Practical Guide to Event Studies

David Novgorodsky Bradley Setzler

Department of Economics, University of Chicago

January 23, 2019

1 Identification with Perfect Control Groups

In this section, we define our parameter of interest in event study designs as well as a set of key assumptions that play a role in its identification. We borrow heavily from the presentation in Abraham and Sun (2018). Throughout, we motivate our approach using an example taken from Fadlon and Nielsen (2017) (FN hereafter) where the outcome is wife's labor supply and the research design leverages variation in the timing of spousal death.

1.1 Notation and Parameter of Interest

Notation: For individual i, denote the outcome in year t by Y_{it} (this is observed), the year at which treatment occurs by E_i (this is observed), and the potential outcomes by $Y_{it}(e)$ (this is not observed unless $E_i = e$). [1]

- Example 1 (FN): Y_{it} = wife's labor supply, E_i = year that husband dies of fatal heart attack.

Parameter of interest: We are interested in the average treatment on the treated, which is $\text{ATT}_t(e) \equiv \mathbb{E}\left[Y_{it}(e) - Y_{it}(\infty) \middle| E_{it} = e\right]$, where $Y_{it}(\infty)$ is the outcome an i would have at t if counterfactually assigned treatment at time ∞ (i.e., never treated). This is the average difference in Y_{it} that is due to being treated at e instead of ∞, among those who are treated at e.

- Example 1 (FN): $\mathbb{E}\left[Y_{it}(1995) - Y_{it}(\infty) \middle| E_{it} = 1995\right]$ is the difference in labor supply at, say, time $t = 1997$ for a wife whose husband died in 1995 versus if her husband had never died.

1.2 Parallel Trends vs No Anticipation

Definition: Parallel Trends is defined as $\mathbb{E}\left[Y_{is}(\infty) - Y_{it}(\infty) \middle| E_i = e\right] = \mathbb{E}\left[Y_{is}(\infty) - Y_{it}(\infty) \middle| E_i = e'\right]$, for all $e \neq e'$ and all $t \neq s$. This says that, for any two observed cohorts e and e', the change over time they would have had in the absence of treatment is the same.

- Example 1 (FN): Parallel trends requires that, for wives whose husbands died in 1995 and wives whose husbands died in 2000, if none of their husbands had died, they would have experienced the same change in mean labor supply between 1995 and 1996.

[1] Athey and Imbens (2018) also begin with a setting where onset dates define potential outcomes, but with a focus on design-based estimators where the source of uncertainty is in the assignment mechanism for onset dates across sample units.

Definition: No Anticipation is defined as $Y_{it}(e) = Y_{it}(\infty)$, for all $t < e$ and for all e. This says that, prior to the onset of treatment, outcomes do not depend on the time at which treatment will occur.

- Example 1 (FN): No anticipation requires that, in 1994, wives of husbands who died in 1995 did not have different labor supply than they would have had if their husbands had never died.

1.3 Differences-in-differences under Parallel Trends and No Anticipation

Let $s < e \leq t < e'$ and consider this estimator:

$$\text{DiD}_{t,s}(e, e') \equiv \left(\mathbb{E}\left[Y_{it}|E_i = e\right] - \mathbb{E}\left[Y_{is}|E_i = e\right]\right) - \left(\mathbb{E}\left[Y_{it}|E_i = e'\right] - \mathbb{E}\left[Y_{is}|E_i = e'\right]\right)$$

which depends only on observable quantities.

Theorem 1: If Parallel Trends and No Anticipation hold, then $\text{DiD}_{t,s}(e, e') = \text{ATT}_t(e)$ if $s < e \leq t < e'$.

Proof of Theorem 1: We start with $\text{ATT}_t(e)$ and show that it equals $\text{DiD}_{t,s}(e, e')$.[2]

1. We add and subtract Y_s^∞:

$$\text{ATT}_t(e) \equiv \mathbb{E}\left[Y_t^e - Y_t^\infty|E = e\right] = \mathbb{E}\left[Y_t^e - Y_s^\infty|E = e\right] - \mathbb{E}\left[Y_t^\infty - Y_s^\infty|E = e\right]$$

2. We apply Parallel Trends:

$$\text{ATT}_t(e) = \mathbb{E}\left[Y_t^e - Y_s^\infty|E = e\right] - \mathbb{E}\left[Y_t^\infty - Y_s^\infty|E = e'\right]$$

3. No Anticipation lets us replace some of these Y^∞ terms with Y^e or $Y^{e'}$ terms:

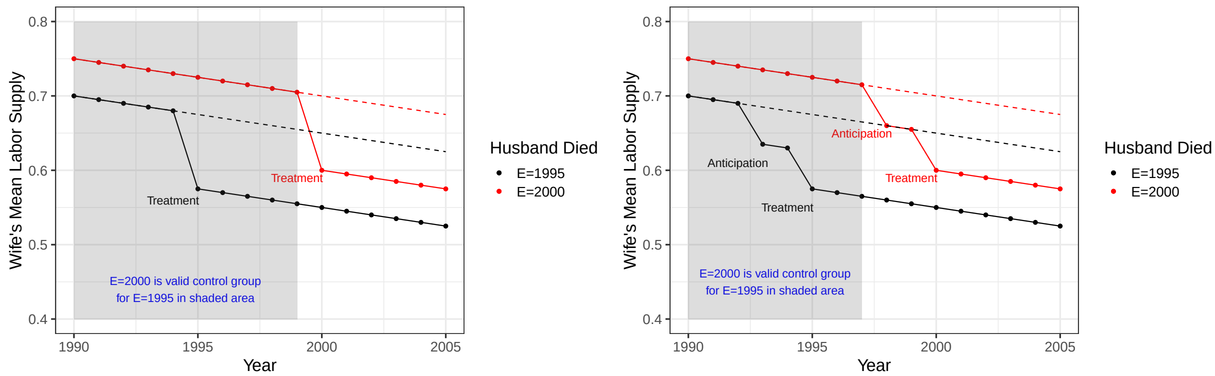
$$\text{ATT}_t(e) = \mathbb{E}\left[Y_t^e - Y_s^e|E = e\right] - \mathbb{E}\left[Y_t^{e'} - Y_s^{e'}|E = e'\right] \equiv \text{DiD}_{t,s}(e, e')$$

