CESifo WORKING PAPERS

7481
2019

January 2019

On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications

Kurt Schmidheiny, Sebastian Siegloch

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7481*
Category 12: Empirical and Theoretical Methods

On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications

Abstract

We discuss important features and pitfalls of panel-data event study designs. We derive the following main results: First, event study designs and distributed-lag models are numerically identical leading to the same parameter estimates after correct reparametrization. Second, binning of effect window endpoints allows identification of dynamic treatment effects even when no never-treated units are present. Third, classic dummy variable event study designs can be naturally generalized to models that account for multiple events of different sign and intensity of the treatment, which are particularly interesting for research in labor economics and public finance.

JEL-Codes: C230, C510, H000, J080.

Keywords: event study, distributed-lag, applied microeconomics, credibility revolution.

Kurt Schmidheiny
University of Basel / Switzerland
kurt.schmidheiny@unibas.ch

Sebastian Siegloch
University of Mannheim / Germany
siegloch@uni-mannheim.de

This version: January 2019
We thank Mona Köhler for excellent research assistance. Samara Gunter, Jesse Shapiro and Juan Carlos Suárez Serrato provided helpful and constructive feedback - thank you very much! Both authors are grateful to the University of California at Berkeley - in particular Enrico Moretti and Emmanuel Saez - for the hospitality during the Academic Year 2017/2018 when most parts of this paper were written. Siegloch is thankful to the German Research Foundation DFG for financing the research stay under the Research Fellowship program (# 361846460).

1 Introduction

The credibility revolution in empirical economics has led researchers to set up more transparent (quasi-)experimental research designs. This shift has increased the policy relevance and the scientific impact of empirical work (Angrist and Pischke, 2010). An important element that enhanced transparency is the visualization of treatment effects and/or identiying assumptions. Ideally, the main empirical take-away of the study can be summarized in one graph. Differences-in-differences (DD) models are particularly popular design as they directly derive from the rationale of experiments and are usually intuitive in terms of the empirical model and the underlying assumptions. A particularly appealing specification, belonging to the family of DD models, is the event study (ES) design. Originating from the finance literature, event study designs have initially been used to analyze the impact of an unanticipated event on stock prices. This rationale has been translated to other fields of applied economics, mostly public finance and labor economics. Here an event is usually defined as a policy change and the outcome is the price or quantity of the market under study. Technically, the outcome variable in a year t is regressed on a set of dummy variables, which indicate when the event has happened relative to t. In many ways, the event study design is the poster child of empirical methods used in the credibility revolution since (i) coefficients can be graphed, (ii) the graphs are very intuitive in the sense that both post-event effects and the identifying assumption of "no pre-event trends" are immediately visible, and (iii) the underlying econometrics are intuitive as they boil down to a simple panel data model where the regressors of interest are a set of non-parametric event indicators which are defined relative to the event. Figure 1 plots the use of event study designs in economics over time. We measure the use by the share of studies mentioning the term "event study" in the Top Five economics journals.[1] While we see a steady increase since 1990, there is sharp increase since 2010. Moreover this increase is mostly driven by the three journals focusing on applied microeconomic work among the Top-Five, i.e. the American Economic Review (AER), the Quarterly Journal of Economics (QJE), and the Journal of Political Economy (JPE).

Despite its intuitive appeal, there is remarkable heterogeneity in how event study designs are implemented in practice. In particular, there is a tendency to refer to the term event study rather loosely. First, there is the distinction between the original event study originating in the finance literature and the adaption used in current applied micro research, mostly in the fields of public and labor economics. Second, within the later field, some researchers employ the term for a simple graph plotting an outcome for the treatment (and sometimes the control) group relative to some event in the spirit of a difference-in-difference graphs. As

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

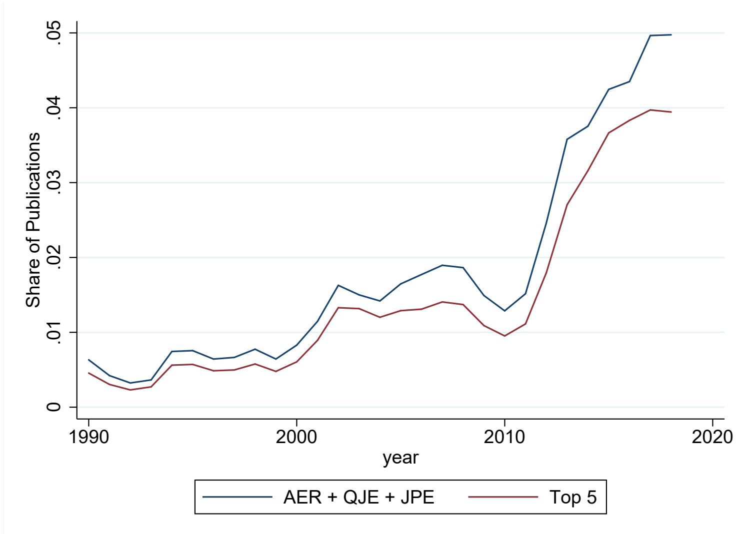

Figure 1: The rise of event studies in economics

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.

