On Event Studies and Distributed-Lags in Two-Way Fixed Effects Models: Identification, Equivalence, and Generalization*

Kurt Schmidheiny   Sebastian Siegloch

November 2020

Abstract

We discuss properties and pitfalls of panel-data event study designs. We derive three main results. First, assuming constant treatment effects before and/or after some event time, also known as binning, is a natural restriction imposed on theoretically infinite effect windows. Binning identifies dynamic treatment effects in the absence of never-treated units and is particularly suitable in case of multiple events. Second, event study designs with binned endpoints and distributed-lag models are numerically identical leading to the same parameter estimates after correct reparametrization. Third, classic dummy variable event study designs can be generalized to models that account for multiple events of different sign and intensity of the treatment, which are common in public and labor economics. We demonstrate the practical relevance of our methodological points in an application studying the effects of unemployment benefit duration on job search effort.

Keywords: event study, distributed-lag, applied microeconomics, credibility revolution
JEL codes: C23, C51, H00, J08

*Kurt Schmidheiny (kurt.schmidheiny@unibas.ch) is affiliated with the University of Basel, CEPR and CESifo; Sebastian Siegloch (sebastian.siegloch@zew.de) is affiliated with ZEW and the University of Mannheim, CEPR, IZA and CESifo. We thank Mona Köhler and Tim Bayer for excellent research assistance. Samara Gunter, Justin McCrary, Jesse Shapiro, Divya Singh and Juan Carlos Suárez Serrato, Tony Strittmatter provided helpful comments – thank you very much. Both authors are grateful to the University of California at Berkeley – in particular Enrico Moretti and Emmanuel Saez – for the hospitality during the Academic Year 2017/2018 when most parts of this paper were written. Siegloch is thankful to the German Research Foundation DFG for financing the research stay under the Research Fellowship program (# 361846460). An early version of this paper circulated as "On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications".

1 Introduction

The credibility revolution in empirical economics has led to more transparent (quasi-)experimental research designs. This shift has increased the policy relevance and the scientific impact of empirical work (Angrist and Pischke, 2010). An important element that enhanced transparency is the visualization of treatment effects and/or identifying assumptions. Differences-in-differences (DD) models are particularly popular in this respect since they are directly connected to the rationale of experiments and the underlying identifying assumptions are intuitive.

The event study (ES) design is the poster child of empirical methods in the DD family since (i) empirical estimates can be plotted, (ii) graphs are very intuitive and immediately show both dynamic post-treatment effects and the identifying assumption of "no pre-event trends", and (iii) the underlying econometrics are straightforward. The empirical specification usually boils down to a simple two-way fixed effects panel data model where the regressors of interest are a set of event indicators which are defined relative to the event. Originating from the finance literature[1], event study designs are now widely used in applied economics, mostly public and labor economics, where an event is usually defined as a policy change whose effects are investigated. Figure 1 plots the use of event study designs in economics over time. We proxy the use by the share of studies mentioning the term "event study" in the Top Five economics journals.[2] While we see a steady increase since 1990, there is sharp increase since 2010. Moreover this increase is mostly driven by the three journals focusing on applied microeconomic work among the Top-Five, i.e. the American Economic Review (AER), the Quarterly Journal of Economics (QJE), and the Journal of Political Economy (JPE).

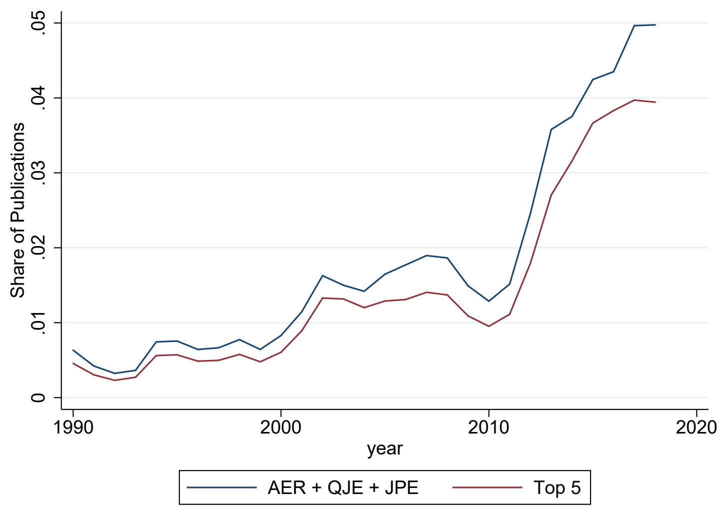
However, the intuitive appeal of event studies and its alleged simplicity entail a risk as it apparently leads researchers to refer to and model event study designs rather loosely. In more than one third of the event study papers published in the AER, QJE or JPE since 2010, no regression equation is specified.

It is the primary purpose of this paper is to clarify the understanding of event study designs both in methodological and practical terms. We show below that the inherent imprecision in specifying event study designs can easily give rise to misspecifications or strong implicit assumptions. We make three main methodological points that are important for applied researchers when setting up event studies. While we derive these three points formally, we choose an intuitive, rather non-technical way of presenting them, in order to make the insights widely accessible to applied researchers. For the same reason, we also discuss various

[1] Dating back to Dolley (1933), see also MacKinlay (1997) for a survey on the financial literature.

[2] More than 80% of the studies mentioning event study designs actually implement one.

Figure 1: The rise of event studies in economics



Notes: This graph plots the three-year moving average of the share of studies mentioning event study designs in top economics journals. We use a 3-year moving average to control for mean reversion. The Top 5 journals are the American Economic Review (AER), the Quarterly Journal of Economics (QJE), the Journal of Political Economy (JPE), Econometrica and the Review of Economic Studies. We report results for AER, QJE and JPE separately as these there journals are known to publish many applied microeconometric studies.

practical implications and pitfalls along the way, making use of simple numerical examples to visualize our claims.

The three points are: First, researchers need to define the window for which the dynamic effects are studied. We call this the *effect window*. While this choice is a practical necessity due to limited data availability, it is not innocuous. Two approaches are commonly used in the literature, and we systematically discuss the underlying assumptions, advantages and caveats. In the first approach, all unit-period observations outside of the effect window are excluded from the estimation sample. This approach requires the presence of never-treated units to identify the dynamic treatment effect. The ultimate question is whether never-treated units can be regarded as a suitable control group. Moreover, dropping observations outside of the effect window is less straightforward in settings where units can receive multiple treatments, which are quite common in public and labor economics. The second, in our view preferable approach is to define the last lag (lead) as open intervals capturing all known events that (will) have happened in the past (future). We refer to this practice as *binning*. We show that binning imposes implicit assumptions that enable to identify dynamic policy effects in the absence of never-treated units. Without binning it would be impossible to separate the dynamic policy effect from secular time effects. Binning is also readily applicable to settings where units might receive multiple treatments (see third contribution below). Most of the most recent empirical literature has put little emphasis on the treatment of endpoints of the effect window, neither formally stating nor explicitly discussing the underlying assumptions. Among the studies published in the AER, QJE or JPE since 2010 that used event studies and specified an empirical model, only 15% provide some information on what has been done at the ends of the effect window. Among those 15%, researcher often do not discuss the modeling choice formally but verbally, which aggravates reproducibility of results and might lead to mistakes when implementing the specific model in a different context. Remarkably, one of the earliest applications of event studies in economics very transparently discussed the role of both never-treated units and binning for the identification of dynamic treatment effects (McCrary, 2007, pp. 334).

Second, we show that event study designs *with binned endpoints* and distributed-lag (DL) models are equivalent. To be precise, the DL model is just a reparametrization of the ES model with binned endpoints; event study estimates can be recovered from DL estimates by properly normalizing the DL model and cumulating the post-treatment and pre-treatment effects away from zero. This isomorphism provides an alternative, more transparent and intuitive way to understand the role of parameter restrictions for identification. The isomorphism also offers the distributed-lag model as an alternative implementation in statistical software which is less error-prone.

Third, we show that the simple event study can be generalized to account for multiple events and/or events of different signs and intensities of the treatment. Such institutional settings are common in public and labor economics; consider, for instance, a sequence of state-level minimum wage or tax reforms of different sign and sizes. However, there was some disagreement in the field of whether event study designs are able to accommodate for settings with multiple events. In the past, DL models were commonly recommended in such settings. We show that the equivalence result between DL and ES models also holds in the general case if and only if endpoints of the effect window are binned. Moreover, we discuss important modeling assumptions that are necessary when applying event studies in environments with multiple, heterogeneous events.

While some of these points may not surprise theoretical econometricians, our characterization of the applied literature shows that more care is needed when specifying the models. Our survey of applied papers in the very top journals shows that many empirical models are not fully specified. While the concrete empirical implementation is usually correct, the incomplete empirical models often serve as references for other studies, which might unknowingly lead to incorrect replications of the original model with strong or implausible implicit identifying assumptions.

In the final part of the paper, we demonstrate the practical relevance of our three contributions, replicating and expanding the study by Baker and Fradkin (2017), henceforth BF2017. In the original paper, the authors suggest a neat way to measure worker search intensity based on Google search data. BF2017 apply their new measure to test whether search intensity responds negatively to increases in the potential benefits duration (PBD) induced by state-level reforms following the Great Recession. While their difference-in-difference estimates clearly show the expected negative relationship, the original event study results are inconclusive. We show that implementing the generalized event study design yields statistically highly significant dynamic effects, which are well in line with the difference-in-difference estimates. Hence, implementing our preferred specification of an event study design strengthens the credibility of the novel measure of search effort suggested by Baker and Fradkin (2017) and provides even stronger support for their key empirical finding that PBD has a negative effect on search effort.

Our paper gives important practical advice to researchers implementing event study designs in ideal conditions. In this sense, our study differs from two recent strands of the literature that investigates event study designs when assumptions of treatment exogeneity or homogeneous treatment effects are not met. The first strand of the literature deals with violations of the common trend assumption. Freyaldenhoven et al. (2019) shows how to extend the standard event study design to account for unobserved confounders generating a pre-treatment trend in the outcomes and still recover the causal effect of the event. Roth (2019) shows that treatment effects can be biased conditional on passing the flat pre-trend test. Malani and Reif (2015) point to the fact that non-flat pre-trends might also be due to anticipation rather than unobserved confounders. The second strand of the current methodological event study literature shows that standard event study specification do not produce average treatment effects if treatment effects are heterogeneous across cohorts (Abraham and Sun, 2018; Callaway and Sant'Anna, 2020; de Chaisemartin and D'Haultfoeuille, 2020a). This point is closely related to a current discussion on how to correctly estimate average treatment effects in difference-in-difference models when the treatment effects are heterogeneous (see Athey and Imbens, 2018; Borusyak and Jaravel, 2017; Callaway and Sant'Anna, 2018; de Chaisemartin and D'Haultfoeuille, 2020b; Goodman-Bacon, 2018a; Gibbons et al., 2019). Notably, none of the cited studies addresses environments with multiple treatments. Our paper abstracts from these two current debates as discussed in more detail in Section 2.3. Throughout the paper, we assume that the common trend assumptions holds and that treatment effects are homogeneous across cohorts and groups.

The remainder of this paper is structured as follows. Section 2 introduces a standard version of an event study design in the simplest institutional environment and discusses how limiting the effect window and binning of endpoints imposes important parameter restrictions.

In Section 3, we show that the event study model is equivalent to a standard distributed-lag model. In Section 4, we generalize the institutional environment and allow for multiple and heterogeneous events across and within units. We show that event study designs can also be used in such settings and discuss the additional adjustments and assumptions that need to be made in these cases. In Section 5, we demonstrate the relevance of our methodological points. We replicate and extend the study by Baker and Fradkin (2017) and further strengthen their results. Section 6 concludes.

2 Standard Event Study Design

In this section, we set up an event study model in the simplest institutional environment. We refer to this model as the standard event study set-up throughout the paper. We use the standard case to highlight the importance of introducing parameter restrictions to identify the model. In Subsection 2.2, we demonstrate that restricting the effect window is a practical necessity in applied work and advocate to bin the endpoints of effect windows. In Subsection 2.3, we show that binning of endpoints enables to overcome an inherent identification problem.

We start our analysis with a standard event study set-up, where each unit $i = 1, ..., N$ receives at most one single treatment at some unit-specific time e_i. All treated units may receive treatment at the same time or treatment may be staggered over time with different units receiving treatment at different points in time.The treatment effect may unfold dynamically over time, but treatment effects are assumed to be homogeneous across cohorts unlike in Abraham and Sun (2018). We seek to estimate the dynamic effects of this treatment on our dependent variable y_{it}, which we observe at different time periods $t = \underline{t}, ..., \bar{t}$.