Definition: Perfect Control Group: We see from the above result that a Perfect Control Group for treated cohort e during the years $t < e'$ is any cohort e' that satisfies both Parallel Trends and No Anticipation. Note: A perfect control group may be a future winner $e' < \infty$ or a never-winner $e' = \infty$, it just depends on the empirical context.

2 Identification with Imperfect Control Groups

An imperfect control group for winners at e is a cohort of winners e' (or never-winners $e' = \infty$) that either fails to satisfy Parallel Trends or fails to satisfy No Anticipation. We now consider various types of imperfect control groups and show conditions under which we can still identify some of the $\text{ATT}_t(e)$ terms.

[2] This proof is essentially an abbreviated version of the proof of Proposition 4 in Abraham and Sun (2018) in the case of a single control cohort e'.



(a) No Anticipation

(b) $k = 2$ Years of Anticipation

Figure 1: Anticipation in Example 1

Notes: Here, we consider the model with $Y_t^\infty = 0.75 - 0.005\,(t-1990)\,, \forall t$ for the $E_i = 2000$ cohort, and $Y_t^\infty = 0.70 - 0.005\,(t-1990)\,, \forall t$ for the $E_i = 1995$ cohort. Clearly, this satisfies the Parallel Trends assumption, since $\mathbb{E}\left[Y_{is}\left(\infty\right) - Y_{it}\left(\infty\right)\middle|E_i = 1995\right] = \mathbb{E}\left[Y_{is}\left(\infty\right) - Y_{it}\left(\infty\right)\middle|E_i = 2000\right] = 0.005\,(t-s)$. In Figure (a), we also impose No Anticipation, so cohort $E = 2000$ is clearly a Perfect Control Group for the $E = 1995$ cohort during the years prior to $t = 2000$. In Figure (b), we relax No Anticipation for $k = 2$ years prior to the event. As discussed in the text, this means that cohort $E = 2000$ is still a valid control group for the $E = 1995$ cohort during the years prior to when anticipation begins in year $t = 1998 = 2000 - k$. The true treatment effect is -0.1. In Figure (b), the anticipation effect is -0.05.

2.1 Parallel Trends Holds, but Anticipation Occurs for the k Years before the Event

2.1.1 What Can Be Identified with No Restrictions on Anticipation?

Suppose that winning is anticipated $k > 0$ years before it happens. The anticipation at t means that $Y_t^{e'} \neq Y_t^\infty$ for $t \geq e' - k$, so the substitution $\mathbb{E}\left[Y_t^\infty | E = e'\right] = \mathbb{E}\left[Y_t^{e'} | E = e'\right]$ is no longer true, so Theorem 1 fails.

Figure 1 shows the following:

- In Figure 1(a), there is no anticipation, so $e' = 2000$ is a valid control group for $e = 1995$ as long as $t < e'$; in particular, $\mathbb{E}\left[Y_t^\infty\middle|E = 2000\right] = \mathbb{E}\left[Y_t\middle|E = 2000\right] = 0.75 - 0.005t$, where $-0.005t$ is the assumed common time trend for all cohorts.
- In Figure 1(b), there is $k = 2$ years of anticipation, so $e' = 2000$ is a valid control group for $e = 1995$ only on $t < e' - k$. We can see that $\mathbb{E}\left[Y_{1999}^\infty | E = 2000\right] = 0.75$, but $\mathbb{E}\left[Y_{1999} | E = 2000\right] = 0.65$ due to anticipation, so clearly $\mathbb{E}\left[Y_{1999} | E = 2000\right] \neq \mathbb{E}\left[Y_{1999}^\infty | E = 2000\right]$, so $E = 2000$ fails as a control group when $t = 1999$.

As illustrated with the shaded area in Figure 1(b), the $E = 2000$ cohort is still a valid control group for the $E = 1995$ cohort during any $t < 1998$, that is, any year prior to when $E = 2000$ begins to anticipate treatment. This makes clear the cost of having an imperfect control group: we can only identify ATT_t for $t < 1998$ in Figure 1(b), whereas we can identify $\text{ATT}_t\left(e\right)$ for $t = 1998, 1999$ as well in Figure 1(a).[3]

2.1.2 What Can Be Identified under Parallel Anticipation?

Finally, we show that, since the anticipation effect is identified for cohort e, we can correct for the anticipation in cohort e' if they both anticipate in the same way.

[3] If anticipation occurs for $k = \infty$ years prior to the event, then $E = 2000$ is never a valid control group for $E = 1995$, as emphasized by Borusyak and Jaravel (2017).

(a) Raw Data

(b) Removing the Linear Trends

Figure 2: Non-Parallel but Linear Trends in Example 1

Notes: Here, we consider the model with $Y_t^{\infty} = 0.70 - 0.01\,(t - 1990)$ for the $E_i = 2000$ cohort, and $Y_t^{\infty} = 0.75 - 0.02\,(t - 1990)$ for the $E_i = 1995$ cohort. The true treatment effect is -0.1. In Figure (b), we replace the $E = 2000$ time-trend $-0.01\,(t - 1990)$ with the $E = 1995$ time-trend $-0.02\,(t - 1990)$ so that it satisfies Parallel Trends.

