Difference-in-Differences Estimators of Intertemporal Treatment Effects

Clément de Chaisemartin   Xavier D'Haultfœuille

Sciences Po   CREST

Stata Symposium

November 9, 2023

Motivation

- We study treatment-effect estimation, using panel of groups (e.g. counties).

- $Y_{g,t}$ may be affected by $D_{g,t}$, but also by $D_{g,t-1}$, $D_{g,t-2}$, etc.

- Standard tool: TWFE reg. E.g. $Y_{g,t} = \alpha_g + \gamma_t + \sum_{l=0}^{L} \beta_l D_{g,t-l} + u_{g,t}$.

- However recent literature (including this paper) shows that TWFE not robust to heterogeneous treatment effects across groups and/or over time.

- Heterogeneity-robust DID estimators have been proposed, but papers that allow lagged D to affect outcome assume binary & absorbing treatment.

- Of the 100 most-cited papers published by AER from 2015 to 2019, 26 estimate a TWFE regression, but only four have a binary-and-staggered treatment. Existing het-robust DID estimators not widely applicable.

This paper:

1. Proposes het-robust DID estimators, in applications where treatment either non-binary and/or non-absorbing, and lagged D may affect outcome.
   - Estimators widely applicable: can be used in any design where some groups keep their period-one treatment for several periods.
   - Computed by `did_multiplegt_dyn` Stata package.

2. Studies three TWFE regressions commonly used to estimate instantaneous and dynamic effects in complex designs: local-projection, distributed lags and event-study with group-specific treatment intensities:
   - All three regressions are non-robust to heterogeneous treatment effects.
   - Local-projection is non-robust even if treatment effect homogeneous!

3. Revisits Favara and Imbs, who study effect of financial liberalization on volume of credit and housing prices:
   - Using LP regressions, authors had found short-lived effects.
   - On the contrary, our het-rob DID estimators indicate very persistent effects, and our decompositions show that estimates of long-run effects based on LP regressions biased towards 0.

Outline

Group-level panel data

- G groups observed at T periods, respectively indexed by g and t.

- Typically, groups = geographical entities (states...) but a group could also just be single individual or firm.

- Estimators below not weighted by $N_{g,t}$, the population of cell (g, t). Just to reduce notational complexity: proposing weighted estimators is a mechanical extension. did_multiplegt_dyn has a weight option.

- Also to reduce notational complexity: we assume panel is balanced. But did_multiplegt_dyn can be used with imbalanced panel, see specific documentation referenced in help file.

Treatment and potential outcomes

- When treatment assigned at (g, t) level: $D_{g,t}$ = treatment of g at t.

- When treatment varies within (g, t) cells (fuzzy design: groups are geographical entities and individuals or firms within the same cell may not all have same treatment): $D_{g,t}$ = average treatment in cell (g, t).

- $\boldsymbol{D}_g = (D_{g,1}, ..., D_{g,T})$: vector stacking g's treatments from period 1 to T, and $\boldsymbol{D} = (\boldsymbol{D}_1, ..., \boldsymbol{D}_G)$: vector stacking the treatments of all groups at every period. We refer to $\boldsymbol{D}$ as the study's design.

- Let $\mathcal{D}$ be the set of values $\boldsymbol{D}_g$ can take.

- For all $(d_1, ..., d_T) \in \mathcal{D}$, let $Y_{g,t}(d_1, ..., d_T)$ denote potential outcome of g at t if $(D_{g,1}, ..., D_{g,T}) = (d_1, ..., d_T)$, and let $Y_{g,t} = Y_{g,t}(\boldsymbol{D}_g)$ denote observed outcome of g at t.

- This dynamic potential outcome framework follows Robins (1986). Allows groups' outcome at t to depend on their past and future treatments.

Conditioning on the design

- When defining our target parameters, we will take the perspective of a social planner, seeking to conduct a cost-benefit analysis comparing groups' actual treatments D to the counterfactual "status-quo" scenario where every group would have kept the same treatment as in period 1 all the time.

- Then, all our analysis is conditional on D, the study's design.

- This implies that our parameters of interest are dictated by the design, rather than chosen by the researcher. A bit similar to LATE of Imbens & Angrist (1994).