a consequence, more than one third of the non-finance papers published in the AER, QJE or JPE since 2010 using event study designs do not specify a regression equation. Another issue which leads to heterogeneity in the implementation of event studies is the treatment of the ends of the effect window, which is the window within which the effect is studied.[2] Among the papers that specified a regression equation, only 15% discussed how they modeled the endpoints, which – as it will turn out below – is important. Moreover, empirical economists differ in how they apply the event study rationale to multiple events and events with different treatment intensities.

The purpose of this paper is to clarify our understanding of event study designs both in methodological and practical terms. We depart from the simplest institutional environment in which event study designs are applied: in a typical panel set-up with a large number of units and relatively few time periods, each unit receives one single-treatment – the event – at a unit-specific point in time. As necessary in most practical applications, we limit the effect window to a finite number of leads and lags, which requires binning the endpoints of the window. In this set-up, we show that event study coefficients are identical to the cumulated lags and leads in a distributed-lag (DL) model. However, the equivalence requires to carefully choose

[2] A more widely used synonym is event window, which as it will turn out, is less accurate.

the corresponding number of leads and lags and the length of the observation window for the dependent variable and the event indicators in the DL and the ES model. Having derived the equivalence between the standard ES design and DL models, we show that the rationale of the equivalence can be easily generalized and adapted to institutional environments, which are more common in public and labor economics. These environments include (i) multiple events, such as a sequence of state-level minimum wage increases, (ii) events of different sign and intensity of the treatment, such as a tax increase in one year and a decrease a couple of years later. While making these methodological points, the paper provides practical advice and highlights pitfalls when setting up event study designs and when comparing the results of ES and DL models. We illustrate our results using simple numerical examples.

This paper adds to the recent methodological literature on event study and distributed-lag models. Freyaldenhoven et al. (2018) suggest 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 (2018) shows that treatment effects can be biased conditional on passing the flat pre-trend test. Abraham and Sun (2018) point to a different form of identification problem arising in the case of staggered treatments with heterogeneous dynamic treatment effects – a similar problem also arises in static diff-in-diff models (Athey and Imbens, 2018; de Chaisemartin and D'Haultfoeuille, 2018). In our set-up, we assume homogeneous treatment effects and abstract from endogeneity concerns. In this respect our study is related to Borusyak and Jaravel (2017) who point to a potential underidentification problem in panel data event study designs where dynamic treatment effects can only be identified up to a linear trend. We show below that restricting the effect window at a certain lead/lag – a practical necessity in many applications – and binning the endpoints alleviates such concerns by introducing parameter restrictions that help to separate trends in the dynamic treatment and secular time effects even in the absence of never treated units. At the same time, we show that choosing the length of the effect window determines whether a unit-year observation is assigned to the treatment or the control group and hence is not innocuous.

The remainder of this paper is structured as follows. Section 2 sets up a standard version of an event study model in the simplest institutional environment and shows that the specification is equivalent to a standard distributed-lag model. In Section 3, we generalize the institutional environment and allow for multiple and heterogeneous events across and within units. We show that event study designs can also be used in these settings, and discuss the additional adjustments and assumptions that need to be made in this case. Section 4 concludes.

2 Standard Event Study Design

In this section, we show that the event study model is equivalent to a distributed-lag model. We start by setting up the event study design in the simplest institutional environment and refer to this model as the standard event study set-up throughout the paper. We then show that this standard event study model is a re-parametrization of the distributed-lag model. On the way, we stress important assumptions that are necessary for the correct specification of both models. In Subsection 2.4, we illustrate all formal claims using a numerical example.