Definition: Parallel Anticipation: Denoting $t' \equiv t + (e' - e)$, we say that two cohorts have parallel anticipation effects if $\mathbb{E}\left[Y_t^e - Y_t^{\infty} \middle| E = e\right] = \mathbb{E}\left[Y_{t'}^{e'} - Y_{t'}^{\infty} \middle| E = e'\right]$.

Theorem 2: If Parallel Trends holds, No Anticipation holds on $e - k < t$, and Parallel Anticipation holds, then

- (a) $\mathrm{DiD}_{t,s}(e, e') = \mathrm{ATT}_t(e)$ on $e - k < t$ for $s < e \leq t < e'$, and,
- (b) $\mathrm{DiD}_{t,s}(e, e') + \mathrm{DiD}_{s,s'}(e, e') = \mathrm{ATT}_t(e)$ for $e - k \leq s < e < t < e'$, where $s' = e - (e' - t)$.

Sketch of Proof: Theorem 2(a) is the same as Theorem 1, just for a smaller range of years. Theorem 2(b) is proven by showing that $\mathrm{DiD}_{t,s}(e, e') = \mathrm{ATT}_t(e) - \mathrm{ATT}_{t-k'}(e)$ for some $k \geq k' > 0$, then we choose $s' = t - k'$ and see that $\mathrm{ATT}_{t-k'}(e) = \mathrm{DiD}_{s,s'}(e, e')$ under Parallel Anticipation, so we just add $\mathrm{DiD}_{s,s'}(e, e')$ to $\mathrm{DiD}_{t,s}(e, e')$ as a bias correction.

2.2 No Anticipation Holds, but Parallel Trends Fails in a Parametric Way

Suppose that the cohort treated at e and the cohort treated at e' would have different trends if never treated. Because there is No Anticipation, any differences in the trends for cohorts e and e' at time $t < e$ must be exactly equal to the deviation from Parallel Trends and implies that we can correct for the differential trends by estimating them in the pre-period. We now show this formally and illustrate it in Figure 2.

Corollary 1: If No Anticipation holds, $\mathbb{E}\left[Y_{is}(\infty) - Y_{it}(\infty) \middle| E_i = e\right] = \mathbb{E}\left[Y_{is} - Y_{it} \middle| E_i = e\right]$ if $s < t < e < e'$.

- **Proof of Corollary 1:** This is simply substituting that $Y_{it}(\infty) = Y_{it}(e)$ if $t < e$ by No Anticipation.

Denote the deviation from Parallel Trends by $\psi_{t,s}(e, e') \equiv \mathbb{E}\left[Y_{is}(\infty) - Y_{it}(\infty) \middle| E_i = e\right] - \mathbb{E}\left[Y_{is}(\infty) - Y_{it}(\infty) \middle| E_i = e'\right]$.

Corollary 2: If No Anticipation holds, then $\psi_{t,s}(e, e') = \mathbb{E}\left[Y_{is} - Y_{it} \middle| E_i = e\right] - \mathbb{E}\left[Y_{is} - Y_{it} \middle| E_i = e'\right]$ if $s < t < e < e'$, where the right-hand side is observable.

- **Proof of Corollary 2:** This is just plugging in Corollary 1 for both e and e' into the definition of $\psi_{t,s}(e, e')$.

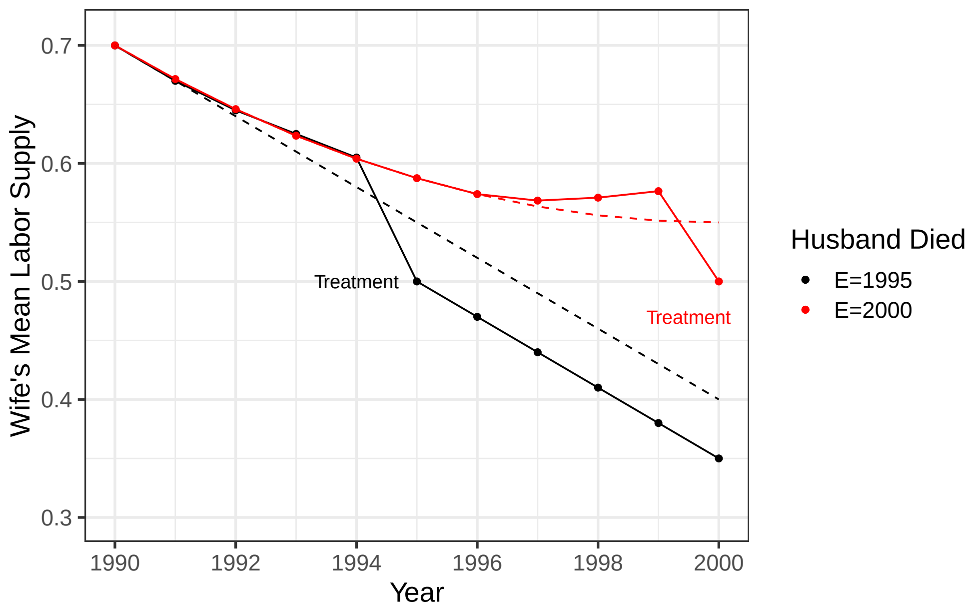


Figure 3: Neither No Anticipation nor Parallel Trends in Example 1

Notes: Suppose the economist only observes the solid lines for years 1990-2000. Here, we consider the model with $Y_t^\infty = 0.70 - 0.03\,(t - 1990)$ for the $E_i = 1995$ cohort, and $Y_t^\infty = 0.70 - 0.03\,(t - 1990) + 0.001\,(t - 1990)^2$ for the $E_i = 2000$ cohort, which are shown with the dashed lines. The anticipation effects are 0.005, 0.015, and 0.025 during the 3 years prior to the event. The true treatment effect is -0.05 (the difference between the black solid and dashed lines), but the DiD estimate (the difference between the solid red and black lines) is about -0.20 in $t = 1999$ (4 times the true effect in magnitude), despite the solid lines coinciding almost perfectly in $t < 1995$.

