Difference in Difference

John Snow in the times of cholera (circa 1850)

- Sometimes it helps to back in time to the earliest application of a method in its simplest form, so we'll talk about John Snow.
- Snow was interested in understanding if cholera was transmitted by water rather than air
- One district in London had changed water source. After the change, cholera deaths decreased, which would suggest that water was the source of infection
- Let's call the houses that changed water source the "treated" group
- One form of evidence would be to do a before and after comparison:
- $E[Y_i|D_i = 1, T_i = 1] - E[Y_i|D_i = 1, T_i = 0]$, where T is 1 if after the change of supplier, Y is an indicator for death and D is our treatment indicator
- The key question, of course, is whether
- $E[Y_i|D_i = 1, T_i = 0] = E[Y_{0i}|D_i = 1, T_i = 1]$
- Is the observed outcome before the change ($T = 0$) a good prediction of the counterfactual after the change ($T = 1$)? Dubious

John Snow in 1850s and cholera

- Remember, if the same people are measured before and after, they are different units in our causal inference framework
- Intuitively, before and after comparisons are valid if nothing else changed at the same time as the treatment; that is, now we need to assume that no other factor X is correlated with T (contemporaneous factors, trends)
- In Snow's case a simple before and after comparison did not solve the problem because the suspicion was that air was a source of contamination.
- What if air changed at the same time as the water supply was changed? How could he "hold air constant"?
- He came up with a clever solution: use control areas and hold air constant by using as controls places where water supply did not change with the catch that these control houses shared the same air
- Snow actually used a salt test to verify water source.

Comparisons

- Let's simplify the notation using only realized outcomes for now: A pre and post comparison of outcome Y for the treated is: $E[Y_{tpost}] - E[Y_{tpre}]$
- We want to compare that difference with the difference in the control:
- $E[Y_{cpost}] - E[Y_{cpre}]$
- So, the estimate of interest is:
- $\Delta DiD = E[Y_{tpost}] - E[Y_{tpre}] - \{[Y_{cpost}] - E[Y_{cpre}]\}$
- If that difference of differences is zero, that would suggest that water is not causing infections
- If that difference is not zero, then there is some evidence that air is not the source of transmission – it doesn't prove that water causes cholera – there could still be other factor, such as physical contact, that could explain transmission

1. DiD方法的本质是排除干扰而非唯一归因。如果该地区在改变水源的同时采取了某种防治政策，那么就无法说明是水源问题导致的霍乱。

- The estimator is a difference-in-differences, hence the name. We need four expected values (four "cells")

Differencing within each group over time

- The Snow example provides some intuition on why this approach works but we need to elaborate to make it clearer
- Comparing before and after the houses (**within** the treated group) that changed water supplier helps us with things that that did not change over time
- For example, cleanliness in each house or genetic factors or the role of sex and race
- All those fixed or constant factors could not explain the before and after change in the outcome since these factors did not change before and after
- Same within the control group

Differencing between groups

- The same logic applies to comparisons between groups
- A comparison between groups is valid if one group is a good conterfactual of the other

1. counterfactual：反事实的

- But we can relax that assumption and argue that we do not need to assume that houses that changed water supplier are the same as those who did not
- We could assume instead that they are in fact different, but the factors that affect cholera mortality between the groups did not change when we do the before and after comparison
- For Snow, the key factor was air. We can think of other important factors in these COVID times: one group was more prone to parties than the other
- As long as we can argue that “party proneness” was the same before and after, we are fine

Two imperfect solutions make a good one

- Note that differences-in-differences is one interesting case in which two imperfect solutions combined make something useful
- A before and after analysis of outcomes is not ideal since it could be hard to argue that nothing else changed
- A comparison between groups that are not comparable is not ideal because other factors could account for differences between the groups
- But when we combine them, a difference-in-differences estimator may give you something right provided the other factors are additive and constant
- It's not a mild assumption, but it can be relaxed by controlling for other covariates (more in a bit)

Key elements

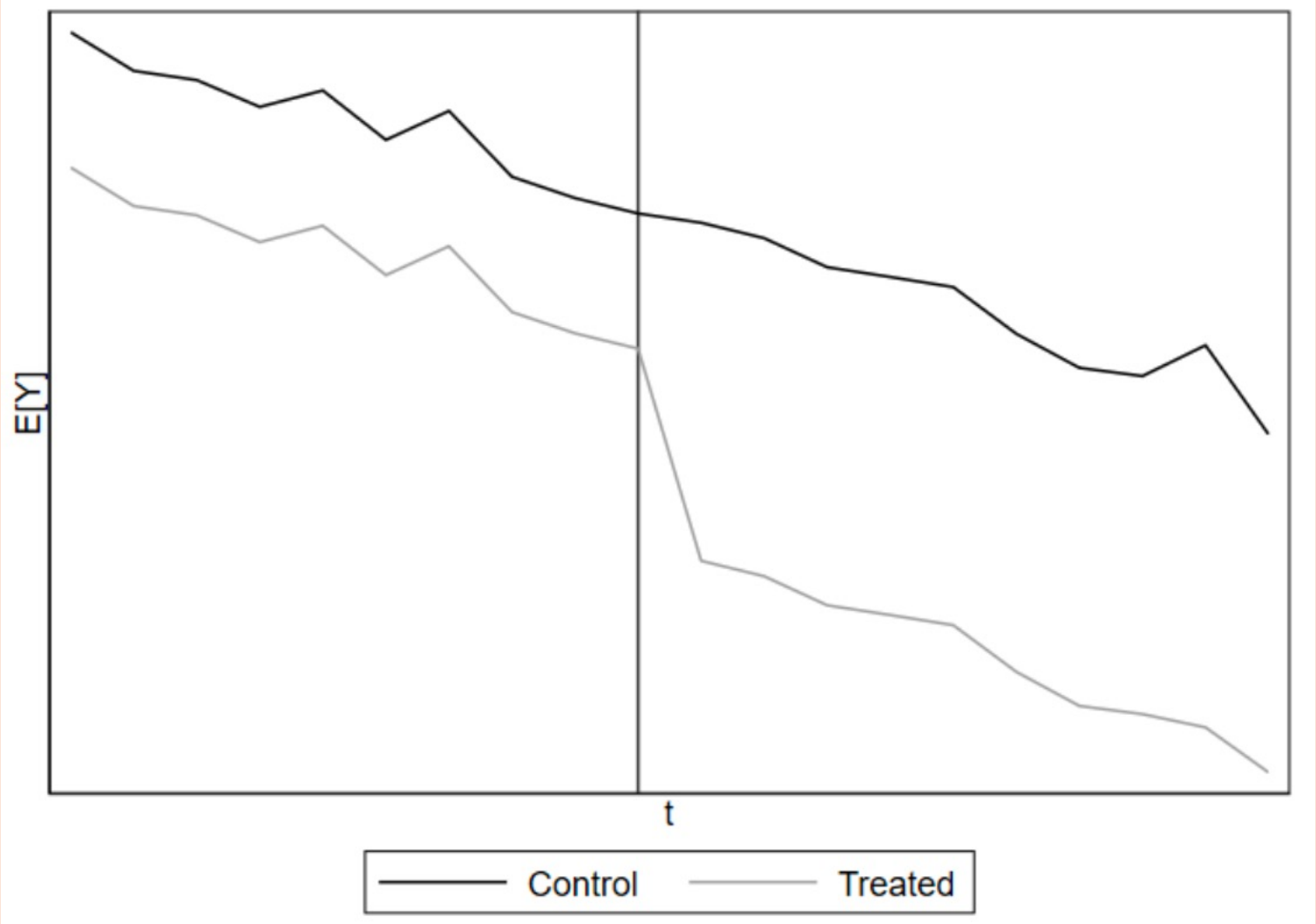
- In the examples that follow the key elements are:

  1. **Time**. In DiD we always have a time: a before and after period. So far, we have talked about only two periods, but we could have multiple periods before and after (more information)
  2. **Comparison groups**: In DiD, one group receives the intervention or is subjected to the policy change only in the post-period. These groups do not need to be comparable
  3. **Fixed factors**: We assume that important factors that explain the outcome Y are fixed during the pre and post periods. If observed, we can control for those factors that could affect trends