- Conditional on D, only groups' potential outcomes are random. Probabilistic statements below are with respect to their joint probability distribution.

Condition on the design for our estimators to be applicable.

Definition 1

(First treatment change) For all g, let $F_g = \min\{t : t \geq 2, D_{g,t} \neq D_{g,t-1}\}$.

Convention: $F_g = T + 1$ if g's treatment never changes.

Design Restriction 1

$\exists (g, g')$ *such that: (i) $D_{g,1} = D_{g',1}$, (ii) $F_g \neq F_{g'}$.*

If groups' period-one treatments are i.i.d. draws from a continuous distribution, $D_{g,1} \neq D_{g',1}$ for all (g, g'), so (i) fails. In our Web Appendix, we extend our estimators to designs where (i) fails.

(ii) requires that there is heterogeneity in the date at which groups change treatment for the first time. (ii) fails if groups' treatment is extremely non-persistent, so that $D_{g,1} \neq D_{g,2}$ and $F_g = 2$ for all g (e.g.: treatment=rainfall).

Commonly-found designs

Design 1

(Binary and staggered treatment) $D_{g,t} = 1\{t \geq F_g\}$, with $F_g \geq 2$.

Design 2

(Binary treatment, groups join and then leave treatment) $D_{g,t} = 1\{E_g \geq t \geq F_g\}$.

Special case of Design 2: $E_g = F_g$, "one-shot-treatment design".

Design 3

(Staggered design with group-specific intensities) $D_{g,t} = I_g 1\{t \geq F_g\}$, $F_g \geq 2$.

Design 4

(Zero treatment at baseline) $D_{g,1} = 0$ *(Designs 1-3 special cases of Design 4)*.

Commonly-found designs

Design 5

(Discrete treatment at baseline) $D_{g,1} \in \{0, 1, ..., K\}$.

No-anticipation

Assumption 1

(No Anticipation) $\forall g$, $\forall (d_1, ..., d_T) \in \mathcal{D}$, $Y_{g,t}(d_1, ..., d_T) = Y_{g,t}(d_1, ..., d_t)$.

Parallel trends

- Let $\mathcal{D}_1^r = \{d : \exists (g, g') \in \{1, ..., G\}^2 : D_{g,1} = D_{g',1} = d, F_g \neq F_{g'}\}$ be set of period-one-treatment values such that two groups with different values of F_g have that period-one treatment.

- For any d in $\mathcal{D}_1^r$ and any t, let $\boldsymbol{d}_t$ denote a $1 \times t$ vector of ds.

- For any k, let $\boldsymbol{D}_{g,1,k}$ be a $1 \times k$ vector whose coordinates are all equal to $D_{g,1}$.

- $Y_{g,t}(\boldsymbol{D}_{g,1,t})$ is g's period-t outcome in a counterfactual where it keeps its period-one treatment till period t. "Status-quo" PO.

Assumption 2

(Parallel trends for the status-quo outcome, conditional on the period-one treatment)
$\forall (g, g')$, if $D_{g,1} = D_{g',1} \in \mathcal{D}_1^r$, then $\forall t \geq 2$,

$$E[Y_{g,t}(\boldsymbol{D}_{g,1,t}) - Y_{g,t-1}(\boldsymbol{D}_{g,1,t-1})|\boldsymbol{D}] = E[Y_{g',t}(\boldsymbol{D}_{g',1,t}) - Y_{g',t-1}(\boldsymbol{D}_{g',1,t-1})|\boldsymbol{D}].$$

- Assumption 2 requires that if two groups have same period-one treatment, then they have same expected evolution of status-quo outcome.

Parallel trends, continued

- Assumption 2: generalization of standard parallel-trends assumption in DID models to our set-up that allows for dynamic effects and for potentially complicated designs where groups may not all be untreated at period one.

- In Design 4, $\mathcal{D}_1^r = \{0\}$, so Assumption 2 only restricts the never-treated potential outcome $Y_{g,t}(\mathbf{0}_t)$, and is similar to identifying assumption considered by CS (2021) and SA (2021).

- Assumption 2 restricts one PO per group, so Assumption 2 does not restrict groups' treatment effects.