In this set-up, the standard event study specification is given for all $t = \underline{t}, ..., \bar{t}$ by:

$$y_{it} = \sum_{j=-\infty}^{\infty} \beta_j d_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it} \tag{1}$$

where $d_{it} = \mathbb{1}[e_i = t]$ is an event indicator that takes the value 1 in the period of the treatment, e_i, and zero otherwise. Unit fixed effects are denoted by μ_i and time fixed effects by θ_t. The parameter β_j is the dynamic treatment effect j time periods after ($j \geq 0$) or before ($j < 0$) the event. All results derived in this paper also hold when the models include additional exogenous control variables X_{it}.

For the sake of clarity, we assume that equation (1) is correctly specified and the data-generating process follows standard panel data assumptions. In particular, we assume effect homogeneity in the parameters β_j, μ_i and θ_t, independence across units i and strict exogeneity of the treatment indicators d_{it}.

2.1 Normalizing the Effects

In a model with individual fixed effect, μ_i, parameters β_j are only identified up to a constant. The intuition behind is straightforward: adding a constant to β_j for all $j = -\infty, ..., \infty$ and subtracting the same constant from the unit fixed effect μ_i for units with an event between $-\infty$ and ∞ does not alter equation (1). This non-identification of the absolute size of the dynamic effects can be solved by setting a reference period to which the effects are compared to:

Remark 1 (Normalization)**.**
Dynamic treatment effects β_j are typically expressed relative to some reference period, for example one period prior to the event. The corresponding coefficient is normalized to zero, e.g. $\beta_{-1} = 0$. In practice, the normalization is implemented by dropping the event indicator for the reference period.

Typically $\beta_{-1} = 0$ is assumed, but any other periods in the effect window, including (binned) endpoints, can serve as a reference period (cf. Remark 2). Choosing a different reference period simply leads to a level shift in the estimates.

2.2 Restricting the Effect Window

In practice, researchers have to impose restrictions on the effect window to implement the event study design since β_j can never be estimated from the infinite past to the infinite future. Hence, the effect window has to be restricted to a finite number of leads and lags. There are two broad way to restrict the effect window in practice. First, it is possible to define the last lag/first lead that should be estimated and exclude all unit-period observations outside of this window from the estimation sample. By construction, (pre)-treatment effects with a longer time horizon are not estimated. As discussed below additional restrictions are necessary to identify dynamic policy effects (cf. Section 2.3). While this restriction is often applied in the case of a single treatment it is less straightforward to be implemented in case of multiple treatment (cf. Section 4). Second, it is possible to define the endpoints of the effect windows as open intervals, which also take into account past (future) treatments beyond the last lag (lead). This second approach goes along with an important, but often unstated assumption about the dynamic treatment effects, which is summarized in the following Remark 2.

Remark 2 (Restricted effect window)**.**
Restricting the effect window to a finite number of leads, $\underline{j}$, and/or lags, $\bar{j}$, requires assumptions about the nature of the effect outside of the window. It is often economically plausible

to assume that treatment effects stay constant before $\underline{j}$ and/or after $\bar{j}$, i.e. $\beta_j = \beta_{\bar{j}}$ for all $j > \bar{j}$ and $\beta_j = \beta_{\underline{j}}$ for all $j < \underline{j}$. These assumptions should be explicitly stated and defended.

Note that restricting the effect window according to Remark 2 can be readily applied in settings with multiple treatments (cf. Section 4).

Applying Remark 2, we rewrite equation (1) as

$$y_{it} = \beta_{\underline{j}} \sum_{j=-\infty}^{\underline{j}} d_{i,t-j} + \sum_{j=\underline{j}+1}^{\bar{j}-1} \beta_j d_{i,t-j} + \beta_{\bar{j}} \sum_{j=\bar{j}}^{\infty} d_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it}$$

which simplifies to our preferred standard event study specification:

$$y_{it} = \sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \mu_i + \theta_t + \varepsilon_{it} \tag{2}$$

with

$$b_{it}^j = \begin{cases} \sum_{s=-\infty}^{\underline{j}} d_{i,t-s} & \text{if} \quad j = \underline{j} \\ d_{i,t-j} & \text{if} \quad \underline{j} < j < \bar{j} \\ \sum_{s=\bar{j}}^{\infty} d_{i,t-s} & \text{if} \quad j = \bar{j}. \end{cases} \tag{3}$$

We refer to coefficients b_{it}^j as binned event indicators, as the indicators at the endpoints, i.e. the maximum lag (lead) take into account all observable past (future) events going beyond the effect window. The definition of endpoints in equation (3) is for example used in Smith et al. (2017) and Fuest et al. (2018). Endpoints can be equivalently defined as $b_{it}^{\underline{j}} = \sum_{s=t-\underline{j}}^{\infty} d_{is}$ and $b_{it}^{\bar{j}} = \sum_{s=-\infty}^{t-\bar{j}} d_{is}$. Another equivalent definition for example used in McCrary (2007) is given by

$$b_{it}^j = \begin{cases} \mathbb{1}[t \leq e_i + j] & \text{if} \quad j = \underline{j} \\ \mathbb{1}[t = e_i + j] & \text{if} \quad \underline{j} < j < \bar{j} \\ \mathbb{1}[t \geq e_i + j] & \text{if} \quad j = \bar{j}\,. \end{cases} \tag{4}$$

By Remark 1, we drop the indicator for the period before the event, b_{it}^{-1} and normalize β_{-1} to zero.

2.2.1 Special cases of effect window restrictions.

As stated in Remark 2, it is possible to only restrict the effect window pre or post treatment. In an extreme case, one could restrict all effects prior to the event and allow the treatment effect to continue indefinitely, i.e. $\underline{j} = -1$, and $\bar{j} \to \infty$ (or equivalently $\beta_j = 0$ for all $j \leq -1$). This identification is also proposed by Borusyak and Jaravel (2017) and de Chaisemartin and D'Haultfoeuille (2020a) assuming no-anticipation and common trends or, equivalently, random timing of events. Pre-trends could be tested for a limited pre-event period by setting a value $\underline{j}$ before -1 (or equivalently $\beta_j = \beta_{\underline{j}}$ for all $j < \underline{j} < -1$). As argued above, not all post-event effects up to infinity can be estimated in practice. The maximum observable post-treatment lag length is given by the maximum number of periods between an observed event and an observed outcome for the same unit. The treatment effect for this maximum observable lag is, however, usually only identified by a few units – in the extreme case by only one. The post-treatment effect window therefore needs further shortening by choosing a maximum estimated lag. Researcher either need to restrict the effect window by either binning at the last estimated lag as in Remark 2 or dropping unit-time observations beyond the last estimated lag.

An even more extreme form of restricting the effect window is to additionally assume that effects constant at and after the event, i.e. $\underline{j} = -1$, and $\bar{j} = 0$. Hence $y_{it} = \beta\, b_{it} + \mu_i + \theta_t + \varepsilon_{it}$ where b_{it} is a dummy variable which takes the value 1 at and after the event. This is a difference-in-differences model with staggered treatment, which may be written as $y_{it} = \beta\, Treat_i \cdot Post_{it} + \mu_i + \theta_t + \varepsilon_{it}$, where $Treat_i$ is a dummy variable indicating whether unit i was treated at some point, and $Post_{it}$ is a dummy variable indicating whether unit i was treated in or before period t. Then, β is the average treatment effect relative to the pre-treatment period under the assumption of homogeneous treatment effects across cohorts and groups.

Another type of restriction, which is sometimes seen in the literature, is to restrict the effect window but without binning of endpoints. Such a model implicitly assumes that treatment effects drop to zero outside of the effect window – an assumption which is typically hard to defend (cf. the replication exercise in Section 5).

2.2.2 Length of the Effect Window

A key question for applied researchers is how to determine the length of the effect window. To first order, this depends on the research question that the model tries to address. Do we expect the effect to materialize quickly? When should they have fully materialized? Are we interested in longer run effects? What is a reasonable pre-treatment period given dependent and treatment variable and endogeneity concerns? Answers to all these questions can be based on theoretical considerations, existing empirical evidence and economic intuition. Second, the choice of the effect window is restricted by data availability. Increasing leads or lags reduces the estimation sample if treatment is not observed for these additional periods. At some point, sample size will be too small to estimate precise effects and/or have meaningful variation. Last, researchers can use the empirical estimates post-estimation to assess whether the choice of size of the effect window was reasonable. By Remark 2, we assume $\beta_j = \beta_{\bar{j}}$ for all $j > \bar{j}$ at the end of the effect window. Hence, $\hat{\beta}_{\bar{j}}$ measures the long-run treatment effect. If treatment effects had fully materialized by $\bar{j}$ after the reform, we would expect estimates leading up to $\hat{\beta}_{\bar{j}}$ to level-of and converge to $\hat{\beta}_{\bar{j}}$. If instead, there is a pronounced drop between $\hat{\beta}_{\bar{j}}$ and $\hat{\beta}_{\bar{j}-1}$, this is an indication that the effect is still unfolding. We summarize these consideration in the following remark.

Remark 3 (Length of the effect window)**.**
Researchers should experiment with different effect windows length (bearing in mind that the estimation sample might change), and assess empirically whether the underlying assumption of restricting the effect window at $\bar{j}$ given in Remark 2 is justified. Practically, researchers can check whether estimates leading up to the endpoint converge towards $\hat{\beta}_{\bar{j}}$.

2.3 Identification

It is important to assure that the model is econometrically identified such that the dynamic effects β_j are distinguished from secular time fixed effects θ_t. Throughout the paper, we understand identification as the purely mechanical recovery of the homogeneous coefficients of interest β and not their causal interpretation as average treatment effects.[3]

2.3.1 Identification with a Never-Treated Group

Assume that there are units which are never treated. These never-treated units will serve as a control group, which uniquely identifies the secular time trends θ_t if there is at least one control group observation for each period t.

In order to identify the dynamic effects β_j, we need to observe at least one treated unit for each lag and lead j of the effect window. No additional identifying assumptions are required in the presence of a never-treated control group.

The existence of a never-treated control group is, however, less trivial than it may appear at first sight. *Observing* a unit not be treated does not imply that this unit is never-treated. A unit could have been treated before or after the observed sample period or even in the yet unrealized future. When treatment effects are allowed to have effects into the infinite future (or affect the infinite past), no unit is known to be never-treated with certainty. There

[3] Recall that we assume effect homogeneity of the parameters β_j, μ_i and θ_t in equation (1), independence across units i and strict exogeneity of the treatment indicators d_{it}. The identification problem discussed in this section could also be avoided by dropping either time effects θ_t or unit fixed effects μ_i. However, dropping either dimension of fixed effects typically leads to omitted variable bias unless the treatment is randomized.

are situations in which some units cannot possibly be treated. For example, in a study on the effects of giving birth to a child, men will constitute a natural never-treated group. But if a unit cannot possibly be treated for its special characteristics, the very same special characteristics are potential confounders. Or in the logic of the Rubin causal model, all units can be manipulated, at least conceptually, so that treatment happens. Hence, relying on a never-treated control group is not a fully convincing identification strategy in most applications.

2.3.2 Identification without a Never-Treated Group

Identification of dynamic treatment effects is more challenging in the absence of a never-treated group. Borusyak and Jaravel (2017) nicely show that with an infinite effect window, $[\underline{j}, \bar{j}] = [-\infty, \infty]$, dynamic effects are only identified up to a linear trend. Formally, we can cast this underidentification problem in our notation as follows: $y_{it} = \sum_{j=-\infty}^{\infty} \beta_j d_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it} = \sum_{j=-\infty}^{\infty} (\beta_j + \lambda j) d_{i,t-j} + (\theta_t - \lambda t) + \tilde{\mu}_i + \varepsilon_{it}$ where $\tilde{\mu}_i = \mu_i + \lambda e_i$.Hence, one can add a linear trend in j, λj, to the dynamic treatment effects and adjust secular time fixed effects θ_t and unit fixed μ_i to maintain the same predicted values of the model.[4] In practice, this underidentification can easily be overlooked as many statistical packages automatically drop regressors in the case of multicollinearity. A non-identified linear trend leads to dropping either one event dummy or one time dummy.