- Trends are key in DiD

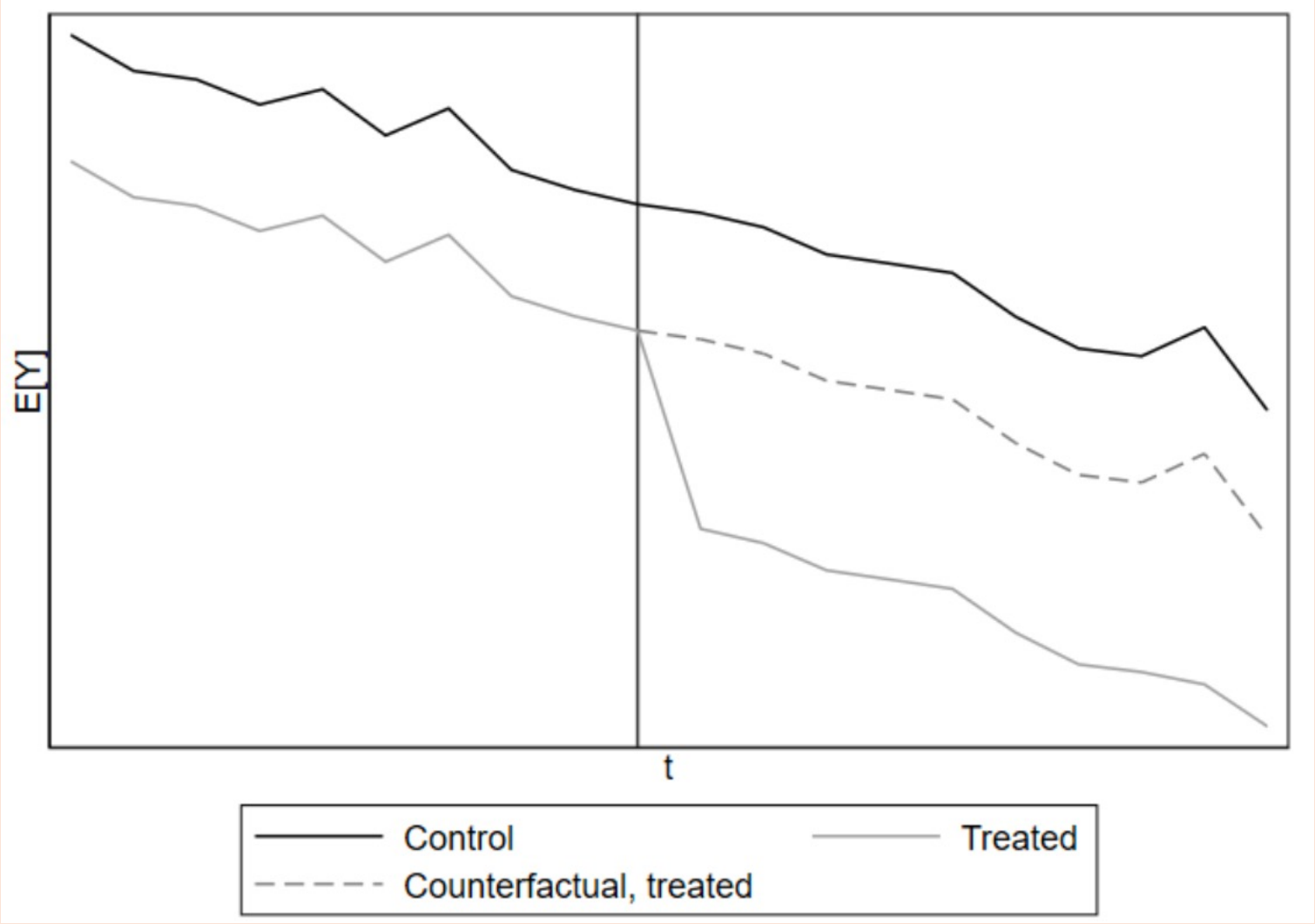
More periods

- This would be in ideal DiD scenario with constant (parallel) trends before the intervention at the time marked by the vertical line (I simulated data)



More periods

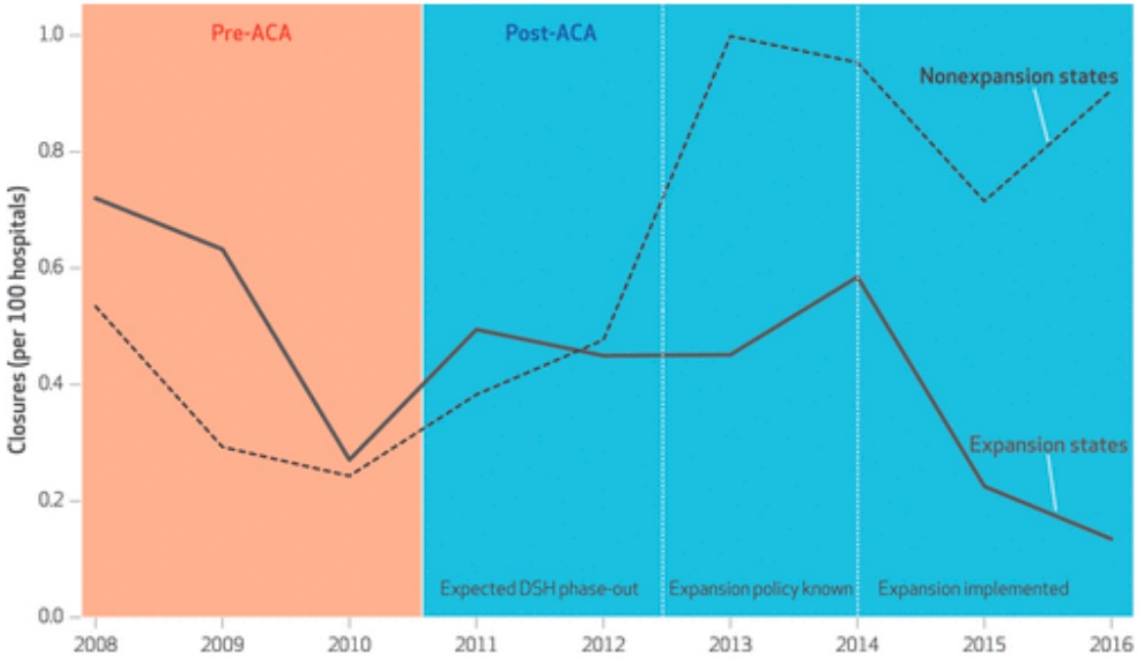
- The counterfactual for the treated is the dashed line: trends would have remained constant over time with no treatment, which is the underlying key assumption



Actual example

- Medicaid expansion and hospital closures (Lindrooth et al., 2018)

Exhibit 2 Unadjusted hospital closure rates by state Medicaid expansion status, 2008–16



SOURCE Authors' analysis of data from the Centers for Medicare and Medicaid Services. NOTE Closures were independently validated from multiple sources.