Outline

- For every g, let

$$T_g = \max_{g':D_{g',1}=D_{g,1}} F_{g'} - 1$$

denote last period where there is still a group with the same period-one treatment as g and whose treatment has not changed since the start of the panel.

- For any g such that $F_g \leq T_g$, and for any $\ell \in \{1, ..., T_g - F_g + 1\}$, let

$$\delta_{g,\ell} = E\left[Y_{g,F_g-1+\ell} - Y_{g,F_g-1+\ell}(D_{g,1}, ..., D_{g,1})|\boldsymbol{D}\right] \quad (1)$$

be expected diff between g's actual outcome at $F_g - 1 + \ell$ and its counterfactual "status quo" outcome. AVSQ effect of g at $F_g - 1 + \ell$.

Estimation of non-normalized AVSQ effects

- Let
$$N_t^g = \#\{g' : D_{g',1} = D_{g,1}, F_{g'} > t\}$$
be number of groups g' with the same period-one treatment as g, and that have kept the same treatment from period 1 to t.

- To estimate $\delta_{g,\ell}$:

$$\text{DID}_{g,\ell} = Y_{g,F_g-1+\ell} - Y_{g,F_g-1} - \frac{1}{N_{F_g-1+\ell}^g} \sum_{g':D_{g',1}=D_{g,1},F_{g'}>F_g-1+\ell} (Y_{g',F_g-1+\ell} - Y_{g',F_g-1}),$$

(2)

DID estimator comparing $F_g - 1$-to-$F_g - 1 + \ell$ outcome evolution of g to groups with same baseline treatment, and that have kept that treatment from period 1 to $F_g - 1 + \ell$.

Lemma 1

If Assumptions 1 and 2 hold, then for every (g,ℓ) such that $1 \leq \ell \leq T_g - F_g + 1$, $E[DID_{g,\ell}|\boldsymbol{D}] = \delta_{g,\ell}$.

Interpretability of AVSQ effects: no-crossing condition

- Assume that g_0 is such that $D_{g_0,1} = 1, D_{g_0,2} = 2, D_{g_0,3} = 0$. Then,

$$\delta_{g_0,2} = E[Y_{g_0,3}(1,2,0) - Y_{g_0,3}(1,1,1)]$$
$$= E[Y_{g_0,3}(1,2,0) - Y_{g_0,3}(1,1,0)] - E[Y_{g_0,3}(1,1,1) - Y_{g_0,3}(1,1,0)]$$

Both effects could be positive but $\delta_{g_0,2}$ negative.

- For all (g,ℓ) such that at $F_g - 1 + \ell$ g has experienced treatments strictly below and strictly above period-one treatment, $\delta_{g,\ell} =$ linear combination, with negative weights, of effects of increasing different treatment lags.

- We assume away the existence of such (g,ℓ)s.

Design Restriction 2

$\forall g \in \{1, ..., G\}$, either $D_{g,t} \geq D_{g,1}$ for every t, or $D_{g,t} \leq D_{g,1}$ for every t.

Restriction 2 implied by Design 4. It also holds automatically when the treatment is binary, or when groups' treatment can only change once.

- When Restriction 2 fails, one can just discard cells (g,t) such that at t, g has experienced treatments strictly below and strictly above its period-one treatment.

- did_multiplegt_dyn drops those cells if drop_larger_lower option specified.

Non-normalized event-study effects

- Let $L = \max_g(T_g - F_g + 1)$ be largest ℓ such that $\delta_{g,\ell}$ can be estimated for one g.

- For every $\ell \in \{1, ..., L\}$, let $N_\ell = \#\{g : F_g - 1 + \ell \leq T_g\}$ be number of groups for which $\delta_{g,\ell}$ can be estimated.

- For all g such that $F_g \leq T$, let $S_g = 1\{D_{g,F_g} > D_{g,1}\} - 1\{D_{g,F_g} < D_{g,1}\}$ be equal to 1 (resp. -1) for groups whose treatment increases (resp. decreases) at F_g.

- Then, let

$$\delta_\ell = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} S_g \delta_{g,\ell}. \tag{3}$$

- Under Restriction 2, for groups with $S_g = -1$, $D_{g,t} \leq D_{g,1}$ for all t, so $\delta_{g,\ell}$ is effect of being exposed to weakly lower treatment dose for ℓ periods. Taking negative of $\delta_{g,\ell}$ for those groups ensures that δ_ℓ is average effect of having been exposed to a weakly larger dose for ℓ periods.