We have shown the the deviation from Parallel Trends is observable if No Anticipation holds. We can then use this to see if it follows a parametric trend, such as linearity. Finally, if we assume that the parametric trend is the same at $t \geq e$ as it is at $t < e$, we can correct for the non-parallel trends on $t \geq e$.

Theorem 3: If No Anticipation holds, then $\text{DiD}_{t,s}(e, e') = \text{ATT}_t(e) + \psi_{t,s}\,(e, e')$.

- **Proof of Theorem 3:** This follows from the proof of Theorem 1, but including the term $\psi_{t,s}\,(e, e')$ in step 2 when subsituting in for the pre-trend.

Figure 2 illustrates how to use the observed pre-trend at time $t < e$ to correct for the non-parallel trend for all t. The same approach can be used to correct for any parametric function $\psi_{t,s}\,(e, e')$. This is an extrapolation approach: we observe the deviation from parallel in the pre-period, extrapolate what the deviation will be in the post-period, then correct the e' cohort so that it becomes parallel.

2.3 Can We Learn Anything if Neither Assumption Holds?

The results in the previous section suggest the following:

- Given No Anticipation, we can test for Parallel Trends and use the data to choose the design;

- Given Parallel Trends, we can test for No Anticipation and use the data to choose the design.

So that is the beginning of the analysis – either we start by believing in Parallel Trends, or we start by believing in No Anticipation, and then we can test/correct for the other. This is where choosing the empirical context is important: just as we look for instruments that are *ex ante* as-good-as-random in IV applications, we look for events that are *ex ante* as-good-as-parallel or as-good-as-unanticipated in event studies.

To illustrate why we need *either* No Anticipation *or* Parallel Trends to get anywhere in an event study, **Figure 3** provides an example where the observed pre-trends "look nice," but both No Anticipation and Parallel Trends fail, amd this causes the DiD estimator to be severely biased. Clearly, we can't use the approaches above to correct for deviations in the pre-trends because there are no deviations in the pre-trends, and we can't correct for anticipation because they match perfectly in the years where anticipation would be expected.

3 Regression Estimators and Stacking Cohorts

3.1 Regressions with Perfect Control Groups

3.1.1 1 Treatment Cohort and 1 Control Cohort

Suppose that any cohort $e' > e$ is a Perfect Control Group for e (that is, it satisfies both Parallel Trends and No Anticipation). Since,

$$\text{DiD}_{t,s}(e,e') \equiv \left(\mathbb{E}\left[Y_{it}|E_i = e\right] - \mathbb{E}\left[Y_{is}|E_i = e\right]\right) - \left(\mathbb{E}\left[Y_{it}|E_i = e'\right] - \mathbb{E}\left[Y_{is}|E_i = e'\right]\right)$$

It is easy to show that, if we estimate the following regression,

$$Y_{it} = \alpha(e,e') + \delta(e,e')1_t + \tau(e,e')1_e + \gamma_{t,s}(e,e')1_t 1_e + \epsilon_{it},\ \textit{for}\ E_i \in \{e,e'\}\ \textit{and}\ s < e \leq t < e'$$

then OLS returns $\gamma_{t,s}(e,e') = \text{DiD}_{t,s}(e,e')$. If e' is a Perfect Control Group for e, it follows that $\gamma_{t,s}(e,e') = ATT_t(e)$.

3.1.2 1 Treatment Cohort and M Control Cohorts

Suppose we have cohorts $e < e' < e''$. Then, we can write,

$$Y_{it} = \alpha(e,e') + \delta(e,e')1_t + \tau(e,e')1_e + \gamma_{t,s}(e,e')1_t 1_e + \epsilon_{it},\ \textit{for}\ E_i \in \{e,e'\}\ \textit{and}\ s < e \leq t < e'$$
$$Y_{it} = \alpha(e,e'') + \delta(e,e'')1_t + \tau(e,e'')1_e + \gamma_{t,s}(e,e'')1_t 1_e + \epsilon_{it},\ \textit{for}\ E_i \in \{e,e''\}\ \textit{and}\ s < e \leq t < e''$$

If both e' and e'' are Perfect Control Groups at the same t, then $\gamma_{t,s}(e,e') = \gamma_{t,s}(e,e'')$. For the years satisfying $s < e \leq t < e' < e''$ (so that e' and e'' are both valid control groups for e), we can recover $\gamma_{t,s}$ in a single regression with both control cohorts by "stacking" them into a single equation as follows:

$$Y_{it} = \alpha(e,\{e',e''\}) + \delta(e,\{e',e''\})1_t + \tau(e,\{e',e''\})1_e + \gamma_{t,s}(e,\{e',e''\})1_t 1_e + \epsilon_{it},\ \textit{for}\ E_i \in \{e,e',e''\}\ \textit{and}\ s < e \leq t < e' < e''$$

