ECONtribute
MARKETS & PUBLIC POLICY

ECONtribute
Discussion Paper No. 201

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

Kurt Schmidheiny Sebastian Siegloch

September 2022 www.econtribute.de

UNIVERSITÄT BONN Universität zu Köln

Funding by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866 is gratefully acknowledged.

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

Kurt Schmidheiny Sebastian Siegloch

September 2022

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 (siegloch@wiso.uni-koelnn.de) is affiliated with the University of Cologne, CEPR, ZEW, 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 this project was started. Siegloch is thankful to the German Research Foundation DFG for financing the research stay under the Research Fellowship program (# 361846460). He is funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany´s Excellence Strategy – EXC 2126/1-390838866. 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). Differences-in-differences (DD) models in general, and event study designs in particular are very popular in this respect as the underlying identifying assumption of parallel post-treatment trends is intuitive and its plausibility can be assessed (though not verified) visually by inspecting whether pre-treatment trends are parallel. Moreover, the intuition of the empirical model is straightforward. The specification usually boils down to the well-known two-way fixed effects (TWFE) 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 typically defined as a policy change whose effects are investigated. Figure 1 plots the use of event study designs in the Top Five economics journals over time.[2] While we see a gradual increase since 1990, there is a surge since 2010 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).

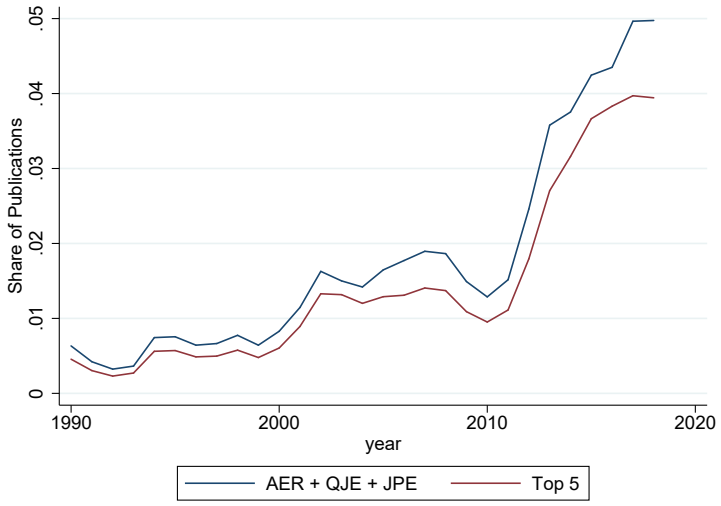
However, the intuitive appeal of event studies entails a risk to model event study designs rather loosely. In more than one third of the event study papers published in the AER, QJE or JPE since 2010, no regression equation is specified. In addition, among the studies that specify an empirical model, only 15% provide some information on the choices that have been made at the ends of the effect window – in many cases only verbally and not formally. We show below that these details of the empirical model often imply identifying assumptions, which lead to misspecification if they are not warranted. The lack of detail also impedes reproducibility and easily lead to mistakes when the methodology is applied in other contexts.

It is the primary purpose of this paper to clarify the understanding of event study designs both in methodological and practical terms. We show below that the inherent imprecision in

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

[2] We proxy the use by the share of studies mentioning the term "event study". 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.

event study specifications can easily give rise to a misspecified model or a model that relies on strong implicit assumptions. We derive three main methodological contributions 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 practical implications and pitfalls along the way, making use of simple numerical examples to visualize our main results.

The three contributions are: First, researchers need to define a range for which the dynamic treatment effects are studied. We call this range 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 dynamic treatment effects. 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*. Interestingly, an early application of event studies in economics formally discusses and implements the binning approach, but few studies refer to it (McCrary, 2007, pp. 334). We show that binning imposes implicit assumptions that enable researchers to identify dynamic treatment effects in the absence of never-treated units. Without binning it would be impossible to separate dynamic policy effects from secular time effects. We show in an application that a failure to bin endpoints leads to biased estimates, while the empirical event study model looks innocuous at first sight.Binning is also readily applicable to settings where units might receive multiple treatments (see third contribution below).

Second, we show that event study (ES) designs *with binned endpoints* and distributed-lag (DL) models with a limited number of leads and lags are equivalent. In a DL model, an outcome is regressed on leads and lags of the treatment variable, i.e. the tax rate instead of an indicator of a tax rate change. We show that the DL model is a reparametrization of the ES model. Binning of endpoints in the ES model thereby corresponds to limiting the number of leads and lags in the DL model. 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. We argue that this isomorphism provides a more transparent and intuitive way to understand the role of parameter restrictions for identification: a distributed lag model can be simply spelled-out, whereas a complete ES specification needs to specify how the endpoints of the effect window are treated (see above). The isomorphism also offers the distributed-lag model as an alternative implementation in statistical software which is less error-prone as the binning of endpoints can be tricky – especially in case of multiple treatments.

Third, we show that the simple event study design can be generalized to account for multiple events and/or events of different sign and intensity of the treatment. Such institutional settings are common in public and labor economics; consider, for instance, a sequence of state-level minimum wage changes or tax reforms of different sign and sizes. Initially, event studies in economics were predominantly used in situations, where a single event, measured by an indicator variable, hits a unit at a unique point in time. In settings with multiple events of different intensity, distributed lag models were frequently used – sometimes in addition to classic event

study models (see, e.g., Suárez Serrato and Zidar, 2016, Fuest et al., 2018).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.

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 practical advice to researchers implementing event study designs. We pay attention to the role of data structure and the type of identifying variation (single vs. multiple treatments of varying intensity) that force researchers to make decisions that are often not discussed in theoretical work, but matter for applied work. In this sense, our paper is similar to the study by Freyaldenhoven et al. (2022) who provide useful guidance for applied researchers when plotting the estimates obtained from event study designs.

We analyze event study designs under ideal conditions in two important dimensions: we assume (i) homogeneous treatment effects across cohorts, and (ii) parallel trends and no anticipation effects. The first condition (i) implies that we do not seek to contribute to an important recent econometric literature that discusses potential bias in two-way fixed effects models with heterogeneous treatment effects that differ across treatment cohorts. Such bias occurs in both event study specifications (Chaisemartin and D'Haultfœuille, 2020a, Sun and Abraham, 2021,

Callaway and Sant'Anna, 2021, Borusyak et al., 2021) and difference-in-difference models (see Gibbons et al., 2019, Chaisemartin and D'Haultfœuille, 2020b, Goodman-Bacon, 2021).[3]The proposed alternative estimators address the problem of heterogeneous treatment effects by relying on a different weighting scheme than the one implied by OLS. There is also some recent work (Callaway et al., 2021, Chaisemartin and D'Haultfœuille, 2020a, 2021) that accounts for either continuous or multiple treatments which are common in public and labor economics.

We nevertheless believe that TWFE will remain the norm or the starting point for some time. First, the new "heterogeneity-robust" estimators are not available in all settings, notably when there are multiple treatments and/or treatments of different intensity. Second, the insights from this paper remain relevant in cases, in which there is no staggered adoption and hence no bias from heterogeneity.[4]Third, the intuitive appeal of the TWFE is one of the reasons for the success of the model in the credibility revolution. Intuitive and transparent models are likely to be preferred, but, of course, the have yield consistent estimates. Hence, in settings where a new "heterogeneity-robust" estimator is available, these estimators should be applied and compared to the TWFE estimates as an informal test for heterogeneous treatment effects (Chaisemartin and D'Haultfœuille, 2020b, Cengiz et al., 2019, Callaway and Sant'Anna, 2021, Sun and Abraham, 2021, Borusyak et al., 2021, Gardner, 2021).In settings with continuous and multiple treatments where no new "heterogeneity-robust" estimator is available, one option is to dichotomize the continuous treatment variable and apply an estimator that can account for multiple treatments.[5]

The second condition (ii) implies that we do not seek to contribute to the important literature discussing fixes to 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. Rambachan and Roth (2022) propose inference methods for event study designs if the parallel trend assumption does not hold exactly. Malani and Reif (2015) point to the fact that

[3] See Chaisemartin and D'Haultfœuille (2022), Baker et al. (2022), and Roth et al. (2022) for recent surveys of the importance of heterogeneous treatment effects and the implications of binning in such contexts.

[4] Covariates can cause problems in the case of treatment heterogeneity even if there is no staggered adoption (Powell, 2021).

[5] See Lichter et al. (2021) and Siegloch et al. (2021) for examples of such a pragmatic approach and Section 4.3 for potential caveats when dichotomizing continuous treatment variables.

non-flat pre-trends might also be due to anticipation rather than unobserved confounders.

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 events of different sign and intensity. 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 contributions by replicating and extending the study by Baker and Fradkin (2017). 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 and advocate to bin the endpoints of effect windows.

We assume a panel, in which each unit $i = 1, ..., N$ is observed over time periods $t = \underline{t}, ..., \bar{t}$.[6] There is a policy variable $T_{i,t}$, which we call *treatment status*. In this section, treatment status is a dummy variable indicating, for instance, if employment protection laws have yet been passed in a certain state or not (Autor, 2003). In the classic event study design, treatment is an absorbing state, i.e. there is a single treatment which is adopted at some unit-specific time E_i and lasts forever. Hence, treatment status changes at most once at some unit-specific time E_i from 0 to 1. *Treatment adoption* is indicated by $\Delta T_{i,t} = T_{i,t} - T_{i,t-1} = \mathbb{1}[E_i = t]$, a dummy variable. Some units i may never be treated. For these never-treated units $\Delta T_{i,t} = 0$ for all t. 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. We seek to estimate the dynamic effects of this treatment on our dependent variable $Y_{i,t}$. Whenever we refer to *dynamic treatment effects*

[6] We assume a balanced panel for convenience but our remarks generalize to panels with unit-specific time windows.

we are talking about the cumulative effect of treatment at a specific time relative to the event. In contrast, we refer to incremental effects as changes in the cumulative treatment effect.