More formal

- We can formalize the previous discussion by assuming that the outcome follows this structure:
- $Y_{it} = c_i + d_t + \delta D_{it} + \eta_{it}$

4. D_{it}**是个体**i在时间t是否受到处理（0-1二元变量）

- This is a causal structural assumption. c are d are variables, not coefficients. η_{it} is an unexplained random cause of variation, where i indexes the unit of observation and t indexes time
- The outcome depends on constant (fixed, time-invariant) factors at the unit of observation level (c_i) and factors that depend on time (d_t) but not on unit of observation *i*

1. confounding variable：混淆变量，指同时与treatment**和**outcome相关的变量。
比如，吸烟（treatment）**肺癌**（outcome），年龄大的人易吸烟，**也易患肺癌**，年龄就是cv

- Think of c and d as confounding variables with a coefficient of 1. They are unobserved effects. We could put a coefficient next to them, but we won't estimate them. We could add more of both, say: $\sum_{j=1}^{w} c_{ij}$ and $\sum_{k=1}^{m} d_{kt}$ so think of c and d representing more than one factor

2. groups分为处理组（水源改变）和对照组（水源不改变）

- δ is the difference between groups and is constant, so we assume homogeneous treatment effects

3. 处理组在干预后的结果变化delta对所有组中个体一致。

Potential outcomes

- Since $Y_{it} = c_i + d_t + \delta D_{it} + \eta_{it}$ represents a causal relationship, they also define potential outcomes:

1. 通过D_{it}**分别取**0，1来得到个体在不同处理状态下的潜在结果。只有一种是被观测到的现实结果，另一种是反事实结果。

- $Y_{1it} = c_i + d_t + \delta + \eta_{it}$
- $Y_{0it} = c_i + d_t + \eta_{it}$
- So $Y_{1it} - Y_{0it} = \delta_{it}$. Note that δ has an index because we could define the difference between potential outcomes at different times
- With two periods, $t \in \{0,1\}$. $t = 0$ is before the intervention (Notation here gets messy with potential outcomes. This notation for potential outcomes would be better: $Y(1)_{it}$ if treated, $Y(0)_{it}$ if not treated)

2. **注意符号的表示方法**

Differencing

- We could do this with potential outcomes or realized outcomes. I'll do it with realized outcomes

(1) Treated group after and before:

$$E[Y_{i1}|D_i = 1] - E[Y_{i0}|D_i = 1] = c_i + d_1 + \delta_1 - (c_i + d_0 + \delta_0)$$
$$= d_1 - d_0 + (\delta_1 - \delta_0)$$
$$= d_1 - d_0 + \delta$$

1. contemporaneous：同时发生的

- The above shows the problem of a simple before and after comparison: it doesn't get rid of contemporaneous factors that depend on time (trends) and would affect an estimate of treatment effects on the treated, measured by the variable d: $d_1 - d_0$
- That could be a measure of air in both periods that changed between t=0 and t=1

(2) Control group after and before:

$$E[Y_{i\,1}|D_i = 0] - E[Y_{i\,0}|D_i = 0] = c_i + d_1 - (c_i + d_0) = d_1 - d_0$$

- The difference of the differences (1)-(2)= $d_1 - d_0 + \delta - (d_1 - d_0) = \delta$
- The **within group** differencing got rid of factor c, the **between group** differencing got rid of d. It would be the same if we had added more factors c's and d's

Notes

- We assumed $E[\eta_{it} | D_{it} =] = E[\eta_{it}]$, that is mean independence
- That assumption could be relaxed by saying that the unobserved error component is mean independent respect to the change before and after ($E[\Delta\eta_i | D_{it}] = E[\Delta\eta_i]$), which is what you would get using "fixed effects" in longitudinal data

1. 理论合理性证明，核心是验证假设（如平行趋势、无干扰等）**是否成立。**

- Separating the justification of the difference-in-difference estimator from estimation is more helpful. We could estimate the four expected values separately, following the logic of estimating treatment effects in previous lectures (we will see a version that combines propensity score with kernel weighting; Heckman, Ichimura, Todd, 1998)
- The constant trend is important, and one that we can't evaluate if we don't have more observations before the intervention took place
- The two-groups, two-periods example is helpful for the intuition but not for evaluating the assumption of constant trends

What about other factors (covariates)?

1. 差分使我们可以不用考虑不随时间（d_t）**或个体**（c_i）变化的变量

- If we follow the logic of differencing, then we do not need to account for any other observed or unobserved constant (fixed, time-invariant) factors
- But we could take into account factors that vary at the unit of observation and by time (time-varying covariates)
- This means that we can extend our notation to condition for a vector of covariates X_{it}, although we will impose some assumptions when we use regression analysis (exogeneity)
- So now we have: $Y_{it} = c_i + d_t + \delta D_{it} + X'_{it}\,\beta + \eta_{it}$

1. covariates：协变量。是指因果模型推断中，除处理变量外，可能对结果产生影响的变量。

2. 协变量中，那些不随时间（d_t）**或个体**（c_i）变化的变量可以通过差分differencing除掉。但随时间和个体都在变化的时变变量time-varying covariates就要通过加入模型中来控制。

The **Stable Unit Treatment Value Assumption (SUTVA)** is a key assumption in causal inference, particularly in the Rubin Causal Model (RCM). It consists of two main components:

1.No Interference: The treatment assigned to one unit does not affect the outcomes of other units. This means that the potential outcome of a given individual depends only on their own treatment status, not on the treatments of others.

2.Consistency of Treatment: The treatment is well-defined, meaning that there are no different versions of the treatment that could lead to different potential outcomes. Formally, if a unit receives treatment $T=t$, then their observed outcome Y is equal to their potential outcome under t, i.e., $Y=Y(t)$.

Why is SUTVA Important?

•Without SUTVA, causal effects become difficult to define because a unit's outcome could depend on the treatment assignments of many other units, leading to a combinatorial explosion of potential outcomes.

•Violations of SUTVA often occur in settings with spillover effects (e.g., social networks, vaccines, or policies where one person's treatment affects others).

Example of a SUTVA Violation:

•If studying the effect of a vaccine, interference could occur if vaccinated individuals reduce disease transmission to unvaccinated individuals, meaning the outcome of an unvaccinated person depends on others' treatment status.

•In such cases, more advanced methods (e.g., partial interference models or network analysis) are needed.

Assumptions

2. 无干扰和处理一致性。无干扰指个体处理状态不影响他人结果，接种疫苗降低未接种疫苗者的感染风险就违反无干扰。处理一致性指treatment的内容对所有人一样。

- We can now state the assumptions following Lechner (2010).

1. 稳定单位处理值假设

1. SUTVA. We always need SUTVA but in DiD it tends to be more relevant. No interference (spillovers) and variation in treatment. Think about this in the context of cholera and Medicaid expansions.

2. Exogeneity: The covariates X are not influenced by the treatment, so we can condition on them.

3. Common trends or constant bias. If the treated had not been treated, both treatment and control groups would have the same trends over time (after controlling for other factors). Constant bias is the same assumption. Treated and control groups are different, but that difference remains constant over time.

- The last assumption could be divided into an assumption about observed parallel trends before the intervention and the idea that “shocks” have a common effect in both groups.
- With two or mode pre-periods, we can test the parallel trend assumption.

Example: Access to healthcare and health outcomes

Does better access to healthcare lead to better health outcomes?

Dataset AED_HEALTHACCESS has data on 1,071 South African children aged 1 to 4years in 54 communities.

In 1993 26 of 54 communities had access to a healthcare clinic.

In 1998 all 54 communities had access to a healthcare clinic.

Outcome y is waz, a weight-for-age z score

Treatment d=1 if have access to a healthcare clinic.

Time t=0 in 1993 (pre-period) and t=1 in 1998 (post-period).

Results: Manual computation

- The following table gives the mean values of waz
  - for the high treated and low treated children
  - before and after the expansion in free healthcare.

	High treated	**Low treated**
Before(1993)	-0.545(n=246)	-0.414(n=325)
After(1998)	0.321(n=212)	-0.069(n=288)
Change over time	0.867	0.345
Difference in differences	0.521	

- High treated: waz increased by 0.867, from -0.545 to 0.321.
- Low treated: waz increased by 0.345, from -0.414 to -0.069.
- DID estimate is 0.867-0.345=0.521.
- This is a very substantial effect.
  - A third of a standard deviation change in waz for this sample.