The stacked regression says that, during the years t in which both e' and e'' are control groups, we can use both

e' and e'' when estimating the parameters. The δ_t will use all 3 cohorts to estimate the year effect; the δ_e cohort will be the average difference at s between cohort e and the other two cohorts (weighted by number of observations for each of the other two cohorts); and $\gamma_{t,s}$ will pick up the DiD estimator of the form,

$$\text{DiD}_{t,s}(e, \{e', e''\}) \equiv \left(\mathbb{E}\left[Y_{it}\middle|E_i = e\right] - \mathbb{E}\left[Y_{is}\middle|E_i = e\right]\right) - \left(\mathbb{E}\left[Y_{it}\middle|E_i = e' \textit{ or } E_i = e''\right] - \mathbb{E}\left[Y_{is}\middle|E_i = e' \textit{ or } E_i = e''\right]\right)$$

where all that has changed is that the conditioning set for the second difference includes both cohorts e' and e''.

Note that, under the assumptions, having more control cohorts improves precision of the estimator by increasing sample size, but does not affect identification. Note also that the control groups available for different t times will vary. For example, if the data is from 2001-2006 and $e = 2003$, then cohorts $e' \in \{2004, 2005, 2006\}$ will be available for $t = 2003$, $e' \in \{2005, 2006\}$ will be available for $t = 2004$, and $e' = 2006$ will be available for $t = 2005$. If the pre-period is $s = 2001$, then all of these control cohorts $e' \in \{2004, 2005, 2006\}$ are available for the pre-period. This means it is important to be careful to use the same cohorts that are available at t when estimating the mean at s; otherwise, the difference $\left(\mathbb{E}\left[Y_{it}\middle|E_i = e' \textit{ or } E_i = e''\right] - \mathbb{E}\left[Y_{is}\middle|E_i = e' \textit{ or } E_i = e''\right]\right)$ will be affected by composition changes between s and t rather than only by the time effects this term is meant to capture. This is done mechanically by including intercepts for each cohort.

3.1.3 L Treatment Cohorts and M Control Cohorts

Notice that, when there are two control cohorts, there are also two treatment cohorts, since the middle cohort can be used both as a treated and control cohort. Let's consider the simplest case where $e = e' - 1 = e'' - 2$ so that e, e', e'' are each in adjacent years. Then, e'' can be used twice as a control group – it can be the control group for e both at $t = e$ and $t = e'$, and as the control group for e' at $t = e'$. Let's focus on $t = e$ and $s < e$:

$$Y_{it} = \alpha(e, \{e', e''\}) + \delta(e, \{e', e''\})1_{t=e} + \tau(e, \{e', e''\})1_e + \gamma_{t=e,s}(e, \{e', e''\})1_{t=e}1_e + \epsilon_{it}, \textit{ for } E_i \in \{e, e', e''\} \textit{ and } t \in \{s, e\}$$
$$Y_{it} = \alpha(e', e'') + \delta(e', e'')1_{t=e+1} + \tau(e', e'')1_{e+1} + \gamma_{t=e+1,s}(e', e'')1_{t=e+1}1_{e'} + \epsilon_{it}, \textit{ for } E_i \in \{e', e''\} \textit{ and } t \in \{s, e+1\}$$

The first equation identifies $ATT_{t=e}(e)$ when pooling the e', e'' cohorts as control groups. The last equation identifies $\gamma_{t=e+1,s}(e', e'') = ATT_{t=e+1}(e')$ using the e'' cohort as the control group. Importantly, the control group in the first equation e' is the treatment group in the second equation.

How can we run a single regression that correctly uses e' as the control group for e but also uses e' as a treatment group where e'' is its control group? We can do this by **stacking the data** with duplicates and introducing the reference cohort variable, r.[4] In the first equation, $r = e$ is the reference cohort for both e' and e''. In the second equation, $r = e'$ is the reference cohort for e''. So there are two copies of the e' observation, but one is coded with $r = e$ and the other is coded with $r = e'$. Then, to estimate $\gamma_{t=e,s}(e, \{e', e''\})$, we fully interact the regression that uses e as the treatment group and e', e'' as the control groups with an indicator $1_{r=e}$, and to estimate $\gamma_{t=e+1,s}(e', e'')$,

[4]Several recent studies use such a stacking approach with various choices of control group, including Fadlon and Nielsen (2017), Deshpande and Li (forthcoming), and Jensen (2018). Borusyak and Jaravel (2017) also suggest a subsample regression approach similar to the stacking approach as a potential solution to negative weighting issues of more typical dynamic event study regression specifications when cohorts have heterogeneous ATT effects. As an alternative, Abraham and Sun (2018) propose a regression approach that does not require stacking, but instead relies on using a single control cohort (the cohort with the latest treatment onset time) to identify all cohort-specific ATT effects.

we fully interact the regression that uses e' as the treatment group and e'' as the control group with an indicator $1_{r=e'}$. Then, we can estimate both regressions simultaneously, since they are fully interacted with the 1_r indicators so that it is equivalent to estimating the two regressions separately. Given the stacked representation with reference cohort interactions, we can consider the following assumption:

Definition: Homogeneous ATT in Event Time: We say the ATT is homogeneous in event time if $\text{ATT}_t(e) = \text{ATT}_{t'}(e') = \overline{\text{ATT}}_{post}$, for all e, e', t, t', where $post \equiv e' - e$ and $t' \equiv t + post$.

