) vs (0,0)



How to make causal inferences with TSCS data

How to make causal inferences with TSCS data

Very carefully

How to make causal inferences with TSCS data

Very carefully

Even under strong assumptions, conditional estimators cannot recover ATHEs.

How to make causal inferences with TSCS data

Very carefully

Even under strong assumptions, conditional estimators cannot recover ATHEs.

Using weights

How to make causal inferences with TSCS data

Very carefully

Even under strong assumptions, conditional estimators cannot recover ATHEs.

Using weights

A fixed effects weighting approach can recover ATHEs and CETs even with unmeasured confounding.