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

$$Y_{i,t} = \sum_{\ell=-\infty}^{\infty} \beta_\ell \Delta T_{i,t-\ell} + \mu_i + \theta_t + \varepsilon_{i,t} \tag{1}$$

where unit fixed effects are denoted by μ_i and period fixed effects by θ_t. The parameter β_ℓ is the dynamic treatment effect ℓ time periods after ($\ell \geq 0$). For the periods before an event ($\ell < 0$), the parameters β_ℓ either capture anticipation effects or – in their absence – are used to test for parallel pre-treatment trends and hence to assess the credibility of the identification strategy. All results derived in this paper also hold when the models include additional exogenous covariates $X_{i,t}$.

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_ℓ, μ_i and θ_t, independence across units i and strict exogeneity of the treatment $T_{i,t}$. Treatment effects are hence assumed to be homogeneous across cohorts in contrast to e.g. Sun and Abraham (2021).

2.1 Normalizing the Effects

In a model with individual fixed effect, μ_i, parameters β_ℓ are only identified up to a constant. The intuition behind is straightforward: adding a constant to β_ℓ for all $\ell = -\infty, ..., \infty$ 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_ℓ are typically expressed relative to some reference period, for example one period prior to treatment adoption. The corresponding coefficient is normalized to zero, e.g. $\beta_{-1} = 0$. In practice, the normalization is implemented by dropping the treatment adoption 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 shifts point estimates of the dynamic effects and affects their standard errors. Generally, the choice of the reference point should depend on the parameter(s) of interest. In the absence of anticipation, policy-relevant post-treatment effects are typically given by normalizing to the immediate pre-treatment period.

2.2 Restricting the Effect Window (Binning)

In practice, researchers have to impose restrictions on the effect window to implement the event study design, since β_ℓ 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 ways 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. In the absence of never-treated units, at least one restriction on dynamic treatment effects in addition to the normalization is needed to identify dynamic treatment effects (cf. Section 2.4 for more details).

Dropping observations outside of the effect window is not straightforward when effect windows of multiple treatments overlap. (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). We argue that this is the preferable approach in many applications. However, it 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, $\underline{\ell}$, and/or lags, $\bar{\ell}$, requires assumptions about the nature of the effect outside of the window. It is often economically plausible to assume that dynamic treatment effects stay constant before $\underline{\ell}$ and/or after $\bar{\ell}$, i.e. $\beta_\ell = \beta_{\bar{\ell}}$ for all $\ell > \bar{\ell}$ and $\beta_\ell = \beta_{\underline{\ell}}$ for all $\ell < \underline{\ell}$. 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). Using Remark 2, we rewrite equation (1) as

$$Y_{i,t} = \beta_{\underline{\ell}} \sum_{\ell=-\infty}^{\underline{\ell}} \Delta T_{i,t-\ell} + \sum_{\ell=\underline{\ell}+1}^{\bar{\ell}-1} \beta_\ell \Delta T_{i,t-\ell} + \beta_{\bar{\ell}} \sum_{\ell=\bar{\ell}}^{\infty} \Delta T_{i,t-\ell} + \mu_i + \theta_t + \varepsilon_{i,t}$$

which simplifies to our preferred standard event study specification:

$$Y_{i,t} = \sum_{\ell=\underline{\ell}}^{\bar{\ell}} \beta_\ell D_{i,t}^{\ell} + \mu_i + \theta_t + \varepsilon_{i,t} \tag{2}$$

with

$$D_{i,t}^{\ell} = \begin{cases} \sum_{s=-\infty}^{\underline{\ell}} \Delta T_{i,t-s} & \text{if} \quad \ell = \underline{\ell} \\ \Delta T_{i,t-\ell} & \text{if} \quad \underline{\ell} < j < \bar{\ell} \\ \sum_{s=\bar{\ell}}^{\infty} \Delta T_{i,t-s} & \text{if} \quad \ell = \bar{\ell}. \end{cases} \tag{3}$$

We refer to the variable $D_{i,t}^{\ell}$ as *binned treatment adoption indicator* which takes into account all observable past (future) treatments beyond the endpoints, i.e. the maximum lag (lead). Note that in practice, the sums at the endpoints given in equation (3) can never be calculated until infinity using observed data but only until the first/last observation of $\Delta T_{i,t}$. We discuss the data requirements that follow in Section 2.2.3. The definition of endpoints in equation (3) is for example used in Smith et al. (2019) and Fuest et al. (2018). Endpoints can be equivalently defined as $D_{i,t}^{\underline{\ell}} = \sum_{s=t-\underline{\ell}}^{\infty} \Delta T_{is}$ and $D_{i,t}^{\bar{\ell}} = \sum_{s=-\infty}^{t-\bar{\ell}} \Delta T_{is}$. Another equivalent definition for example used in McCrary (2007) is given by

$$D_{i,t}^{\ell} = \begin{cases} \mathbb{1}[t \leq E_i + \ell] & \text{if} \quad \ell = \underline{\ell} \\ \mathbb{1}[t = E_i + \ell] & \text{if} \quad \underline{\ell} < \ell < \bar{\ell} \\ \mathbb{1}[t \geq E_i + \ell] & \text{if} \quad \ell = \bar{\ell}\,. \end{cases} \tag{4}$$

2.2.1 Length of the Effect Window

From the previous section it follows that it is a key question for applied researchers how to set the length of the effect window. First and foremost, this choice depends on the research question

that the model tries to address. When do we expect the effects to 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 see also Section 2.2.3). At some point, the 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 $\beta_\ell = \beta_{\bar{\ell}}$ for all $\ell > \bar{\ell}$ at the end of the effect window. Hence, $\hat{\beta}_{\bar{\ell}}$ measures the long-run dynamic treatment effect. If treatment effects had fully materialized by $\bar{\ell}$ after the reform, we would expect estimates leading up to $\hat{\beta}_{\bar{\ell}}$ to level-off and converge to $\hat{\beta}_{\bar{\ell}}$. A pronounced drop between $\hat{\beta}_{\bar{\ell}}$ and $\hat{\beta}_{\bar{\ell}-1}$ is instead an indication that dynamic effects are still unfolding.[7] We summarize these points 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 $\bar{\ell}$ given in Remark 2 is justified. Practically, researchers can check whether estimates leading up to the endpoint converge towards $\hat{\beta}_{\bar{\ell}}$.

2.2.2 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. $\underline{\ell} = -1$ (or equivalently $\beta_\ell = 0$ for all $\ell \leq -1$). But clearly, such a model does not allow to estimate and visualize pre-trends.

The most extreme form of restricting the effect window is to bin both at the first pre-treatment

[7] Freyaldenhoven et al. (2022) provide a similar discussion and propose a formal test for what they call the leveling-off at the end of the effect window. However, the model selection procedure affects the sampling properties of the post-model-selection estimator, which invalidates the usual inference (Leeb and Pötscher, 2005).

and the first post-treamtent period, i.e. $\underline{\ell} = -1$, and $\bar{\ell} = 0$. In this case, the binned treatment indicator $D^{\ell}_{i,t}$ simplifies to $T_{i,t}$. The empirical model is given by $Y_{i,t} = \beta\, T_{i,t} + \mu_i + \theta_t + \varepsilon_{i,t}$. Note that this is a standard difference-in-differences model with staggered treatment and two-way fixed effects.

Another type of restriction, which is sometimes seen in the literature (cf. Baker and Fradkin, 2017 discussed in Section 5), is to restrict the effect window but without binning of endpoints. If observations outside of the effect window are not dropped, such a model implicitly assumes that dynamic treatment effects drop to zero outside of the effect window While such an assumption is typically hard to defend, it is often only stated implicitly by not spelling out the event study specification completely, i.e. not specifying what happens at the endpoints of the effect window.

2.2.3 Data Requirements

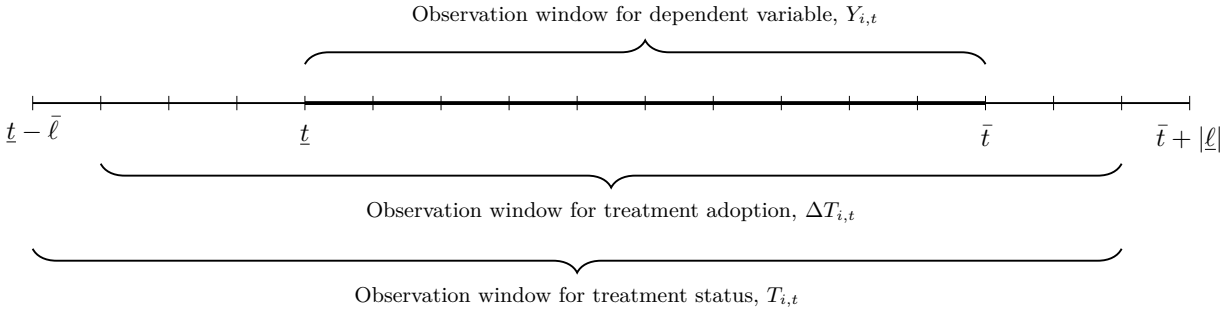
Due to the leads and lags of the effect window, we need to observe the treatment variable for a longer observation window than the dependent variable. Restrictions on the effect window determine the requirements on data availability for dependent variable $Y_{i,t}$, treatment status $T_{i,t}$ and event indicator $\Delta T_{i,t}$. In the following remark, we summarize the data requirements for a given observation window of the dependent variable.