To impose homogeneity in our regression representation, we want $\gamma_{t=e,s}(e, \{e', e''\}) = \gamma_{t=e+1,s}(e', e'')$, but we still want to allow for $\delta(e, \{e', e''\}) \neq \delta(e', e'')$ and $\tau(e, \{e', e''\}) \neq \tau(e', e'')$. This is easy in our reference-cohort representation: we simply drop the interaction with 1_r with each of the γ terms so that the regression will force $\gamma_{t=e,s}(e, \{e', e''\}) = \gamma_{t=e+1,s}(e', e'')$. This is equivalent to a GMM estimator where we estimate the first regression separately from the second regression, but impose a penalty on deviations from $\gamma_{t=e,s}(e, \{e', e''\}) = \gamma_{t=e+1,s}(e', e'')$.

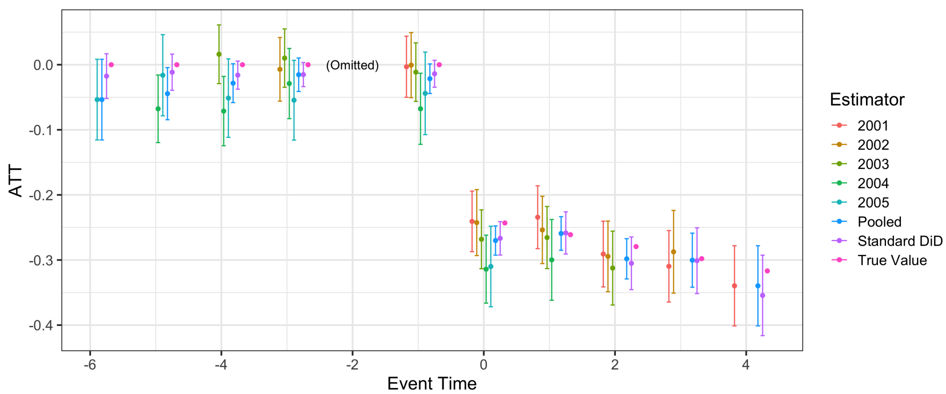
This extends to the general case: we can always account for L treatment cohorts and M control cohorts by reframing it as each 1 treatment cohort matched to as many control cohorts as possible for each outcome time t, then fully interacting each of these regressions with the treatment reference cohort indicator 1_r. We present the general case in **Figure 4**. Here, we simulate data from a model with dynamic ATTs (across event time) that are homogeneous (across cohorts). The Perfect Control Group assumption is satisfied. We compare estimates when allowing for heterogeneity across cohorts, the Pooled estimate when imposing homogeneity in the estimator, and also compare these to the standard DiD estimator that does not use stacking.

We see that the true values are well recovered by the Pooled and Standard DiD estimators, with some evidence that the Pooled standard errors are tighter due to using more control observations in the central event times. By contrast, the cohort-specific estimates are substantially more noisy with larger standard errors, illustrating the benefit of imposing homogeneity when the true model is homogeneous.

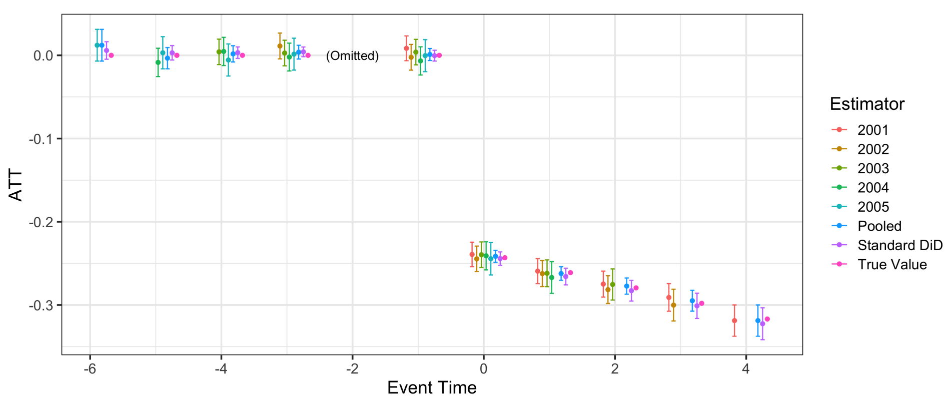
3.2 Regressions with Imperfect Control Groups

3.2.1 L Treatment Cohorts and M Control Cohorts: No Anticipation Holds, but Trends are not Parallel

Suppose that we are in the case where for all cohorts, No Anticipation holds, however Parallel Trends does not hold. As demonstrated in Figure 2, if the precise form of the deviation from parallel trends, $\psi_{t,s}(e, e')$, is known, we could directly remove the bias due to differing parallel trends from each observation of unit i at time t and proceed as in Section 3.1.3. However, in practice, $\psi_{t,s}(e, e')$ is unknown. Nonetheless, for a given parametric form, we may use the pre-treatment observations $t < e$ for each cohort to identify the cohort-specific components of the trend (under the assumption that the pre-treatment trend continues unchanged post-treatment). Then, by residualizing the outcome Y_{it} with respect to the recovered trend component, we again regain Parallel Trends and are back in the case in Section 3.1.3. For example, suppose that the underlying deviation from parallel trends is cohort-specific linear trends. For exposition, let t_{min} denote each unit i's earliest calendar time observation. To adjust for cohort-specific linear trends, we first restrict the sample to observations $t < e$ and consider the following regression:



(a) Small N



(b) Large N

Figure 4: Stacked Regression in Example 1

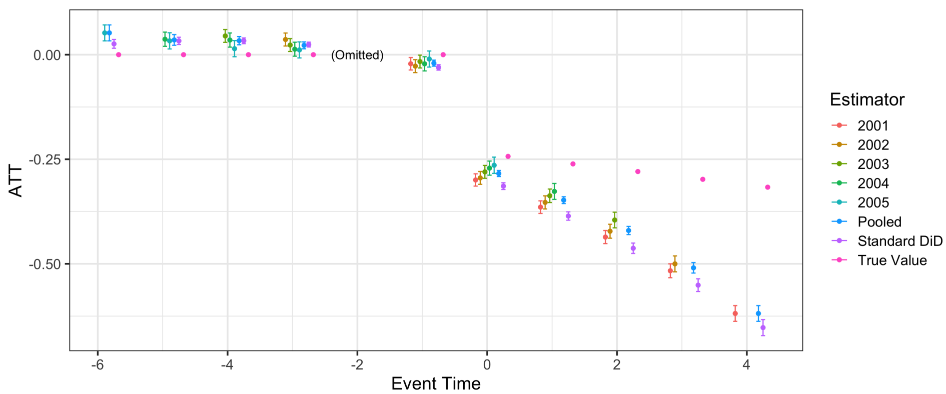
Notes: This figure presents estimates using the stacked regression described above for small and large data simulations. Each point estimate uses the maximum possible treatment and control observations. Results are presented without imposing homogeneity (each treatment cohort has its own estimate) and imposing homogeneity ("Pooled"). Finally, the standard DiD regression is presented which does not stack data. The true value is provided for comparison. Note that the true model has homogeneous ATT effects across cohorts.

$$Y_{it} = \alpha + \psi_e 1_e(t - t_{min}) + \epsilon_{it}, \; for \;\; t < e$$

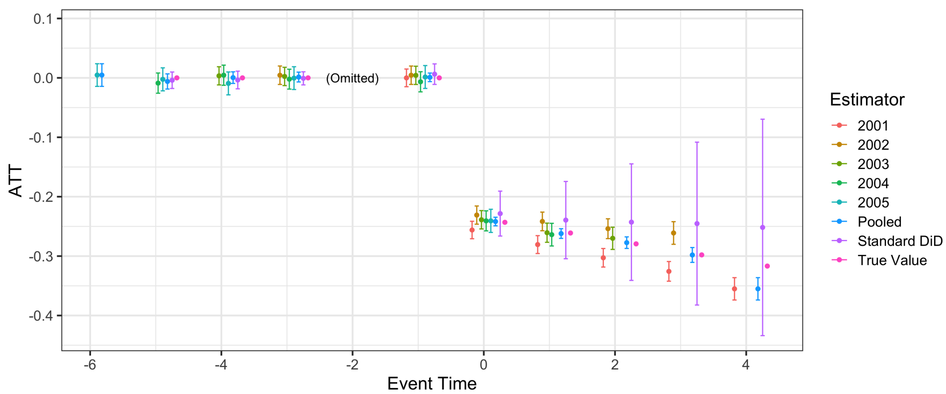
Next, we use the recovered slope parameters ψ_e from the above regression to construct the trend contribution to each Y_{it}, $\tilde{Y_{it}} = \psi_e(t - t_{min})$. Finally, we produce residualized outcome data $\hat{Y_{it}} \equiv Y_{it} - \tilde{Y_{it}}$. Under the joint assumptions that the deviation from parallel trends takes a cohort-specific linear form and that this trend persists post-treatment, $\hat{Y_{it}}$ would now satisfy both No Anticipation and Parallel Trends and we can proceed with the approach in Section 3.1.3.[5]

Figure 5 demonstrates the correction for non-parallel trends with stacked data. In Figure 5(a), we show how biased the estimators are when Parallel Trends fails in a linear way. In Figure 5(b), we correct the stacked estimators using the two-step approach where we fit time-trends on pre-treatment data and then extrapolate those

[5]We ignore discussion of inference, however, given that this strategy would rely on pre-estimated pre-trend parameters, one would want standard errors on estimated ATT parameters to reflect this pre-estimation.



(a) Linear Trends, No Correction



(b) Linear Trends, Corrected

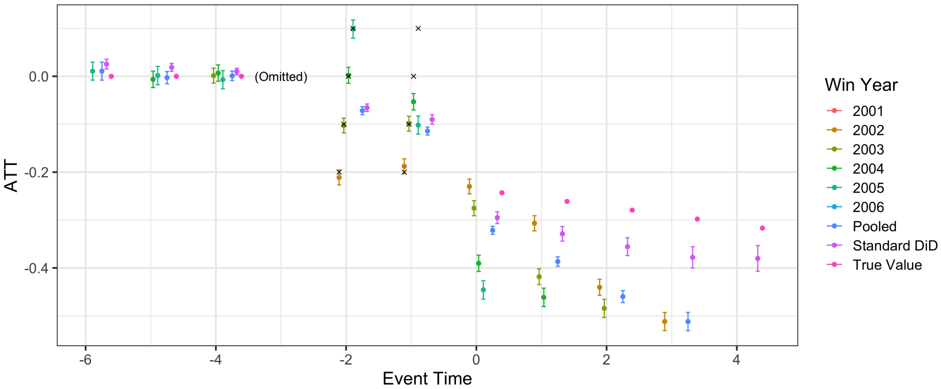
Figure 5: Stacked Regression with Linear (Non-parallel) Trends in Example 1

Notes: The true model has linear non-parallel trends. This figure presents estimates using the stacked regression described above (a) not corrected for non-parallel trends, and (b) correcting linearly for non-parallel trends. Each point estimate uses the maximum possible treatment and control observations. Results are presented without imposing homogeneity (each treatment cohort has its own estimate) and imposing homogeneity ("Pooled"). Finally, the standard DiD regression is presented which does not stack data. The true value is provided for comparison. Note that the true model has homogeneous ATT effects across cohorts. We use the two-step residualization method in each estimator except for Standard DiD, where we follow standard practice by controlling for linear cohort trends in the regression.