In the following, we show that restricting the effect window as proposed in Remark 2 introduces restrictions that allow separately identifying dynamic effects, β_j, and secular time trends, θ_t. Restricting the effect window as in Remark 2 leads to our equation (2), $y_{it} = \sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \mu_i + \theta_t + \varepsilon_{it}$. In this model, adding a linear trend λj to the restricted number of dynamic treatment effects does not offset adding a linear trend to the secular time trend for the observations outside of the effect window, i.e. if $t < e_i - |\underline{j}| + 1$ or $t > e_i + \bar{j}$:[5]

[4] The underidentification problem arises if all units are treated at some point and $\sum_{j=-\infty}^{\infty} j\, d_{i,t-j} = \sum_{j=-\infty}^{\infty} j\, \mathbb{1}[t = e_i + j] = t - e_i$ for all units i and all time periods t.

[5] As

$$y_{it} = \left(\sum_{j=\underline{j}}^{\bar{j}} (\beta_j + \lambda j) d_{i,t-j} \right) + (\theta_t - \lambda t) + \mu_i + \lambda e_i + \varepsilon_{it}$$

$$= \begin{cases} \left[\sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \theta_t + \mu_i + \varepsilon_{it} \right] + \lambda \left[\underline{j} - t + e_i \right] & \text{if } t < e_i - |\underline{j}| \\ \sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \theta_t + \mu_i + \varepsilon_{it} & \text{if } e_i - |\underline{j}| \leq t \leq e_i + \bar{j} \\ \left[\sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \theta_t + \mu_i + \varepsilon_{it} \right] + \lambda \left[\bar{j} - t + e_i \right] & \text{if } t > e_i + \bar{j} \end{cases} \tag{5}$$

$$\neq \sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \theta_t + \mu_i + \varepsilon_{it} \tag{6}$$

Treated observations outside the event window serve as a control group and help to pin down secular time trends. Restricting the effect window may even produce units i for which we only observe the outcome $|\underline{j}|$ or more periods before the event. The observed outcomes of such units will all be affected by the same constant effect $\beta_{\underline{j}}$. Hence all changes over time can be attributed to period fixed effects and orthogonal noise. The analogous argument holds if the outcome is only observed $\bar{j}$ or more periods after the event. Hence, binning of one or both endpoints allows for identification with or without the presence of never-treated units. We summarize this in the following remark.

Remark 4 (Identification and restricted effect window).
Unit-period observations outside of the effect window serve as control group observations. The length of the effect window hence directly affects identification and helps to separately identify dynamic treatment and secular time fixed effects.

In Appendix B, we present intuitive and highly stylized examples demonstrating how identification is achieved technically. To summarize, the model is econometrically identified if two conditions are fulfilled: (i) for each lag/lead j, there is at least one unit i with an observation j periods after/before the event; (ii) for at least one endpoint ($\underline{j}$ or $\bar{j}$) observed for some unit i in some period t, there is at least one other unit $\ell \neq i$, which is outside of its effect window in the same period t. Condition (ii) is automatically satisfied in the presence of at least one never-treated unit. Condition (i) identifies all other effects either from a direct comparison with a control group or from an iterative comparison of effects. The identified endpoint allows backing out all other treatment effects and all time fixed effects iteratively – akin to the econometric identification in staggered treatment difference-in-differences designs.

Borusyak and Jaravel (2017) propose an alternative to overcome the underidentification issue in event studies with infinite effect windows by dropping a second event indicator from the regression. It depends on the empirical analysis at hand whether binning endpoints or restricting a second parameter is preferable in single-treatment environments. If units can receive multiple treatment, it is natural to bin the endpoints of the effects.

3 Event Studies and Distributed-lag Models

In this section, we show that event study (ES) models with binned endpoints and distributed-lag (DL) models yield identical parameter estimates. In Subsection 3.1, we formally demonstrate under which restrictions ES and DL models yield equivalent dynamic treatment effects. In Subsection 3.2, we discuss the practical implications of this isomorphism and argue that DL models are easier to implement and less-error prone. In Appendix Section A.1, we illustrate all formal claims using a simple numerical example.

3.1 Equivalence

We start by showing the equivalence of event study and distributed-lag models in the general case without binning of endpoints. Taking first differences of the standard event study specification given in equation (1), we can rewrite the event study specification into a distributed-lag model:

$$\begin{aligned}
\Delta y_{it} &= y_{it} - y_{i,t-1} \\
&= \sum_{j=-\infty}^{\infty} \beta_j d_{i,t-j} - \sum_{j=-\infty}^{\infty} \beta_j d_{i,t-1-j} + \phi_t + \Delta\varepsilon_{it} \\
&= \sum_{j=-\infty}^{\infty} \beta_j d_{i,t-j} - \sum_{j=-\infty}^{\infty} \beta_{j-1} d_{i,t-j} + \phi_t + \Delta\varepsilon_{it} \\
&= \sum_{j=-\infty}^{\infty} \gamma_j \, d_{i,t-j} + \phi_t + \Delta\varepsilon_{it} \\
&= \sum_{j=-\infty}^{\infty} \gamma_j \Delta x_{i,t-j} + \phi_t + \Delta\varepsilon_{it} \qquad (7)
\end{aligned}$$

where $\gamma_j = \beta_j - \beta_{j-1}$ and $\phi_t = \theta_t - \theta_{t-1}$ are time fixed effects and the event indicator $d_{i,t-j}$ is the first difference $\Delta x_{i,t} = x_{it} - x_{i,t-1}$ of what we call the treatment status x_{it} . In the standard case with a single binary treatment, the treatment status x_{it} is a dummy variable with an arbitrary constant as initial value, for example zero, that increases by 1 if an event occurred in period$t$. Parameters γ_j are the incremental changes of the treatment effects β_j, measuring the slope of treatment effects from one time period to the next. The distributed-lag specification in equation (7) is the first difference of the following distributed-lag specification in levels

$$y_{it} = \sum_{j=-\infty}^{\infty} \gamma_j \, x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it} \qquad (8)$$

where μ_i denotes unit fixed effects. Note that the distributed-lag specification is either a regression of levels on levels (eq. 8) or of changes on changes (eq. 7) while the event-study specification is a regression of levels on (binned) changes (eq. 2).

We proved the equivalence between event study and distributed-lag models in the general case without restricting the effect window. Next, we show that the equivalence between ES and DL models for restricted effect windows holds *only* if endpoint are binned as in

Remark 2. The distributed-lag parameters γ_j are related to the event study parameters β_j by $\gamma_j = \beta_j - \beta_{j-1}$. Binning the upper endpoint, $\beta_j = \beta_{\bar{j}}$ for all $j > \bar{j}$, is therefore equivalent to assuming that $\gamma_j = 0$ for all $j > \bar{j}$; for the lower endpoint $\beta_j = \beta_{\underline{j}}$ for all $j < \underline{j}$ is equivalent to $\gamma_j = 0$ for all $j \leq \underline{j}$. The event study model with restricted effect window between $\underline{j}$ and $\bar{j}$ and binned endpoints

$$y_{it} = \sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \mu_i + \theta_t + \varepsilon_{it} \tag{9}$$

is therefore equivalent to a distributed-lag specification with $\bar{j}$ lags and $|\underline{j}| - 1$ leads

$$y_{it} = \sum_{j=\underline{j}+1}^{\bar{j}} \gamma_j x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it} \tag{10}$$

Without binning, ES and DL specifications are based on different parameter restrictions and yield different parameter estimates. We summarize this result in the following remark:

Remark 5 (Equivalence of Event Study and Distributed-Lag Model).
The event study specification with binned endpoints at $j = \bar{j}$ and $j = \underline{j}$ as specified in equations (2) and (4) is equivalent to a distributed lag models with $\bar{j}$ lags and $|\underline{j}| - 1$ leads as given by equation (10).

It is important to note that distributed-lag coefficients measure treatment effect changes, such that one fewer lead has to be estimated: we include leads and lags running from γ_j from $\underline{j} + 1$ (not $\underline{j}$ as in the event study design) to $\bar{j}$. Then event study parameters β_j can be calculated from the distributed-lag parameters γ_j by using the difference equation $\beta_j = \beta_{j-1} + \gamma_j$. The starting point for this difference equation is given by the normalization in Remark 1. Normalizing to one period prior to the effect, i.e. $\beta_{-1} = 0$, treatment effects β_j can be uniquely recovered as

$$\beta_j = \begin{cases} -\sum_{k=j+1}^{-1} \gamma_k & \text{if} \quad j \leq -2 \\ 0 & \text{if} \quad j = -1 \\ \sum_{k=0}^{j} \gamma_k & \text{if} \quad j \geq 0. \end{cases} \tag{11}$$

We summarize this result in the following remark:

Remark 6 (Recovery of treatment effect from the distributed-lag model).
Dynamic event study treatment effects β_j are recovered from distributed-lag parameters γ_j as cumulative effects starting from a reference period, typically the period prior to the effect, according to equation (11).

As in the event study model, we need a normalization in the distributed-lag model since parameters β_j are only identified up to a constant due to the individual fixed effect μ_i (cf. Remark 1). Equation (11) shows how to recover the dynamic treatment effects β_j as the sums of distributed-lag parameters γ_j. Concretely, for post-treatment effects $j > -1$, we intuitively cumulate upwards: $\beta_j = \beta_{j-1} + \gamma_j$ with $\beta_{-1} = 0$. Importantly, for pre-treatment effect $j \leq -1$, we cumulate *downwards* with a *negative* sign: $\beta_j = \beta_{j-1} - \gamma_{j-1}$ with $\beta_{-1} = 0$. For instance, $\beta_{-2} = -\gamma_{-1}$; we *must not* assume $\gamma_{-1} = 0$.[6]

3.2 Practical Implications

While the derivation of equivalence results of Remark 5 is mathematically straightforward, it has useful practical implications. We briefly discuss the most important ones in the following subsection.

Model choice. The estimates of the parameters from a event study design with binned endpoints and corresponding cumulative distributed-lag parameters are numerically equivalent. The choice of the model is therefore purely a question of convenience, yet there are some practical (dis-)advantages for both models to be discussed below. Most importantly in our view, the equivalence implies that if one regards the distributed-lag model with a limited number of leads and lags as a sensible econometric model, binning of the endpoints in event study designs must be equally sensible.

Binning vs. cumulating. In the event study design, treatment variables have to be binned at the endpoints of the effect window according to equation (3). Consequently, the event study model delivers direct estimates of the dynamic treatment effects and therefore readily interpretable parameters. In contrast, the coefficients from the distributed-lag model $\gamma = [\gamma_{\underline{j}+1}, ..., \gamma_{\overline{j}}]'$ have to be cumulated following equation (11) to obtain the event study parameters $\beta = [\beta_{\underline{j}}, ..., \beta_{-2}, \beta_0, ...\beta_{\overline{j}}]'$. This linear transformation transfers the statistical properties (consistency and asymptotic normality) of $\widehat{\gamma}$ to the calculated $\widehat{\beta}$. Standard errors of $\widehat{\beta}_j$ can be calculated from the variances and covariances of the vector $\widehat{\gamma}$ by the usual formula for linear combinations and are identical to the direct event study estimates.

Data Requirements The isomorphism between ES and DL models is also insightful when thinking about the data requirements to estimate the model. In order to include observation y_{it} in the sample, the distributed lag model given in equation (10) reveals that treatment

[6] In our example 1 with effect window from $\underline{j} = -3$ to $\overline{j} = 4$, the coefficients are $\beta_{-3} = -(\gamma_{-1} + \gamma_{-2})$, $\beta_{-2} = -\gamma_{-1}$, $\beta_{-1} = 0$, $\beta_0 = \gamma_0$, $\beta_1 = \gamma_0 + \gamma_1$, $\beta_2 = \gamma_0 + \gamma_1 + \gamma_2$, $\beta_3 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3$, $\beta_4 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$.

status x_{it} needs to be observed from period $(t - \bar{j})$ to $(t + |\underline{j}| - 1)$. Assuming that y_{it} is observed from $\underline{t}$ to $\bar{t}$, we need to observe the treatment status x_{it} from $(\underline{t} - \bar{j})$ to $(\bar{t} + |\underline{j}| - 1)$ if all observations of the dependent variable y_{it} are to be included in the estimation sample. In case we observe event indicators $d_{it} = x_{it} - x_{i,t-1}$, we only need to observe the treatment dummy d_{it} from $(\underline{t} - \bar{j} + 1)$ to $(\bar{t} + |\underline{j}| - 1)$.[7]

Clearly, the same data requirements have to apply in the equivalent ES model, but they are less obvious. The binned event indicators defined in equation (3) seem to require treatment dummies d_{it} to be observed from $\underline{t} - \bar{j}$ to $\bar{t} + |\underline{j}|$, i.e. two more periods than in the distributed-lag specification. However, whether an event happened in $\underline{t} - \bar{j}$ or $\bar{t} + |\underline{j}|$ is irrelevant for the event study estimates. To see this, assume there was an event in $\underline{t} - \bar{j}$. Then all observations $y_{i\underline{t}}, ..., y_{i\bar{t}}$ would be treated with a distance of $\bar{j}$ periods. So all observations will be affected by the constant effect $\beta_{\bar{j}}$ which is absorbed by the unit fixed effects μ_i. The analogous argument holds for $\bar{t} + |\underline{j}|$.