Remark 4 (Data requirements)**.**
For a given balanced panel of the dependent variable from $[\underline{t}, \bar{t}]$ *and a limited effect window* $[\underline{\ell}, \bar{\ell}]$,[8] *we need to observe treatment adoption from* $\underline{t} - \bar{\ell} + 1$ *to* $\bar{t} + |\underline{\ell}| - 1$. *If treatment adoption is derived from changes in treatment status we need to observe treatment status from* $\underline{t} - \bar{\ell}$ *to* $\bar{t} + |\underline{\ell}| - 1$. *The following figure visualizes the required width of the observation window for a given limited effect window.*

To understand the intuition behind Remark 4, it is first important to note that an event that happens before $\underline{t}$, i.e. the first data year of the dependent variable, can affect the outcome like any other event happening between $\underline{t}$ and $\bar{t}$ and needs to be taken into account. Likewise, events happening after $\bar{t}$ might affect pre-treatment effects. This rationale prompts the second question

[8] We assume an identical observation window for the dependent variable for all units i for notational simplictiy. Remark 4 also holds for unit-specific observation windows $[\underline{t}_i, \bar{t}_i]$ which leads to unit-specific observation windows for treatment status, $[\underline{t}_i - \bar{\ell}, \bar{t}_i + |\underline{\ell}| - 1]$, and treatment adoption, $[\underline{t}_i - \bar{\ell} + 1, \bar{t}_i + |\underline{\ell}| - 1]$.

Figure 2: Data requirements



of how long before $\underline{t}$ and after $\bar{t}$ we need to observe events. By Remark 2, dynamic treatment effects are constant beyond the endpoints of the effect window $\bar{\ell}$ and $\underline{\ell}$. Hence, we need to observe events at least from $\underline{t} - \bar{\ell}$ to $\bar{t} + |\underline{\ell}|$. Remark 4 states that it suffices to observe events in one fewer year at each end of the data window, hence from $\underline{t} - \bar{\ell} + 1$ to $\bar{t} + |\underline{\ell}| - 1$. To see this, consider a case where the event takes place at $\underline{t} - \bar{\ell}$. The binned treatment adoption indicator $D_{i,t}^{\bar{\ell}}$ will be equal to one for this unit for all t. Conditional on unit fixed effects, this unit cannot be used to identify treatment effects. An analogous argument applies at the other end of the data window. The binned treatment adoption indicator in eq. (3) is therefore in practice implemented as

$$D_{i,t}^{\ell} = \begin{cases} \sum_{s=\underline{t}-\bar{\ell}+1}^{\underline{\ell}} \Delta T_{i,t-s} & \text{if} \quad \ell = \underline{\ell} \\ \Delta T_{i,t-\ell} & \text{if} \quad \underline{\ell} < j < \bar{\ell} \\ \sum_{s=\bar{\ell}}^{\bar{t}+|\underline{\ell}|-1} \Delta T_{i,t-s} & \text{if} \quad \ell = \bar{\ell}. \end{cases} \tag{5}$$

The top left panel of the numerical example shown in Appendix A.1 shows the construction of the binned treatment indicator in practice. The matrix immediately demonstrates the need for normalization (cf. Remark 1) as all row sums in the left matrix are equal to one. Binning also implies that if the reform had happened on or before $t = \underline{t} - \bar{\ell} = 2000 - 4 = 1996$ (rather than in $t = 2005$ as assumed in the example), $D_{i,t}^{\bar{\ell}} = D_{i,t}^{4} = 1$ for all t from 2000 to 2010. In this case of an always-treated unit, $D_{i,t}^{\bar{\ell}} = D_{i,t}^{4}$ is a constant and its effect is absorbed in the unit fixed effect μ_i. By the analogous argument, events on or after $t = \bar{t} + |\bar{\ell}| = 2010 + 3 = 2013$ imply $D_{i,t}^{\underline{\ell}} = D_{i,t}^{-3} = 1$ for all t from 2000 to 2010 whose effect is absorbed in μ_i. It therefore suffices to know all events from time period $\underline{t} - \bar{\ell} + 1 = 1997$ to $t = \bar{t} + |\bar{\ell}| - 1 = 2012$ to estimate the model, see Remark 4.

In the case of an absorbing single treatment, missing data can be filled. If unit i is observed to adopt treatment in period E_i, then we know that there was no treatment adoption in any period $t < E_i$ or $t > E_i$ and hence $\Delta T_{i,t} = 0$ for all $t \neq E_i$. This assumption is implicitly made when the binned treatment adoption indicators are constructed by equation (4) rather than (3). For units that are never observed to adopt treatment, we may assume that treatment is not adopted outside the observed window including the yet unknown future, either (see the discussion on assuming never-treated units in Section 2.4.1). This assumption is made when equation (4) is used and E_i is set to ∞ for units which are not observed to have been treated.

The data requirements in Remark 4 can also be derived from the equivalent distributed-lag specification described in Section 3. However, they are immediately spotted in the estimation equations (10) and (11) without the involved discussion of the endpoints here. The data requirements described for the standard event study design also apply to the generalized design in Section 4.

2.3 Estimation

The dynamic treatment effects β_ℓ in the standard event study model of equation (2) can be estimated with ordinary least squares including two-way fixed effects for units and time periods. This is a regression of the dependent variable $Y_{i,t}$ on the binned treatment adoption indicators with leads and lags

$$Y_{i,t} = \sum_{\ell=\underline{\ell}}^{-2} \beta_\ell D_{i,t}^\ell + \sum_{\ell=0}^{\bar{\ell}} \beta_\ell D_{i,t}^\ell + \mu_i + \theta_t + \varepsilon_{i,t} \ , \tag{6}$$

where we drop $D_{i,t}^{-1}$ as the reference period. As the treatment adoption indicator $D_{i,t}^\ell = \Delta T_{i,t}$ is the first difference of the treatment status $T_{i,t}$ for $\underline{\ell} < \ell < \bar{\ell}$, this is a regression of levels on (binned) changes.

Alternatively, dynamic cumulative treatment effects β_ℓ can be estimated as a regression in first differences without unit fixed effects

$$\Delta Y_{i,t} = \sum_{\ell=\underline{\ell}}^{-2} \beta_\ell \Delta D_{i,t}^\ell + \sum_{\ell=0}^{\bar{\ell}} \beta_\ell \Delta D_{i,t}^\ell + \phi_t + \Delta\varepsilon_{i,t} \ , \tag{7}$$

where $\phi_t = \theta_t - \theta_{t-1}$ and where we again drop $D_{i,t}^{-1}$ as the reference period. As $\Delta D_{i,t}^{\ell} = \Delta\Delta T_{i,t}$ for $\underline{\ell} < \ell < \bar{\ell}$, this is a regression of changes on changes of (binned) changes. Estimation of the event study model in first differences may therefore look somewhat disturbing and is rarely estimated in practice. However, the equivalent distributed-lag specification shown later in equation (11) in Section 3 is numerically equivalent to equation 7 and commonly applied.

Under standard assumptions of the fixed effects regression model, i.e. effect homogeneity, independence across units i and strict exogeneity of the treatment adoption indicator $\Delta T_{i,t}$, the estimated regression coefficients $\widehat{\beta}_\ell$ are consistent and asymptotically normally distributed (see e.g. Wooldridge, 2002 or Cameron and Trivedi, 2005).

For a cohort which experiences treatment at the same point in time and in the presence of never-treated units, two-way fixed effects estimates for this cohort can be interpreted as average treatment effects (ATT) when individual treatment effects are heterogeneous. Sun and Abraham (2021) call these effects cohort-specific average treatment effects (CATT) and show that they are consistently estimated by two-way fixed effects assuming parallel trends in baseline outcomes and no anticipatory behavior prior to treatment. The interpretation holds for effects at a specific lag/lead but also for binned endpoints. It therefore also holds for the standard difference-in-differences (DiD) model which is the most extreme form of binning. The estimated ATTs for the endpoints are averages over individual heterogeneity *as well as over all lags/leads* beyond the binned endpoint.

If, however, treatment adoption is staggered, the recent literature on heterogeneous treatment effects in event studies has shown that the two-way fixed effects estimates of β_ℓ cannot be interpreted as average treatment effects (Chaisemartin and D'Haultfœuille, 2020a,b, Sun and Abraham, 2021, Callaway and Sant'Anna, 2021, Borusyak et al., 2021, Gibbons et al., 2019, Goodman-Bacon, 2021). Even if individual effects were homogenous (as we assume throughout this paper), the two-way fixed effect estimates of binned endpoints cannot be interpreted as ATTs when the true dynamic treatment effects are not constant beyond the endpoints. Imagine now that the dynamic treatment effects have not fully materialized and increase beyond the endpoint. In this situation, the estimate of the binned endpoint will correctly pick up the level of the effects at and beyond the endpoint. But the increasing trend around the level cannot

be picked up by the the single parameter and is instead picked up by the secular period effects. These biased secular period effects, in turn, pin down the counterfactual for other treated units whose lags/leads of outcomes are still within the effect window. Consequently, all estimated dynamic treatment effects are biased because two-way fixed effects estimation fails to correctly separate dynamic causal effects from time trends when treatment is staggered. This problem is more severe in the absence of never-treated units, where the secular period effects are only identified by the restrictions imposed by binning (see Section 2.4.2 and Remark 5). We discuss in Section 3.2 that estimation in first differences does not solve this problem.