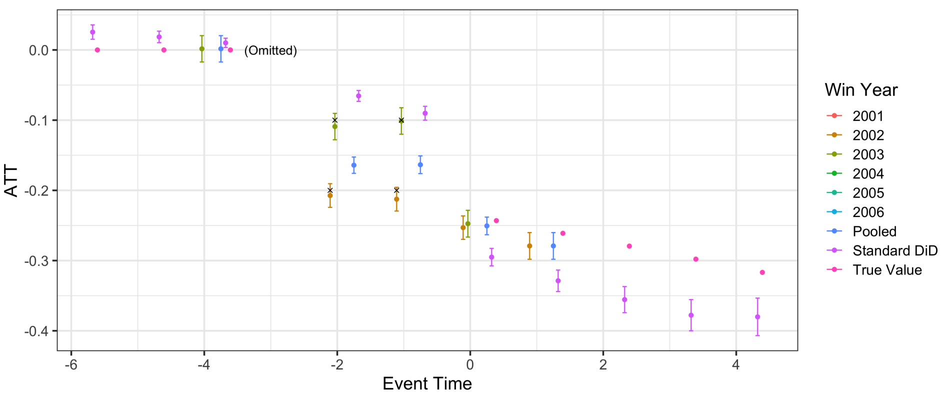
trends on post-treatment time periods in order to residualize out the time trends. In the Standard DiD estimator, we include linear cohort-specific time-trends in the regression. We see that the corrections generally perform well, though standard DiD is affected by large standard errors.

3.2.2 L Treatment Cohorts and M Control Cohorts: Parallel Trends Holds, but there is Anticipation

Suppose that we are in the case where for all cohorts, Parallel Trends holds, but No Anticipation does not hold; that is, for each treated cohort with onset time e, individuals begin adjusting their outcome in anticipation of treatment as of period $e-k$, $k>0$. As discussed in Section 2.1.1, for a given cohort e, anticipation is analogous to moving the treatment onset time for all later cohorts k periods into the future. This mapping suggests how we adjust the approach in Section 3.1.3 for a given anticipation period - when stacking control observations for a given reference cohort $r=e$, only use $E_i \in \{e' \,|\, e' > e+k\}$. For example, consider the case with with five adjacent cohorts, i.e., $e = e'-1 = e''-2 = e'''-3 = e''''-4$ and suppose $k=2$, meaning that all cohorts begin adjusting their outomc in anticipation of being treated two periods prior to treatment onset. Unlike in Section 3.1.3, for $t=e$, we could no



(a) Anticipation, No Correction



(b) Anticipation, Corrected

Figure 6: Stacked Regression with 2 Years of Anticipation in Example 1

Notes: The true model has 2 years of anticipation, and anticipation is heterogeneous across cohorts. This figure presents estimates using the stacked regression described above (a) not corrected for 2 years of anticipation, and (b) correcting for 2 years of anticipation by restricting control groups. Each point estimate uses the maximum possible treatment and control observations. Results are presented without imposing homogeneity (each treatment cohort has its own estimate) and imposing homogeneity ("Pooled"). Finally, the standard DiD regression is presented which does not stack data. The true value is provided for comparison. Note that the true model has homogeneous ATT effects across cohorts.

longer include $\{e', e''\}$ in the set of control groups. Thus, for the reference cohort $r = e$, the underlying regression for the $r = e$ subsample would now be

$$Y_{it} = \alpha(e, \{e''', e''''\}) + \delta(e, \{e''', e''''\})1_{t=e} + \tau(e, \{e''', e''''\})1_e + \gamma_{t=e,s}(e, \{e''', e''''\})1_{t=e}1_e + \epsilon_{it},$$
$$\textit{for } E_i \in \{e, e''', e''''\} \textit{ and } t \in \{s, e\}$$

Otherwise, the approach to recover the identified ATT effects proceeds as in Section 3.1.3.

In **Figure 6**, we introduce $k = 2$ periods of anticipation. Figure 6(a) demonstrates the bias in the estimators when not accounting for this anticipation. Note that, as long as the reference event time is prior to anticipation (event time -3 in the figure), the standard DiD performs well with anticipation. In Figure 6(b), we correct for anticipation in the stacked estimators by limiting which control cohorts are used to be at least two cohorts ahead. As a result, we cannot estimate as many effects across event times. Note: Standard DiD would be able to correct for anticipation if anticipation were homogeneous across cohorts, as Standard DiD implicitly includes a correction for homogeneous anticipation.

4 Code

We wrote a software package in R, called eventStudy, which makes everything here extremely easy. For information on how to install and use the package, see:

https://github.com/setzler/eventStudy

References

Abraham, S. and L. Sun (2018): "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects," *Working paper.*

Athey, S. and G. W. Imbens (2018): "Design-based Analysis in Difference-In-Differences Settings with Staggered Adoption," *Working paper.*

Borusyak, K. and X. Jaravel (2017): "Revisiting Event Study Designs with an Application to the Estimation of the Marginal Propensity to Consume," *Working paper.*

Deshpande, M. and Y. Li (forthcoming): "Who Is Screened Out? Application Costs and the Targeting of Disability Programs," *American Economic Journal: Economic Policy.*

Fadlon, I. and T. H. Nielsen (2017): "Family Labor Supply Responses to Severe Health Shocks," *Working paper.*

Jensen, A. (2018): "Loaded but Lonely: Housing and Saving Responses to Spousal Death in Old Age," *Working paper.*