A frequent practical problem is that we observe treatment status from $(\underline{t} - \bar{j})$ to $(\bar{t} + |\underline{j}| - 1)$. Let us assume that we observe treatment status for unit i only from $(\underline{t} - \bar{j} + 2)$. Not observing treatment in the data does not imply that the unit was not treated in $(\underline{t} - \bar{j} + 2)$, $(\underline{t} - \bar{j} + 1)$ or $(\underline{t} - \bar{j})$. If we can rule out that treatment happened in these periods for unit i, for instance because we analyze a setting where treatment can only happen once and we know it happened later, we can impute the missing treatment observations with zero. If we cannot rule this out, for instance because we are in setting with multiple treatments (see Section 4), the respective observations have to be dropped from the sample.

Fixed effect vs. first difference estimator. Both the event study model in equation (2) and the distributed-lag model in equation (10) are panel data models including unit and time effects. The parameters β or γ can be estimated either with standard fixed effects estimation in levels or in first differences. Both estimators are consistent and asymptotically normal under standard assumptions for panel data models. In finite samples, the estimates obtained with the fixed effect estimator differ from the ones obtained with the first difference estimator.

The deviation between fixed effects and first difference estimation is small if the dynamic nature of the effect is modeled correctly, i.e. if the effect is truly constant $\bar{j}$ periods after the event. If, however, the true treatment effect continues to unfold beyond $\bar{j}$, fixed effects and first difference estimates can differ strongly. As an example, assume that the true treatment effect is negligible at and shortly after treatment and only materializes after several post-

[7] The starting value for the treatment status $x_{i,\underline{t}-\bar{j}}$ can be set to an arbitrary number, typically zero as this constant will be absorbed by the unit fixed effects μ_i.

treatment periods. Further assume, that the researchers includes too few post-treatment parameters to capture the full treatment effect. In this case, the first difference estimator will be close to zero while the fixed effect estimator will pick up some average of the delayed response. Both estimator will clearly underestimate the true long-term response. In contrast, if the number of lags is specified such that the treatment effect has fully materialized within the effect window (cf. Remarks 2 and 3), both the fixed effects and the first difference estimator will correctly pick up the dynamic effects and correctly estimate the long-run effect.

When choosing between fixed effects and first difference, it is important to keep the equivalence of Remark 5 in mind. While it might not seem intuitive to estimate an event study specification in first differences, note that it is completely natural to estimate the numerically equivalent distributed-lag specification in first differences.

4 Generalized Event Study Design

In many applications, treatment may occur repeatedly and be of different intensities across units and/or time. In this section, we show that the standard event study design can be generalized to accommodate institutional set-ups where multiple events with known but varying treatment intensity take place. We formally derive the generalized event study in Subsection 4.1. Subsection 4.2 discusses four typical institutional environments in which the generalized event study can be applied. Appendix Section A.2 provides an empirical numerical example visualizing such a case.

4.1 Set-up and Equivalence in the Generalized Design

In the following, we set up a generalized event study design that can be used in case of multiple events of identical intensity, single events with varying treatment intensity, and multiple events of different intensities. The set-up also nests the standard event study design set up in Section 2 as a special case.

In the generalized design, the treatment variable is defined as the change in the treatment status $\Delta x_{it} = x_{it} - x_{i,t-1}$. In other word, the treatment variable measures the exact size of a change in a certain policy variable (e.g. a tax rate) from unit $t-1$ to unit t. In contrast, the treatment variable in the standard design is a dummy indicating that any change in the policy variable happened.

It is easy to see that the equivalence between event study designs and distributed-lag models shown in Section 3.1 also holds in the general case. The standard event study design with infinite event windows shown in equation (1) becomes:

$$y_{it} = \sum_{j=-\infty}^{\infty} \beta_j \Delta x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it} \tag{12}$$

where $\Delta x_{it} = x_{it} - x_{i,t-1}$.

Taking first differences of equation (12) and rewriting yields the distributed-lag model

$$\Delta y_{it} = \sum_{j=-\infty}^{\infty} \gamma_j \Delta x_{i,t-j} + \phi_t + \Delta\varepsilon_{it} \tag{13}$$

where $\gamma_j = \beta_j - \beta_{j-1}$. The distributed-lag model in levels is given by

$$y_{it} = \sum_{j=-\infty}^{\infty} \gamma_j \, x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it}. \tag{14}$$

The event study specification given in equation (12) is a regression of levels (y_{it}) on changes (Δx_{it}) which may look disturbing. However, it is derived from the equivalent distributed-lag model in levels which is a completely intuitive regression of levels (y_{it}) on levels (x_{it}). The event study specification just takes care of the re-parametrization and directly delivers the cumulative effects β_j rather than the incremental effects γ_j.

When restricting the effect window to $\bar{j}$ periods after and $\underline{j}$ before the event, the generalized event study in levels is given by:

$$y_{it} = \sum_{j=\underline{j}}^{\bar{j}} \beta_j c_{it}^j + \mu_i + \theta_t + \varepsilon_{it} \tag{15}$$

where binned treatment variables c_{it}^j are easily generated analogously to the definition for binned treatment dummies in (3):

$$c_{it}^j = \begin{cases} \sum_{s=-\infty}^{\underline{j}} \Delta x_{i,t-s} & \text{if} \quad j = \underline{j} \\ \Delta x_{i,t-j} & \text{if} \quad \underline{j} < j < \bar{j} \\ \sum_{s=\bar{j}}^{\infty} \Delta x_{i,t-s} & \text{if} \quad j = \bar{j}. \end{cases} \tag{16}$$

Note that the more common definition in (4) cannot be generalized, which is why we prefer the more versatile event indicator definition given in equations (3) or 16. The analogous distributed lag model is

$$y_{it} = \sum_{j=-\underline{j}+1}^{\bar{j}} \gamma_j \, x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it}. \tag{17}$$

Remark 1 on normalization, Remarks 2 and 4 on restricting the event window and the practical implications in Section 3.2 on estimating the event study vs. the distributed-lag models also hold in the general case.

Importantly, estimating dynamic treatment effects using the generalized event study only produces unbiased estimates under a linearity and additivity assumption, which is summarized in the following remark.

Remark 7 (Applicability of the Generalized Event Study Design).
Assuming that the treatment effect is proportional to the observed treatment intensity, the generalized event study described by equations (15) and (16) delivers consistent estimates of the dynamic treatment effect. Treatment effects can also be estimated using a distributed-lag models as specified in equation (14).

Note that it may be the case that the proportionality assumption of Remark 7 is violated. For example, treatments with opposite signs may have very asymmetric effects on the outcome (see, e.g. Fuest et al., 2018; Benzarti et al., 2020). It is therefore always advisable to test for symmetric effects. Even if treatments of opposite signs do not have symmetric effects, it is still valuable to exploit treatment variation within the set of positive and negative treatments (see application in Section 5).

As the generalized event study specification incorporates the intensity of treatment, estimated effects can be interpreted as the effect of a one-unit increase akin to the interpretation in a generalized differences-in-differences model. This way, event study estimates can be used to infer long-term effects on an intuitive scale.

4.2 Typical Cases and Applications

In this subsection, we discuss typical cases of the generalized event study design and provide selected examples from recently published applications.

Case 1: Single Events of Identical Intensity. This is the standard case discussed in Section 2.

Case 2: Multiple Events of Identical Intensity. Consider the case in which events of identical intensity take place repeatedly for a unit. Using definition (16), this implies that $\Delta x_{it} = d_{it}$ is an event dummy that takes value 1 in *any* period where an event took place and 0 in other periods (see Appendix C.1 for a numerical example). Few analyses have applied event studies in such an institutional context (see Dube et al., 2011, for an exception). However, many institutional set-ups, such as hospital admissions or firm switches, fit the model. Sometimes, only the first of potentially many events is considered in a standard event study framework as developed in Section 2. This approach leads to inconsistent estimates unless the second and subsequent events are known to have no additional effect at all.

Case 3: Single Events of Varying Treatment Intensity. Next, consider the case where each unit receives one treatment, but treatment intensity s_i differs across units, hence $\Delta x_{it} = d_{it} s_i$ in definition (16). A numerical example is given in Appendix C.2. This case is quite frequently applied as it fits an institutional setting where a shock at some aggregate level hits units at a disaggregate level with different intensities (see, e.g., Alsan and Wanamaker, 2018; Charles et al., 2018; Clemens et al., 2018; Goodman-Bacon, 2018b). Many applications of this type formally refer to the standard event study model but discuss generalization and treatment of endpoints only verbally if at all.

Case 4: Multiple Events of Different Intensities and Direction. Last, we consider the most general case, developed in Section 4.1, in which events may occur multiple times per unit and their treatment intensity differs both across individuals and across events. A numerical example is given in Appendix Section A.2. There are many settings that fit this model, such as multiple tax changes or minimum wage hikes, and correspondingly many applications. Traditionally, the respective models were framed as distributed-lag models rather than event study designs (Suárez Serrato and Zidar, 2016; Drechsler et al., 2017; Fuest et al., 2018).

A special case is when events have a different direction. Assume that d_{it} is a variable that takes the value 1 in periods with a "positive" treatment, value -1 in periods with a "negative" treatment and value 0 in periods without a treatment. The parameter β_j estimates the average effect j periods after the event of all "positive" treatments and – with reversed sign – all "negative" treatments. In other words, the effects of "positive" and "negative" treatments are assumed symmetric with opposing signs. A typical example would be the introduction of a new law in some period and the abolition of the law in some later period, or the opening and closing of plants across regions.

4.3 Dichotomizing treatment variables.

A common alternative empirical specification used when treatment effects are of different sizes is to dichotomize treatment variables and use a dummy variable that is only switched on for large events (see, e.g., Simon, 2016; Fuest et al., 2018). However, the parameter estimates of such a dichotomization are harder to interpret both in magnitude and direction. To see this, consider the following case: each unit is treated once, there are two types of treatment: a small reform $d^s_{it} = 1$ or a large reform $d^\ell_{it} = 2$; treatments are distributed randomly in time and treatment effects are linear in the intensities of the reform. Ignoring small events and applying the standard event dummy set-up yields $d^s_{it} = 0$, $d^\ell_{it} = 1$. In this case, units with small reforms become part of the the control group although they respond to the reform. This induces a bias in the time fixed effects and thereby also in the treatment coefficients. Depending on the elasticity of the treatment effect with respect to the reform intensity, the share of large vs. small reforms and the size of the effect window, it is possible that estimates in the model only using the large reforms can be larger, smaller or identical to the model using all reforms. A possible fix for this ambiguity is to exclude units with small events from the sample, in which case, the model is, however, estimated on a different and possibly selective sample. Moreover, the dichotomization of the treatment variable eliminates valuable information which could otherwise be used to identify the magnitude of the effect.

5 Application

In this section, we demonstrate the relevance of the results derived in Sections 2 to 4 by replicating and extending the study by Baker and Fradkin (2017) (BF2017).[8]We will particularly focus on the importance of restricting the effect window and on the power of the generalized event study design.

BF2017 makes an important contribution to the literature on search models and unemployment insurance (UI) by proposing a novel way to measure job search effort using Google Search data. Job search is a key parameter in theoretical search and matching models but it is notoriously difficult to quantify and measure precisely. The proposed Google Job Search Index (GJSI) is a convenient and broadly applicable way to operationalize job search in empirical studies. In the last part of the study, BF2017 apply their novel measure and test whether job search behavior responds to changes of potential benefit duration (PBD)). Theoretically, we would expect a negative effect of extended PBD on search behavior.

Empirically, the authors exploit variation in unemployment insurance generosity across US states and time, and regress the Search Index on PBD in a state-month panel. They first estimates a simple differences-in-differences model (reported in Table 7 of their paper), in which they regress GJSI (in logs) on PBD (in weeks) controlling for state and time fixed effects, state-specific quadratic time trends, state-level total unemployment (second order polynomial) and the fraction of the population in the labor force. The results clearly indicate the expected negative effect of potential benefit duration on job search. In the preferred specification (4), they find a highly significant estimate of -0.00207, which implies that a

[8] Replication code is available at https://doi.org/10.7910/DVN/LXMYV6

ten week increase in UI benefits leads to 2.07 % drop in aggregate job search.