2.4 Identification

It is important to assure that the model is econometrically identified such that the dynamic treatment effects β_ℓ are disentangled from secular time fixed effects θ_t. Throughout the paper, we understand by identification the identifiability of the regression coefficients β_ℓ in a given sample, i.e. the design matrix is of full column rank.[9] In particular, we do not mean the identification of average treatment effects in the population. Recall that we assume effect homogeneity of the parameters β_ℓ (see also the discussion in the introduction).

2.4.1 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 treatment effects β_ℓ, we need to observe at least one treated unit for each lag and lead j of the effect window. The number of leads and lags can be purely based on data availability; there is no need to restrict the effect window to overcome the underidentification problem discussed in Subsection 2.4.2.

The existence of a never-treated control group is, however, less trivial than it may appear at first sight. *Observing* that a unit is not treated does not imply that this unit is never-treated. A unit could have been treated outside of the observation window – even in the yet unrealized future. When treatment effects are allowed to have effects into the infinite future (or affect the

[9] 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 treatment is randomized.

infinite past), no unit is known to be never-treated with certainty. There 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 strategy in many applications as it raises the question whether units that are never treated are comparable to units that receive treatments.

2.4.2 Identification without a Never-Treated Group

Identification of dynamic treatment effects is more challenging in the absence of a never-treated group. Borusyak et al. (2021) show that with an infinite effect window, $[\underline{\ell}, \bar{\ell}] = [-\infty, \infty]$, dynamic effects are only identified up to a linear trend. Formally, we can cast this underidentification problem in our notation as follows: $Y_{i,t} = \sum_{\ell=-\infty}^{\infty} \beta_\ell \Delta T_{i,t-\ell} + \mu_i + \theta_t + \varepsilon_{i,t} = \sum_{\ell=-\infty}^{\infty} \tilde{\beta}_\ell \Delta T_{i,t-\ell} + \tilde{\theta}_t + \tilde{\mu}_i + \varepsilon_{i,t}$ where $\tilde{\beta}_\ell = \beta_\ell + \lambda \cdot \ell$, $\tilde{\theta}_t = \theta_t - \lambda \cdot t$ and $\tilde{\mu}_i = \mu_i + \lambda \cdot E_i$. Hence, one can add a linear trend $\lambda \cdot \ell$ 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.[10] This violates the assumption of full rank of the design matrix. 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_ℓ, and secular time trends, θ_t. Restricting the effect window as in Remark 2 leads to our equation (2), $Y_{i,t} = \sum_{\ell=\underline{\ell}}^{\bar{\ell}} \beta_\ell D_{i,t}^\ell + \mu_i + \theta_t + \varepsilon_{i,t}$. In this model, adding a linear trend $\lambda \cdot \ell$ 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 $t < E_i - |\underline{\ell}| + 1$ or $t > E_i + \bar{\ell}$. [11]

[10] The underidentification problem arises because all units are treated at some point in the absence of never-treated units and hence $\sum_{\ell=-\infty}^{\infty} \ell \, \Delta T_{i,t-\ell} = \sum_{\ell=-\infty}^{\infty} \ell \, \mathbb{1}[t = E_i + \ell] = t - E_i$ for all units i and all time periods t. This underidentification problem also holds in the presence never-treated units when unit- or group-specific linear trends are included in the model. Wolfers (2006) shows that it is challenging to identify the treatment effect in the standard difference-in-differences specification, which is an extreme form of binning (see Section 2.2.2).

[11] Adding a linear trend to both dynamic effects and to secular time trends leads to

In the absence of never-treated units, observations of treated units outside the effect window are used to estimate the secular time trends. These observations therefore serve as the control group for observations of units which are within the effect window at the same period. Restricting the effect window may even produce units i for which we only observe the outcome $|\underline{\ell}|$ or more periods before the event. The observed outcomes of such units will all be affected by the same constant effect $\beta_{\underline{\ell}}$. Hence all changes over time can be attributed to secular time fixed effects and orthogonal noise. The analogous argument holds if the outcome is only observed $\bar{\ell}$ 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 5 (Identification and restricted effect window)**.**
In the absence of both never-treated units and unit-period observations outside of the effect window, dynamic treatment effects and secular time fixed effects are not separately identified. Binning of one or both endpoints generates unit-period observations outside of the effect window, which 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.

As an alternative to binning, identification is also achieved by dropping observations outside of the effect window and setting another dynamic treatment effect β_ℓ to zero, e.g. $\beta_{-2} = 0$ in addition to $\beta_{-1} = 0$ (as proposed in earlier versions of Borusyak et al., 2021). This assumption affects the shape of the estimated pre-trends as well as the estimated secular time trends. Clearly, this identifying assumption must be defended by a convincing a priori argument.

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. The derivation of this equivalence results mainly

$$
\begin{aligned}
Y_{i,t} &= \left(\sum\nolimits_{\ell=\underline{\ell}}^{\bar{\ell}} (\beta_\ell + \lambda \cdot \ell) D_{i,t}^{\ell}\right) + (\theta_t - \lambda \cdot t) + \mu_i + \lambda \cdot E_i + \varepsilon_{i,t} \\
&= \begin{cases}
\left[\sum_{\ell=\underline{\ell}}^{\bar{\ell}} \beta_\ell D_{i,t}^{\ell} + \theta_t + \mu_i + \varepsilon_{i,t}\right] + \lambda \left[\underline{\ell} - t + E_i\right] & \text{if } t < E_i - |\underline{\ell}| \\
\sum_{\ell=\underline{\ell}}^{\bar{\ell}} \beta_\ell D_{i,t}^{\ell} + \theta_t + \mu_i + \varepsilon_{i,t} & \text{if } E_i - |\underline{\ell}| \leq t \leq E_i + \bar{\ell} \\
\left[\sum_{\ell=\underline{\ell}}^{\bar{\ell}} \beta_\ell D_{i,t}^{\ell} + \theta_t + \mu_i + \varepsilon_{i,t}\right] + \lambda \left[\bar{\ell} - t + E_i\right] & \text{if } t > E_i + \bar{\ell}
\end{cases}
\end{aligned}
$$

serves two purposes. First, there is still some confusion about the two types of parametrizations and their similarities. Second, the isomorphism helps to build intuition about data requirements and identification of the underlying model. In Subsection 3.1, we formally demonstrate under which restrictions ES and DL models yield equivalent dynamic treatment effects. In Subsection 3.4, we discuss the practical implications of this isomorphism and argue that DL models are easier to implement and less-error prone. In Appendix 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 with infinite effect window. 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:

$$\begin{aligned}
\Delta Y_{i,t} &= Y_{i,t} - Y_{i,t-1} \\
&= \sum_{\ell=-\infty}^{\infty} \beta_\ell \Delta T_{i,t-\ell} - \sum_{\ell=-\infty}^{\infty} \beta_\ell \Delta T_{i,t-1-j} + \phi_t + \Delta\varepsilon_{i,t} \\
&= \sum_{\ell=-\infty}^{\infty} \beta_\ell \Delta T_{i,t-\ell} - \sum_{\ell=-\infty}^{\infty} \beta_{\ell-1} \Delta T_{i,t-\ell} + \phi_t + \Delta\varepsilon_{i,t} \\
&= \sum_{\ell=-\infty}^{\infty} \gamma_\ell \, \Delta T_{i,t-\ell} + \phi_t + \Delta\varepsilon_{i,t} \ , \qquad (8)
\end{aligned}$$

where $\gamma_\ell = \beta_\ell - \beta_{\ell-1}$ and $\phi_t = \theta_t - \theta_{t-1}$ are time fixed effects. Parameters γ_ℓ are the incremental changes of the dynamic (cumulative) treatment effects β_ℓ, measuring the slope of treatment effects from one time period to the next. The distributed-lag specification in equation (8) is the first difference of the following distributed-lag specification in levels

$$Y_{i,t} = \sum_{\ell=-\infty}^{\infty} \gamma_\ell \, T_{i,t-\ell} + \mu_i + \theta_t + \varepsilon_{i,t} \ , \qquad (9)$$

where μ_i denotes unit fixed effects. Note that the distributed-lag specification is either a regression of levels on levels (equation 9) or of changes on changes (equation 8) while the event-study specification is a regression of levels on (binned) changes (equation 6) or a regression of changes on changes of (binned) changes (equation 7).

At first sight, it might not seem intuitive to estimate an event study specification in first differences since the explanatory variable is the second difference of the treatment status, i.e. $\Delta^2 T_{i,t} = \Delta T_{i,t} - \Delta T_{i,t-1} = T_{i,t} - 2\Delta T_{i,t-1} + T_{i,t-2}$. At the same time, it is completely natural to estimate the numerically equivalent distributed-lag specification in first differences where both the dependent and the explanatory variables are first differences. Hence, it is important to keep the equivalence in mind and freely choose the estimator that is most appropriate for the question under study (see also Remark 6 and Section 3.4.)

We have proven 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 if and only if endpoints of the effect window are binned as in Remark 2. The distributed-lag parameters γ_ℓ are related to the event study parameters β_ℓ by $\gamma_\ell = \beta_\ell - \beta_{j-1}$. Binning the upper endpoint, $\beta_\ell = \beta_{\bar{\ell}}$ for all $j > \bar{\ell}$, is therefore equivalent to assuming that $\gamma_\ell = 0$ for all $j > \bar{\ell}$; for the lower endpoint $\beta_\ell = \beta_{\underline{\ell}}$ for all $j < \underline{\ell}$ is equivalent to $\gamma_\ell = 0$ for all $j \leq \underline{\ell}$. The event study model with restricted effect window between $\underline{\ell}$ and $\bar{\ell}$ and binned endpoints in equation (2)

$$Y_{i,t} = \sum_{\ell=\underline{\ell}}^{\bar{\ell}} \beta_\ell D_{i,t}^{\ell} + \mu_i + \theta_t + \varepsilon_{i,t}$$

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

$$Y_{i,t} = \sum_{\ell=\underline{\ell}+1}^{\bar{\ell}} \gamma_\ell T_{i,t-\ell} + \mu_i + \theta_t + \varepsilon_{i,t} \; . \tag{10}$$

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 6 (Equivalence of Event Study and Distributed-Lag Model)**.**
The event study specification with binned endpoints at $\ell = \bar{\ell}$ and $\ell = \underline{\ell}$ as specified in equations (2) and (3) is equivalent to a distributed lag models with $\bar{\ell}$ lags and $|\underline{\ell}| - 1$ leads as given by equation (10).

