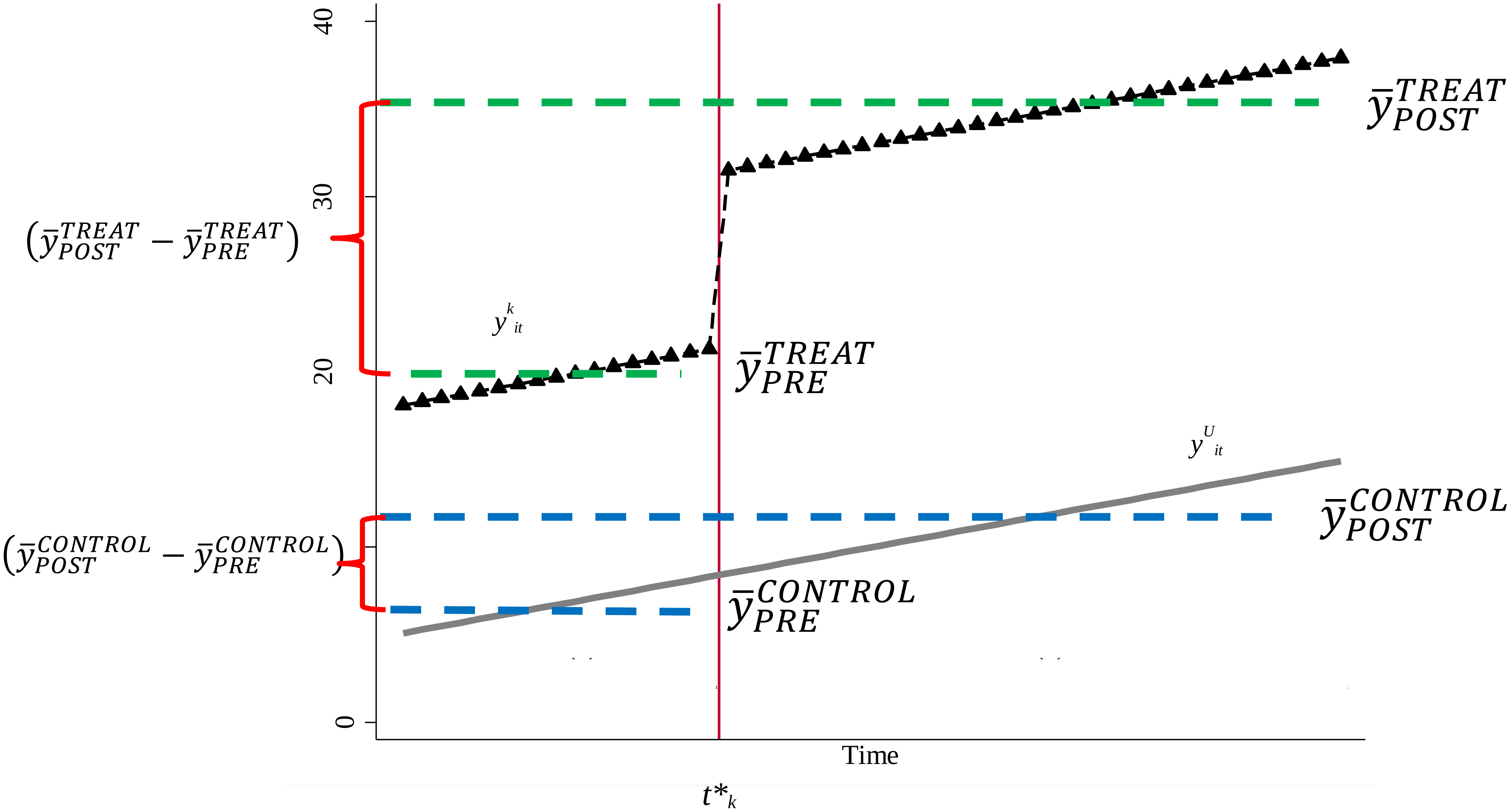
Difference-in-Differences With Variation in Treatment Timing

EMPE – Mres TSE

What is Difference-in-Differences?

What is Difference-in-Differences?



What is Difference-in-Differences?

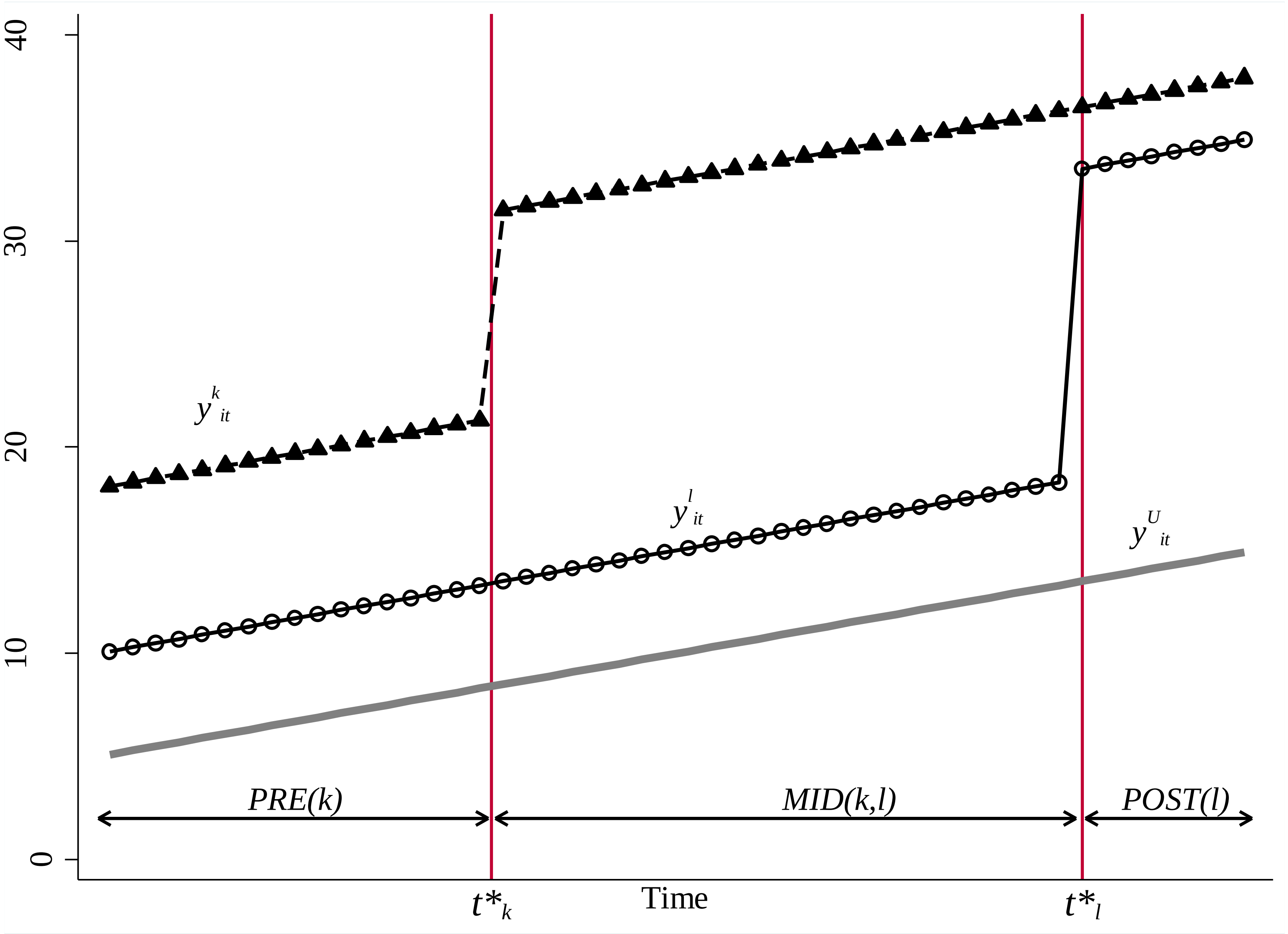
$$\hat{\beta}^{DD} = \left(\bar{y}_{POST}^{TREAT} - \bar{y}_{PRE}^{TREAT}\right) \text{-} \left(\bar{y}_{POST}^{CONTROL} - \bar{y}_{PRE}^{CONTROL}\right)$$

$$y_{it} = \gamma_0 TREAT_i + \gamma_1 POST_t + \hat{\beta}^{DD} TREAT_i POST_t + u_{it}$$

Variation in Timing

Treatment turns on at different times t^*

Variation in Timing



Variation in Timing

D_{it} turns on at different times t^*

a. Federalism
b. Case-by-case judicial enforcement
c. Sub-federal funding process
d. Natural disasters

2014/2015 AER/QJE/JPE/ReStud/JHE/JDE
published 93 DD papers:

49% had timing variation

Two-Way Fixed Effects Estimator

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

Unit fixed effects

Time fixed effects

Treatment dummy

What is $\widehat{\beta}^{DD}$?

What is $\hat{\beta}^{DD}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

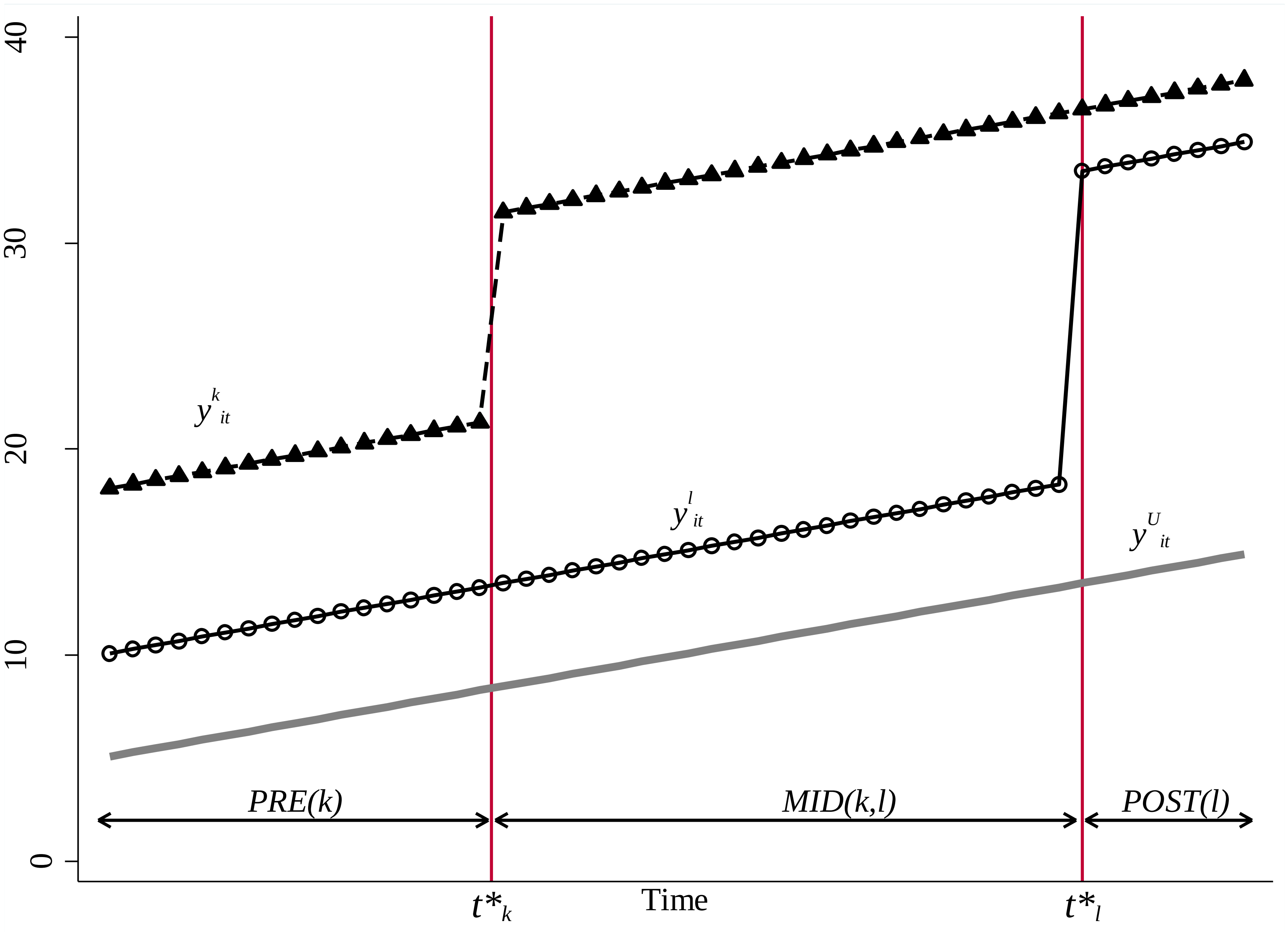
1. Partial out fixed effects (Frisch-Waugh):

$$\tilde{D}_{it} = \left(D_{it} - \bar{\bar{D}}\right) - \left(\bar{D}_i - \bar{\bar{D}}\right) - \left(\bar{D}_t - \bar{\bar{D}}\right)$$
$$\tilde{y}_{it} = (y_{it} - \bar{\bar{y}}) - (\bar{y}_i - \bar{y}) - (\bar{y}_t - \bar{\bar{y}})$$