Results: Regression computation

- Again, greater access to health clinics increased waz by 0.521
- Since the treatment was at the community level, use cluster-robust standard errors with clustering on community
  - The standard error is 0.236 whereas heteroskedastic-robust s.e. is 0.194.

```
. * Diff-in-diff - no controls and cluster-robust standard errors
. reg waz postXhigh post hightreat, vce(cluster idcommunity) noheader
                          (Std. err. adjusted for 54 clusters in idcommunity)
```

waz	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
postXhigh	.5216188	.2352991	2.22	0.031	.0496685	.993569
post	.3450874	.1371018	2.52	0.015	.070096	.6200788
hightreat	-.1310593	.1968084	-0.67	0.508	-.525807	.2636884
_cons	-.4141846	.1151423	-3.60	0.001	-.6451308	-.1832384

Further analysis

- A richer and better model
  - Controls for community by adding fixed effects for each community
  - Controls for each individual by adding regressors such as parental education and household income
- For child i in community c
  - $y_{ic} = \beta_1 + \beta_2 t_i + \beta_3 d_i + \beta_4 t_i\, d_i + \gamma_c + \beta_5 x_{ic} + \ldots + u_i$.

```
. * D in D with fixed effects for community and individual controls
. reg waz postXhigh post hightreat i.idcommunity ///
> fedu medu hhsizep lntotminc immuniz nonclinic, ///
> vce(cluster idcommunity) noheader
note: 242.idcommunity omitted because of collinearity.
                          (Std. err. adjusted for 54 clusters in idcommunity)
```

waz	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
postXhigh	.6428807	.2710993	2.37	0.021	.0991243	1.186637
post	-.6807024	.3487963	-1.95	0.056	-1.380299	.0188944
hightreat	-.2911247	.2360665	-1.23	0.223	-.7646142	.1823648

Stata didregress command

- Stata didregress command defines the treatment variable to be d_{it} =1or d_{it} =0
  - ➢ This is $d_i \times t_i$ in the previous notation
  - ➢ i.e. postXhigh in the current example (not hightreat)
- With group and time effects and control variables we give command

```
. didregress (waz fedu medu hhsizep lntotminc immuniz nonclinic) (postXhigh), ///
> group(idcommunity) time(post)

Treatment and time information

Time variable: post
Control:       postXhigh = 0
Treatment:     postXhigh = 1
```

	Control	Treatment
Group		
idcommunity	29	25
Time		
Minimum	0	1
Maximum	0	1

Further Details

- We get the same ATET and standard error as the earlier regress command.

Difference-in-differences regression — Number of obs = **1,071**
Data type: Repeated cross-sectional

(Std. err. adjusted for **54** clusters in **idcommunity**)

waz	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
ATET						
postXhigh (1 vs 0)	**.6428807**	**.2710993**	**2.37**	**0.021**	**.0991243**	**1.186637**

Note: ATET estimate adjusted for covariates, group effects, and time effects.

Testing parallel trends

- We will see two ways of testing parallel trends. We need at least two pre-period time points:

  1. Using the pre-period data only: test if changes in trends are the same in both groups.

  2. Placebo tests: using lags and leads, past treatment predicts future outcome, but future treatment should not predict present changes in the outcome.

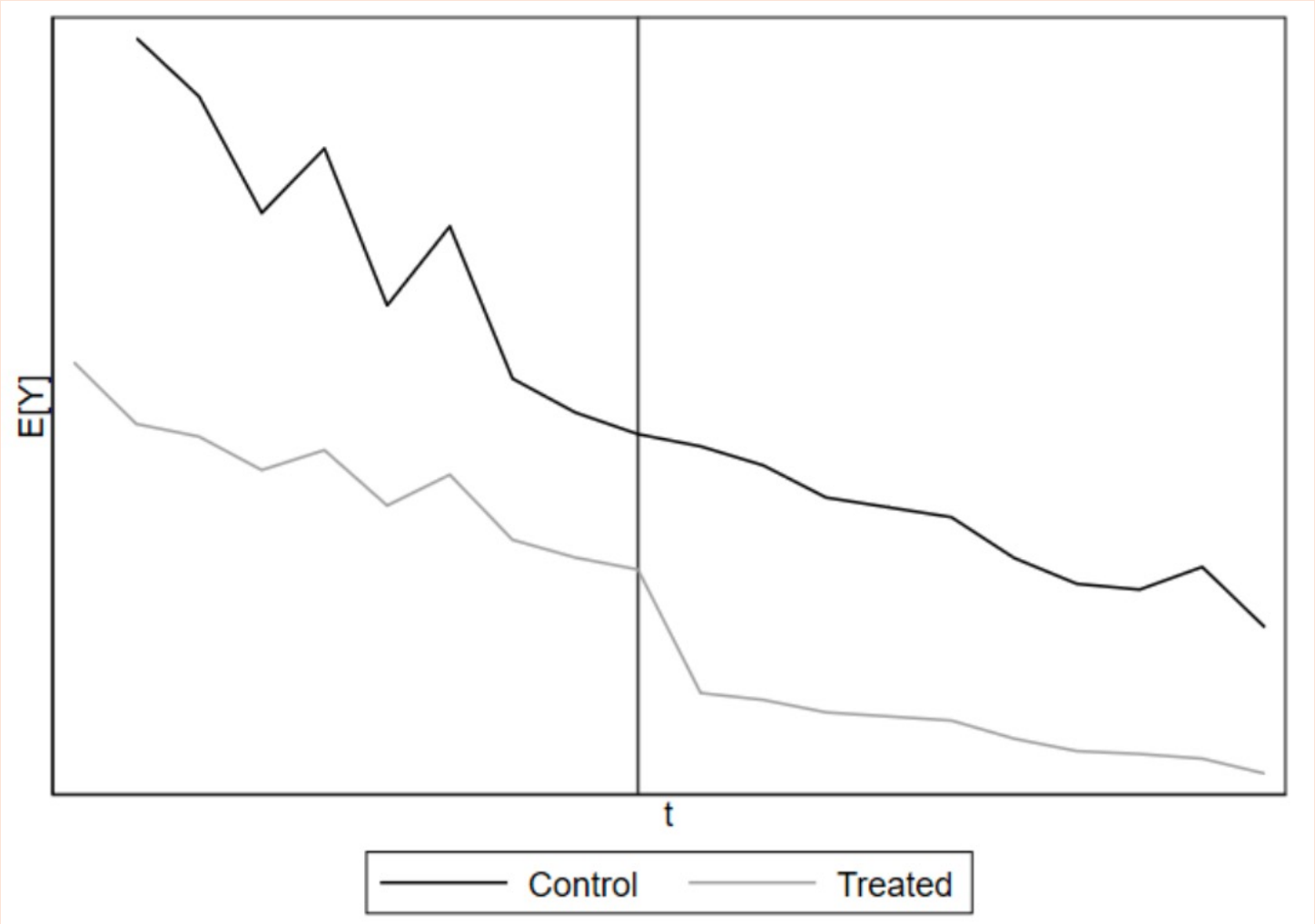
- The first approach is much more intuitive. With two pre-periods only, say, $t = -1$ and $t = 0$ the test is whether the change $E[Y_{t=0} - Y_{t=-1}]$ is the same in the treatment and control groups. The relevant test is an interaction, so it follows the parametric structure of DiD models

- The second one is less intuitive and it estimates more parameters. We need to use lag and lead variables for the treatment indicator.

- One aspect of the parallel trend assumption is annoying: if it holds in Y then it doesn't hold in retransformations like $\log(y)$ (a monotonic transformation)