An implication of Remark 6 is 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.

3.2 Estimation of the distributed-lag model

The dynamic incremental treatment effects γ_ℓ of the distributed-lag model in equation (10) can estimated with ordinary least squares including two-way fixed effects for units and time periods. This is a regression of the dependent variable $Y_{i,t}$ in levels on the treatment status in levels with leads and lags. Alternatively, the dynamic incremental treatment effects γ_ℓ can be estimated as a regression in first differences without unit fixed effects

$$\Delta Y_{i,t} = \sum_{\ell=\underline{\ell}+1}^{\bar{\ell}} \gamma_\ell \Delta T_{i,t-\ell} + \phi_t + \Delta\varepsilon_{i,t} \ , \tag{11}$$

where $\phi_t = \theta_t - \theta_{t-1}$. This is a regression of changes on changes.

Estimation in both levels and first differences is consistent as discusses in Subsection 2.3 when the model is correctly specified. However, the incremental effect at the endpoint of the effect window $\gamma_{\bar{\ell}}$ will pick up all effects after ℓ periods when estimated in levels while it only picks up the incremental effect after exactly ℓ periods when estimated in first differences. When the effect window is too short and dynamic effects have hence not fully materialized after ℓ periods, the dynamic effect $\beta_{\bar{\ell}}$ at the endpoint often shows a pronounced spike when estimated in levels but not when estimated in first differences. This is exemplified in the application in Section 5 and demonstrated in Figure 6. This feature also leads to the often observed difference between fixed effects and first difference estimates in standard difference-in-difference models if the treatment effect builds up over time.

As already discussed in Section 2.3, dynamic treatment effects are not consistently estimated by two-way fixed effects regression if treatment adoption is staggered and the effects are not constant beyond the endpoints. This finding also holds when the binned model is estimated in first differences. This is because changes in the outcome variable beyond the endpoint are fully attributed to the secular period effects. As above, these biased secular time fixed effects will lead to biases when pinning down the counterfactual for all incremental effects. As in levels, the problem is more severe in the absence of never-treated units, where the secular period effects are

only identified by the restrictions imposed by binning (see Section 2.4.2 and Remark 5).

3.3 Recovery of cumulative treatment effects from the distributed-lag model

The distributed-lag coefficients γ_ℓ measure incremental changes of the dynamic (cumulative) treatment effects. Hence, we can construct these dynamic cumulative treatment effects from the estimated incremental effects. Note that we need to estimate one fewer incremental effect γ_ℓ than event study coefficient β_ℓ since we always normalize one dynamic treatment effect to zero (see Remark 1 and Remark 7 below). Thus, we include DL leads and lags γ_ℓ running from $\underline{\ell}+1$ (not $\underline{\ell}$ as in the event study design) to $\bar{\ell}$. Then event study parameters β_ℓ can be calculated from the distributed-lag parameters γ_ℓ by using the difference $\gamma_\ell = \beta_\ell - \beta_{\ell-1}$. 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_ℓ can be uniquely recovered recursively as $\beta_\ell = \beta_{\ell-1} + \gamma_\ell$ for $\ell = 1, 2, ..., \bar{\ell}$ and as $\beta_\ell = \beta_{\ell+1} - \gamma_{\ell+1}$ for $\ell = -1, -2, ..., \underline{\ell}$:

$$\beta_\ell = \begin{cases} -\sum_{k=\ell+1}^{-1} \gamma_k & \text{if} \quad \ell \leq -2 \\ 0 & \text{if} \quad \ell = -1 \\ \sum_{k=0}^{\ell} \gamma_k & \text{if} \quad \ell \geq 0\,. \end{cases} \tag{12}$$

We summarize this result in the following remark:

Remark 7 (Recovery of dynamic (cumulative) treatment effects from a distributed-lag model)**.** *Dynamic treatment effects β_ℓ are recovered from distributed-lag parameters γ_ℓ as cumulative effects starting from a reference period, typically the period prior to the effect, according to equation (12).*

As in the event study model, we need a normalization in the distributed-lag model since parameters β_ℓ are only identified up to a constant due to the individual fixed effect μ_i (cf. Remark 1). Equation (12) shows how to recover the dynamic treatment effects β_ℓ as the sums of distributed-lag parameters γ_ℓ. Concretely, for post-treatment effects $\ell > -1$, we intuitively cumulate upwards: $\beta_\ell = \beta_{\ell-1} + \gamma_\ell$ with $\beta_{-1} = 0$. Importantly, for pre-treatment effect $j \leq -1$, we cumulate *downwards* with a *negative* sign: $\beta_{\ell-1} = \beta_\ell - \gamma_\ell$ with $\beta_{-1} = 0$. For instance,

$\beta_{-2} = -\gamma_{-1}$; we *must not* assume $\gamma_{-1} = 0$.[12]

3.4 Practical Implications

While the derivation of equivalence results of Remark 6 is mathematically straightforward, it has useful practical implications. We briefly discuss the most important ones in the following subsection.

Model choice. As dynamic treatment effects estimated from an event study design and dynamic treatment effects recovered from a distributed-lag model are numerically equivalent, the choice of the empirical model is purely a question of convenience.

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 $\gamma = [\gamma_{\underline{\ell}+1}, ..., \gamma_{\bar{\ell}}]'$ have to be cumulated following equation (12) to obtain the event study parameters $\beta = [\beta_{\underline{\ell}}, ..., \beta_{-2}, \beta_0, ...\beta_{\bar{\ell}}]'$. This linear transformation transfers the statistical properties (consistency and asymptotic normality) of $\widehat{\gamma}$ to the calculated $\widehat{\beta}$. Standard errors of $\widehat{\beta}_\ell$ 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 because obviously the same data requirements have to apply to both the ES and DL model. The DL specification with $\bar{\ell}$ lags and $|\underline{\ell}| - 1$ leads given in equation (10) immediately reveals that we need to observe treatment status $T_{i,t}$ from $(\underline{t} - \bar{\ell})$ to $(\bar{t} + |\underline{\ell}| - 1)$ if all observations of the dependent variable $Y_{i,t}$ are to be included in the estimation sample. In case we observe treatment adoption $\Delta T_{i,t} = T_{i,t} - T_{i,t-1}$, we only need to observe it from $(\underline{t} - \bar{\ell} + 1)$ to $(\bar{t} + |\underline{\ell}| - 1)$.

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

Fixed effect vs. first difference estimator. We have shown that the ES and the DL model can both be estimated 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. If the dynamic nature of the effect is modeled correctly, i.e. if the effect is truly constant $\bar{\ell}$ periods after the event, the two estimators will be numerically very close. If, however, the true dynamic treatment effect continues to unfold beyond $\bar{\ell}$, fixed effects and first difference estimates can differ strongly. As an example, assume that the true treatment effect is negligible in the short-run and only materializes after several post-treatment periods. Further assume, that the researcher includes too few post-treatment parameters to capture the full dynamic treatment effects. 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 dynamic treatment effects have 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 estimate the long-run effect.

4 Generalized Event Study Design

In many applications, treatment may occur repeatedly and be of different intensity 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. Appendixs A.2, A.3 and A.4 provide numerical examples visualizing these cases.

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 intensity. 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 status $T_{i,t}$ does no longer need to be a dummy variable but can be a continuous measure. The change of the treatment status $\Delta T_{i,t} = T_{i,t} - T_{i,t-1}$ can therefore also be a continuous measure of the change in the intensity of the treatment. We call $\Delta T_{i,t}$ *scaled treatment adoption* in the generalized design. Scaled treatment adoption $\Delta T_{i,t}$ measures the exact size of a change in a certain policy variable (e.g. a tax rate) from unit $t-1$ to unit t.

The event study specification in equation (1) does also apply to the generalized interpretation of $T_{i,t}$ and $\Delta T_{i,t}$. The equivalence between event study and distributed-lag model therefore holds in the generalized design exactly as in the standard events study specification. As in the standard event study design, the equivalence also holds when the effect window is restricted through binning of the endpoints. Importantly, the generalization does only work with the definition of the binned treatment adoption measure $D^{\ell}_{i,t}$ in equation (3) and not with the more common definition in equation (4). We therefore call this variable *binned treatment adoption measure* rather than binned treatment adoption indicator in the generalized design.

Remark 1 on normalization, Remarks 2 and 5 on restricting the event window and the practical implications discussed in Section 3.4 on estimating the event study vs. the distributed-lag models also hold in the general case.

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

Remark 8 (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 (2) and (3) delivers consistent estimates of the dynamic treatment effects. Numerically identical treatment effects can also be obtained from a distributed-lag model as specified in equation (9).

Note that it may be the case that the proportionality assumption of Remark 8 is violated. For example, treatments with opposite signs may have 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 the 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; a numerical example is given in Appendix A.1.

Case 2: Multiple Events of Identical Intensity. Consider the case in which events of identical intensity take place repeatedly for a unit. This implies that treatment adoption $\Delta T_{i,t} = T_{i,t} - T_{i,t-1}$ is a dummy variable that takes value 1 in *any* period where an event took place and 0 in other periods. Treatment status $T_{i,t}$ increases accordingly by one unit with every event (see Appendix A.2 for a numerical example).