2.1 Set-up

We start our analysis with a standard event study set-up, where each unit i receives at most one single treatment at unit-specific time e_i. We intend to estimate the effect of this treatment on our dependent variable y, which we observe at different time periods $t = \underline{t}, ..., \bar{t}$. We call $[\underline{t}, ..., \bar{t}]$ the observation window for the dependent variable. As the treatment effect is allowed to vary over time, we are interested in studying its dynamics over a window ranging from $\underline{j} < 0$ periods prior to the event to $\bar{j} \geq 0$ after the event. We refer to this window as the effect window. In this set-up, the standard event study specification is given for all $t = \underline{t}, ..., \bar{t}$ by:

$$y_{it} = \sum_{j=\underline{j}}^{\bar{j}} \beta_j b_{it}^j + \mu_i + \theta_t + \varepsilon_{it} \tag{1}$$

where μ_i is a unit fixed effect, θ_t is a time fixed effect. In the standard set-up, b_{it}^j is a treatment indicator for an event happening $j \in [\underline{j}, \bar{j}]$ periods away from t, which is commonly defined as

$$b_{it}^j = \begin{cases} \mathbb{1}[t \leq e_i + j] & \text{if} \quad j = \underline{j} \\ \mathbb{1}[t = e_i + j] & \text{if} \quad \underline{j} < j < \bar{j} \\ \mathbb{1}[t \geq e_i + j] & \text{if} \quad j = \bar{j}\,. \end{cases} \tag{2}$$

Treatment indicators b_{it}^j are binned at the endpoints, indicating if treatment for unit i at time t happened $\bar{j}$ *or more* periods ago, or $\underline{j}$ periods *or more* into the future. We rewrite definition (2) and propose the following generalized, more versatile (cf. Section 3.3.2) and arguably more intuitive definition for all $t = \underline{t}, ..., \bar{t}$:[3]

$$b_{it}^{j} = \begin{cases} \sum_{s=t-\underline{j}}^{\bar{t}-\underline{j}-1} d_{is} & \text{if} \quad j = \underline{j} \\ d_{i,t-j} & \text{if} \quad \underline{j} < j < \bar{j} \\ \sum_{s=\underline{t}-\bar{j}+1}^{t-\bar{j}} d_{is} & \text{if} \quad j = \bar{j}. \end{cases} \tag{3}$$

In the standard set-up, d_{it} is an event dummy that takes the value 1 in the year of the treatment, e_i, and zero otherwise. Section 2.4 shows the construction of the binned treatment indicator in a numerical example.

Remark 1 (Limited effect window)**.**
Following the standard, we limit the effect window to $\underline{j} < 0$ periods prior to the event and $\bar{j} \geq 0$ after the event, i.e. the effect is assumed to stay constant before and after this effect window. This is equivalent to an effect window from $-\infty$ to ∞ and assuming $\beta_j = \beta_{\underline{j}}$ for all $j < \underline{j}$ and $\beta_j = \beta_{\bar{j}}$ for all $j > \bar{j}$. This is achieved by binning the treatment indicator.

Remark 1 implies that the standard event study model relies on additional but often economically plausible parameter restrictions. These restrictions reduce the number of parameters to be estimated and thereby alleviate potential underidentification problems (see Section 2.3).

Due to the leads of lags of the treatment indicator in the event study model, information on the treatment needs to be observed for a longer observation window than for the dependent variable. In the following remark, we summarize the data requirements for a given observation window of the dependent variable.

Remark 2 (Data requirements)**.** *For a given balanced panel of the dependent variable from $[\underline{t}, \bar{t}]$ and an effect window $[\underline{j}, \bar{j}]$, we need to observe events from $\underline{t} - \bar{j} + 1$ to $\bar{t} + |\underline{j}| - 1$. If events are derived from changes in policy variables, i.e. treatment status, we need to observe treatment status from $\underline{t} - \bar{j}$ to $\bar{t} + |\underline{j}| - 1$.*

To understand the intuition behind Remark 2, 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, we should account for events that happen after $\bar{t}$ if we want to test for pre-trends.

[3] Note that the definition of endpoints in (3) is equivalent to other recently applied variants. In our notation, the specification of Fuest et al. (2018) is given by $b_{it}^{\underline{j}} = \sum_{s=t}^{\bar{t}-1} d_{i,s-\underline{j}}$ and $b_{it}^{\bar{j}} = \sum_{s=\underline{t}+1}^{t} d_{i,s-\bar{j}}$. Likewise, it can be rewritten as $b_{it}^{\underline{j}} = \sum_{s=t-(\bar{t}-\underline{j}-1)}^{\underline{j}} d_{i,t-s}$ and $b_{it}^{\bar{j}} = \sum_{s=\bar{j}}^{t-(\underline{t}-\bar{j}+1)} d_{i,t-s}$, which is equivalent to Smith et al. (2017) using $\underline{t}, \bar{t}$ instead of the implicitly assumed negative and positive infinity.