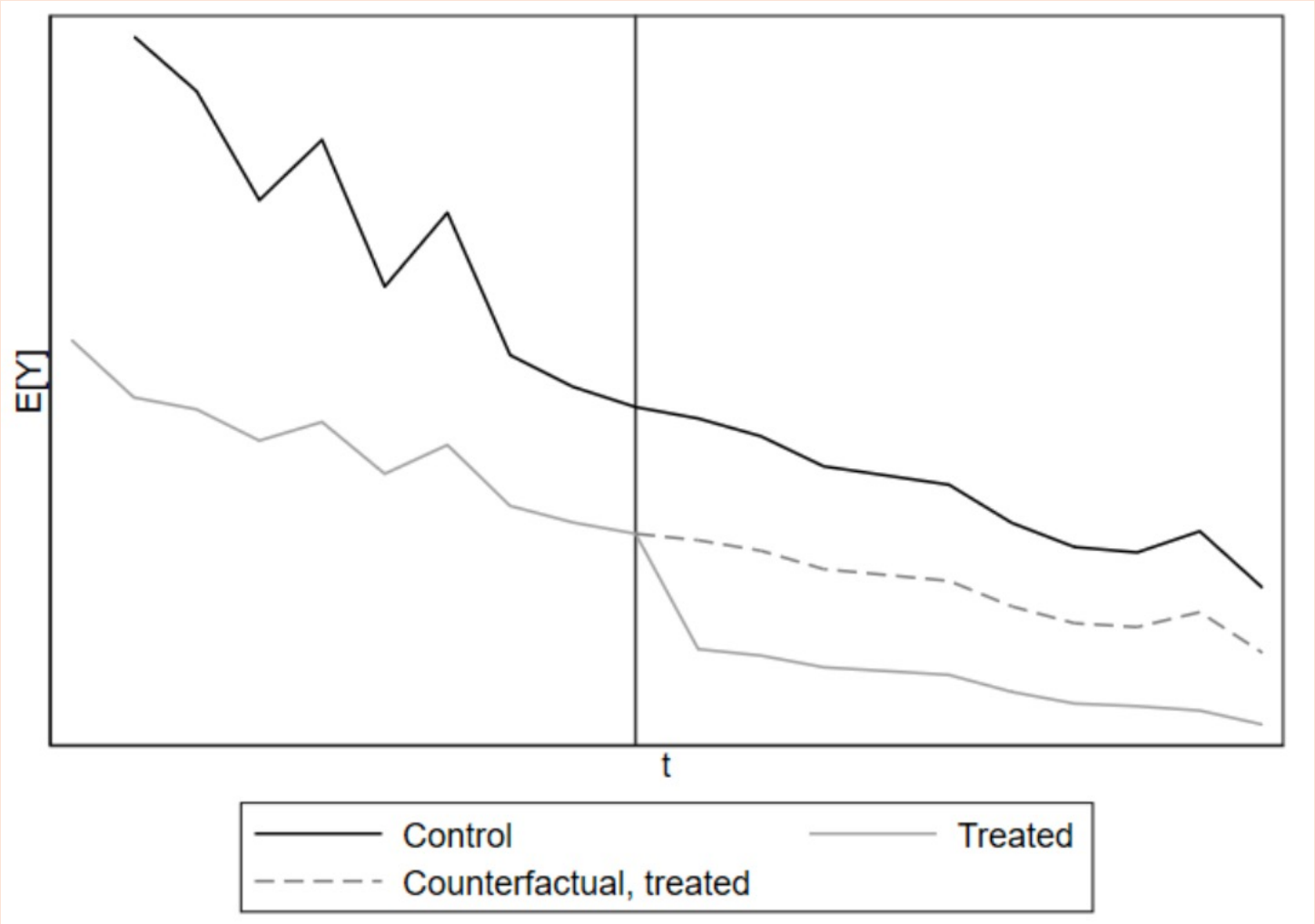
Yikes?

- The perfect example of parallel trends doesn't hold in the log scale. I simply took the log of Y in the simulated data



Yikes? II

- Look at the counterfactual



Parametric regression

1. **重点。**DiD的计量模型！！！

- We can estimate models like:
- $Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 P_t + \beta_3 (D_i \times P_t) + \epsilon_{it}$
- $Y_{it} = \beta_0 + \beta_1 D_i + \beta_2 P_t + \beta_3 (D_i \times P_t) + \mathbf{X}'_{it}\,\beta + \epsilon_{it}$
- The unit of analysis i could be a person or state. P is an indicator that equals 1 after the change. Note the index *it* carefully. Treatment D doesn't depend on time (a person i is treated or control in both periods), while P depends on time but it's the same by person.
- The first model is a saturated model; we will get four predicted means.
- We are going to come back to the structure of the data, which is reflected in the index, because it's important for inference and can be confusing.

Person id	D	P	Y
101	1	0	100
101	1	1	120
104	0	0	90
104	0	1	92
.	.	.	.
.	.	.	.

Model interpretation

- Why is that model a difference-in-difference? (Using the version without covariates; with covariates we need to hold them constant)

  E [Y] Treated in post period: $\beta_0 + \beta_1 + \beta_2 + \beta_3$

  E [Y] Treated in pre period: $\beta_0 + \beta_1$

  (1) Difference treated post - pre: $\beta_2 + \beta_3$

  E [Y] Control in post period: $\beta_0 + \beta_2$

  E [Y] Control in pre period: β_0

  (2) Difference control post - pre: β_2

  Difference of differences (1)-(2): $\Delta_{DiD} = \beta_3$

- **Caution**: Interacted models are not difference-in-differences **research designs**, but interactions with dummy variables are difference-in-differences

This is not a DiD research design

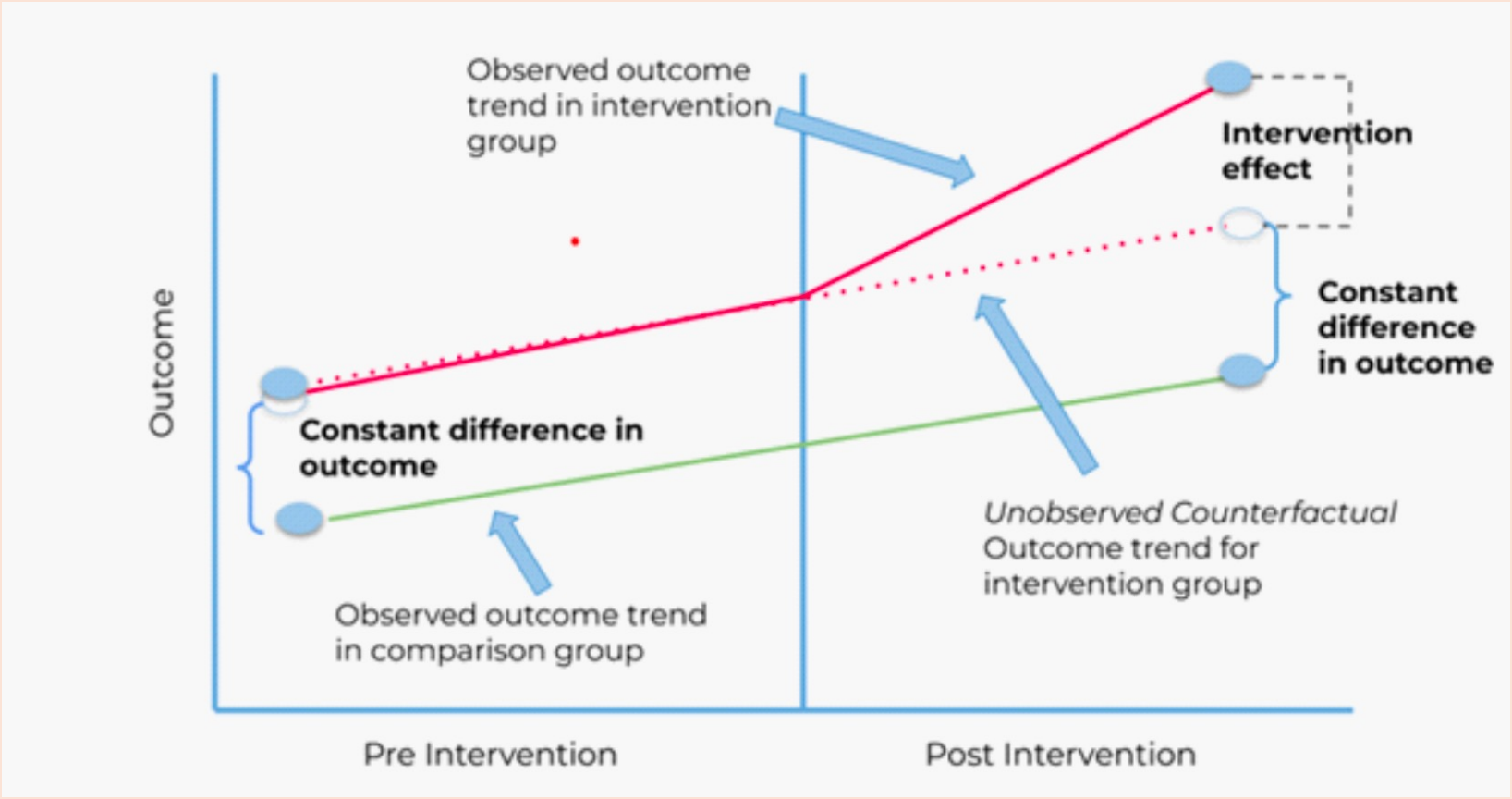
- Suppose you are interested in evaluating the effects of a new preventive care benefit offered at no cost to patients (e.g. Medicare's annual wellness visit, AWV) on race disparities
- The outcome is a measure of preventive care. The AWV became available in 2011. For simplicity, let's say that you want to compare Whites vs other races. You could estimate the following model:

$$PC_{it} = \beta_0 + \beta 1 White_i + \beta_2 Post2011_t + \beta_3 Post2011_t * White_i + \eta_{it}$$