- For every $\ell \in \{1, ..., L\}$, let

$$\text{DID}_\ell = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} S_g \text{DID}_{g,\ell}. \tag{4}$$

DID_ℓ conditionally unbiased for δ_ℓ under Assumptions 1, 1, and 2.

Interpreting non-normalized event-study effects

- In Design 1,

$$\delta_\ell = \frac{1}{N_\ell} \sum_{g : F_g - 1 + \ell \le T_g} E\left[Y_{g, F_g - 1 + \ell}(\mathbf{0}_{F_g - 1}, \mathbf{1}_\ell) - Y_{g, F_g - 1 + \ell}(\mathbf{0}_{F_g - 1 + \ell}) | \boldsymbol{D}\right].$$

- Outside of Design 1, interpretation of δ_ℓ more complicated:
  - E.g.: in Design 2 ($D_{g,t} = 1\{E_g \ge t \ge F_g\}$), for all g such that $F_g - 1 + \ell > E_g$,

  $$\delta_{g,\ell} = E\left[Y_{g, F_g - 1 + \ell}(\mathbf{0}_{F_g - 1}, \mathbf{1}_{E_g - F_g + 1}, \mathbf{0}_{F_g - 1 + \ell - E_g}) - Y_{g, F_g - 1 + \ell}(\mathbf{0}_{F_g - 1 + \ell}) | \boldsymbol{D}\right],$$

  effect of having been treated for $E_g - F_g + 1$ periods, $F_g - 1 + \ell - E_g$ periods before outcome is measured: number and recency of treatment periods generating $\delta_{g,\ell}$ vary across groups, complicating interpretation of δ_ℓ.

- δ_ℓ can be interpreted as average effect of having been exposed to a weakly higher treatment for ℓ periods, but very "reduced-form".

Comparison with existing estimators

- When groups all initially untreated, DID_ℓ = estimator obtained by redefining treatment as an indicator equal to one if group g has ever received a non-zero treatment at t, and computing ES estimators of CS (2021).

- When groups' period-one treatment varies, redefining treatment as $1\{$group g's treatment has ever changed at $t\}$ and computing the estimators of CS, as was for instance done in East et al (AER, 2023), does not yield estimators numerically equivalent to DID_ℓ.

- DID_ℓ only compares switchers and non-switchers with same period-one treatment, whereas "naively-extended" CS estimator compares switchers and non-switchers with different period-one treatments.

- Accordingly, naively-extended CS relies on unconditional parallel trends assumption for status-quo outcome.

- Stronger than Assumption 2.

- When combined with standard parallel-trends on groups' never-treated outcome, implies that lagged treatments cannot affect the outcome! Very strong.

Normalized AVSQ effects

- Let

$$\delta_{g,\ell}^D = \sum_{k=0}^{\ell-1} (D_{g,F_g+k} - D_{g,1}) \tag{5}$$

  be difference between total treatment dose received by group g from F_g to $F_g - 1 + \ell$, and total dose would have received in status-quo.

- Then, let

$$\delta_{g,\ell}^n = \frac{\delta_{g,\ell}}{\delta_{g,\ell}^D}. \tag{6}$$

  $\delta_{g,\ell}^n$: normalized AVSQ effect of group g at $F_g - 1 + \ell$.

- $\mathrm{DID}_{g,\ell}/\delta_{g,\ell}^D$ is conditionally unbiased for $\delta_{g,\ell}^n$.

$\delta_{g,\ell}^n$: average of effects of current treat and $\ell - 1$ first lags

- For $k \in \{0, ..., \ell - 1\}$, let

$$s_{g,\ell,k} = E\left[Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, \underline{D_{g,F_g-1+\ell-k}}, \boldsymbol{D}_{g,1,k})\right.$$
$$\left. - Y_{g,F_g-1+\ell}(\boldsymbol{D}_{g,1,F_g-1}, D_{g,F_g}, ..., D_{g,F_g-1+\ell-k-1}, \underline{D_{g,1}}, \boldsymbol{D}_{g,1,k})|\boldsymbol{D}\right]$$
$$/(D_{g,F_g-1+\ell-k} - D_{g,1})$$

be slope of PO of g at $F_g - 1 + \ell$ wrt its kth treatment lag (the underlined term), when that lag is switched from $D_{g,1}$ to actual value $D_{g,F_g-1+\ell-k}$.