Few analyses have applied event studies in such an institutional context (see Dube et al., 2011, for a notable exception). However, many institutional set-ups, such as hospital admissions or firm switches, fit the case. 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.

Case 3: Single Events of Varying Treatment Intensity. Next, consider the case where each unit receives one treatment in period E_i, but treatment intensity s_i differs across units. Hence, scaled treatment adoption is $\Delta T_{i,t} = \mathbb{1}[t = E_i] \cdot s_i$. A numerical example is given in Appendix A.3.

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 intensity (see, e.g., Alsan and Wanamaker, 2018, Charles et al., 2018, Clemens et al., 2018, Goodman-Bacon, 2018).

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 Sign and Intensity. Last, we consider the most general case, in which events may occur multiple times per unit and their treatment intensity differs. A numerical example is given in Appendix A.4.

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, Brülhart et al., 2022). Note that the notion of events becomes less pertinent the more of such events happens. In the extreme case, treatment may change each period and $T_{i,t}$ becomes a constantly changing continuous variable.

4.3 Dichotomizing treatment variables.

A common alternative empirical specification used when treatments have different intensity is to dichotomize treatment variables and use a dummy variable that is only switched on for relatively large events (see, e.g., Simon, 2016, Fuest et al., 2018). However, the parameter estimates of such a dichotomization may be difficult 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 $\Delta T^S_{i,t} = 1$ and a large reform $\Delta T^L_{i,t} = 2$; treatments are distributed randomly in time and treatment effects are linear in the intensity of the reform. Ignoring small events and applying the standard event dummy set-up yields $\Delta T^S_{i,t} = 0$, $\Delta T^L_{i,t} = 1$. In this case, units with small reforms become part of 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. In such a specification, estimated dynamic treatment effects can be larger, smaller or identical to a specification that considers all reforms with their intensity. 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 selected 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).[13] We will particularly focus on the importance of restricting the effect window and on the advantages 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 estimate 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 report a highly significant estimate of -0.00207, which implies that a 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:

$$\ln GJSI_{i,t} = \sum_{j=-3}^{4} \beta_\ell \Delta T_{i,t-\ell} + w'_{i,t}\xi + \mu_i + \theta_t + \varepsilon_{i,t}, \tag{13}$$

where $GJSI_{st}$ is the the natural logarithm of the Google Job Search Index in state i and period t

[13] Replication code is available in the Harvard Dataverse at https://doi.org/10.7910/DVN/LXMYV6

(year-month), $\Delta T_{i,t-\ell}$ is an indicator variable that indicates whether PBD in state i was changed $j \in [-3, 4]$ month before or after t without binning endpoints. Parameter μ_i captures state fixed effects and θ_t denotes period fixed effects. The vector $w_{i,t}$ 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 intensity 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, ΔT_{it} 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 ΔT_{it} is switched on for decreases of 7 weeks or more. In the respective models, the event indicator ΔT_{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.[14]

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 3, we replicate the main event study results for large increases and large decreases on the two respective samples estimating equation (13). Our results are identical to the original version. Unlike results from the differences-in-differences model, event study estimates do not point to a strong negative relationship between search effort and PBD. However, the results depicted in Panel A of Figure 3 are based on strong implicit assumptions and parameter restrictions embodied in equation (13), which speak directly to our main points raised in the previous sections. While the empirical model looks like a classic

[14] 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.

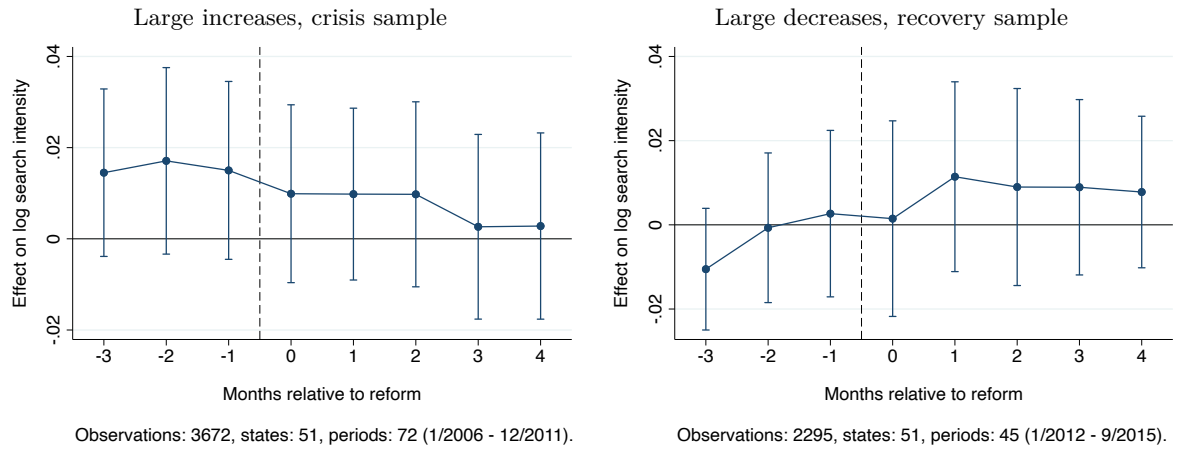
event study design and therefore innocuous at first sight, event indicators $\Delta T_{i,t}$ are not binned at the endpoints (cf Remark 2) and no coefficient is normalized to zero (cf Remark 1). This implies that dynamic treatment effects are implicitly normalized to be zero four and more periods before the event as well as five and more periods after the event, i.e. $\beta_\ell = 0$ 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 5). In contrast, binning of endpoints assumes that the effect builds up over 4 years and stays constant thereafter, an assumption more in line with the theoretical priors.

Next, we estimate equation (13) 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 (12). The two methods are equivalent and lead to identical parameter estimates and standard errors as shown in Remark 6. The choice of the estimation method is purely a question of convenience as explained in Section 3.4. Panel B of Figure 3 shows results with binned endpoints and normalized pre-event period. Other than 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, so the parallel trends assumption cannot be rejected 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 the right graph of Panel B of Figure 3.

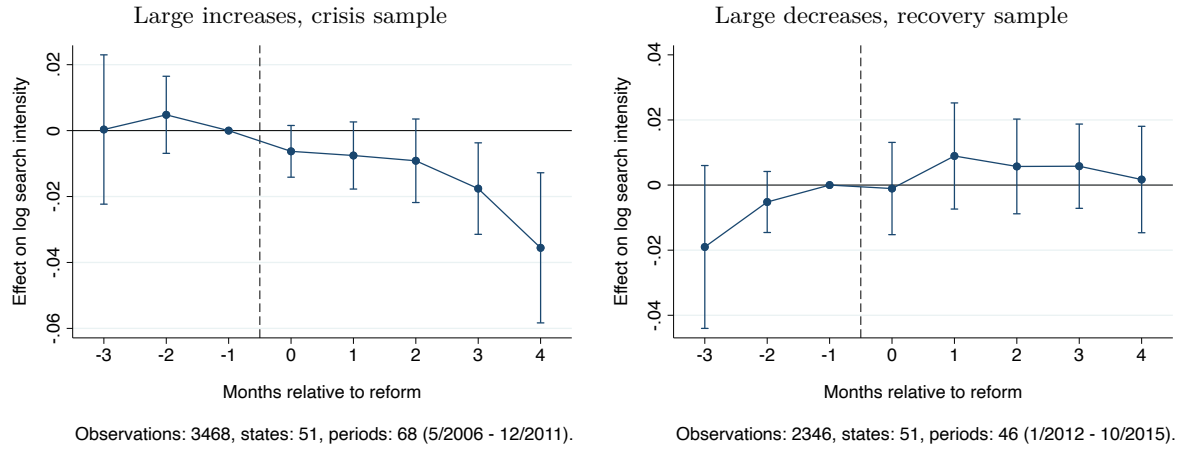
Note that the number of observations differs between Panels A and B of Figure 3. This is

Figure 3: 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.

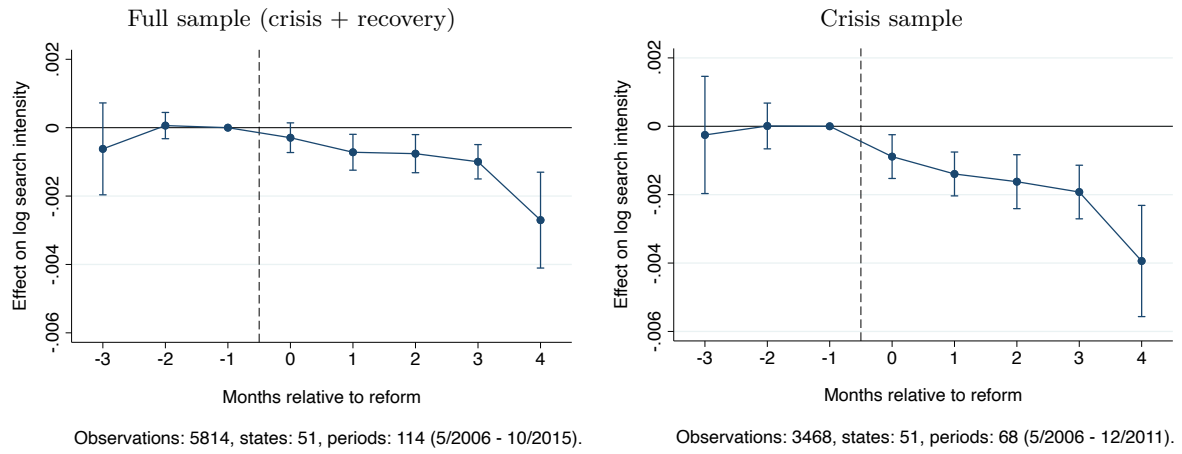
due to dropped observations from missing treatment information (see the paragraph on data requirements in Section 3.4). For increases, i.e. the crisis sample, the dependent variable is observed from 1/2006 to 12/2011. For the last month (12/2011), we are able to generate all leads up until $j = -3$ as we observe treatment status until 12/2015. However, we can calculate the first binned endpoint for a specification with four lags $b^4_{i,t}$ only in 5/2006. Consequently, our sample is $\underline{\ell} = 4$ 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 $\bar{\ell} - 1 = 2$ periods. The sample is automatically reduced to the correctly shortened sample when the distributed-lag model is estimated as we discuss in Section 3.4. 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 only switches on for large reforms. While Panel B of Figure 3 shows that binning endpoints leads to convincing event study coefficients, which match the difference-in-differences estimates, the zero-one models does 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 Section 4.2 for Case 4, all events are scaled with the respective treatment intensity, i.e. the changes in PBD of different magnitudes. The resulting left graph in Figure 4 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).