- β_3 is a difference of differences. If not zero, then disparities after the AWV are different that disparities before
- But this is a before and after analysis. There is no comparison group; all races were exposed to the treatment –the AWV. The outcome is a comparison. It just happens that the parametric model structure is the same as a simple DiD design

Parallel Trends Assumption for Causality

- In order for differences in differences to have a causal interpretation we need to assume that the change over time in the outcome, in the absence of treatment, is the same for treated and untreated groups.



Parallel Trends Assumption (continued)

- Notation has three components
  - time: t = pre or post
  - treatment: d=1 if treated and d=0 if not treated
  - Potential outcome: Y(1) if treated and Y(0) if not treated
    - For each person we can only observe one of Y(1) or Y(0).
- Define
  - $Y_{post}(1)$ = post-treatment Y if treated
  - $Y_{post}(0)$ = post-treatment Y if not treated.
- We want the average treatment effect on the treated = expected outcome if treated - expected outcome if not treated, for those who are treated
  $=E[Y_{post}(1)|d=1] - E[Y_{post}(0)|d=1]$

  1. d代表组别；括号里的数字代表是否经过处理。post period**二者一般相等，**所以不等时就是反事实。
  - but $Y_{post}(0)|d=1$ is not observed.

Parallel Trends Assumption (continued)

- The difference in difference estimate is
  - $\hat{\gamma} = \Delta\bar{y}$ for the treated $-\Delta\bar{y}$ for the not treated.
  - $\hat{\gamma} = \left(\bar{y}_{post,d=1} - \bar{y}_{pre,d=1}\right) - \left(\bar{y}_{post,d=0} - \bar{y}_{pre,d=0}\right)$
- This is an estimate of

$$\gamma = \left\{E\left[Y_{post}(1)|d=1\right] - E\left[Y_{pre}(0)|d=1\right]\right\}$$
$$-\left\{E\left[Y_{post}(0)|d=0\right] - E\left[Y_{pre}(0)|d=0\right]\right\}$$

  - Since we observe Y(0) in all cases except we observe Y(1) for the treated in the post period.

Parallel Trends Assumption (continued)

- Add and subtract the unobserved $E\left[Y_{post}(0)|d=1\right]$

$$\gamma = \left\{E\left[Y_{post}(1)|d=1\right] - E\left[Y_{pre}(0)|d=1\right]\right\}$$
$$-\left\{E\left[Y_{post}(0)|d=0\right] - E\left[Y_{pre}(0)|d=0\right]\right\}$$
$$+E\left[Y_{post}(0)|d=1\right] - E\left[Y_{post}(0)|d=1\right]$$

- Rearrange

$$\gamma = \left\{E\left[Y_{post}(1)|d=1\right] - E\left[Y_{post}(0)|d=1\right]\right\}$$
$$+\left\{E\left[Y_{post}(0)|d=1\right] - E\left[Y_{pre}(0)|d=1\right]\right\}$$
$$-\left\{E\left[Y_{post}(0)|d=0\right] - E\left[Y_{pre}(0)|d=0\right]\right\}$$

- So $\gamma = \left\{E\left[Y_{post}(1)|d=1\right] - E\left[Y_{post}(0)|d=1\right]\right\}$ under the parallel-trends assumption that

$$\left\{E\left[Y_{post}(0)|d=1\right] - E\left[Y_{pre}(0)|d=1\right]\right\}$$
$$= \left\{E\left[Y_{post}(0)|d=0\right] - E\left[Y_{pre}(0)|d=0\right]\right\}$$

Parallel Trends Assumption (continued)

- Note that if the parallel trends assumption holds in level, it will not hold in logs (and vice-versa).
- So, we have to use the appropriate scaling of the outcome.

Differences-in-Differences: Multiple Time Periods

- Consider individual i in state s in year t, and the treatment of interest d_{st} occurs at the state-year level.
- Then we estimate the **two-way fixed effects model**:

$$y_{ist} = \phi_s + \gamma_t + \alpha d_{st} + \beta_1 \times X_{ist} + \cdots + u_{ist}$$

  - where: ϕ_s = state-specific fixed effects; γ_t = time-specific fixed effects
- **Key Assumption: Parallel Trends**
  - The time trend each period is the same for each state.
    - γ_t is the same for each state (rather than γ_{st}).
  - This is partly testable in some applications using pre-treatment data.
- **Inference:** Standard errors are clustered at the **state (s)** level.
  - This leads to the **"few clusters" problem** if there are few clusters.
- OLS estimation is straightforward, but interpretation is difficult if treatment is **staggered** (occurs at different times for different states).
  - This is an area of **current academic research**.

Example: From Stata Documentation

- **Example:**
  - y (outcome) = *satis* (Patient satisfaction score)
  - d (treatment) = *procedure* (=1 if treated)
  - s (group) = hospital (there are 46 hospitals)
  - t (time) = month (7 months: January to July)
  - i = individual
- Treatment begins in **April** for **18 out of the 46 hospitals**.

Stata didregress command: Treatment effect is 0.84789

```
. didregress (satis)(procedure), group(hospital) time(month)

Treatment and time information

Time variable: month
Control:       procedure = 0
Treatment:     procedure = 1
```

	Control	Treatment
Group		
hospital	**28**	**18**
Time		
Minimum	**1**	**4**
Maximum	**1**	**4**

```
Difference-in-differences regression                 Number of obs = 7,368
Data type: Repeated cross-sectional

                             (Std. err. adjusted for 46 clusters in hospital)
```

satis	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
ATET						
procedure (New vs Old)	**.8479879**	**.0321121**	**26.41**	**0.000**	**.7833108**	**.912665**