This rationale prompts the second question of how long before $\underline{t}$ and after $\bar{t}$ do we need to observe events at the minimum. From Remark 1, we know that treatment effects are assumed to be constant after $\bar{j}$ and before $\underline{j}$ periods. Hence, we need to observe events at least from $\underline{t} - \bar{j}$ to $\bar{t} + |\underline{j}|$. By Remark 2, it, however, suffices to observe events in one fewer year at each end of the data window, hence from $\underline{t} - \bar{j} + 1$ to $\bar{t} + |\underline{j}| - 1$. To see this, consider a case where the event takes place at $\underline{t} - \bar{j}$. Due to the binning, the treatment indicator $b_{it}^{\bar{j}}$ 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.

So far, we assumed to observe events directly. In many applications, an event is, however, defined by the researcher as a change in policy variables, such as tax rates or minimum wages. We call such policy variables the *treatment status* x_{it}. In the standard set-up, treatment status is a dummy variable that changes from 0 to 1 at the time of the event. If we generate event dummies from changes in policy states, the observation window for the treatment status needs to be observed for an additional period in the beginning, hence from $\underline{t} - \bar{j}$ onwards. Figure 2.1 visualizes the required width of the observation window for a given limited effect window.

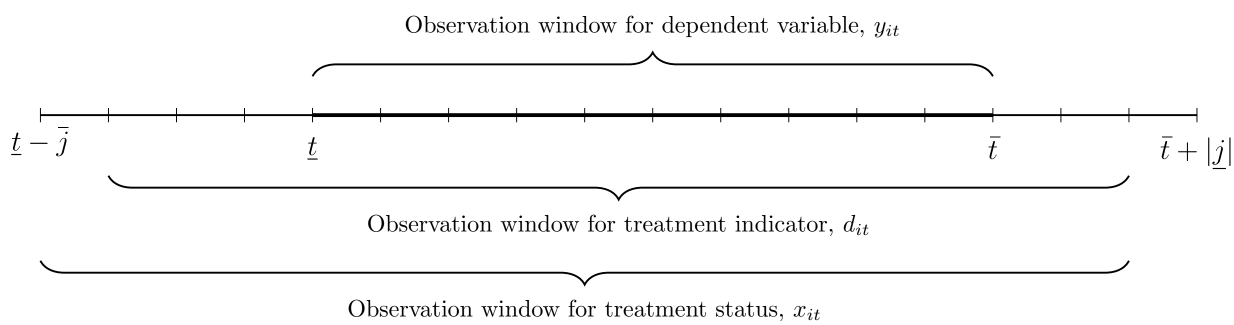

Figure 2: Data requirements

Remark 3 (Standardization)**.**
Treatment indicators b_{it}^{j} sum up to one over all j for treated units or to zero for non-treated units.[4] *The parameters β_j are therefore only identified up to a constant due to the individual fixed effect μ_i. Thus at least one coefficient β_j needs to be fixed as a standardization. We follow the standard and drop the pre treatment indicator b_{it}^{-1} from the regression, standardizing its coefficient β_{-1} to zero.*

[4] The binned event indicators b_{it}^{j} are perfectly multicollinear with the individual effect μ_i because for all $t = \underline{t}, ..., \bar{t}$: $\sum_{j=\underline{j}}^{\bar{j}} b_{it}^{j} = 1$ if $\underline{t} - \bar{j} \leq e_i \leq \bar{t} + |\underline{j}|$ and $\sum_{j=\underline{j}}^{\bar{j}} b_{it}^{j} = 0$ if $e_i < \underline{t} - \bar{j}$ or $e_i > \bar{t} + |\underline{j}|$. For non-treated as well not-yet treated and always-treated units (see below) units all indicators are zero for all t.

After the standardization $\beta_{-1} = 0$, the parameter β_j is the effect j time periods after the event compared to the level one period before the event.