In a next step, the authors analyze the dynamics of the relationship by implementing an event study design. We recast their preferred event study model in our notation as:

$$\ln GJSI_{it} = \sum_{j=-3}^{4} \beta_j d_{i,t-j} + w_{it}'\xi + \mu_i + \theta_t + \varepsilon_{it}, \tag{18}$$

where $GJSI_{st}$ is the the natural logarithm of the Google Job Search Index in state i and period t (year-month), $d_{i,t-j}$ is an indicator variable that indicates whether PBD in state i was changed $j \in [-3, 4]$ month before or after t. Parameter μ_i captures state fixed effects and θ_t denotes period fixed effects. The vector w_{it} captures state-year specific covariates. BF2017 control for the number of unemployment insurance claims in state i and period t (month-year) divided by state population.

Changes in PBD happen frequently and with different intensities across US states over time. The authors analyze PBD increases and decreases in separate regressions and for different time windows. Increases of PBD mainly occurred during the Great Recession up to 2011 while, decreases occurred thereafter. BF2017 consequently investigate the effects of PBD increases using data from January 2006 to December 2011 and the effects of PBD decreases using data from January 2012 to December 2015; we refer to the former as the "crisis sample" and the latter as the "recovery sample". For both increases and decreases, BF2017 only focus on large changes. For increases, d_{st} is equal to 1 if PBD in state i and period t (year-month) has increased by 13 weeks or more; for decreases, the dummy d_{it} is switched on for decreases of 7 weeks or more. In the respective models, the event indicator d_{it} is zero if (i) no change happened, (ii) a change of the same sign but with smaller absolute size occurred, or (iii) the state adjusted PBD in the respectively opposite direction. The results from these specifications are presented in specifications (3) and (5) of BF2017-Table 8 and BF2017-Figure 4.[9]

In BF2017's sample, states experience up to five large increases in the crisis and seven large decreases in the recovery sample. In Panels A and B of Figure 2, we replicate the main event study results for large increases and large decreases on the two respective samples estimating equation (18). Our results are identical to the original version. Unlike the results from differences-in-differences model, the BF2017 event study estimates do not point to a strong negative relationship between search effort and PBD. However, the results depicted

[9] In columns (1), (2), (4) of Baker and Fradkin (2017)'s Table 8, the authors estimate different specifications, in which they focus on the largest single change observed within a state, exclude observations when other changes happen within this largest event's window and/or match control state-time-periods for the respective largest changes without any PBD decrease. While we replicate the results in our programs posted online, we only focus on Baker and Fradkin (2017)'s preferred models here.

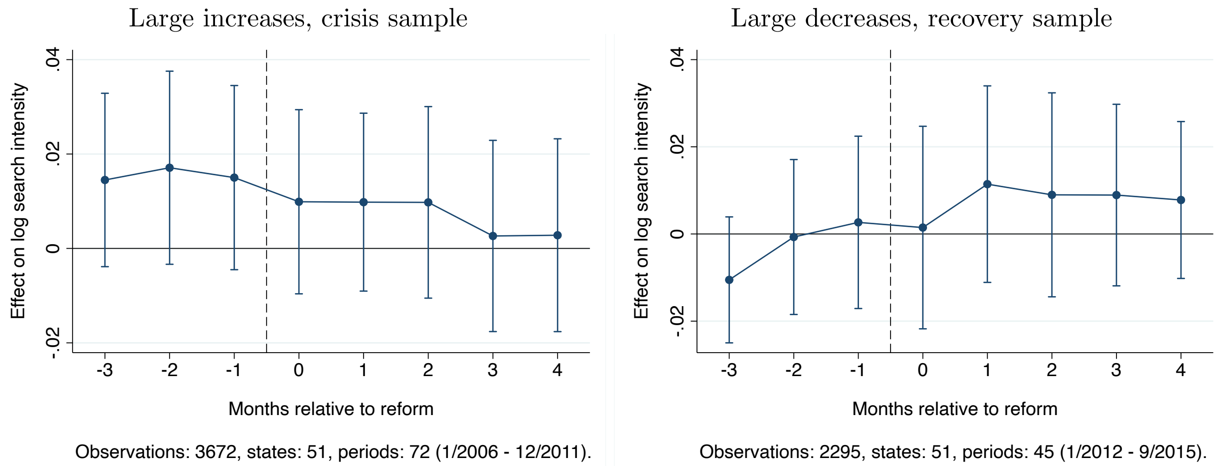
in Panel A of Figure 2 are based on strong implicit assumptions and parameter restrictions embodied in equation 18, which speak directly to our main points raised in the previous sections. While the empirical model looks like a classic event study design and therefore innocuous at first sight, event indicators d_{it} are not binned at the endpoints (cf Remark 2) and no coefficient is normalized to zero (cf Remark 1). This implies that treatment effects are implicitly normalized to be zero four and more periods before the event as well as five and more periods after the event, i.e. $\beta_j = 0$ for all $j \leq -4$ and for all $j \geq 5$. In particular, the assumption $\beta_5 = 0$ is very strong since it assumes that the effect builds up over 4 years and then immediately drops to zero (cf. Remark 4). In contrast, binning of endpoints assumes that the effect builds up over 4 years and stays constant thereafter which is an assumption more in line with the theoretical priors.

Next, we estimate equation (18) as an event study model with restrictions suggested in Sections 2.2 and 4. We bin endpoints according to Remark 2 and we normalized the pre-event coefficient $\beta_{-1} = 0$ according to Remark 1. As events can occur several times per state in our application, this leads to Case 2 "multiple events of identical intensity" in our Section 4.2. The β-coefficients can be estimated by creating binned treatment indicators at the endpoints $j = -3$ and $j = 4$ according to equation (3). Alternatively, γ-coefficients can be estimated in a distributed-lag model with 4 lags and 2 (not 3) leads and β-coefficients can be recovered according to equation (11). The two methods are equivalent and lead to identical parameter estimates and standard errors as shown in Remark 5. The choice of the estimation method is purely a question of convenience as explained in Section 3.2. Panel B of Figure 2 shows results with binned endpoints and normalized pre-event period. Different from the original results in Panel A, large increases of potential benefit duration (PBD) have a negative effect on job searches building up over 4 months and becoming statistically significant at the 5%-level 3 and 4 months after the increase. The long-term effect is estimated as -0.036 (s.e. = 0.012), i.e. a fall in job searches by 3.6% for every large increase in potential benefit duration by 13 weeks or more. There are no significant effects prior to the large increase in PBD indicating that the parallel trends assumption is satisfied prior to the treatment. Hence, the estimated dynamic treatment effects are fully consistent with the simple difference-in-differences estimation. In contrast, the large decreases occurring during the recovery period after the Great Recession do not seem to have a systematic effect on search intensity as shown in subgraph B2.

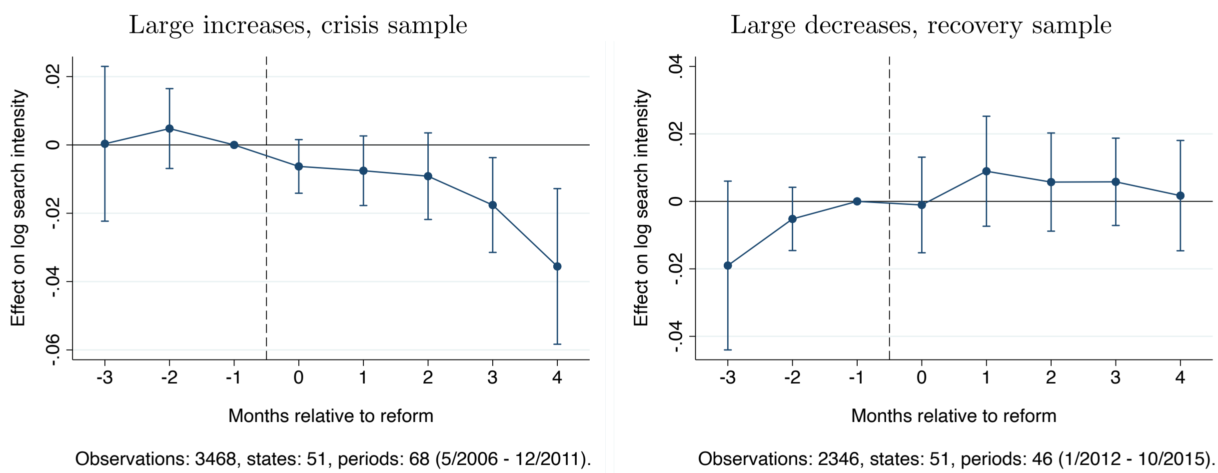
Note that the number of observations differs between Panels A and B of Figure 2. This is due to dropped observations from missing treatment information (see the paragraph on data requirements in Section 3.2). For increases, i.e. the crisis sample, the dependent variable is observed from 1/2006 to 12/2011. For the last month (12/2011), we are able to gener-

Figure 2: Baseline Results and the Role of Binning

Panel A: No binning and no normalization at -1 (Baker and Fradkin, 2017)



Panel B: Binning and normalization at -1 (own calculations)



Notes: The figure replicates and extends the main event study estimates reported in (Baker and Fradkin, 2017), BF2017. The graphs show point estimates and 95%-confidence intervals based on standard errors clustered by states. Graphs in Panel A replicate the estimates reported in specifications (3) and (5) of BF2017-Table 8 and plotted in the two panels of BF2017-Figure 4. The left graph in Panel A plots the dynamic effect of large increase (at least 13 weeks) in potential benefit duration (PBD) on log search intensity as measured with the newly proposed Google Job Search Index (GJSI). States that experiences no changes in a certain months or smaller changes, including negative ones are in the control group. The right graph in Panel A shows the analogous results for large PBD decreases (at least 7 weeks). Panel B extends the original specifications by binning endpoints of the effect window according to Remark 2 and by normalizing the effect at the pre-event period to zero according to Remark 1. All models are estimated in levels with state and time fixed effects.

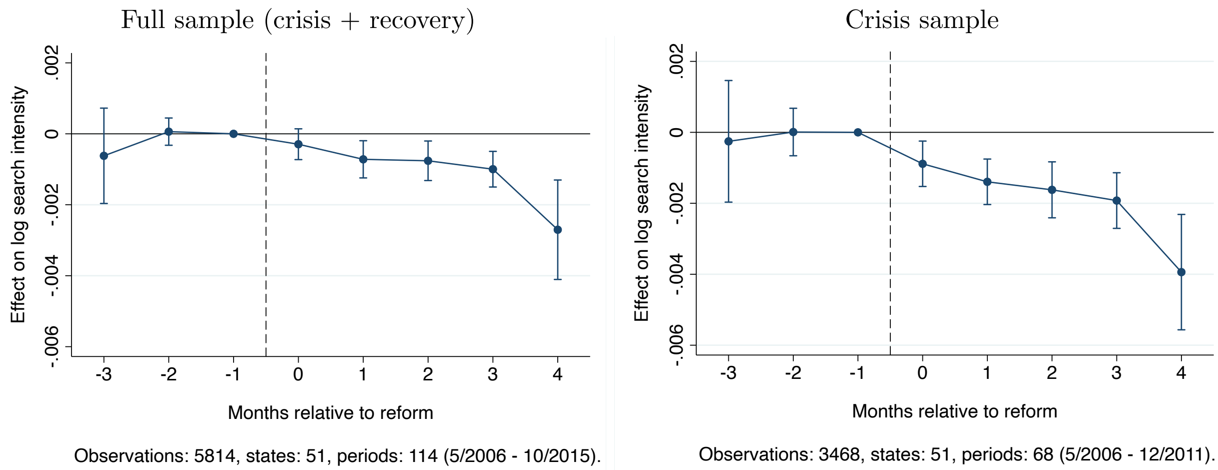
ate all leads up until $j = -3$ as we observe treatment status until 12/2015. However, we can calculate the first binned endpoint for a specification with four lags b^4_{it} only in 5/2006. Consequently, our sample is $\underline{j} = 4$ periods shorter and $4 \cdot 51 = 204$ observations smaller. An analogous argument applies for the decrease specification and the corresponding recovery sample. Here, BF2017 use observations of the dependent variable from 1/2012 to 12/2015. Given that we observe treatment status from 1/2006, we can generate all lags at time 1/2012. However, we cannot generate all leads in 12/2015. We have to shorten our estimation sample by $\bar{j} - 1 = 2$ periods. The sample is automatically reduced to the correctly shortened sample when the distributed-lag model is estimated as we discuss in Section 3.2. By estimating the models on the respective larger samples, Baker and Fradkin (2017) implicitly assume that there are no changes in the PBS prior to 1/2006 and after 12/2015, which might be true, but would need to be demonstrated or at least explicitly assumed.