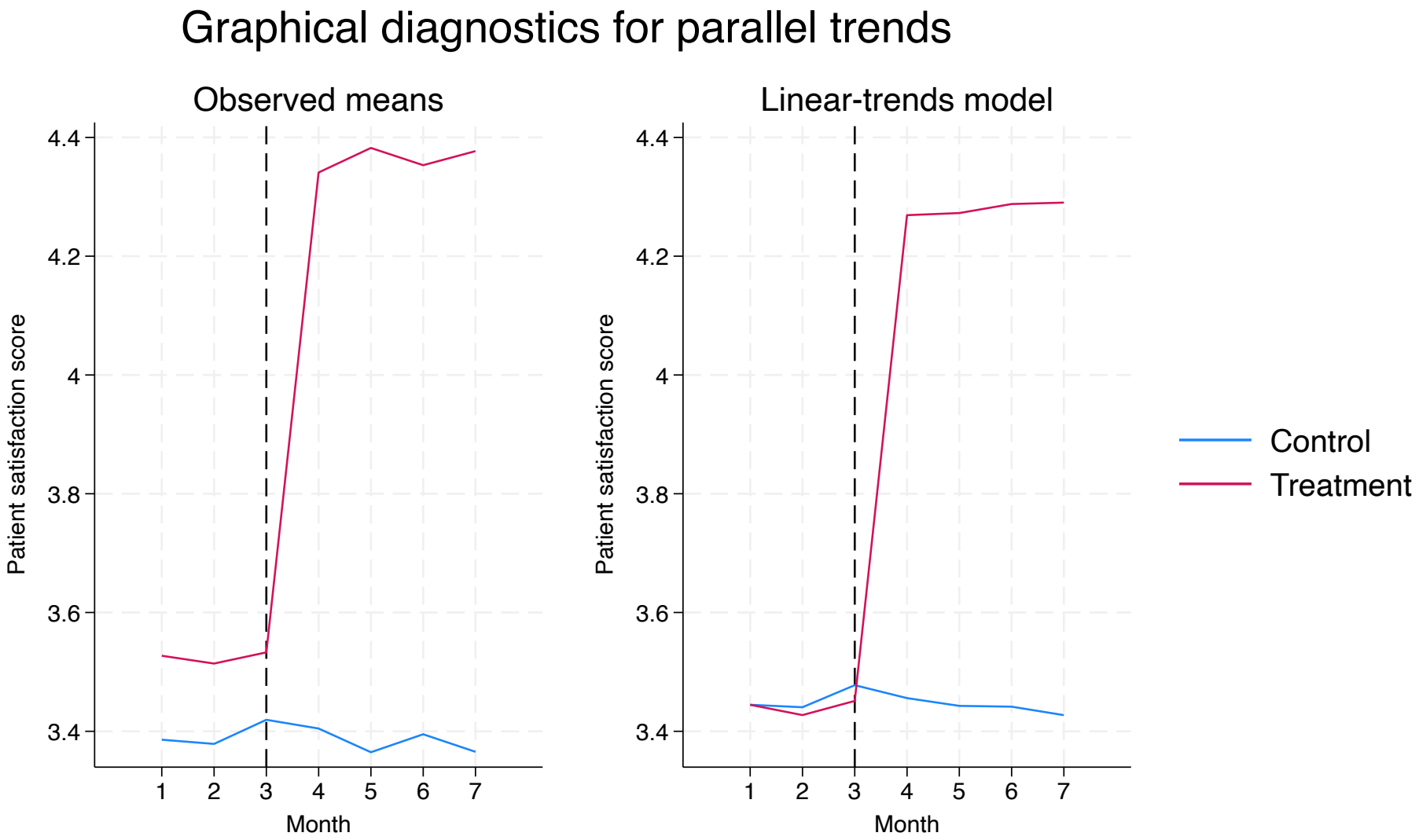
```
Note: ATET estimate adjusted for group effects and time effects.
```

```
. * The following gives the same results as didregress
. regress satis procedure i.hospital i.month, vce(cluster hospital)

Linear regression                               Number of obs     =      7,368
                                                F(6, 45)          =          .
                                                Prob > F          =          .
                                                R-squared         =     0.5333
                                                Root MSE          =     .72384

                                    (Std. err. adjusted for 46 clusters in hospital)
------------------------------------------------------------------------------
             |               Robust
       satis |      Coef.   Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
   procedure |   .8479879   .0321121    26.41   0.000     .7833108    .912665
```

STATA command:

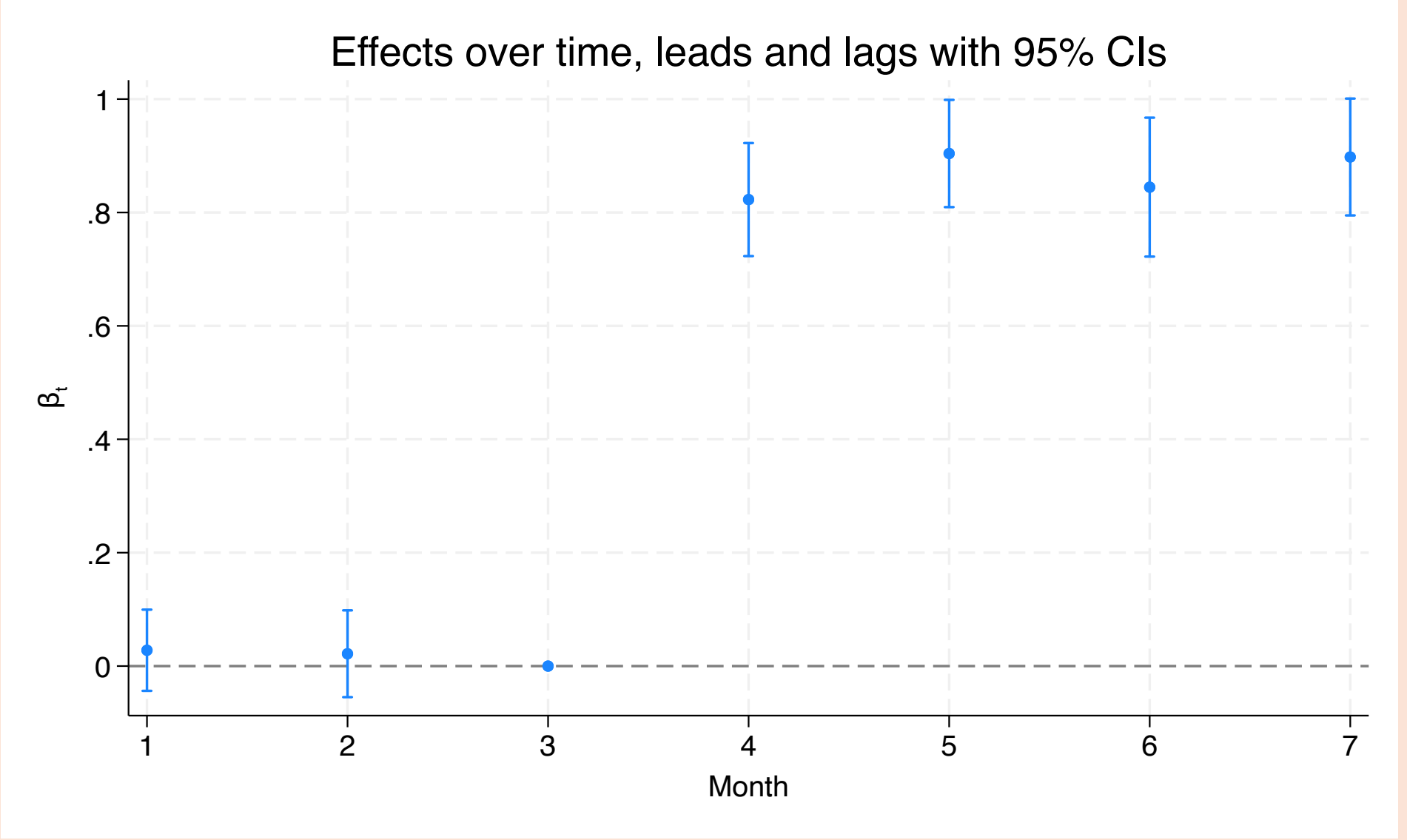
. qui didregress (satis) (procedure), group(hospital) time(month)

. estat trendplots



```
. qui didregress (satis) (procedure), group(hospital) time(month)

. estat ptrends

Parallel-trends test (pretreatment time period)
H0: Linear trends are parallel

F(1, 45) =   0.55
Prob > F = 0.4615
```

```
. qui didregress (satis) (procedure), group(hospital) time(month)

. estat grangerplot
```



```
. qui didregress (satis) (procedure), group(hospital) time(month)

. estat granger

Granger causality test
H0: No effect in anticipation of treatment

F(2, 45) =   0.33
Prob > F = 0.7239
```

Panel data did

. xtset classid

Panel variable: classid (balanced)

. xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)

Treatment and time information

Time variable: **year**
Control: **gotpatent = 0**
Treatment: **gotpatent = 1**

	Control	Treatment
Group		
classid	**6912**	**336**
Time		
Minimum	**1875**	**1919**
Maximum	**1875**	**1919**