- For any $k \in \{0, ..., \ell - 1\}$, let

$$w_{g,\ell,k} = \frac{D_{g,F_g-1+\ell-k} - D_{g,1}}{\delta_{g,\ell}^D}.$$

Under Restriction 2, $w_{g,\ell,k} \geq 0$. Moreover, $\sum_{k=0}^{\ell-1} w_{g,\ell,k} = 1$.

Lemma 2

$\delta_{g,\ell}^n = \sum_{k=0}^{\ell-1} w_{g,\ell,k} s_{g,\ell,k}$.

Normalized event-study effects

- Let $\delta_\ell^D = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} |\delta_{g,\ell}^D|$ and

$$\delta_\ell^n = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} \frac{|\delta_{g,\ell}^D|}{\delta_\ell^D} \delta_{g,\ell}^n. \tag{7}$$

- We have:

$$\delta_\ell^n = \frac{\delta_\ell}{\delta_\ell^D}. \tag{8}$$

- One can show that

$$\text{DID}_\ell^n := \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} \frac{|\delta_{g,\ell}^D|}{\delta_\ell^D} \frac{\text{DID}_{g,\ell}}{\delta_{g,\ell}^D} \tag{9}$$

is conditionally unbiased for δ_ℓ^n.

- did_multiplegt_dyn computes normalized effects, if normalized option specified.

Interpretation of δ_ℓ^n

- Follows from Lemma 2 that δ_ℓ^n is weighted average of effects of groups' current and $\ell - 1$ first treatment lags on outcome.

- The total weight assigned by δ_ℓ^n to the effect of the kth-lag (for $0 \leq k \leq \ell - 1$) is equal to

$$w_{\ell,k} = \frac{1}{N_\ell} \sum_{g:F_g-1+\ell \leq T_g} \frac{|D_{g,F_g-1+\ell-k} - D_{g,1}|}{\delta_\ell^D}.$$

  In one-shot treatment designs, $w_{\ell,\ell-1} = 1$.

- When groups' treatment can only change once, $w_{\ell,k} = 1/\ell$, so $\ell \mapsto w_{\ell,k}$ is decreasing for $\ell \geq k + 1$: δ_ℓ^n assigns less weight to the effect of recent treatments when ℓ increases.

- One always has $w_{1,0} = 1$ and $w_{\ell,0} \leq 1$ for $\ell \geq 2$: δ_1^n only averages effects of groups' current treatment, δ_ℓ^n also averages effects of lags for $\ell \geq 2$.

- We recommend reporting $k \mapsto w_{\ell,k}$: document which lags contribute most to δ_ℓ^n.

Extensions (Stata implementation)

- Cost benefit ratio, average total effect per unit of treatment.
- Inference, conditional on design.
- Placebo estimators to test parallel trends/no anticipation (placebo)
- Continuous treatment at period one (continuous, coming up soon, command can already be tweaked to get non-normalized effects in that case).
- Controlling for covariates (controls).
- Group-specific linear trends (trends_lin, coming up soon).
- Supergroup-specific trends (trends_nonparam).
- Estimating heterogeneous effects (hetx, coming up soon).

Outline

Favara and Imbs, 2015

- In 1994, the Interstate Banking and Branching Efficiency Act (IBBEA) allowed US Banks to operate across state borders without formal authorization from state authorities.

- Initially, all states still imposed 4 types of restrictions, but several states deregulated later, by lifting heterogeneous number of restrictions at heterogeneous times.

- Treatment: number of restrictions lifted, ranging from 0 to 4.

- Use 1994-to-2005 county-level data to estimate the effect of the number of regulations lifted on the growth of mortgages originated by banks, and on the growth of houses prices.

- Local projection regressions. To estimate effect on growth rate of loan volume, regress, for every $\ell \in \{1, ..., 9\}$, $\Delta \ln(L_{g,t-1+\ell})$, the log growth rate of loans in county g in year $t - 1 + \ell$, on county and year FEs, and $D_{g,t}$ (number of deregulations in county g and year t).