In their event study, BF2017 follow standard practice and dichotomize the changes in the PBD into a zero-one treatment dummy, which is only switched on for large reforms. While Panel B of Figure 2 shows that binning endpoints leads to convincing event study coefficients, which match the difference-in-differences estimates, the zero-one models do not use all available information. First, increases and decreases are estimated in two separate models (and samples). Second, smaller changes are ignored and used as control group observations, i.e. untreated observations. In the following, we therefore estimate a generalized event study design of Case 4 that exploits all available variation. Moreover, we estimate the model on the full sample, merging the "crisis" sample (1/2006 – 12/2011) and the "recovery" sample (1/2012 – 2015).

As described in equations (15) and (16), all events are scaled with the respective treatment intensities, i.e. the changes in PBD of different magnitudes. The resulting left graph in Figure 3 shows a strong and more precisely estimated negative effect of potential benefit duration (PBD) on job search effort ($GJSI$). Pre-trends are reasonably flat and never significantly different from zero, which corroborates the parallel trend assumption of the research design. As expected confidence bands are much tighter as this specification uses all available variation in the data to identify the policy effects. In terms of magnitude, a 10-week increase in potential benefit duration leads to a decrease in log job search activity of -0.027 (s.e. = 0.007), i.e. 2.7%, after 4 months. Conventionally, the estimates of the generalized event study design are measured on the same scale as simple difference-in-differences model and can be readily compared (see below for more details).

Merging crisis and the recovery samples is not per se the right thing to do. The generalized event study relies on the assumption that treatment effects are proportional to observed treatment intensities as stated in Remark 7. In the context of the replication, the remark

Figure 3: Generalized Event Study Design

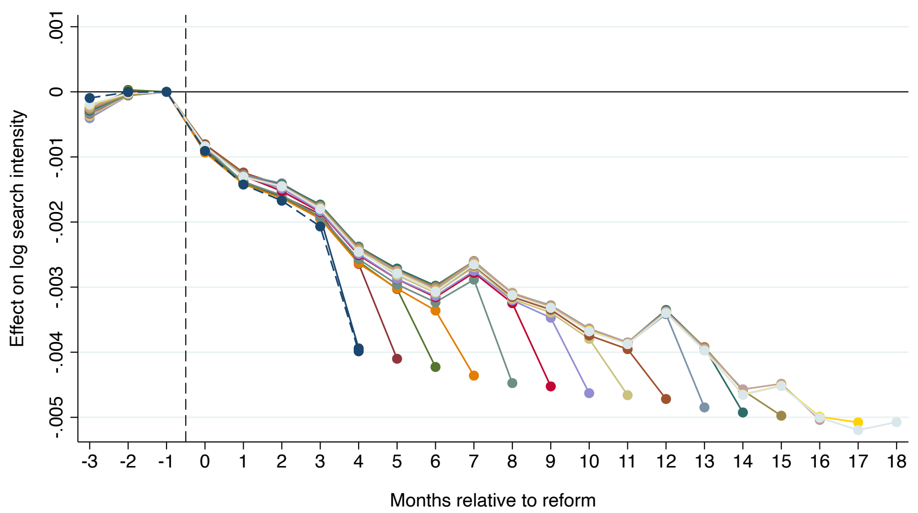


Notes: The figure plots the results when applying the generalized event study as defined by equations (15) and (16) to the setting in (Baker and Fradkin, 2017). The graphs shows the dynamic effect of an increase the potential benefit duration by one week on log search intensity as measured by Google Job Search Index (GJSI). 95% confidence intervals are plotted.

implies symmetry between increases and decreases. It is crucial to test these assumptions, e.g. by separating between treatments of different signs (see, e.g., Fuest et al., 2018; Benzarti et al., 2020) and/or splitting by clear-cut time periods as done by BF2017. Panel B of Figure 2 has already pointed to asymmetric effects, with increases in PBD leading to a strong and significant negative effect in search intensity, while decreases in PBD show no effect. For this reason, we also estimate the generalized event study model on the crisis sample only, where mainly increases occurred. The right graph in Figure 3 shows that effects are stronger when focusing only on the crisis sample and pre-trends become even flatter. Hence, there are good reasons to follow BF2017 and analyze the crisis and the recovery sample separately – either because increases and decreases of PBD have asymmetric effects or because treatment effects are different during crisis and recovery period or both. We make the crisis sample our baseline sample for the remainder of the analysis.

Next, we study the role of determining the length of the effect window. By Remark 2, binning of endpoints comes along with the assumption that treatment effects have fully materialized after $\bar{j}$ periods. In Figure 3, we see that treatment effects are still on the decline four months after the reform. Moreover, the slope of the event study graph becomes steeper between lag 3 and lag 4. As argued in Remark 3, this is an indication that treatment effects have not fully materialized within the effect window and that the assumption of Remark 2 might not hold. We explore this in the following.

Figure 4: Varying the Effect Window (Crisis Sample)

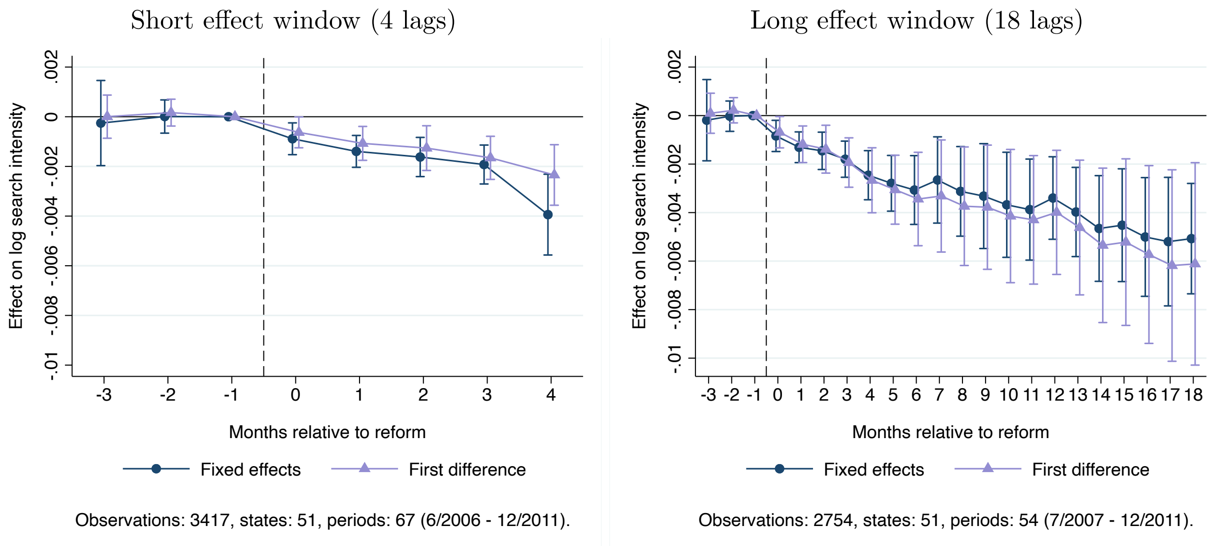


Notes: The figure plots the results when applying the generalized event study as defined by equations (15) and (16) to the setting in (Baker and Fradkin, 2017). The graphs shows the dynamic effect of an increase the potential benefit duration by one week on log search intensity as measured by Google Job Search Index (GJSI) for specifications with a varying number of lags. The dashed blue estimate the standard effect window with four lags in the estimation sample of the specification with the longest effect window of 18 lags. Confidence intervals are omitted.

One procedure to determine the length of the effect window is to simply increase the number of lags until the treatment effect flattens out. However, this approach comes at a cost as it will often reduce sample size and precision. Nonetheless, we re-estimate the generalized event study design given in equations (15) and (16) gradually increase $\bar{j}$ to one and a half years (18 months). Results are presented in Figure 4. The figure suggests that treatment effects have fully materialized approximately after 16 months. As a result, the long-run effect of PBD on search intensity is around -0.005 (0.001). This effect is higher than the DiD estimate of -0.002 because DiD is an average of the smaller short-run effects and the larger long-run effects.

While increasing the length of the effect window may be possible in some applications, data restrictions and sample size might prevent researchers from reaching the point at which treatment effects have fully materialized. An alternative check to assess whether the effect window is long enough is to compare estimates from a model specified in levels and estimated with unit fixed effects with estimates from a model estimated in first differences. At the end-point of the effect window, the first difference model only accounts for the change happening from $\bar{j}-1$ to $\bar{j}$, while the fixed effects model takes into account a weighted average of the remaining changes. As a result, coefficients from the fixed effects and the first difference specification will deviate if the effect has not fully materialized within the given effect win-

Figure 5: Fixed Effects vs. First Differences (Crisis Sample)



Notes: The figure plots the results when applying the generalized event study as defined by equations (15) and (16) to the setting in (Baker and Fradkin, 2017). The graphs shows the dynamic effect of an increase the potential benefit duration by one week on log search intensity as measured by Google Job Search Index (GJSI) for specifications estimated in levels with a fixed effects model (circle) and in first differences (triangle). 95% confidence intervals are plotted.

dow. This pattern is nicely demonstrated in Figure 5, which shows a clear deviation between first difference and fixed-effects estimates for a short (Panel A) but smaller differences for a longer effect window (Panel B). Clearly, in case the effect window is too short and treatment effects unfold monotonically, the long-run estimates will be biased toward zero. [10]

6 Conclusion

This paper makes three interrelated methodological points, which are important to bear in mind when setting up event study designs in economics. The points are valid in general, and might be particularly helpful when applying the event study technique to settings in public and labor economics with multiple policy shocks of different intensities.

First, researchers need to define an effect window, i.e. the window within which the effect is studied. While this choice is a practical necessity due to limited data availability, it is far from being innocuous. Setting the number of leads and lags to a finite number, practically

[10] Note that there is no ex ante prediction on whether the effect of the first-difference or fixed effects model should be smaller or larger. As sample size grows both models should eventually yield identical estimates.

requires to define the last lag (lead) as an open interval capturing all known events that (will) have happened in the past (future). We refer to this practice as binning. We show that binning affects which unit-period observations are assigned to treatment or control group and thus directly affects the identifying assumption. At the same time, binning introduces important parameter restrictions, which help to identify the model econometrically.

Second, we demonstrate that event study designs and distributed-lag models are equivalent. To be precise, the distributed-lag model is a reparametrization of an event study *with binned endpoints*. Event study estimates can be recovered from distributed-lag models by cumulating the post-treatment and pre-treatment effects away from zero. We use this isomorphism to reinforce the necessity and importance of limiting the effect window properly and critically discuss the plausibility of alternative parameter restrictions used in the literature. The distributed-lag model is in our view also less error-prone in the practical implementation.

Third, we generalize the simple event study with single event dummy events to account for multiple events and/or events of different sign and intensity of the treatment. We show that the event study methodology is perfectly applicable to such environments and that the equivalence between event study and distributed-lag models also holds in the general case. We point to the necessary underlying assumptions and briefly discuss where generalized event study designs could be implemented in light of current empirical research.

In a final part of the paper, we demonstrate the practical relevance of our three methodological points replicating and discussing the event study in Baker and Fradkin (2017).

References

Abraham, S. and L. Sun (2018). Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects. *Working Paper, arXiv:1804.05785*.

Alsan, M. and M. Wanamaker (2018). Tuskegee and the Health of Black Men. *Quarterly Journal of Economics 133*(1), 407–455.

Angrist, J. D. and J.-S. Pischke (2010). The Credibility Revolution in Empirical Economics: How Better Research Design Is Taking the Con out of Econometrics. *Journal of Economic Perspectives 24*(2), 3–30.

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

Baker, S. R. and A. Fradkin (2017). The Impact of Unemployment Insurance on Job Search: Evidence from Google Search Data. *The Review of Economics and Statistics 99*(5), 756–768.

Benzarti, Y., D. Carloni, J. Harju, and T. Kosonen (2020). What Goes Up May Not Come Down: Asymmetric Pass through of Value Added Taxes. *Journal of Political Economy forthcoming.*

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

Callaway, B. and P. H. Sant'Anna (2018). Difference-in-differences with multiple time periods and an application on the minimum wage and employment. *arXiv e-print 1803.09015*.

Callaway, B. and P. H. C. Sant'Anna (2020). Difference-in-Differences with Multiple Time Periods. *arXiv:1803.09015*.