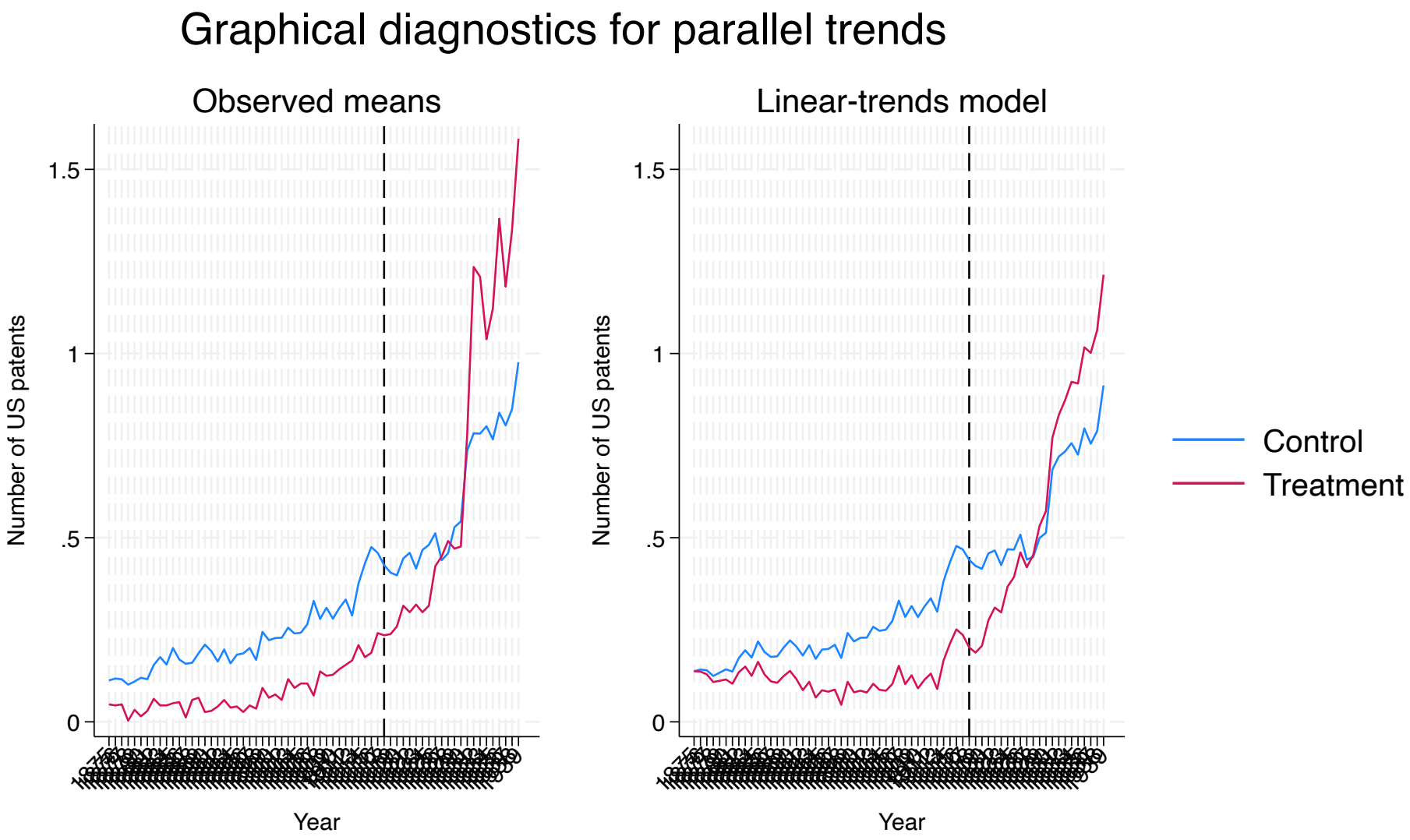
Difference-in-differences regression — Number of obs = **471,120**
Data type: Longitudinal

(Std. err. adjusted for **7,248** clusters in **classid**)

uspatents	Coefficient	Robust std. err.	t	P>|t|	[95% conf.	interval]
ATET						
gotpatent (Patent vs None)	**.150516**	**.0356081**	**4.23**	**0.000**	**.0807137**	**.2203183**

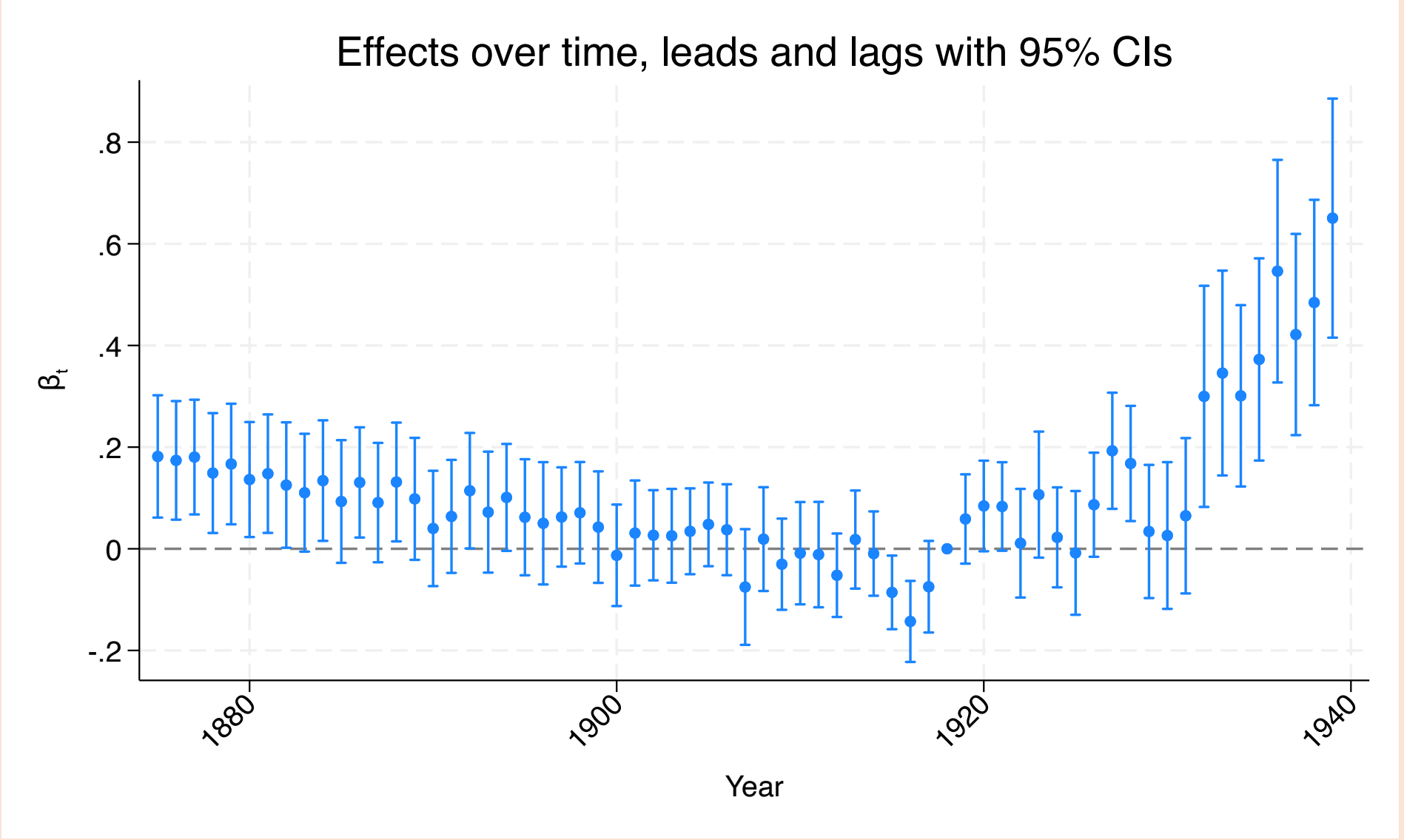
Note: ATET estimate adjusted for covariates, panel effects, and time effects.

```
. qui xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)
. estat trendplots
```



```
. qui xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)

. estat ptrends

Parallel-trends test (pretreatment time period)
H0: Linear trends are parallel

F(1, 7247) =  21.69
  Prob > F = 0.0000


.
end of do-file
```

```
. qui xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)
. estat grangerplot
```



```
. qui xtdidregress (uspatents fpatents) (gotpatent), group(classid) time(year)

. estat granger

Granger causality test
H0: No effect in anticipation of treatment

F(43, 7247) =   6.69
  Prob > F = 0.0000
```