2. Calculate univariate coefficient by brute force:

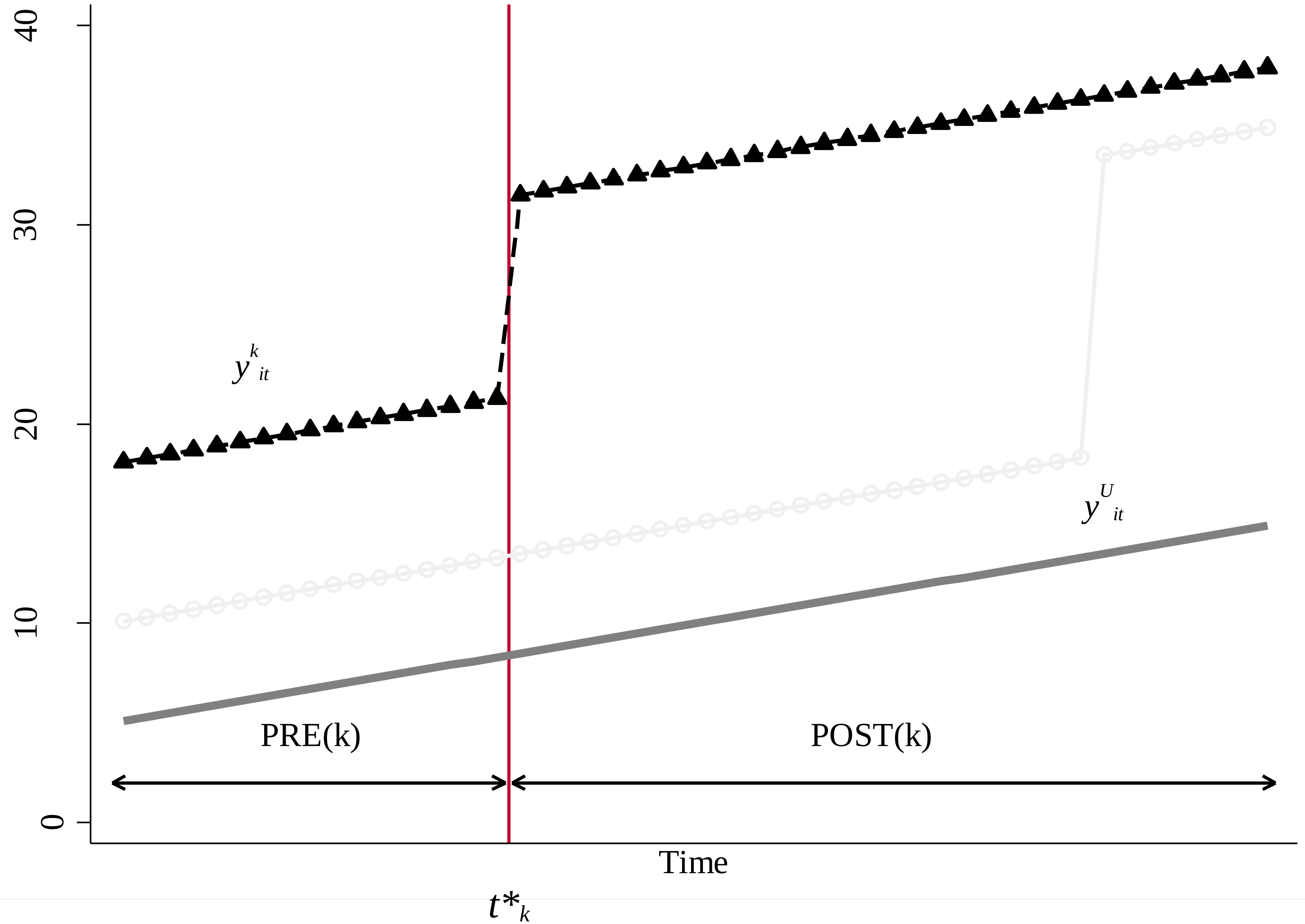
$$\hat{\beta}^{DD} = \frac{\widehat{cov}(\tilde{D}_{it}, \tilde{y}_{it})}{\hat{V}(\tilde{D}_{it})} = \frac{\frac{1}{NT}\sum_i \sum_t (y_{it} - \bar{\bar{y}})\left(D_{it} - \bar{\bar{D}}\right)}{\frac{1}{NT}\sum_i \sum_t \left(D_{it} - \bar{\bar{D}}\right)}$$

$\widehat{\beta}^{DD}$?