Charles, K. K., E. Hurst, and M. Notowidigdo (2018). Housing Booms and Busts, Labor Market Opportunities, and College Attendance. *American Economic Review 108*(10), 2947–2994.

Clemens, M. A., E. G. Lewis, and H. M. Postel (2018). Immigration Restrictions as Active Labor Market Policy: Evidence from the Mexican Bracero Exclusion. *American Economic Review 108*(6), 1468–1487.

de Chaisemartin, C. and X. D'Haultfoeuille (2020a). Difference-in-Differences Estimators of Intertemporal Treatment Effects. *arXiv:2007.04267*.

de Chaisemartin, C. and X. D'Haultfoeuille (2020b). Two-way fixed effects estimators with heterogeneous treatment effects. *American Economic Review 110*(9), 2964–2996.

Dolley, J. C. (1933). Characteristics and Proce- dure of Common Stock Split-Ups. *Harvard Business Review 11*(316-326).

Drechsler, I., A. Savov, and P. Schnabl (2017). The Deposits Channel of Monetary Policy. *The Quarterly Journal of Economics 132*(4), 1819–1876.

Dube, A., E. Kaplan, and S. Naidu (2011). Coups, Corporations, and Classified Information. *Quarterly Journal of Economics 126*(3), 1375–1409.

Freyaldenhoven, S., C. Hansen, and J. M. Shapiro (2019). Pre-event Trends in the Panel Event-study Design. *American Economic Review 109*(9).

Fuest, C., A. Peichl, and S. Siegloch (2018). Do higher corporate taxes reduce wages? micro evidence from germany. *American Economic Review 108*(2), 393–418.

Gibbons, C. E., J. C. Suárez Serrato, and M. B. Urbancic (2019). Broken or Fixed Effects? *Journal of Econometric Methods 8*(1), 1–12.

Goodman-Bacon, A. (2018a). Difference-in-Differences with Variation in Treatment Timing. *NBER Working Paper No. 25018*.

Goodman-Bacon, A. (2018b). Public Insurance and Mortality: Evidence from Medicaid Implementation. *Journal of Political Economy 126*(1), 216–262.

MacKinlay, A. C. (1997). Event Studies in Economics and Finance. *Journal of Economic Literature 35*(1), 13–39.

Malani, A. and J. Reif (2015). Interpreting pre-trends as anticipation: Impact on estimated treatment effects from tort reform. *Journal of Public Economics 124*, 1–17.

McCrary, J. (2007). The Effect of Court-Ordered Hiring Quotas on the Composition and Quality of Police. *The Amercian Economic Review 97*(1), 318–353.

Roth, J. (2019). Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends. *Working Paper*.

Simon, D. (2016). Does Early Life Exposure to Cigarette Smoke Permanently Harm Childhood Welfare? Evidence from Cigarette Tax. *American Economic Journal: Applied Economics 8*(4), 128–159.

Smith, M., D. Yagan, O. Zidar, and E. Zwick (2017). Capitalists in the Twenty-First Century. *Working Paper*.

Suárez Serrato, J. C. and O. Zidar (2016). Who Benefits from State Corporate Tax Cuts? A Local Labor Markets Approach with Heterogeneous Firms. *The Amercian Economic Review 106*(9), 2582–2624.

Appendix A Numerical Examples

A.1 The Standard Case

In the following, we illustrate the standard event study design set-up, discussed in Section 2 using a simple numerical example. The example also demonstrates the equivalence result between event study and distributed lag models summarized in Remark 5.

Example A.1. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{j} = -3$ *to* $\bar{j} = 4$*. For unit* i*, the single event takes place at* $e_i = 2005$*.*

In example A.1, the explanatory variables of the event study model in levels (equation 9) and in first differences are visualized by the following matrices, respectively.

t	b_{it}^{-3}	b_{it}^{-2}	b_{it}^{-1}	b_{it}^{0}	b_{it}^{1}	b_{it}^{2}	b_{it}^{3}	b_{it}^{4}	Δb_{it}^{-3}	Δb_{it}^{-2}	Δb_{it}^{-1}	Δb_{it}^{0}	Δb_{it}^{1}	Δb_{it}^{2}	Δb_{it}^{3}	Δb_{it}^{4}
2000	1	0	0	0	0	0	0	0								
2001	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	0	1	0	0	0	0	0	0	-1	1	0	0	0	0	0	0
2004	0	0	1	0	0	0	0	0	0	-1	1	0	0	0	0	0
2005	0	0	0	1	0	0	0	0	0	0	-1	1	0	0	0	0
2006	0	0	0	0	1	0	0	0	0	0	0	-1	1	0	0	0
2007	0	0	0	0	0	1	0	0	0	0	0	0	-1	1	0	0
2008	0	0	0	0	0	0	1	0	0	0	0	0	0	-1	1	0
2009	0	0	0	0	0	0	0	1	0	0	0	0	0	0	-1	1
2010	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0

The following matrices visualize the explanatory variables of the distributed-lag model (eq. 10) applied to Example A.1, again in levels and first-differences respectively.

t	$x_{i,t+2}$	$x_{i,t+1}$	x_{it}	$x_{i,t-1}$	$x_{i,t-2}$	$x_{i,t-3}$	$x_{i,t-4}$	$\Delta x_{i,t+2}$	$\Delta x_{i,t+1}$	Δx_{it}	$\Delta x_{i,t-1}$	$\Delta x_{i,t-2}$	$\Delta x_{i,t-3}$	$\Delta x_{i,t-4}$
2000	0	0	0	0	0	0	0							
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	1	0	0	0	0	0	0	1	0	0	0	0	0	0
2004	1	1	0	0	0	0	0	0	1	0	0	0	0	0
2005	1	1	1	0	0	0	0	0	0	1	0	0	0	0
2006	1	1	1	1	0	0	0	0	0	0	1	0	0	0
2007	1	1	1	1	1	0	0	0	0	0	0	1	0	0
2008	1	1	1	1	1	1	0	0	0	0	0	0	1	0
2009	1	1	1	1	1	1	1	0	0	0	0	0	0	1
2010	1	1	1	1	1	1	1	0	0	0	0	0	0	0

Note how the event study model with effects up to $\bar{j} = 4$ years after event and $|\underline{j}| = 3$ years before the event corresponds to a distributed-lag model with $\bar{j} = 4$ lags and $|\underline{j}| - 1 = 2$ leads. Also notice that the right matrix becomes a zero matrix if the event takes place on or before 1996 and on or after 2013. Hence, again only information of events between 1997 and 2012 is necessary to estimate the model. The four matrices can also be used to verify the condition that allow deriving our Result 5: $b_{it}^j = d_{i,t-j} = \Delta x_{i,t-j}$ and $b_{i,t-1}^j = d_{i,t-j-1} = \Delta x_{i,t-j-1}$ for $\underline{j} = -3 < j < \bar{j} = 4$ as well as $\Delta b_{it}^{\underline{j}} = \Delta b_{it}^{-3} = -d_{i,t-\underline{j}-1} = -d_{i,t-2} = -\Delta x_{i,t-2}$ and $\Delta b_{it}^{\bar{j}} = \Delta b_{it}^4 = d_{i,t-\bar{j}} = d_{i,t-4} = \Delta x_{i,t-4}$.

In example A.1, the event study effects are calculated according to equation (11) from the distributed-lag/lead coefficients as $\beta_{-3} = -(\gamma_{-1} + \gamma_{-2})$, $\beta_{-2} = -\gamma_{-1}$, $\beta_{-1} = 0$, $\beta_0 = \gamma_0$, $\beta_1 = \gamma_0 + \gamma_1$, $\beta_2 = \gamma_0 + \gamma_1 + \gamma_2$, $\beta_3 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3$, $\beta_4 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$.

A.2 The General Case

In the following subsection, we present a brief generic numerical example that features the general case derived in Section 4.

Example A.2. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{j} = -3$ *to* $\bar{j} = 4$. *For individual i, one event of intensity* $d_{i,2003} = 0.2$ *takes place in* 2003*, another event of intensity* $d_{i,2004} = -0.1$ *in* 2004 *and yet another event of intensity* $d_{i,2006} = 0.3$ *in* 2006*; there are no events in the other years.*

The following four matrices show the explanatory variables for the event study in levels b_{it}^j and in first differences Δb_{it}^j, as well as for the distributed-lag model in levels, $x_{it} = x_{it} + \Delta x_{i,t-1}$ with initial value $x_{i,t-\bar{j}} = 0$, and in first differences, $\Delta x_{it} = d_{it}$:

t	c_{it}^{-3}	c_{it}^{-2}	c_{it}^{-1}	c_{it}^{0}	c_{it}^{1}	c_{it}^{2}	c_{it}^{3}	c_{it}^{4}	Δc_{it}^{-3}	Δc_{it}^{-2}	Δc_{it}^{-1}	Δc_{it}^{0}	Δc_{it}^{1}	Δc_{it}^{2}	Δc_{it}^{3}	Δc_{it}^{4}
2000	0.4	0	0	0	0	0	0	0								
2001	0.2	0.2	0	0	0	0	0	0	-0.2	0.2	0	0	0	0	0	0
2002	0.3	-0.1	0.2	0	0	0	0	0	0.1	-0.3	0.2	0	0	0	0	0
2003	0.3	0	-0.1	0.2	0	0	0	0	0	0.1	-0.3	0.2	0	0	0	0
2004	0	0.3	0	-0.1	0.2	0	0	0	-0.3	0.3	0.1	-0.3	0.2	0	0	0
2005	0	0	0.3	0	-0.1	0.2	0	0	0	-0.3	0.3	0.1	-0.3	0.2	0	0
2006	0	0	0	0.3	0	-0.1	0.2	0	0	0	-0.3	0.3	0.1	-0.3	0.2	0
2007	0	0	0	0	0.3	0	-0.1	0.2	0	0	0	-0.3	0.3	0.1	-0.3	0.2
2008	0	0	0	0	0	0.3	0	0.1	0	0	0	0	-0.3	0.3	0.1	-0.1
2009	0	0	0	0	0	0	0.3	0.1	0	0	0	0	0	-0.3	0.3	0
2010	0	0	0	0	0	0	0	0.4	0	0	0	0	0	0	-0.3	0.3

t	$x_{i,t+2}$	$x_{i,t+1}$	x_{it}	$x_{i,t-1}$	$x_{i,t-2}$	$x_{i,t-3}$	$x_{i,t-4}$	$\Delta x_{i,t+2}$	$\Delta x_{i,t+1}$	Δx_{it}	$\Delta x_{i,t-1}$	$\Delta x_{i,t-2}$	$\Delta x_{i,t-3}$	$\Delta x_{i,t-4}$
2000	0	0	0	0	0	0	0							
2001	0.2	0	0	0	0	0	0	0.2	0	0	0	0	0	0
2002	0.1	0.2	0	0	0	0	0	-0.1	0.2	0	0	0	0	0
2003	0.1	0.1	0.2	0	0	0	0	0	-0.1	0.2	0	0	0	0
2004	0.4	0.1	0.1	0.2	0	0	0	0.3	0	-0.1	0.2	0	0	0
2005	0.4	0.4	0.1	0.1	0.2	0	0	0	0.3	0	-0.1	0.2	0	0
2006	0.4	0.4	0.4	0.1	0.1	0.2	0	0	0	0.3	0	-0.1	0.2	0
2007	0.4	0.4	0.4	0.4	0.1	0.1	0.2	0	0	0	0.3	0	-0.1	0.2
2008	0.4	0.4	0.4	0.4	0.4	0.1	0.1	0	0	0	0	0.3	0	-0.1
2009	0.4	0.4	0.4	0.4	0.4	0.4	0.1	0	0	0	0	0	0.3	0
2010	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0	0	0	0	0	0	0.3

Appendix B Identification

In the following, we present intuitive examples that demonstrate how identification is achieved. The empirical model with an effect window from $\underline{j} = -2$ to $\bar{j} = 1$ is described by equation (7), hence

$$\Delta y_{it} = \gamma_{-1} d_{i,t-1} + \gamma_0 d_{i,t} + \gamma_1 d_{i,t+1} + \theta_t + \Delta\varepsilon_{it}.$$

There are no unit fixed effects and there is no constant in this regression, so no time fixed effect has to be dropped for identification. Moreover, the examples in this appendix reveal that identification is most easily studied in the first difference version of the distributed-lag specification.

Consider the following seven examples:

Example B.1 (identified). Unit 1 is treated in $t = 2$, unit 2 is not treated, panel from $\underline{t} = 0$ to $\bar{t} = 3$.