Figure 4: Generalized Event Study Design

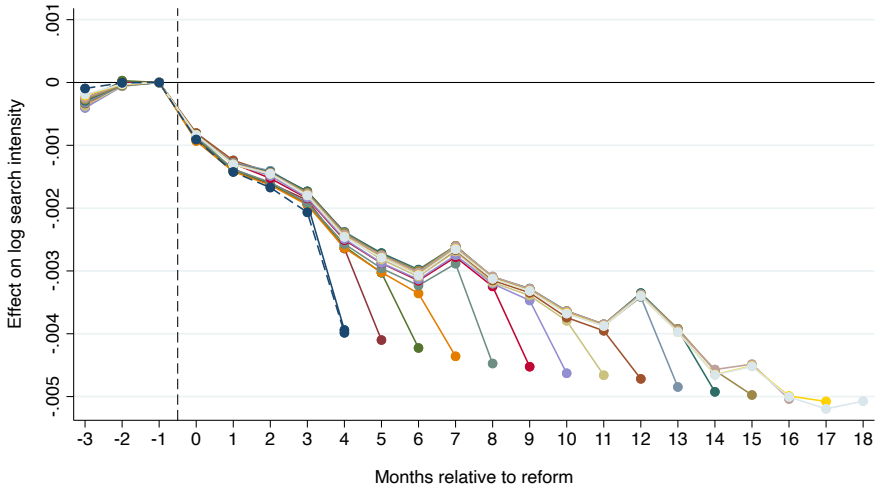


Notes: The figure plots the results when applying the generalized event study with binned endpoints and normalization at -1 to the setting in (Baker and Fradkin, 2017). The graphs shows the dynamic effect of an increase the potential benefit duration by one week on log search intensity as measured by Google Job Search Index (GJSI). 95% confidence intervals are plotted.

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 intensity as stated in Remark 8. In the context of the replication, the remark implies symmetry between increases and decreases. It is crucial to test these assumptions, e.g. by separating between treatments of different sign (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 3 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 4 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 dynamic treatment effects have fully materialized after $\bar{\ell}$ periods. In Figure 4, we see that effects are still on the decline four months

Figure 5: Varying the Effect Window (Crisis Sample)



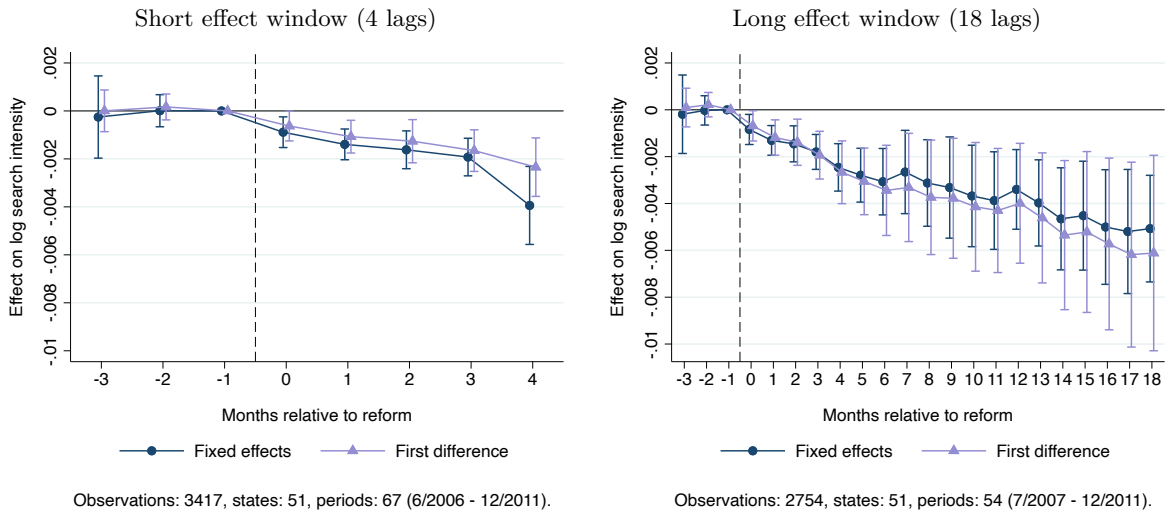
Notes: The figure plots the results when applying the generalized event study with binned endpoints and normalization at -1 to the setting in (Baker and Fradkin, 2017). The graphs shows the dynamic treatment effects of an increase the potential benefit duration by one week on log search intensity as measured by Google Job Search Index (GJSI). In the original contribution, the authors assumed a maximum lag length of 4 months. In this graph, we extend the effect window to up to 18 months. As the model with 18 lags is estimated on a smaller sample, we also provide estimates of the model with 4 lags on this shorted sample (dashed blue line). Confidence intervals are omitted.

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 dynamic 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.

One procedure to determine the length of the effect window is to simply increase the number of lags until the dynamic treatment effects flatten out. However, this approach comes at a cost as it will often reduce sample size and precision. This model-selection procedure also affects the inference (cf. footnote 7). Nonetheless, we re-estimate the generalized event study design gradually increasing $\bar{\ell}$ to one and a half years (18 months). Results are presented in Figure 5. The figure suggests that dynamic 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 the DiD estimate 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 dynamic treatment effects have fully materialized. An alternative check to assess whether the effect

Figure 6: Fixed Effects vs. First Differences (Crisis Sample)



Notes: The figure plots the results when applying the generalized event study to the setting in (Baker and Fradkin, 2017). The graphs shows the dynamic treatment effects 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.

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 (cf. Section 3.2). At the endpoint of the effect window, the first difference model only accounts for the change happening from $\bar{\ell} - 1$ to $\bar{\ell}$, 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 window. This pattern is nicely demonstrated in Figure 6, 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 dynamic treatment effects unfold monotonically, the long-run estimates will be biased toward zero.[15]

6 Conclusion

This paper makes three interrelated methodological contributions, which are important to bear in mind when setting up event study designs in economics. The points are valid in general, and

[15] 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.

might be particularly helpful when applying the event study technique to settings in public and labor economics with multiple policy shocks of different intensity.

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 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 design *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 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 treatment intensity. 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 contributions replicating and extending the event study design in Baker and Fradkin (2017).

References

Alsan, M. and M. Wanamaker (2018). "Tuskegee and the Health of Black Men". *Quarterly Journal of Economics* 133 (1), pp. 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), pp. 3–30.

Autor, D. (2003). "Outsourcing at Will: The Contribution of Unjust Dismissal Doctrine to the Growth of Employment Outsourcing". *Journal of Labor Economics* 21 (1), pp. 1–42.

Baker, A. C., D. F. Larcker, and C. C. Y. Wang (2022). "How Much Should We Trust Staggered Difference-In-Differences Estimates". *Journal of Financial Economics* 2, pp. 370–395.

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), pp. 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., X. Jaravel, and J. Spiess (2021). "Revisiting Event Study Designs: Robust and Efficient Estimation". *mimeo*.

Brülhart, M., J. Gruber, M. Krapf, and K. Schmidheiny (2022). "Behavioral Responses to Wealth Taxes: Evidence from Switzerland". *American Economic Journal: Economic Policy* 4 (14).

Callaway, B., A. Goodman-Bacon, and P. H. C. Sant'Anna (2021). "Difference-in-Differences with a Continuous Treatment". *arXiv:2107.02637*.

Callaway, B. and P. H. C. Sant'Anna (2021). "Difference-in-Differences with Multiple Time Periods". *Journal of Econometrics* 225 (2), pp. 200–230.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: Methods and Applications*. New York: Cambridge University Press.

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). "The Effect of MinimumWages on Low-Wage Jobs". *The Quarterly Journal of Economics* 134 (4), pp. 1405–1454.

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

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

– (2021). "Two-way Fixed Effects Regressions with Several Treatments". *arXiv:2012.10077*.

– (2022). "Two-Way Fixed Effects and Differences-in-Differences with Heterogeneous Treatment Effects: A Survey". *NBER Working Paper 29691*.

Charles, K. K., E. Hurst, and M. Notowidigdo (2018). "Housing Booms and Busts, Labor Market Opportunities, and College Attendance". *American Economic Review* 108 (10), pp. 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), pp. 1468–1487.

Dolley, J. C. (1933). "Characteristics and Procedure of Common Stock Split-Ups". *Harvard Business Review* (11), pp. 316–326.