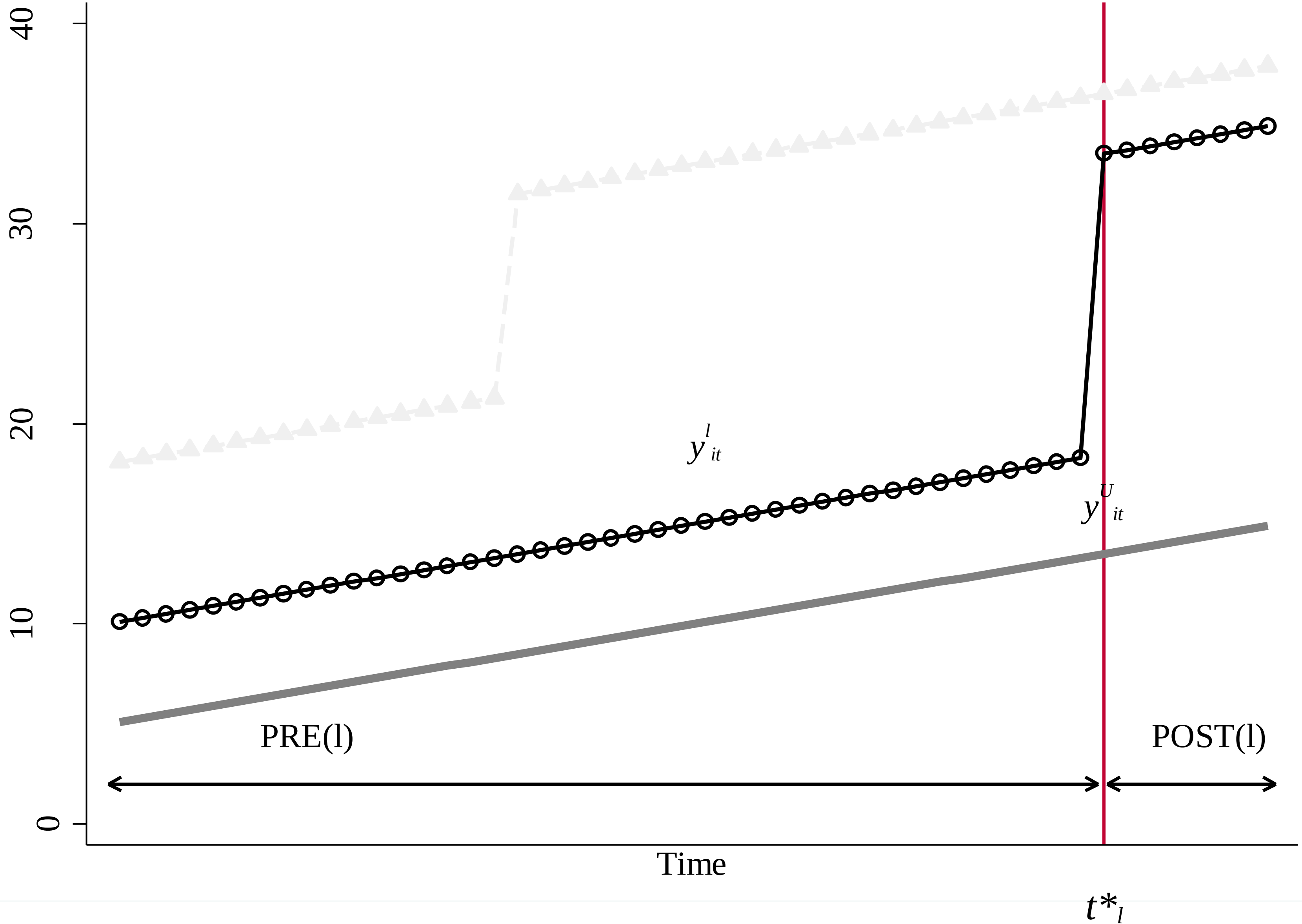
$$\widehat{\beta}^{DD}_{kU}$$

A. Early Group vs. Untreated Group

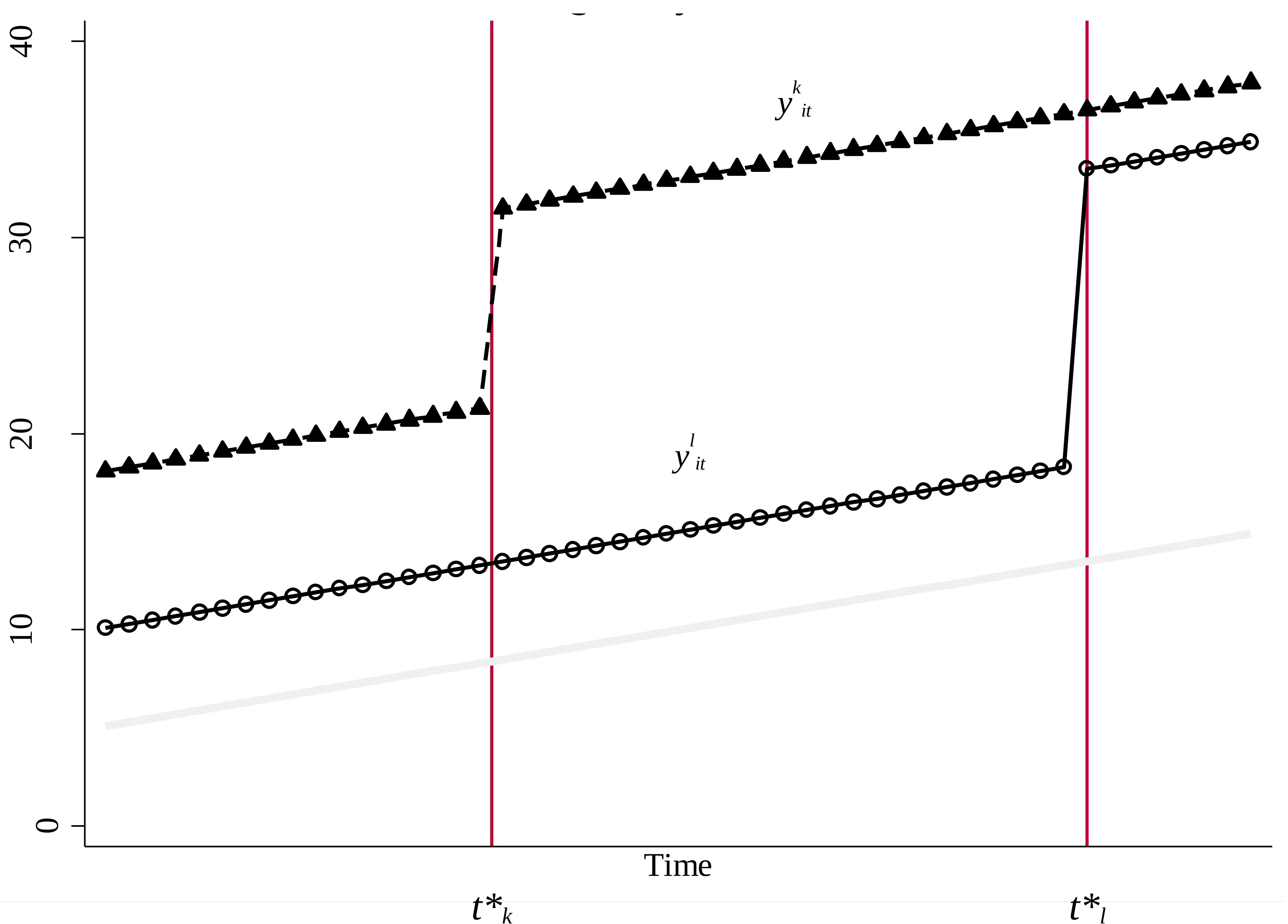


$$\widehat{\beta}^{DD}_{\ell U}$$

B. Late Group vs. Untreated Group

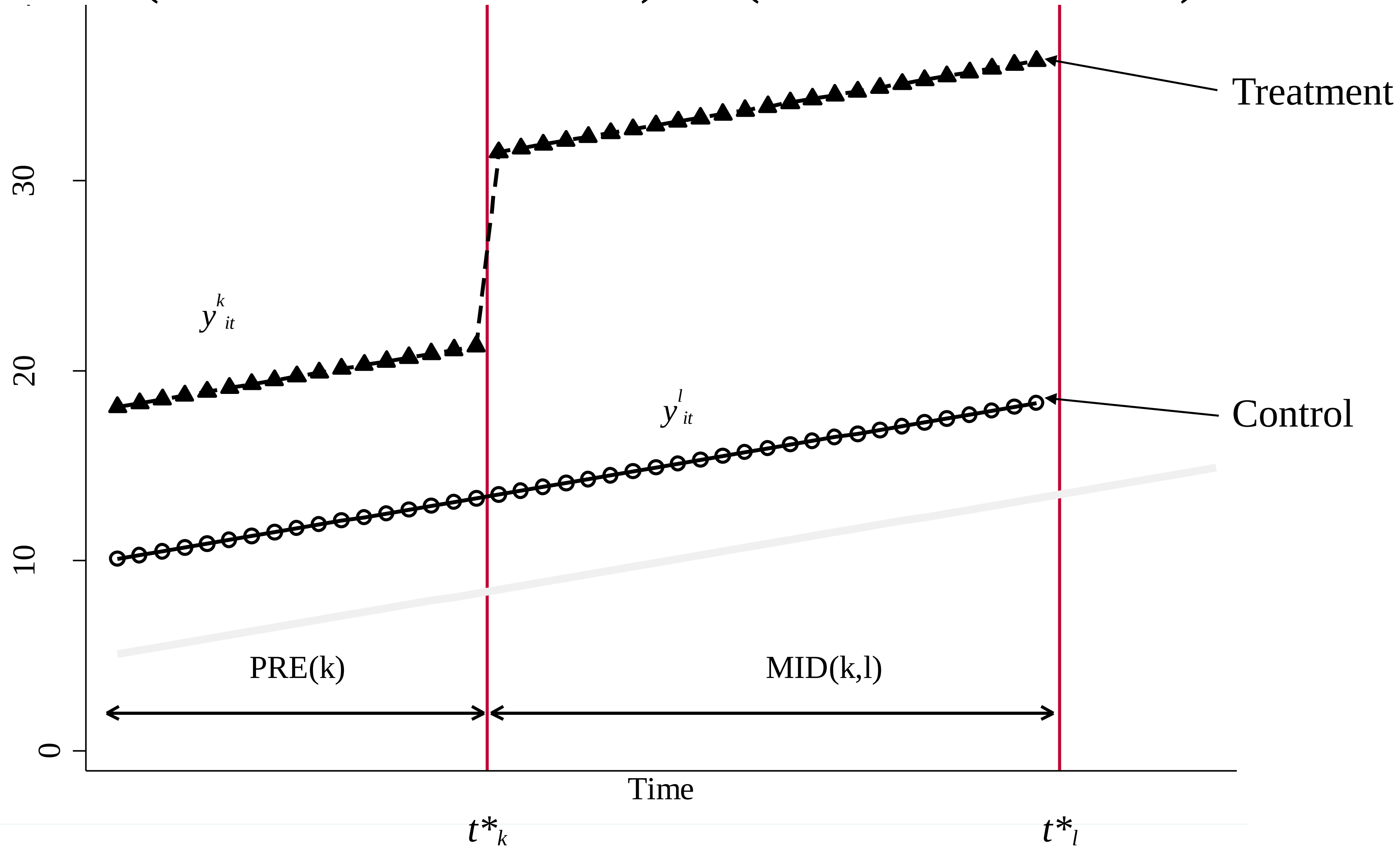


Two-Group Timing-Only Estimator ($\widehat{\beta}^{DD}_{k\ell}$)



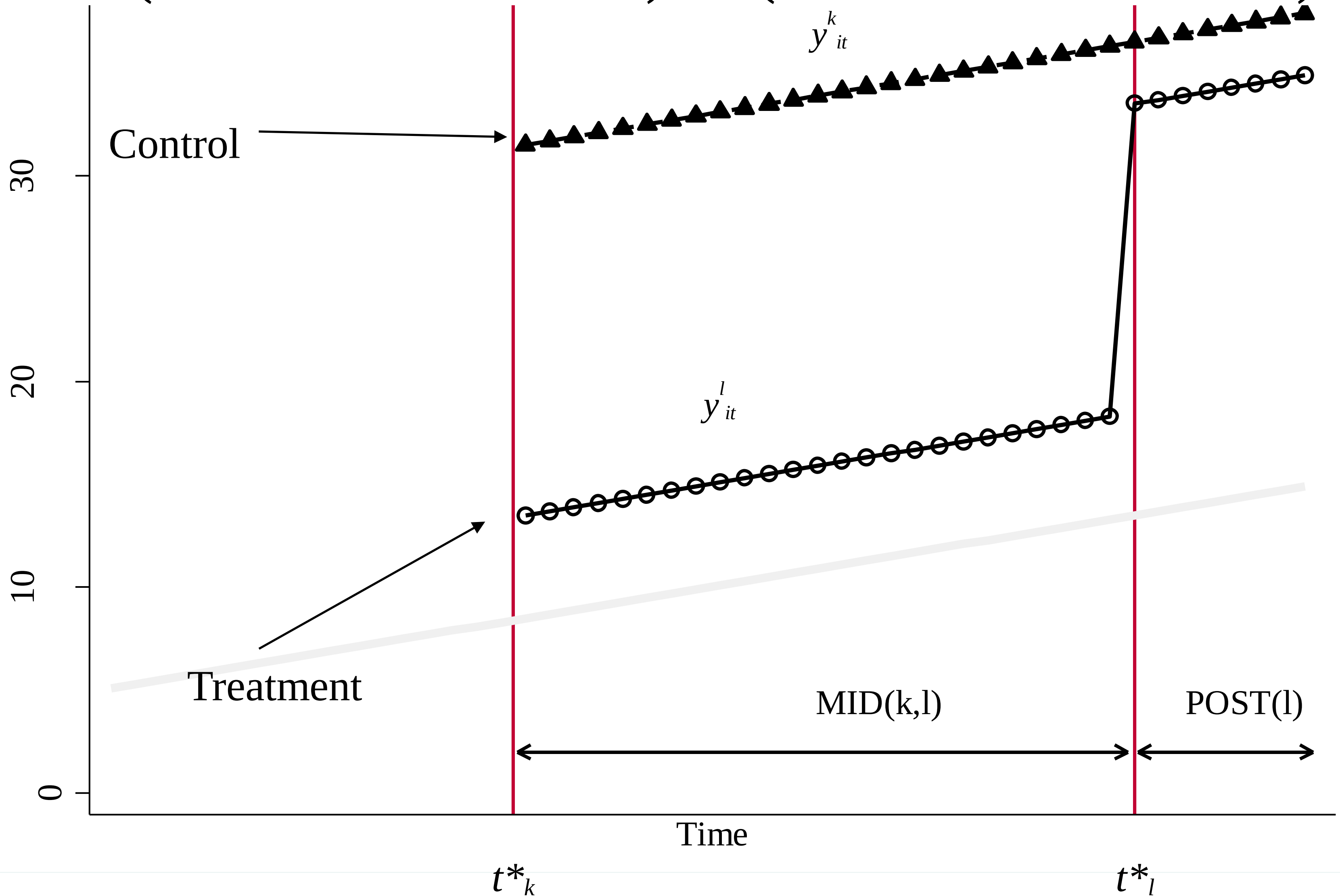
$$\widehat{\boldsymbol{\beta}}^{\boldsymbol{DD,k}}_{\boldsymbol{k\ell}}$$

$$\hat{\beta}^{DD,k}_{k\ell} = \left(\bar{y}_k^{MID(k,\ell)} - \bar{y}_\ell^{MID(k,\ell)}\right) - \left(\bar{y}_k^{PRE(k)} - \bar{y}_\ell^{PRE(k)}\right)$$



$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{k\ell}}^{\boldsymbol{DD,\ell}}$$

$$\hat{\beta}_{k\ell}^{DD,\ell} = \left(\bar{y}_{\ell}^{POST(\ell)} - \bar{y}_{k}^{POST(\ell)}\right) - \left(\bar{y}_{\ell}^{MID(k,\ell)} - \bar{y}_{k}^{MID(k,\ell)}\right)$$



What is $\widehat{\boldsymbol{\beta}}^{\boldsymbol{DD}}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

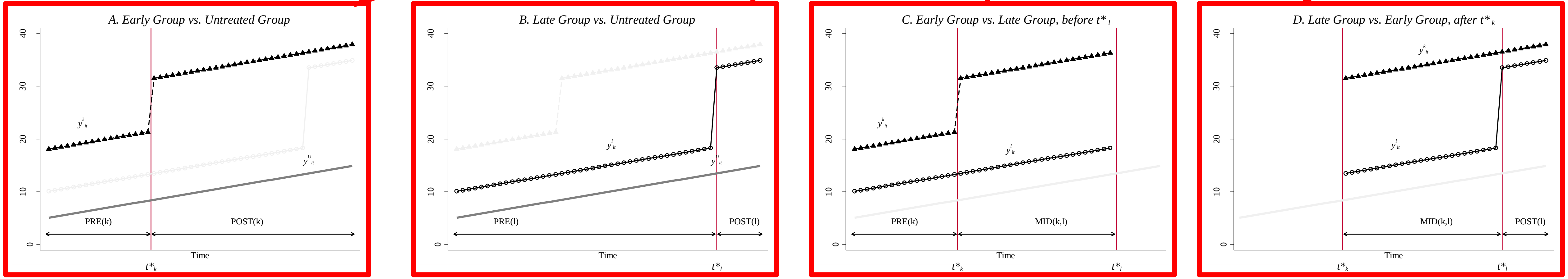
$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}_{kU}^{DD} + s_{\ell U}\hat{\beta}_{\ell U}^{DD} + \left[s_{k\ell}^{k}\hat{\beta}_{k\ell}^{DD,k} + s_{k\ell}^{\ell}\hat{\beta}_{k\ell}^{DD,\ell}\right]$$

What is $\hat{\beta}^{DD}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU}\boxed{\hat{\beta}^{DD}_{kU}} + s_{\ell U}\boxed{\hat{\beta}^{DD}_{\ell U}} + \left[s^{k}_{k\ell}\boxed{\hat{\beta}^{DD,k}_{k\ell}} + s^{\ell}_{k\ell}\boxed{\hat{\beta}^{DD,\ell}_{k\ell}}\right]$$



What is $\widehat{\beta}^{DD}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{DD}_{kU} + s_{\ell U}\hat{\beta}^{DD}_{\ell U} + \left[s^{k}_{k\ell}\hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell}\hat{\beta}^{DD,\ell}_{k\ell}\right]$$

$$s_{kU} = \frac{(n_k + n_U)^2 n_{kU}(1 - n_{kU})\bar{D}_k(1 - \bar{D}_k)}{V(\tilde{D}_{it})}$$

Sample size²

$$s^{k}_{k\ell} = \frac{\left((n_k + n_\ell)(1 - \bar{D}_\ell)\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell}\frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}{V(\tilde{D}_{it})}$$

$$s^{\ell}_{k\ell} = \frac{\left((n_k + n_\ell)\bar{D}_k\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}\frac{\bar{D}_\ell}{\bar{D}_k}}{V(\tilde{D}_{it})}$$

What is $\widehat{\beta}^{DD}$?

$$y_{it} = \alpha_i + \alpha_t + \hat{\beta}^{DD} D_{it} + u_{it}$$

For three groups:

$$\hat{\beta}^{DD} = s_{kU}\hat{\beta}^{DD}_{kU} + s_{\ell U}\hat{\beta}^{DD}_{\ell U} + \left[s^{k}_{k\ell}\hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell}\hat{\beta}^{DD,\ell}_{k\ell}\right]$$

$$s_{kU} = \frac{(n_k + n_U)^2 n_{kU}(1 - n_{kU})\bar{D}_k(1 - \bar{D}_k)}{V(\tilde{D}_{it})}$$

$$s^{k}_{k\ell} = \frac{\left((n_k + n_\ell)(1 - \bar{D}_\ell)\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{1 - \bar{D}_\ell}\frac{1 - \bar{D}_k}{1 - \bar{D}_\ell}}{V(\tilde{D}_{it})}$$

$$s^{\ell}_{k\ell} = \frac{\left((n_k + n_\ell)\bar{D}_k\right)^2 n_{k\ell}(1 - n_{k\ell})\frac{\bar{D}_k - \bar{D}_\ell}{\bar{D}_k}\frac{\bar{D}_\ell}{\bar{D}_k}}{V(\tilde{D}_{it})}$$

Subsample variance of treatment

Difference-in-Differences Decomposition Theorem

Assume that there are $k = 1, \ldots, K$ groups of treated units ordered by treatment time t_k^* and one control group, U, which does not receive treatment in the data. The share of units in group k is n_k, and the share of periods that group k spends under treatment is $\overline{D}_k$. The DD estimate from a two-way fixed effects model is a weighted average all two-group DD estimators:

$$\hat{\beta}^{DD} = \sum_{k \neq U} s_{kU}\, \hat{\beta}_{kU}^{DD} + \sum_{k \neq U} \sum_{\ell > k} \left[s_{k\ell}^{k} \hat{\beta}_{k\ell}^{DD,k} + s_{k\ell}^{\ell} \hat{\beta}_{k\ell}^{DD,\ell} \right]$$

With weights equal to:

$$s_{kU} = \frac{(n_k + n_U)^2\, \hat{V}_{kU}^{D}}{V\left(\widetilde{D}_{it}\right)}$$

$$s_{k\ell}^{k} = \frac{\left((n_k + n_\ell)(1 - \overline{D}_\ell)\right)^2 \hat{V}_{k\ell}^{D,k}}{V\left(\widetilde{D}_{it}\right)}$$

$$s_{k\ell}^{\ell} = \frac{\left((n_k + n_\ell)\overline{D}_k\right)^2 \hat{V}_{k\ell}^{D,\ell}}{V\left(\widetilde{D}_{it}\right)}$$

$\sum_{k \neq U} s_{kU} + \sum_{k \neq U} \sum_{\ell > k} \left[s_{k\ell}^{k} + s_{k\ell}^{\ell} \right] = 1.$

What can we do with this?