Stata implementation of our estimators

```
ssc install did_multiplegt_dyn

net get did_multiplegt_dyn

use favara_imbs_did_multiplegt_dyn.dta, clear
```
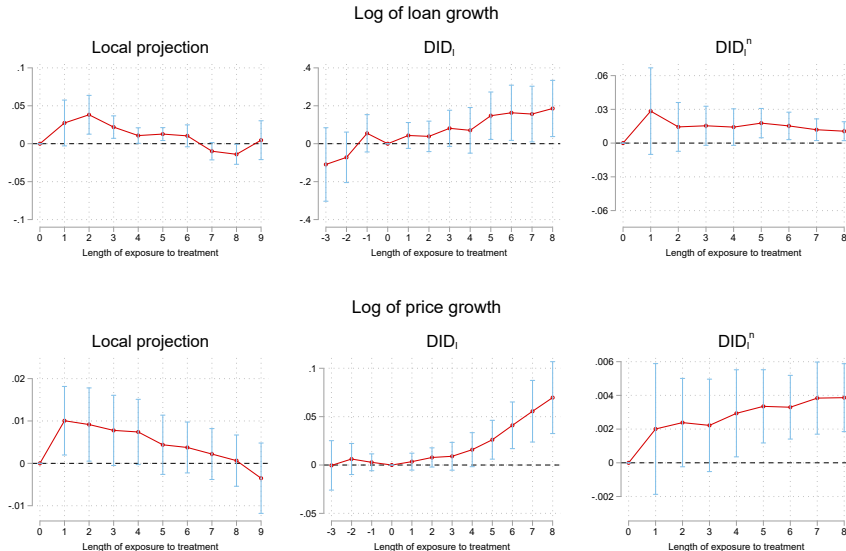
Non-normalized event-study estimators:

```
did_multiplegt_dyn Dl_vloans_b county year inter_bra, effects(8)
placebo(3) cluster(state_n)
```

Normalized event-study estimators:

```
did_multiplegt_dyn Dl_vloans_b county year inter_bra, effects(8)
placebo(3) cluster(state_n) normalized same_switchers
effects_equal
```

A split second later (thanks Mélitine and Doulo!):

Log of loan growth

Local projection · DID_l · DID_l^n

Log of price growth

Local projection · DID_l · DID_l^n

Decompositions of LP regressions

- We use one of our results to decompose LP regression coefficients.

- $\widehat{\beta}_{lp,1}$ is a weighted sum of 7,626 effects $\delta_{g,k}/\bar{D}_{g,k}$, where 4,670 effects are weighted positively and 2,956 effects are weighted negatively, and where positive and negative weights respectively sum to 1.067 and -0.125.

- $\widehat{\beta}_{lp,1}$, supposed to measure effect of one year of exposure, contaminated by effects of other exposure lengths: weights on effects of one year of exposure sum to 0.294, weights on other exposure lengths sum to 0.648.

- $\widehat{\beta}_{lp,2}$ is a weighted sum of 7,626 effects, where 4,424 effects are weighted positively and 3,202 effects are weighted negatively, and where positive and negative weights sum to 1.085 and -0.584. Weights sum to much less than one. Then, even if $\delta_{g,k}/\bar{D}_{g,k}$ does not vary across g or k, $\widehat{\beta}_{lp,2}$ severely biased towards zero.

- $\widehat{\beta}_{lp,2}$ contaminated by effects of other lengths of exposure than two years. Most effects weighted positively, except for effects of one year of exposure.

- Intuitively, groups with $D_{g,t} = 0, D_{g,t+1} > 0$ used as "control groups" by $\widehat{\beta}_{lp,2}$ ($Y_{g,t+1}$ on $D_{g,t}$), whereas treated at $t + 1$.

- Results similar for $\widehat{\beta}_{lp,3}$, except that weights only sum to 0.069.

- For $\widehat{\beta}_{lp,4}$, weights sum to -0.018. Even if $\delta_{g,k}/\bar{D}_{g,k}$ constant across g and k, $E\left[\widehat{\beta}_{lp,4}\right]$ of a different sign than the treatment effect.

Thank you!