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

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

Freyaldenhoven, S., C. Hansen, J. Pérez Pérez, and J. M. Shapiro (2022). "Visualization, identification, and estimation in the linear panel event-study design". *Advances in Economics and Econometrics: Twelfth World Congress (forthcoming)*.

Freyaldenhoven, S., C. Hansen, and J. M. Shapiro (2019). "Pre-event Trends in the Panel Event-study Design". *The 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), pp. 393–418.

Gardner, J. (2021). "Two-stage differences in differences". *Working Paper*.

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

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

Goodman-Bacon, A. (2021). "Difference-in-Differences with Variation in Treatment Timing". *Journal of Econometrics* 225 (2), pp. 254–277.

Leeb, H. and B. M. Pötscher (2005). "Fairness and Redistribution". *Econometric Theory* 21, pp. 21–59.

Lichter, A., M. Löffler, I. Isphording, T.-V. Nguyen, F. Pöge, and S. Siegloch (2021). "Profit Taxation, R&D Spending, and Innovation". *CEPR Discussion Paper No. 16702*.

MacKinlay, A. C. (1997). "Event Studies in Economics and Finance". *Journal of Economic Literature* 35 (1), pp. 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, pp. 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), pp. 318–353.

Powell, D. (2021). "The Labor Supply Consequences of the Opioid Crisis". *RAND Discussion Paper*.

Rambachan, A. and J. Roth (2022). "A More Credible Approach to Parallel Trends". *Review of Economic Studies (forthcoming)*.

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

Roth, J., P. H. C. Sant'Anna, A. Bilinski, and J. Poe (2022). "What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature". *Journal of Econometrics (forthcoming)*.

Siegloch, S., N. Wehrhöfer, and T. Etzel (2021). "Direct, Spillover and Welfare Effects of Regional Firm Subsidies". *CEPR Discussion Paper No. 16129*.

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), pp. 128–159.

Smith, M., D. Yagan, O. Zidar, and E. Zwick (2019). "Capitalists in the Twenty-First Century". *The Quarterly Journal of Economics* 134 (4), pp. 1675–1745.

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), pp. 2582–2624.

Sun, L. and S. Abraham (2021). "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects". *Journal of Econometrics* 225 (2), pp. 175–199.

Wolfers, J. (2006). "Did Unilateral Divorce Laws Raise Divorce Rates? A Reconciliation and New Results". *The American Economic Review* 96 (5), pp. 1802–1820.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Cambridge: MIT Press.

Appendix A Numerical Examples

In Appendix A, we illustrate various event study design set-ups discussed in the paper using simple numerical examples. We start in subsection A.1 with the standard case of a single binary treatment presented in Section 2 of the paper. We then show different cases of the generalized event study set-up developed in Section 4 of the paper: multiple event of identical intensity (Subsection A.2), single event of varying treatment intensity (Subsection A.3) and multiple event of varying treatment intensity (Subsection A.4).

The examples visualize the data structure in the respective set-ups and show how binning works in practice. They also demonstrate the equivalence result between event study and distributed lag models summarized in Remark 6.

A.1 Single Event with Binary Treatment

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

In example A.1, the explanatory variables of the event study model in levels (equation 6) and in first differences (equation 7) are visualized by the following matrices, respectively.

t	$D_{i,t}^{-3}$	$D_{i,t}^{-2}$	$D_{i,t}^{0}$	$D_{i,t}^{1}$	$D_{i,t}^{2}$	$D_{i,t}^{3}$	$D_{i,t}^{4}$	$\Delta D_{i,t}^{-3}$	$\Delta D_{i,t}^{-2}$	$\Delta D_{i,t}^{0}$	$\Delta D_{i,t}^{1}$	$\Delta D_{i,t}^{2}$	$\Delta D_{i,t}^{3}$	$\Delta D_{i,t}^{4}$
2000	1	0	0	0	0	0	0							
2001	1	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
2003	0	1	0	0	0	0	0	-1	1	0	0	0	0	0
2004	0	0	0	0	0	0	0	0	-1	0	0	0	0	0
2005	0	0	1	0	0	0	0	0	0	1	0	0	0	0
2006	0	0	0	1	0	0	0	0	0	-1	1	0	0	0
2007	0	0	0	0	1	0	0	0	0	0	-1	1	0	0
2008	0	0	0	0	0	1	0	0	0	0	0	-1	1	0
2009	0	0	0	0	0	0	1	0	0	0	0	0	-1	1
2010	0	0	0	0	0	0	1	0	0	0	0	0	0	0

The following matrices visualize the explanatory variables of the distributed-lag model in levels (equation (10)) and in first differences (equation (11)), respectively.

t	$T_{i,t+2}$	$T_{i,t+1}$	$T_{i,t}$	$T_{i,t-1}$	$T_{i,t-2}$	$T_{i,t-3}$	$T_{i,t-4}$	$\Delta T_{i,t+2}$	$\Delta T_{i,t+1}$	$\Delta T_{i,t}$	$\Delta T_{i,t-1}$	$\Delta T_{i,t-2}$	$\Delta T_{i,t-3}$	$\Delta T_{i,t-4}$
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 $\bar{\ell} = 4$ years after event and $|\underline{\ell}| = 3$ years before the event corresponds to a distributed-lag model with $\bar{\ell} = 4$ lags and $|\underline{\ell}| - 1 = 2$ 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.

In example A.1, the dynamic treatment effects β_ℓ are calculated according to equation (12) from the incremental changes of the dynamic treatment effects γ_ℓ in the distributed-lag specification 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 Multiple Events of Identical Intensity

Example A.2. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{\ell} = -3$ *to* $\bar{\ell} = 4$*. For individual* i*, a first event takes place at* 2004 *and a second at* 2006*.* $\ell = -1$ *is taken as the reference period.*

The explanatory variables for the event study in levels and in first differences are

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

The explanatory variables of the distributed-lag model in levels and in first differences are

t	$T_{i,t+2}$	$T_{i,t+1}$	$T_{i,t}$	$T_{i,t-1}$	$T_{i,t-2}$	$T_{i,t-3}$	$T_{i,t-4}$	$\Delta T_{i,t+2}$	$\Delta T_{i,t+1}$	$\Delta T_{i,t}$	$\Delta T_{i,t-1}$	$\Delta T_{i,t-2}$	$\Delta T_{i,t-3}$	$\Delta T_{i,t-4}$
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

A.3 Single Events of Varying Treatment Intensity

Example A.3. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{\ell} = -3$ *to* $\bar{\ell} = 4$*. For individual* i*, the single treatment of intensity* $s_i = 0.1$ *is adopted at* $E_i = 2005$*.* $\ell = -1$ *is taken as the reference period.*

The explanatory variables for the event study in levels and in first differences are

t	$D_{i,t}^{-3}$	$D_{i,t}^{-2}$	$D_{i,t}^{0}$	$D_{i,t}^{1}$	$D_{i,t}^{2}$	$D_{i,t}^{3}$	$D_{i,t}^{4}$	$\Delta D_{i,t}^{-3}$	$\Delta D_{i,t}^{-2}$	$\Delta D_{i,t}^{0}$	$\Delta D_{i,t}^{1}$	$\Delta D_{i,t}^{2}$	$\Delta D_{i,t}^{3}$	$\Delta D_{i,t}^{4}$
2000	0.1	0	0	0	0	0	0							
2001	0.1	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
2003	0	0.1	0	0	0	0	0	0.1	0.1	0	0	0	0	0
2004	0	0	0	0	0	0	0	0	-0.1	0	0	0	0	0
2005	0	0	0.1	0	0	0	0	0	0	0.1	0	0	0	0
2006	0	0	0	0.1	0	0	0	0	0	-0.1	0.1	0	0	0
2007	0	0	0	0	0.1	0	0	0	0	0	-0.1	0.1	0	0
2008	0	0	0	0	0	0.1	0	0	0	0	0	-0.1	0.1	0
2009	0	0	0	0	0	0	0.1	0	0	0	0	0	-0.1	0.1
2010	0	0	0	0	0	0	0.1	0	0	0	0	0	0	0

The corresponding explanatory variables of the distributed-lag model in levels and in first differences are

t	$T_{i,t+2}$	$T_{i,t+1}$	$T_{i,t}$	$T_{i,t-1}$	$T_{i,t-2}$	$T_{i,t-3}$	$T_{i,t-4}$	$\Delta T_{i,t+2}$	$\Delta T_{i,t+1}$	$\Delta T_{i,t}$	$\Delta T_{i,t-1}$	$\Delta T_{i,t-2}$	$\Delta T_{i,t-3}$	$\Delta T_{i,t-4}$
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

A.4 Multiple Events of Varying Treatment Intensity

Example A.4. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{\ell} = -3$ *to* $\bar{\ell} = 4$*. For individual* i*, one treatment of intensity* 0.2 *is adopted in* 2003*, another treatment of intensity* -0.1 *is adopted in* 2004 *and yet another treatment of intensity* 0.3 *is adopted in* 2006*; there are no changes in treatment in other years. Without any influence on the estimation results, we assume an arbitrary initial value of treatment as* $T_{i,\underline{t}} = T_{i,2000} = 0$*.* $\ell = -1$ *is taken as the reference period.*

The following four matrices show the explanatory variables for the event study in levels and in first differences:

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

t	$T_{i,t+2}$	$T_{i,t+1}$	$T_{i,t}$	$T_{i,t-1}$	$T_{i,t-2}$	$T_{i,t-3}$	$T_{i,t-4}$	$\Delta T_{i,t+2}$	$\Delta T_{i,t+1}$	$\Delta T_{i,t}$	$\Delta T_{i,t-1}$	$\Delta T_{i,t-2}$	$\Delta T_{i,t-3}$	$\Delta T_{i,t-4}$
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