1. Describe sources of variation
(more on this later)

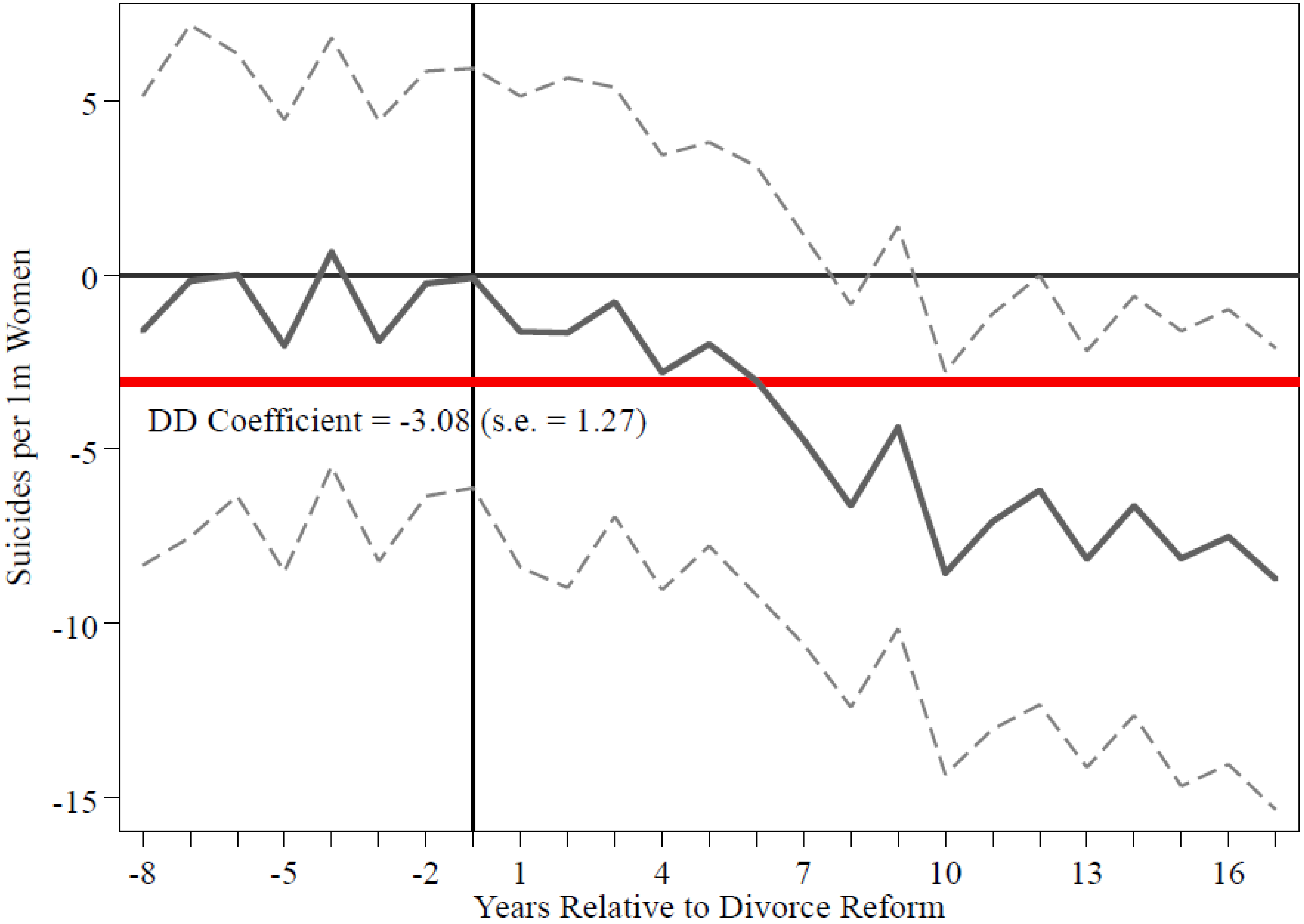
BARGAINING IN THE SHADOW OF THE LAW: DIVORCE LAWS AND FAMILY DISTRESS*

Betsey Stevenson and Justin Wolfers

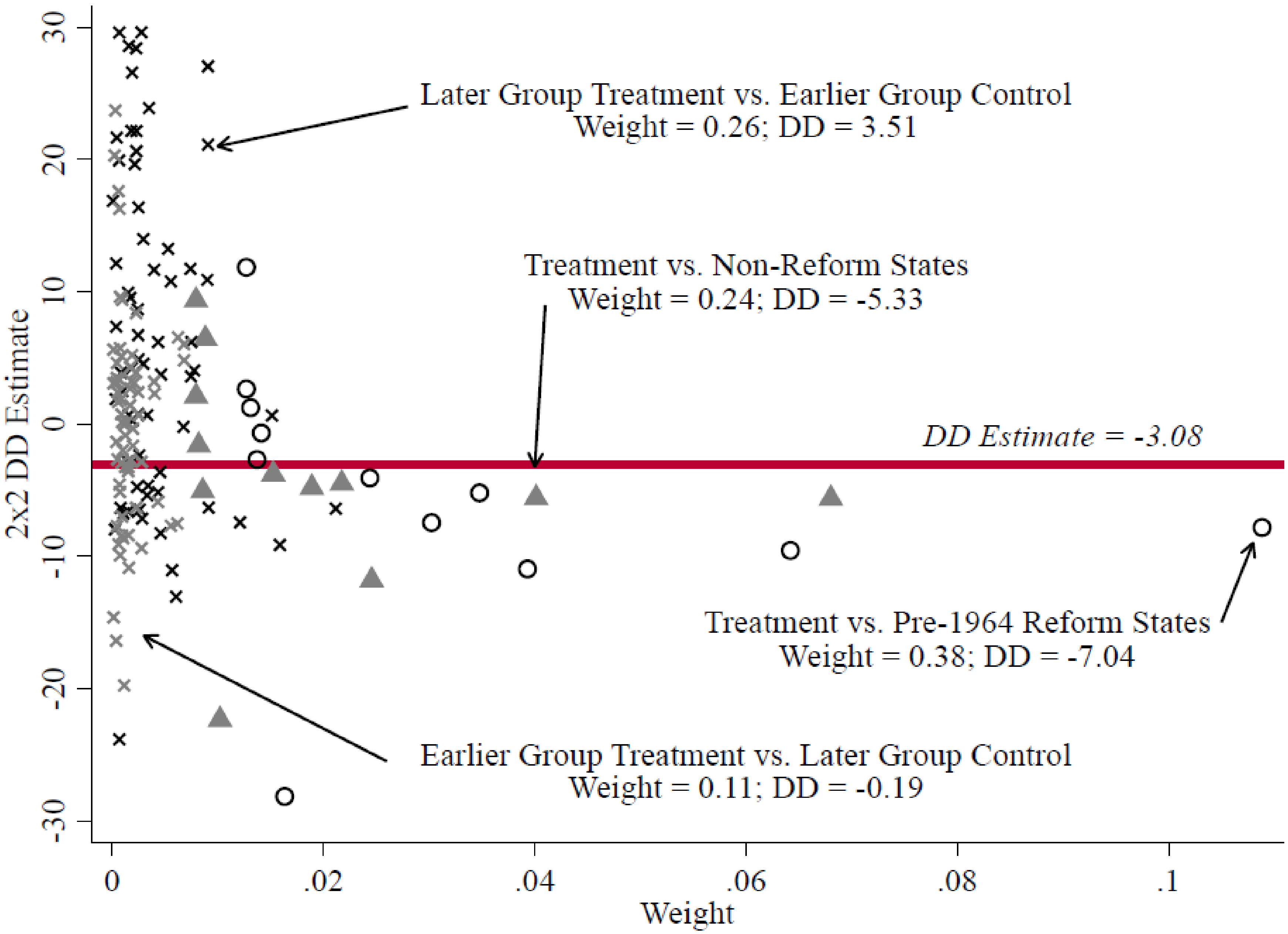
This paper exploits the variation occurring from the different timing of divorce law reforms across the United States to evaluate how unilateral divorce changed family violence and whether the option provided by unilateral divorce reduced suicide and spousal homicide.

$$Suicide\ rate_{s,t} = \sum_k \beta_k Unilateral^k_{s,t} + \sum_s \eta_s State_s + \sum_t \lambda_t Year_t + Controls_{s,t} + \varepsilon_{s,t}.$$

Example: Stevenson and Wolfers (QJE 2006)



Graphing the Decomposition: Divorce Example



2. Connect $\widehat{\beta}^{DD}$ and potential outcomes

Potential Outcomes in Each 2x2

(1)

$$ATT_k(W) \equiv E\left[Y_{it}^1 - Y_{it}^0 \middle| k, t \in W\right]$$

Aggregate of "group-time ATT" (Callaway and Sant'Anna 2018)

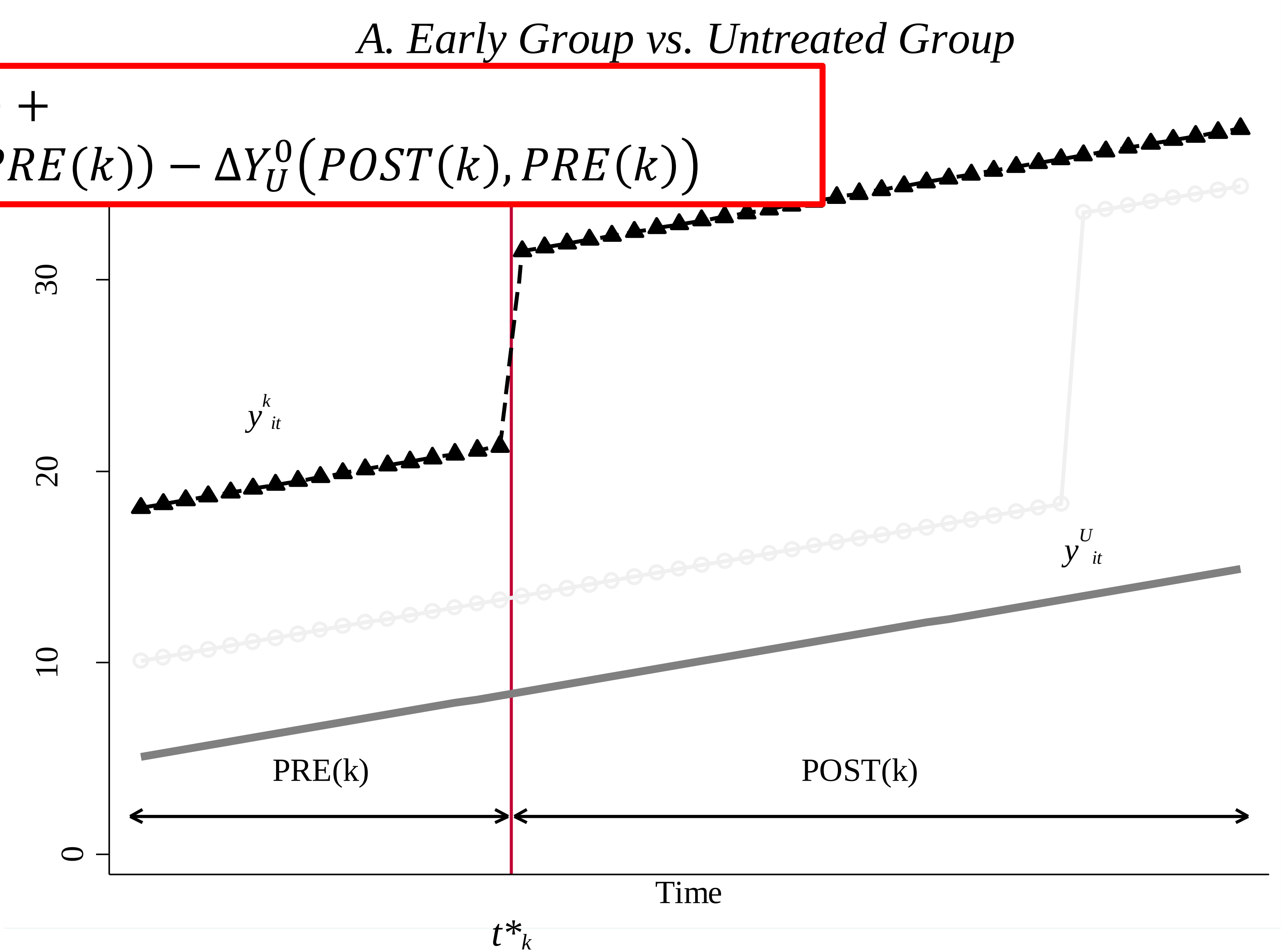
(2)

$$\Delta Y_k^h(A,B) \equiv E\left[Y_{it}^h \middle| k, B\right] - E\left[Y_{it}^h \middle| k, A\right], \qquad h = 0,1$$