The matrix of explanatory variables in Example B.1 is given by

t	i	$t1$	$t2$	$t3$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	.	.	
1	1	1	0	0	1	0	0	← observation of Δy_1 one period before event
2	1	0	1	0	0	1	0	← observation of Δy_1 at event
3	1	0	0	1	0	0	1	← observation of Δy_1 one period after event
0	2	.	.	.	.	.	.	
1	2	1	0	0	0	0	0	← control for Δy_1 one period before event
2	2	0	1	0	0	0	0	← control for Δy_1 at event
3	2	0	0	1	0	0	0	← control for Δy_1 after period before event

This is the example given in Borusyak and Jaravel (2017). The non-treated unit pins down the time fixed effects, which thereby can be separated from the dynamic treatment effects. The matrix of explanatory variables has full rank.

Example B.2 (not identified). Both units are treated in $t = 2$, panel from $\underline{t} = 0$ to $\bar{t} = 3$.

t	i	$t1$	$t2$	$t3$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	.	.	
1	1	1	0	0	1	0	0	← observation of Δy_1 one period before event
2	1	0	1	0	0	1	0	← observation of Δy_1 at event
3	1	0	0	1	0	0	1	← observation of Δy_1 one period after event
0	2	.	.	.	.	.	.	
1	2	1	0	0	1	0	0	← observation of Δy_1 one period before event
2	2	0	1	0	0	1	0	← observation of Δy_1 at event
3	2	0	0	1	0	0	1	← observation of Δy_1 one period after event

Clearly, the model in Example B.2 is not identified. Treatment and time effects cannot be separated. This can be remedied if we shift the treatment of one unit by one year.

Example B.3 (identified). Unit 1 treated in $t = 2$, unit 2 treated in $t = 3$, panel from $\underline{t} = 0$ to $\bar{t} = 3$.

t	i	$t1$	$t2$	$t3$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	.	.	
1	1	1	0	0	1	0	0	← observation of Δy_1 one period before event
2	1	0	1	0	0	1	0	← observation of Δy_1 at event
3	1	0	0	1	0	0	1	← observation of Δy_1 one period after event
0	2	.	.	.	.	.	.	
1	2	1	0	0	0	0	0	← control for Δy_1 one period before event
2	2	0	1	0	1	0	0	
3	2	0	0	1	0	1	0	

Example B.3 demonstrates the main intuition behind the identification when binning endpoints. The staggered treatment enables to pin down one time fixed effects for unit 2 and $t = 1$. If $t1$ is identified, we can back out d_{t-1} for unit 1, then $t2$ for unit 2, and so on. For such an iterative procedure it is necessary that we observe all event indicators in the data window, they do not have to observable completely for one unit.

Example B.4 (identified). Unit 1 treated in $t = 2$, unit 2 treated in $t = 4$, panel from $\underline{t} = 0$ to $\bar{t} = 3$.

t	i	$t1$	$t2$	$t3$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	.	.	
1	1	1	0	0	0	1	0	← observation of Δy_1 at event
2	1	0	1	0	0	0	1	← observation of Δy_1 one period after event
3	1	0	0	1	0	0	0	← control for Δy_2 one period before event
0	2	.	.	.	.	.	.	
1	2	1	0	0	0	0	0	← control for Δy_1 at event
2	2	0	1	0	0	0	0	← control for Δy_1 one period after event
3	2	0	0	1	1	0	0	← observation of Δy_2 one period before event

Again, we can iteratively separate event from time effects even though we do not observe a full set of event effects for a single unit. However, it is important that we observe at least one endpoint in a year t where the other unit is not treated.

Example B.5 (not identified). Unit 1 treated in $t = -1$, unit 2 treated in $t = 4$, panel from $\underline{t} = 0$ to $\bar{t} = 3$.

t	i	$t1$	$t2$	$t3$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	.	.	
1	1	1	0	0	0	0	1	← observation of Δy_1 one period after event
2	1	0	1	0	0	0	0	
3	1	0	0	1	0	0	0	← control for Δy_2 one period before event
0	2	.	.	.	.	.	.	
1	2	1	0	0	0	0	0	← control for Δy_1 one period after event
2	2	0	1	0	0	0	0	
3	2	0	0	1	1	0	0	← observation of Δy_2 one period before event

Here, identification is not achieved. The matrix of explanatory variables has rank 5, as e.g., $d_{t+1} = t_1 - t_3 - d_{t-1}$. The effect one period before and one period after the event are identified but the effect at the event is not observed for any unit.

Example B.6 (not identified). Unit 1 treated in $t = 1$, unit 2 treated in $t = 3$, panel from $\underline{t} = 0$ to $\bar{t} = 3$.

t	i	$t1$	$t2$	$t3$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	.	.	
1	1	1	0	0	0	1	0	$\leftarrow$ observation of Δy_1 at event
2	1	0	1	0	0	0	1	$\leftarrow$ observation of Δy_1 one period after event
3	1	0	0	1	0	0	0	$\leftarrow$ control for Δy_2 at event
0	2	.	.	.	.	.	.	
1	2	1	0	0	0	0	0	$\leftarrow$ control for Δy_1 at event
2	2	0	1	0	1	0	0	$\leftarrow$ observation of Δy_2 one period before event
3	2	0	0	1	0	1	0	$\leftarrow$ observation of Δy_2 at event

Here, identification is not achieved. The matrix of explanatory variables has rank 5, as e.g., $d_{t-1} = t2 - d_{t+1}$. Iterative identification is not possible. The reason is that only two endpoints in the data window are observed in the same year ($t = 2$).

Example B.7 (identified). Unit 1 treated in $t = 0$, unit 2 treated in $t = 1$, unit 3 treated in $t = 2$, unit 4 not treated, panel from $\underline{t} = 0$ to $\bar{t} = 1$.

t	i	$t1$	d_{t+1}	d_t	$d_{t\text{-}1}$	
0	1	.	.	.	.	
1	1	1	1	0	0	$\leftarrow$ observation of Δy_1 one period before event
0	2	.	.	.	.	
1	2	1	0	1	0	$\leftarrow$ observation of Δy_2 at event
0	3	.	.	.	.	
1	3	1	0	0	1	$\leftarrow$ observation of Δy_3 one period after event
0	4	.	.	.	.	
1	4	1	0	0	0	$\leftarrow$ control for $\Delta y_{1,t-1}$, $\Delta y_{2,t}$, $\Delta y_{3,t+1}$

All three dynamic effects are directly identified in direct comparison to a never-treated unit. The matrix of explanatory variables is full rank.

Appendix C More Numerical Examples

C.1 Multiple Events of Identical Intensity

Example C.2. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{j} = -3$ *to* $\bar{j} = 4$*. For individual* i*, a first event takes place at* 2004 *and a second at* 2006*.*

The explanatory variables for the event study in levels, $c_{it}^{j} = b_{it}^{j}$, and in first differences, $\Delta c_{it}^{j} = \Delta b_{it}^{j}$, are

t	c_{it}^{-3}	c_{it}^{-2}	c_{it}^{-1}	c_{it}^{0}	c_{it}^{1}	c_{it}^{2}	c_{it}^{3}	c_{it}^{4}	Δc_{it}^{-3}	Δc_{it}^{-2}	Δc_{it}^{-1}	Δc_{it}^{0}	Δc_{it}^{1}	Δc_{it}^{2}	Δc_{it}^{3}	Δc_{it}^{4}
2000	2	0	0	0	0	0	0	0								
2001	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	1	1	0	0	0	0	0	0	-1	1	0	0	0	0	0	0
2003	1	0	1	0	0	0	0	0	0	-1	1	0	0	0	0	0
2004	0	1	0	1	0	0	0	0	-1	1	-1	1	0	0	0	0
2005	0	0	1	0	1	0	0	0	0	-1	1	-1	1	0	0	0
2006	0	0	0	1	0	1	0	0	0	0	-1	1	-1	1	0	0
2007	0	0	0	0	1	0	1	0	0	0	0	-1	1	-1	1	0
2008	0	0	0	0	0	1	0	1	0	0	0	0	-1	1	-1	1
2009	0	0	0	0	0	0	1	1	0	0	0	0	0	-1	1	0
2010	0	0	0	0	0	0	0	2	0	0	0	0	0	0	-1	1

The explanatory variables of the distributed-lag model in levels, $x_{it} = x_{it} + \Delta x_{i,t-1} = x_{it} + d_{i,t-1}$ with $x_{i,t-\bar{j}} = 0$, and in first differences, $\Delta x_{it} = d_{it}$, are

t	$x_{i,t+2}$	$x_{i,t+1}$	x_{it}	$x_{i,t-1}$	$x_{i,t-2}$	$x_{i,t-3}$	$x_{i,t-4}$	$\Delta x_{i,t+2}$	$\Delta x_{i,t+1}$	Δx_{it}	$\Delta x_{i,t-1}$	$\Delta x_{i,t-2}$	$\Delta x_{i,t-3}$	$\Delta x_{i,t-4}$
2000	0	0	0	0	0	0	0							
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	1	0	0	0	0	0	0	1	0	0	0	0	0	0
2003	1	1	0	0	0	0	0	0	1	0	0	0	0	0
2004	2	1	1	0	0	0	0	1	0	1	0	0	0	0
2005	2	2	1	1	0	0	0	0	1	0	1	0	0	0
2006	2	2	2	1	1	0	0	0	0	1	0	1	0	0
2007	2	2	2	2	1	1	0	0	0	0	1	0	1	0
2008	2	2	2	2	2	1	1	0	0	0	0	1	0	1
2009	2	2	2	2	2	2	1	0	0	0	0	0	1	0
2010	2	2	2	2	2	2	2	0	0	0	0	0	0	1

C.2 Single Events of Varying Treatment Intensity

Example C.3. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{j} = -3$ *to* $\bar{j} = 4$*. For individual* i*, the single event of intensity* $d_i = 0.1$ *takes place at* $e_i = 2005$*.*

The explanatory variables for the event study in levels, $c^j_{it} = b^j_{it} \times s_i$, and in first differences, $\Delta c^j_{it} = \Delta b^j_{it} \times \Delta s_i$, are

t	c^{-3}_{it}	c^{-2}_{it}	c^{-1}_{it}	c^{0}_{it}	c^{1}_{it}	c^{2}_{it}	c^{3}_{it}	c^{4}_{it}	Δc^{-3}_{it}	Δc^{-2}_{it}	Δc^{-1}_{it}	Δc^{0}_{it}	Δc^{1}_{it}	Δc^{2}_{it}	Δc^{3}_{it}	Δc^{4}_{it}
2000	0.1	0	0	0	0	0	0	0								
2001	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	0.1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	0	0.1	0	0	0	0	0	0	0.1	0.1	0	0	0	0	0	0
2004	0	0	0.1	0	0	0	0	0	0	-0.1	0.1	0	0	0	0	0
2005	0	0	0	0.1	0	0	0	0	0	0	-0.1	0.1	0	0	0	0
2006	0	0	0	0	0.1	0	0	0	0	0	0	-0.1	0.1	0	0	0
2007	0	0	0	0	0	0.1	0	0	0	0	0	0	-0.1	0.1	0	0
2008	0	0	0	0	0	0	0.1	0	0	0	0	0	0	-0.1	0.1	0
2009	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	-0.1	0.1
2010	0	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0	0

The corresponding explanatory variables of the distributed-lag model in levels, $x_{it} = x_{it} + \Delta x_{i,t-1} = x_{it} + d_{it} \times s_i$ with $x_{i,t-\bar{j}} = 0$, and in first differences $\Delta x_{it} = d_{it} \times s_i$, are

t	$x_{i,t+2}$	$x_{i,t+1}$	x_{it}	$x_{i,t-1}$	$x_{i,t-2}$	$x_{i,t-3}$	$x_{i,t-4}$	$\Delta x_{i,t+2}$	$\Delta x_{i,t+1}$	Δx_{it}	$\Delta x_{i,t-1}$	$\Delta x_{i,t-2}$	$\Delta x_{i,t-3}$	$\Delta x_{i,t-4}$
2000	0	0	0	0	0	0	0							
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2003	0.1	0	0	0	0	0	0	0.1	0	0	0	0	0	0
2004	0.1	0.1	0	0	0	0	0	0	0.1	0	0	0	0	0
2005	0.1	0.1	0.1	0	0	0	0	0	0	0.1	0	0	0	0
2006	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0.1	0	0	0
2007	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0.1	0	0
2008	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0.1	0
2009	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0.1
2010	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0	0	0	0	0	0	0