Trend in potential outcome b/w periods (ie. "pre" and "post")

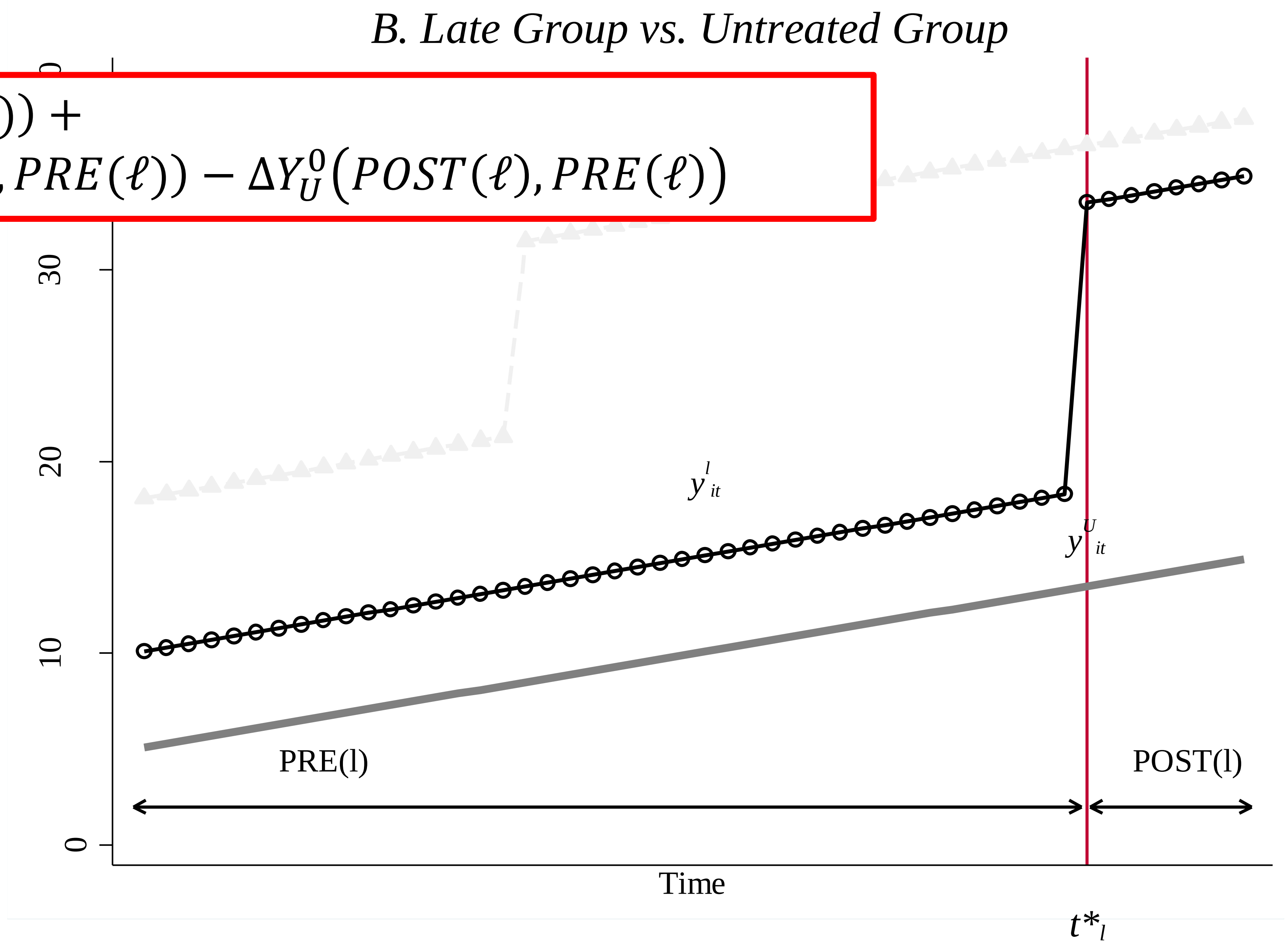
$$\widehat{\beta}^{DD}_{kU}$$

$$ATT_k(POST(k)) + \Delta Y^0_k(POST(k), PRE(k)) - \Delta Y^0_U\big(POST(k), PRE(k)\big)$$



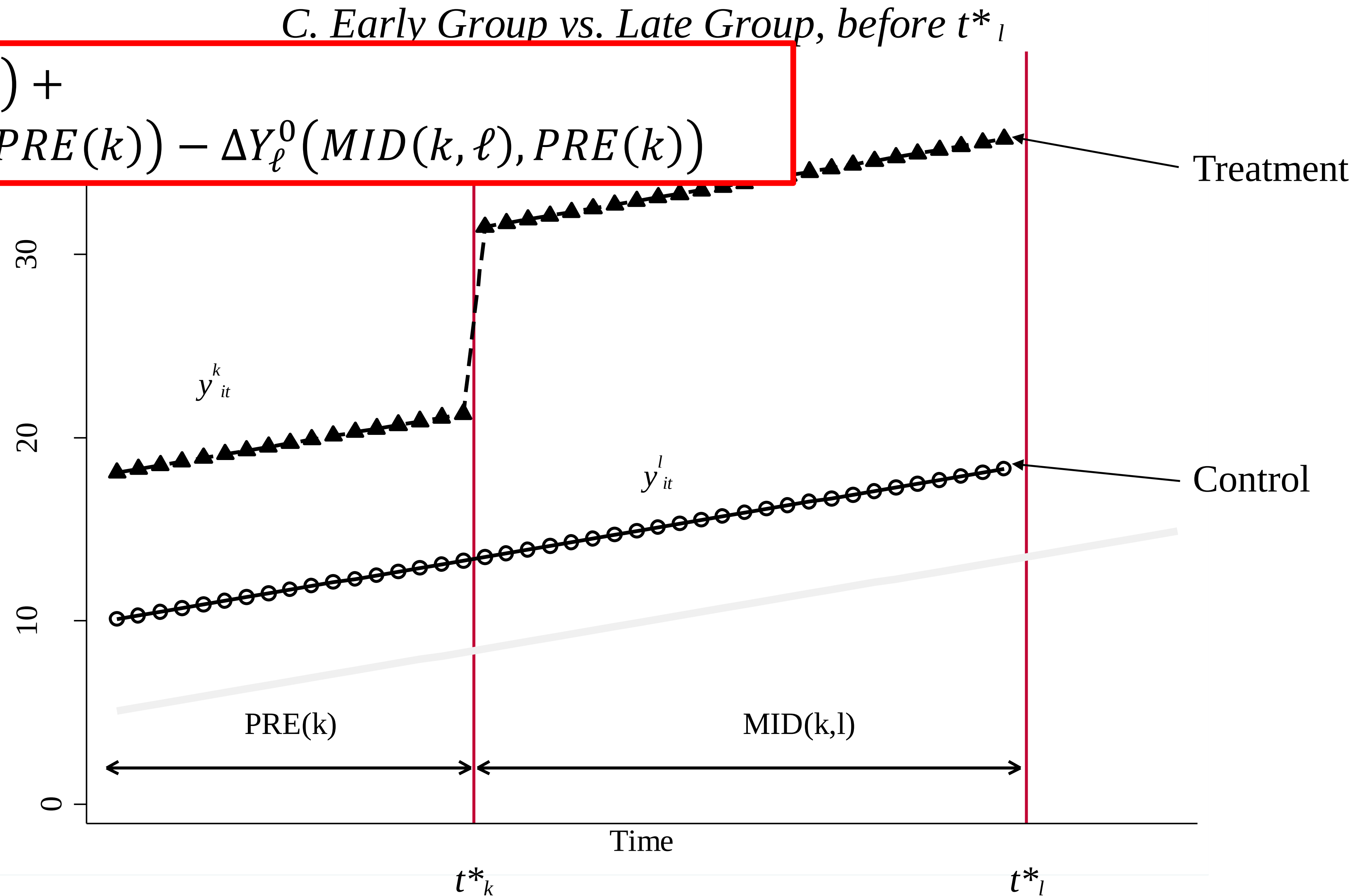
$$\widehat{\beta}^{DD}_{\ell U}$$

$$ATT_\ell(POST(\ell)) +$$
$$\Delta Y^0_\ell(POST(\ell), PRE(\ell)) - \Delta Y^0_U\big(POST(\ell), PRE(\ell)\big)$$



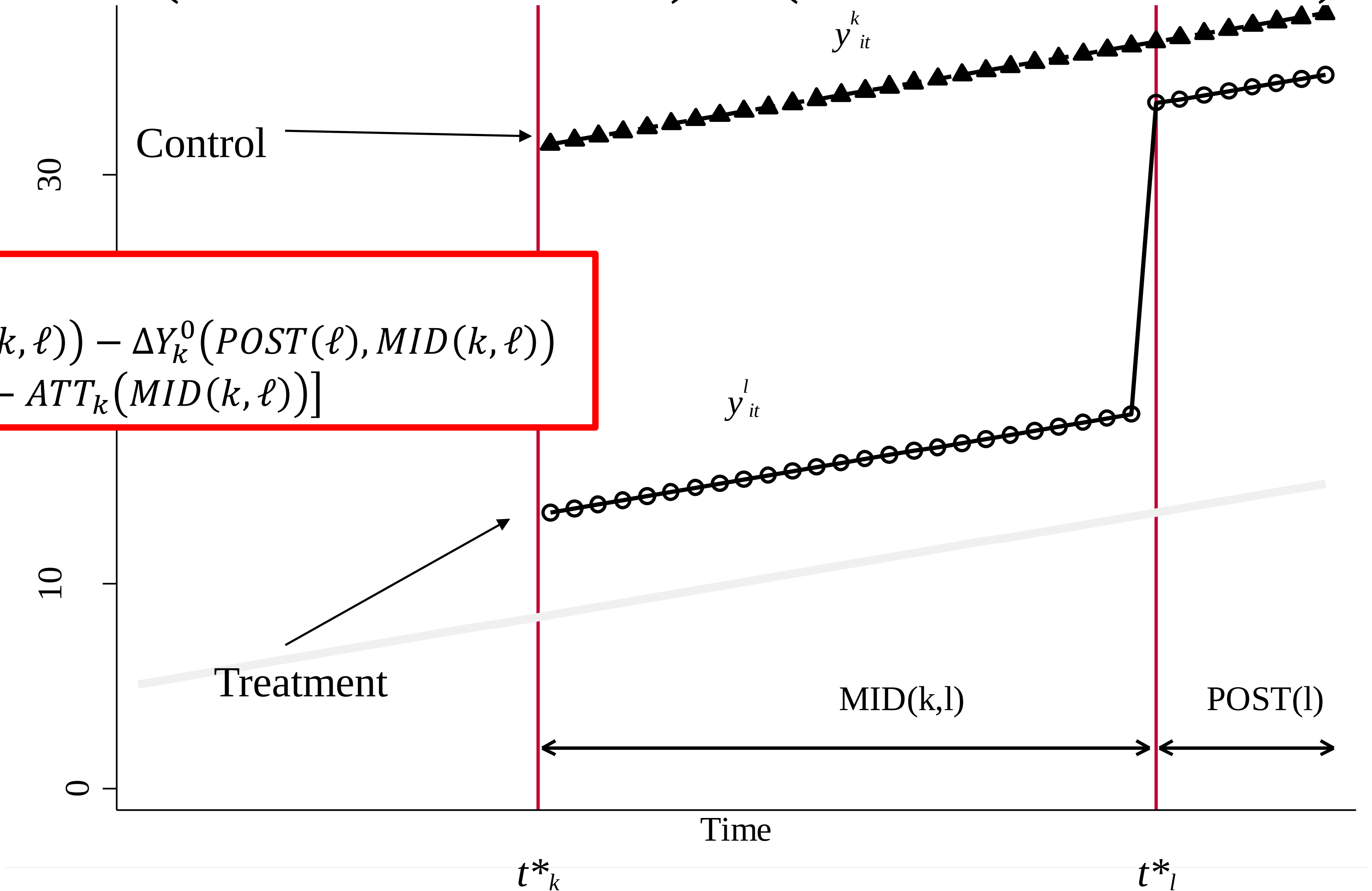
$$\widehat{\beta}_{k\ell}^{DD,k}$$

$$ATT_k\left(MID(k,\ell)\right) + \Delta Y_k^0\left(MID(k,\ell), PRE(k)\right) - \Delta Y_\ell^0\left(MID(k,\ell), PRE(k)\right)$$



$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{k\ell}}^{\boldsymbol{DD,\ell}}$$

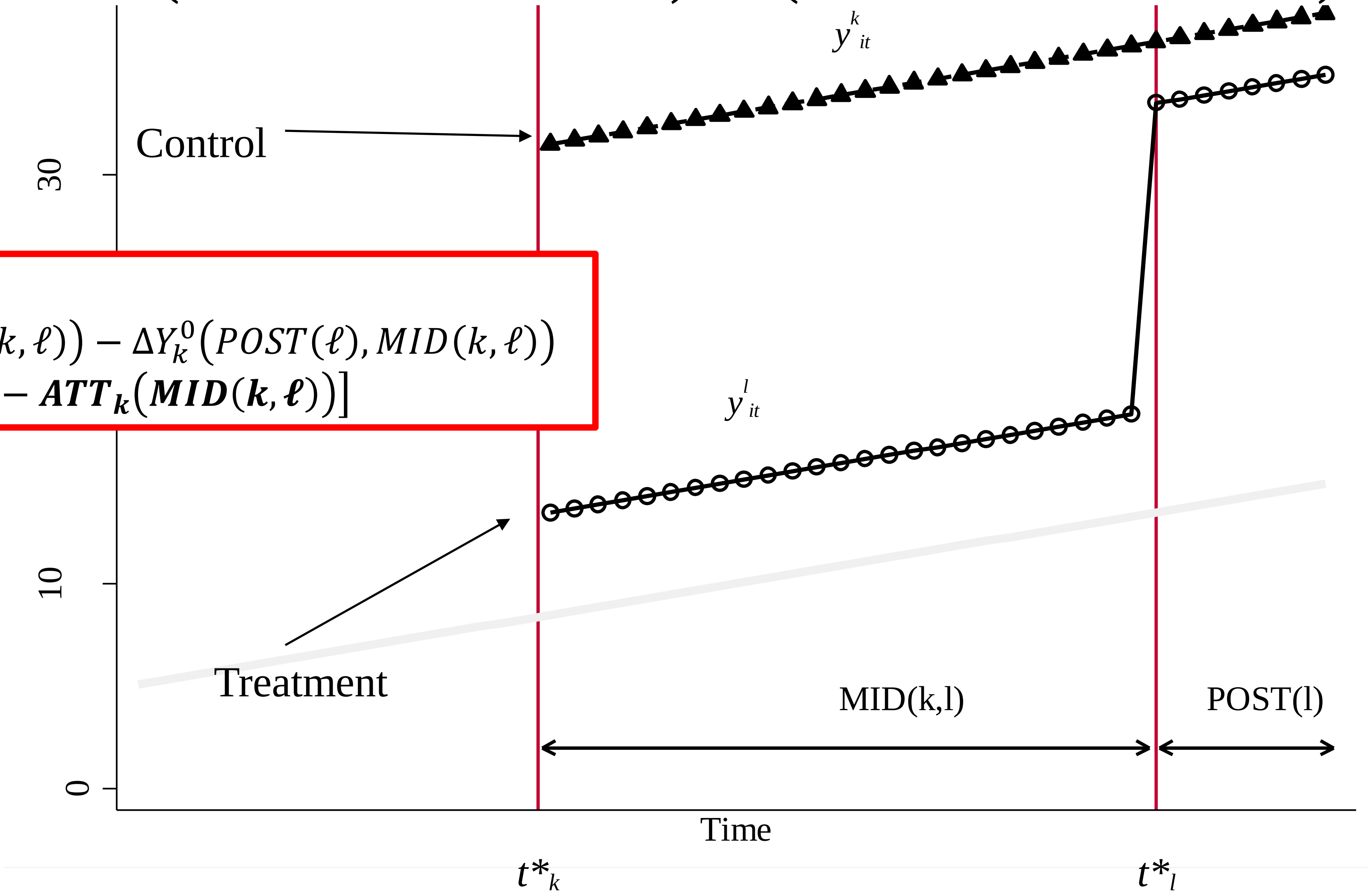
$$\hat{\beta}_{k\ell}^{DD,\ell} = \left(\bar{y}_{\ell}^{POST(\ell)} - \bar{y}_{k}^{POST(\ell)}\right) - \left(\bar{y}_{\ell}^{MID(k,\ell)} - \bar{y}_{k}^{MID(k,\ell)}\right)$$



$$\begin{aligned}
&ATT_{\ell}(POST(\ell)) + \\
&\Delta Y_{\ell}^{0}(POST(\ell), MID(k,\ell)) - \Delta Y_{k}^{0}(POST(\ell), MID(k,\ell)) \\
&- \left[ATT_{k}(POST(\ell)) - ATT_{k}(MID(k,\ell))\right]
\end{aligned}$$

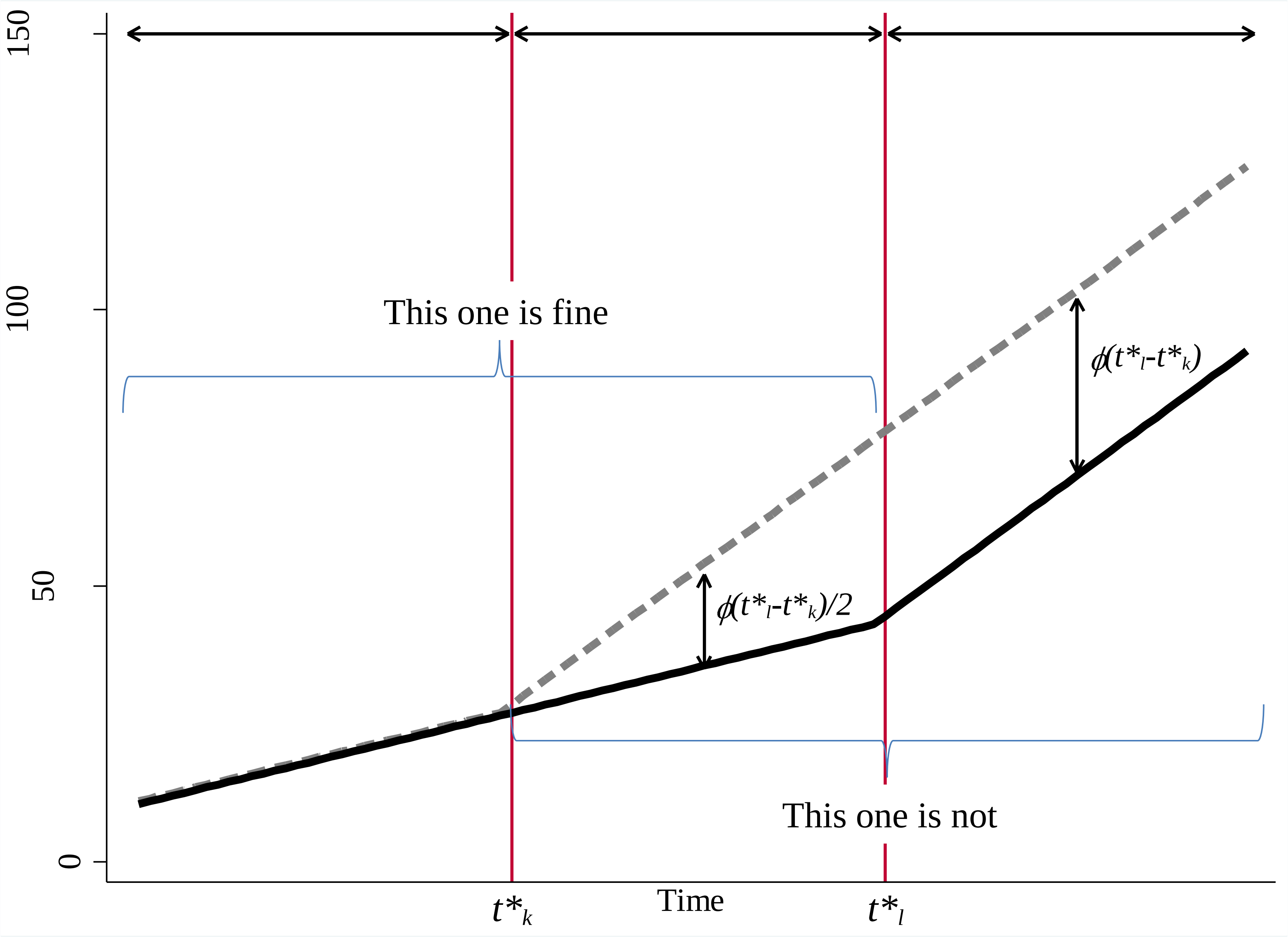
$$\widehat{\boldsymbol{\beta}}_{\boldsymbol{k\ell}}^{\boldsymbol{DD,\ell}}$$

$$\hat{\beta}_{k\ell}^{DD,\ell} = \left(\bar{y}_{\ell}^{POST(\ell)} - \bar{y}_{k}^{POST(\ell)}\right) - \left(\bar{y}_{\ell}^{MID(k,\ell)} - \bar{y}_{k}^{MID(k,\ell)}\right)$$



$$ATT_{\ell}\left(POST(\ell)\right) +$$
$$\Delta Y_{\ell}^{0}\left(POST(\ell), MID(k,\ell)\right) - \Delta Y_{k}^{0}\left(POST(\ell), MID(k,\ell)\right)$$
$$- \left[\boldsymbol{ATT_{k}\left(POST(\ell)\right) - ATT_{k}\left(MID(k,\ell)\right)}\right]$$

Time-Varying Homogeneous Effects



3. Define the parameter we get

Overlap with:
Borusyak and Jaravel
de Chaisemartin and D’Haultfoeuille
Abraham and Sun
Strezhnev
Imai and Kim
Athey and Imbens

Intuition for how DD weights heterogeneity

$$\hat{\beta}^{DD} = \sum_{k \neq U} s_{kU}\, \hat{\beta}^{DD}_{kU} + \sum_{k \neq U} \sum_{\ell > k} \left[s^{k}_{k\ell} \hat{\beta}^{DD,k}_{k\ell} + s^{\ell}_{k\ell} \hat{\beta}^{DD,\ell}_{k\ell} \right]$$

Under "identical trends" each 2x2 term ID's an ATT:

$$ATT_k(POST(k))$$

Or

$$ATT_k(MID(k,\ell))$$

Or*

$$ATT_\ell(POST(\ell))$$

*There's an issue here—will return to it.

Constant Homogeneous Effects

If all ATTs are the same across groups and over time:

$$plim\ \hat{\beta}^{DD} = \sum_{k \neq U} \sigma_{kU}\, ATT + \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT + \sigma_{k\ell}^{\ell} ATT \right]$$

$$= ATT$$

Two-way FE DD gives a thing we understand!

Constant Heterogeneous Effects

If ATTs differ across groups but not over time:

$$plim\ \hat{\beta}^{DD} = \sum_{k \neq U} \sigma_{kU}\, ATT_k + \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT_k + \sigma_{k\ell}^{\ell} ATT_\ell \right]$$

$$= \sum_{k \neq U} \underbrace{ATT_k \left(\overbrace{\sigma_{kU} + \sum_{j=1}^{k-1} \sigma_{jk}^{k} + \sum_{j=k+1}^{K} \sigma_{kj}^{k}}^{\text{wts on DDs where k is the treatment group} \equiv \boldsymbol{w}_{\boldsymbol{k}}^{\boldsymbol{T}}} \right)}_{\text{Not generally proportional to } n_k \rightarrow\ plim\ \hat{\beta}^{DD} \neq ATET}$$

Constant Heterogeneous Effects

If ATTs differ across groups but not over time:

$$plim\ \hat{\beta}^{DD} = \sum_{k \neq U} \sigma_{kU}\, ATT_k + \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT_k + \sigma_{k\ell}^{\ell} ATT_\ell\right]$$

"Variance-Weighted Average Treatment Effect" ™

$$VWATT \equiv \sum_{k \neq U} ATT_k \left(\sigma_{kU} + \sum_{j=1}^{k-1} \sigma_{jk}^{k} + \sum_{j=k+1}^{K} \sigma_{kj}^{k} \right)$$

Time-Varying Homogeneous Effects

VWATT

$$
plim\ \hat{\beta}^{DD} = \boxed{\sum_{k \neq U} \sigma_{kU}\, ATT\big(POST(k)\big) + \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT\big(MID(k,\ell)\big) + \sigma_{k\ell}^{\ell} ATT\big(POST(k)\big)\right]}
$$

$$
+ \sum_{k \neq U} \sum_{\ell > k} \sigma_{k\ell}^{\ell} \left[ATT_k\big(POST(\ell)\big) - ATT_k\big(MID(k,\ell)\big)\right]
$$

Time-Varying Homogeneous Effects

VWATT

$$plim\ \hat{\beta}^{DD} = \boxed{\sum_{k \neq U} \sigma_{kU}\, ATT\big(POST(k)\big) + \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT\big(MID(k,\ell)\big) + \sigma_{k\ell}^{\ell} ATT\big(POST(k)\big)\right]}$$
$$+ \sum_{k \neq U} \sum_{\ell > k} \sigma_{k\ell}^{\ell} \left[ATT_k\big(POST(\ell)\big) - ATT_k\big(MID(k,\ell)\big)\right]$$

Borusyak and Jaravel:

average of treatment effects that severely overweighs short-run effects

Here's why: Most 2x2 "post" windows are small.

Time-Varying Homogeneous Effects

$$plim\ \hat{\beta}^{DD} = \sum_{k \neq U} \sigma_{kU}\, ATT\big(POST(k)\big)$$

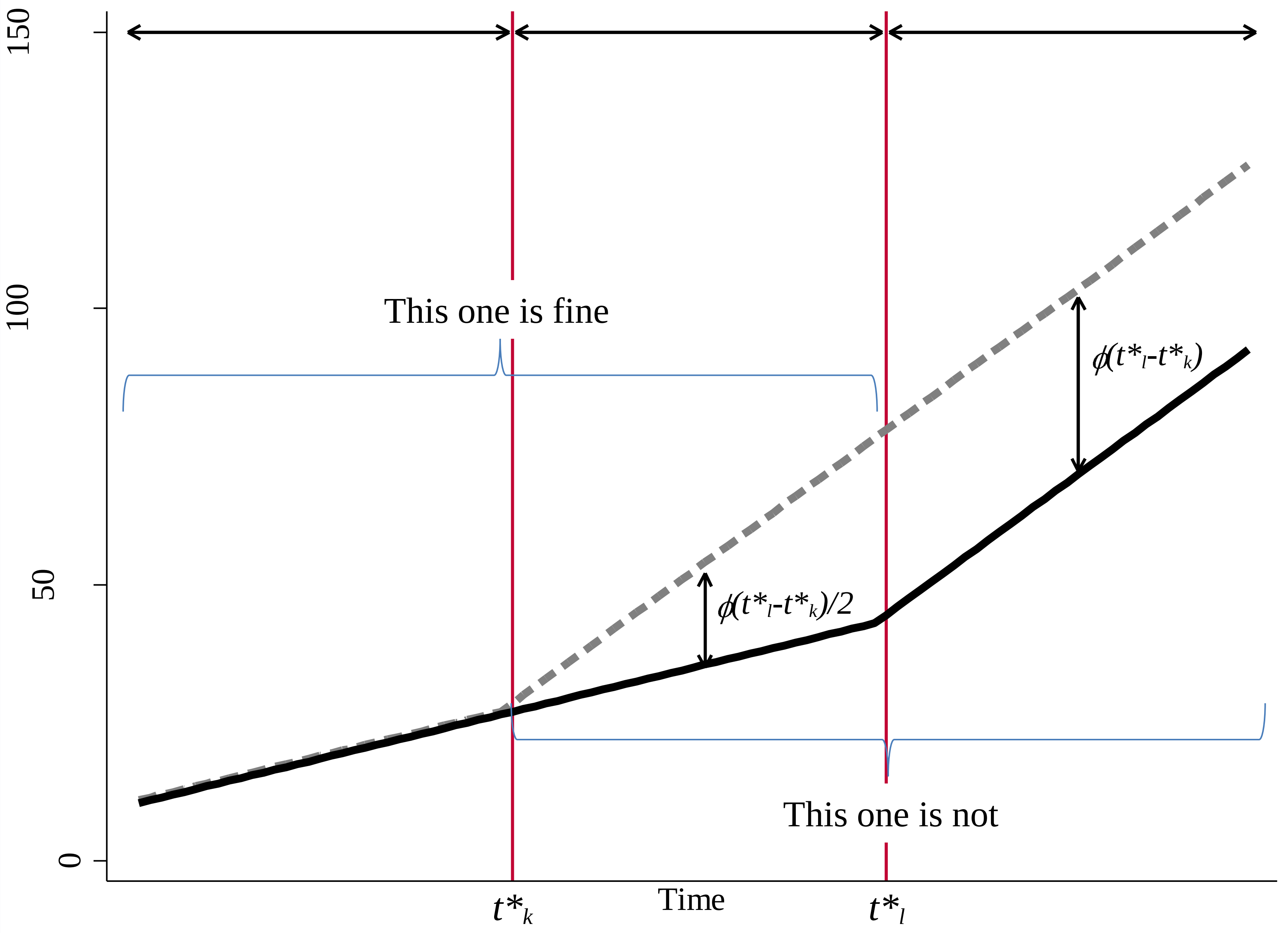
$$+ \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT\big(MID(k,\ell)\big) + \sigma_{k\ell}^{\ell} ATT\big(POST(k)\big)\right]$$

$$+ \underbrace{\sum_{k \neq U} \sum_{\ell > k} \sigma_{k\ell}^{\ell} \left[ATT_k\big(POST(\ell)\big) - ATT_k\big(MID(k,\ell)\big)\right]}_{\text{But you don't even get VWATT}}$$

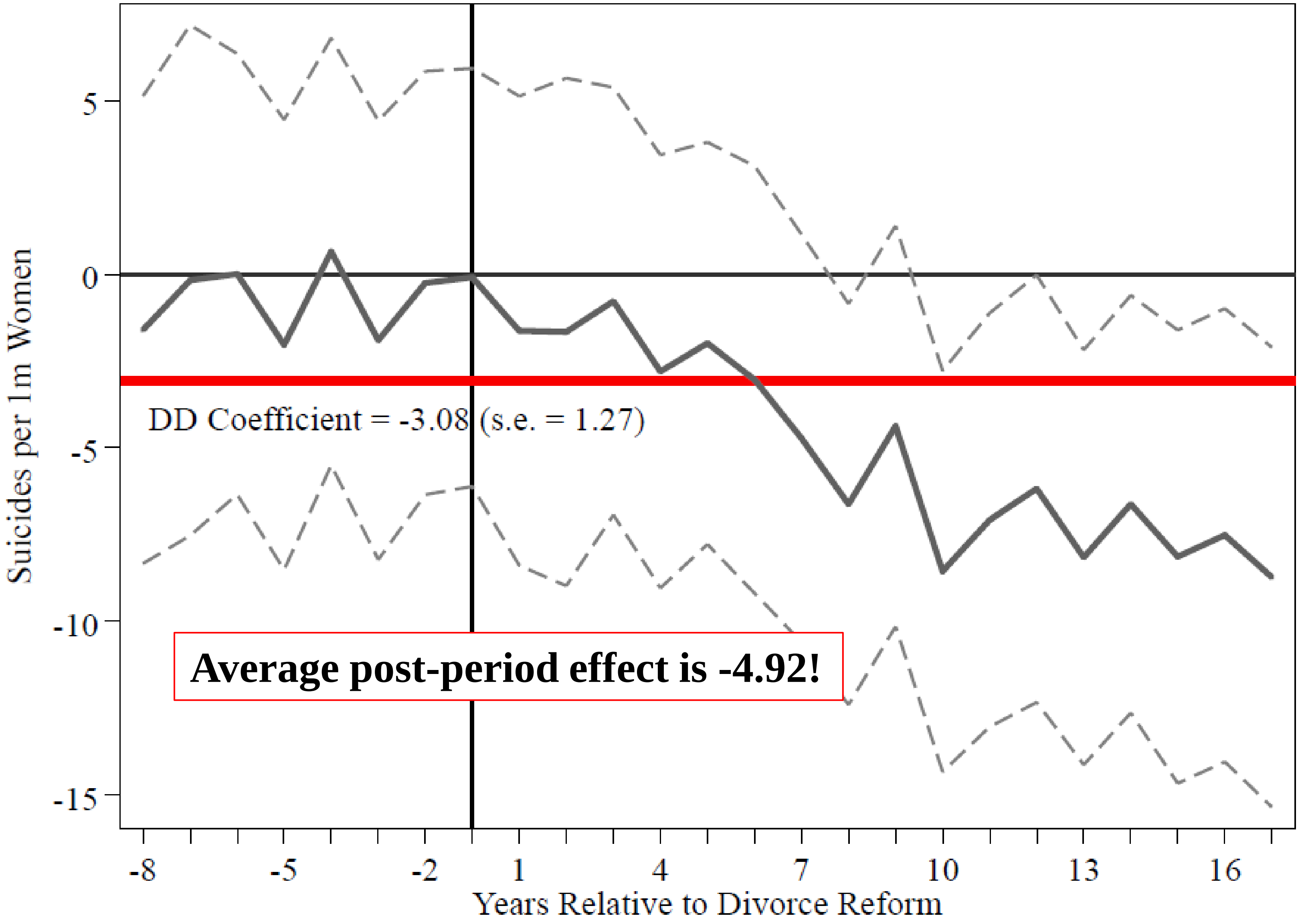
Time-Varying Homogeneous Effects

$$
\begin{aligned}
plim\ \hat{\beta}^{DD} = & \sum_{k \neq U} \sigma_{kU}\, ATT\big(POST(k)\big) \\
& + \sum_{k \neq U} \sum_{\ell > k} \left[\sigma_{k\ell}^{k} ATT\big(MID(k,\ell)\big) + \sigma_{k\ell}^{\ell} ATT\big(POST(k)\big)\right] \\
& \underbrace{+ \sum_{k \neq U} \sum_{\ell > k} \sigma_{k\ell}^{\ell} \left[ATT_k\big(POST(\ell)\big) - ATT_k\big(MID(k,\ell)\big)\right]}_{\boxed{\text{But you don't even get VWATT}}} \Big\} \boxed{\Delta ATT}
\end{aligned}
$$

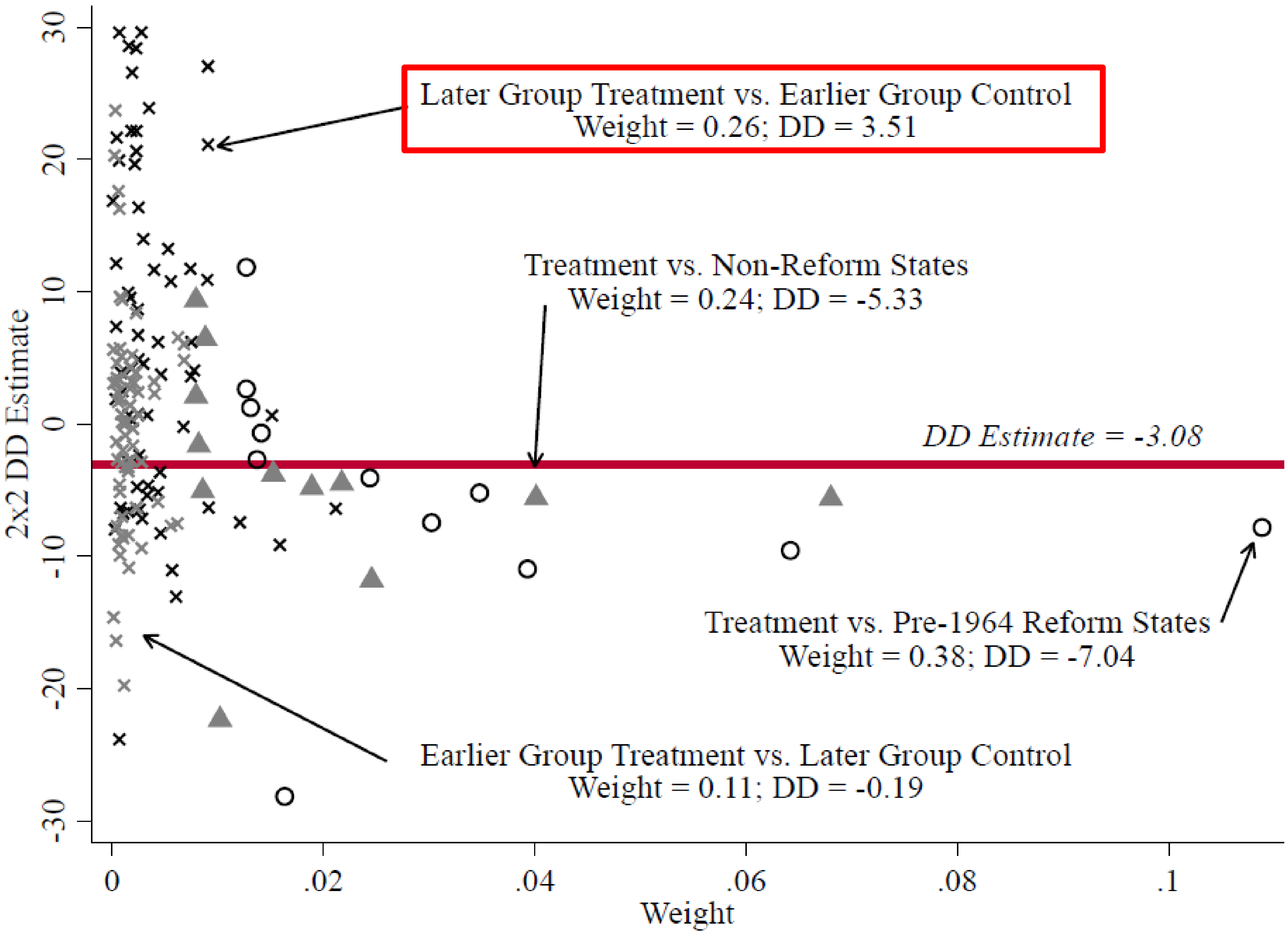
Time-Varying Homogeneous Effects



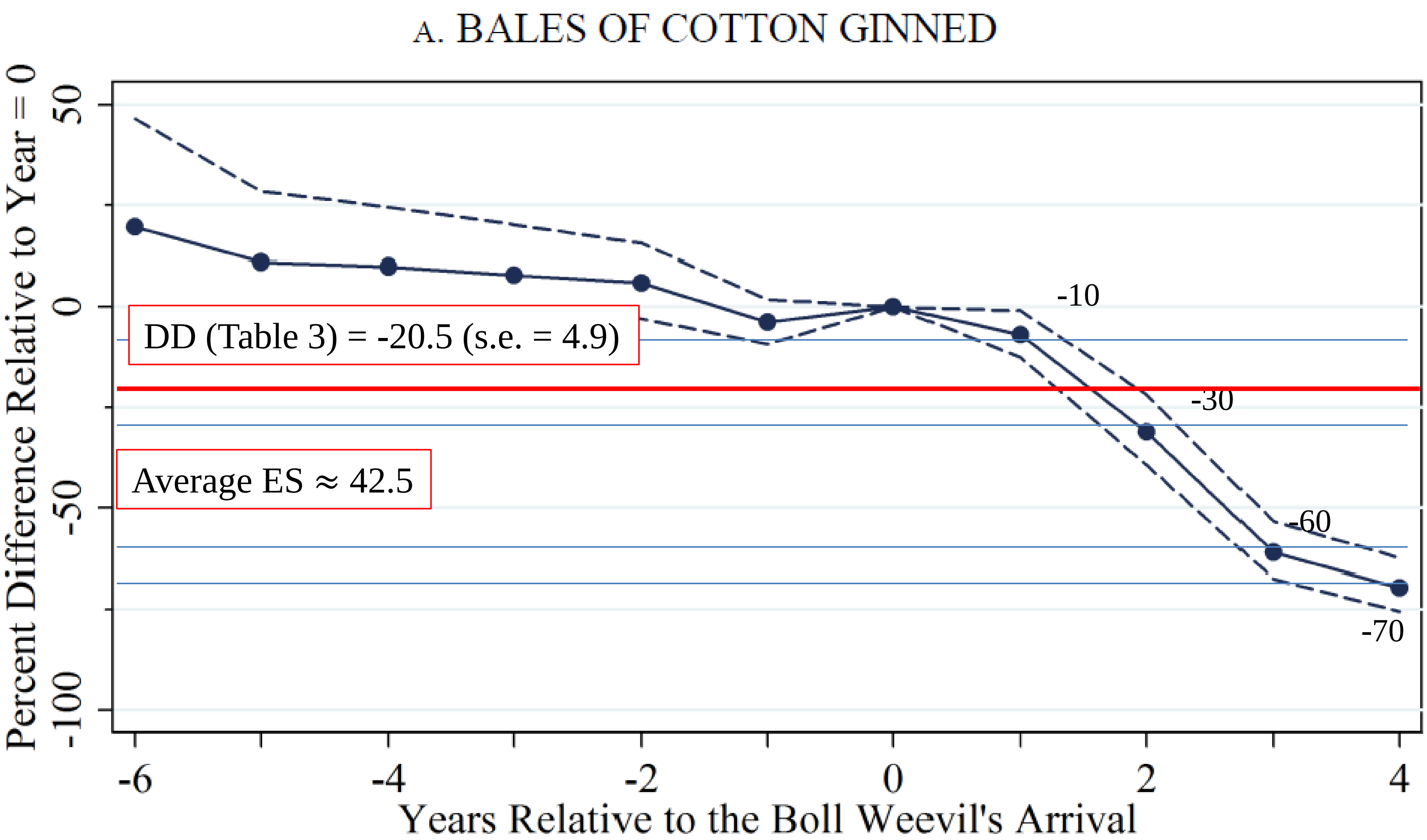
Example: Stevenson and Wolfers (2006)



Graphing the Decomposition: Divorce Example



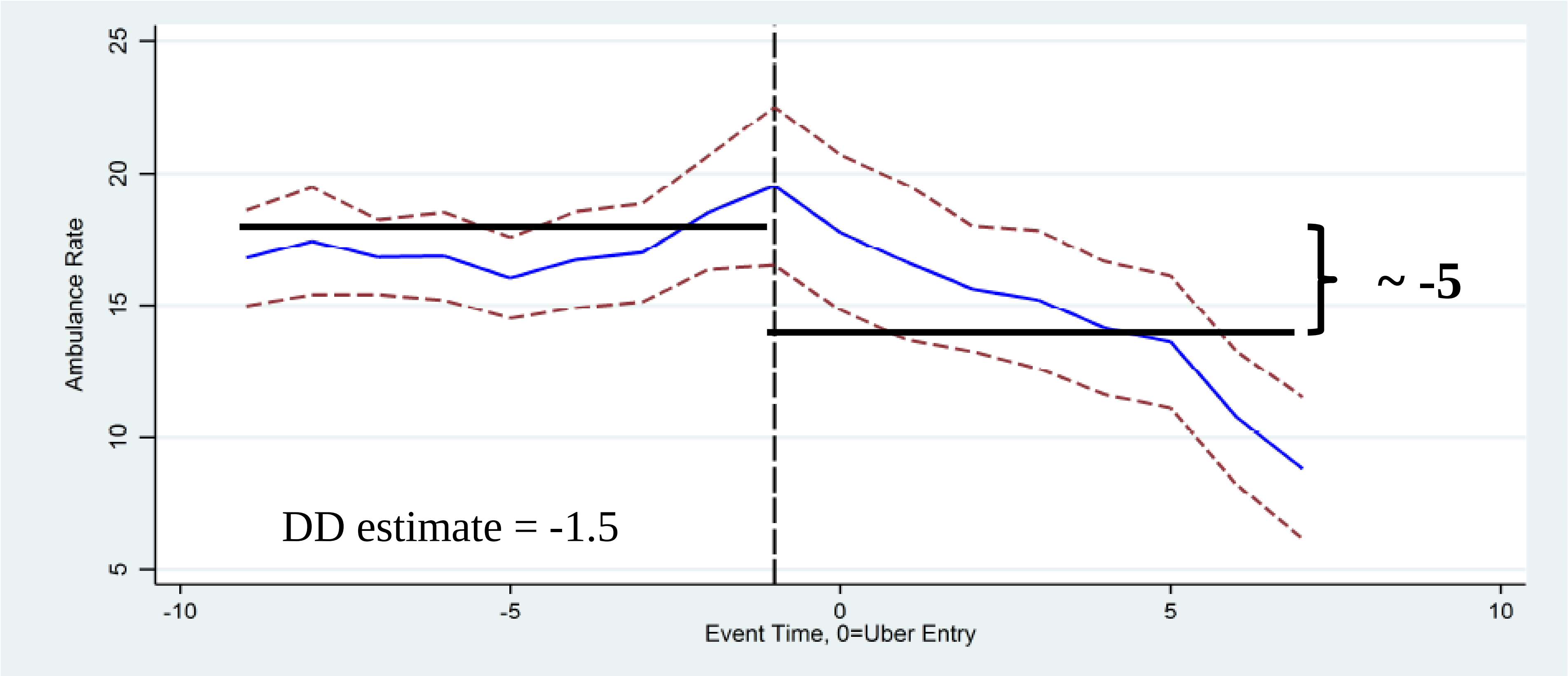
Baker (2015)



A. BALES OF COTTON GINNED
Percent Difference Relative to Year = 0
50
0
-50
-100
-6
-4
-2
0
2
4
Years Relative to the Boll Weevil's Arrival
DD (Table 3) = -20.5 (s.e. = 4.9)
Average ES ≈ 42.5
-10
-30
-60
-70

Moskatel and Slusky (2017)

Figure 2: Event Study



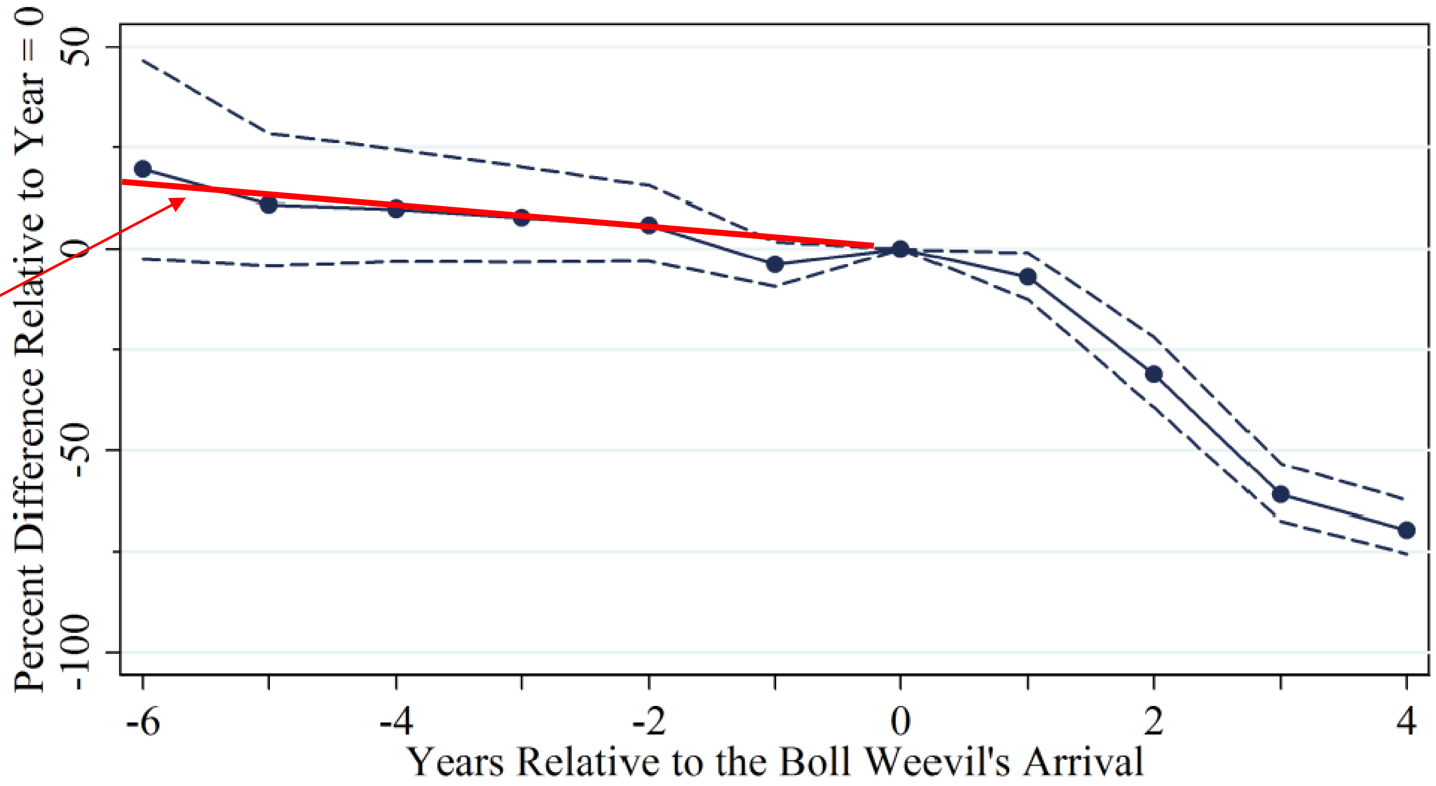
Notes: The dashed red lines show the 95% confidence interval. The vertical dashed black line shows the last quarter when Uber was not available at least some of the time.

Is event-study OK?

Abraham and Sun:

under treatment effect heterogeneity, we can have a spurious non-zero positive lead coefficient even when there is no pretrend.

A. BALES OF COTTON GINNED



Maybe this isn't even real!

Conclusion

1. Two way FE can be problematic:

a) Groups in the middle of the panel weight up their respective 2x2 TE via the variance weighting

b) Decomposition highlights the strange role of panel length when using TWFE

c) Different choices about panel length change both the 2x2 and the weights based on variance of treatment

d) Variation in TE over time can cause problems in some of the 2x2 DIDs

2. Many other alternative estimators that can be used instead

   a. De Chaisemartin and D'Haultfoeuille (AER 2020): « Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects"
   b. Callaway and Sant'Anna (JoE 2020): « Difference-in-Differences with multiple time periods"
   c. Athey and Imbens (JoE 2021) :'Design-based analysis in Difference-In-Differences settings with staggered adoption"
   d. Borusyak and Jaravel (2021): « Revisiting Event Study Designs »
   e. Abraham and Sun (JoE 2020): « Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects"
   f. Strezhnev (2018):« Semiparametric Weighting Estimators for Multi-Period Difference-in-Differences Designs"

See https://andrewcbaker.netlify.app/2019/09/25/difference-in-differences-methodology/ for a short summary

Also https://asjadnaqvi.github.io/DiD/

Next week

- Wooldridge, Jeffrey M. "Control function methods in applied econometrics." Journal of Human Resources 50.2 (2015): 420-445.
- Cornelissen et al (Labor Economics, 2016): "From LATE to MTE: Alternative methods for the evaluation of policy interventions"
- Heckman and Vytlacil (Econometrica, 2005): "Structural Equations, Treatment Effects, and Econometric Policy Evaluation"