2.2 Equivalence

Result 1 shows how the standard event study specification in equation (1) can be transformed into an equivalent distributed-lag model:

Result 1 (Equivalence of Event Study and Distributed-Lag Model)**.**
The standard event study specification with effect window from $\underline{j}$ to $\bar{j}$ for all $t = \underline{t}, ..., \bar{t}$

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

and binned event indicator defined in (3) is equivalent to a distributed-lag specification with $\bar{j}$ lags and $|\underline{j}|$-1 leads for all $t = \underline{t}, ..., \bar{t}$

$$y_{it} = \sum_{j=\underline{j}+1}^{\bar{j}} \gamma_j x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it} \tag{4}$$

where the explanatory variable x_{it} is a dummy variable with initial value $x_{i,\underline{t}-\bar{j}} = 0$ that changes from 0 to 1 at an event, i.e. $x_{it} = x_{i,t-1} + d_{it}$ for all $t = \underline{t} - \bar{j} + 1, ..., \bar{t} + |\underline{j}| - 1$. Given the standardization $\beta_{-1} = 0$, the event study effects β can be recovered from the distributed-lag coefficient γ as

$$\beta_j = \begin{cases} -\sum_{k=j+1}^{-1} \gamma_k & \text{if} \quad \underline{j} \leq j \leq -2 \\ 0 & \text{if} \quad j = -1 \\ \sum_{k=0}^{j} \gamma_k & \text{if} \quad 0 \leq j \leq \bar{j} \end{cases} \tag{5}$$

Proof: see Appendix A.

Result 1 encompasses various implications. Equation (5) shows that event study coefficients β are numerically identical to the cumulative sum of the distributed-lag coefficients γ_j. When looking at post-treatment, that is $j \geq 0$, the cumulative sum of γ_j for $k = 0, ..., j$ needs to be calculated. For pre-treatment, $j < 0$, we standardize the cumulated distributed-lag coefficients analogously to the event study design by setting $\gamma_{-1} = 0$ (cf. Remark 3). We then generate the cumulative sum for the pre treatment period by summing distributed-lag coefficients *away* from zero (minus one) for each $j <= -2$.

The equivalence of event study and distributed-lag model with the re-parametrization in (5) also holds in first differences:

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

$$= \sum_{j=\underline{j}+1}^{\bar{j}} \gamma_j \Delta x_{i,t-j} + \theta_t + \Delta\varepsilon_{it} \tag{7}$$

where $\Delta x_{i,t-j} = d_{i,t-j}$.

The distributed-lag specification in equation (4) immediately reveals that its estimation requires to observe the event dummy d_{it} from period $(\underline{t} - \bar{j} + 1)$ to $(\bar{t} + |\underline{j}| - 1)$ as in Remark 2. Note that the distributed-lag specification is either a regression of levels on levels (equation 4) or of changes on changes (equation 7) while the event-study specification is a regression of levels on (binned) changes (equation 1).[5]

Small differences in the notation are very important and distinguish distributed-lag and event study models. In the event study, variables b have to be defined, which is indicated by the superscript index j. In the distributed-lag model j is part of the time index and we need no further definition of the distributed-lag variables. Vice versa, we can directly interpret the coefficients of the event study design β_j, while we need to cumulate the γ coefficients of distributed-lag model to learn about the dynamic treatment effects.

Remark 4 (Deriving event study effects from the distributed-lag estimates).
Estimates for the event study effects β can be derived from the distributed-lag estimates $\widehat{\gamma}$ according to equation (5). This linear transformation transfers the statistical properties (consistency and asymptotic normality) of $\widehat{\gamma}$ to the calculated $\widehat{\beta}$. Standard errors of $\widehat{\beta}$ can be calculated from the variances and covariances of the vector $\widehat{\gamma}$ by the usual formula for linear combinations. Point estimates and standard errors of directly estimating $\widehat{\beta}$ in the event study model or of indirectly estimating $\widehat{\beta}$ in the corresponding distributed-lag model are numerically identical.

Remark 5 (Interpreting the estimated parameters).
The models do not impose any restriction on the effect size across units. It is possible that the event for unit i at time t has a different effect than the event happening in unit l at time s. The coefficient β_j, for all $j \geq 0$, is the average effect of the treatment j periods after the event. Likewise, β_j, for all $j < 0$, is the average pre-treatment effect.

[5] With an *infinite* effect window, the event study specification is just a regression of levels on changes in the treatment status, i.e. a dummy variable for an event: $y_{it} = \sum_{j=-\infty}^{\infty} \beta_j \mathbb{1}[t = e_i + j] + \mu_i + \theta_t + \varepsilon_{it} = \sum_{j=-\infty}^{\infty} \gamma_j x_{i,t-j} + \mu_i + \theta_t + \varepsilon_{it}$ where $\Delta x_{it} = d_{it} = \mathbb{1}[t = e_i + j]$ and $\gamma_j = \beta_j - \beta_{j-1}$.

2.3 Estimation and Identification

The event study model in equation (1) and the distributed-lag model in equation (4) are panel data models with a large number of units $i = 1, ..., N$ over few time periods $t = \underline{t}, ..., \bar{t}$. Both models include unit and time effects. The parameters β and γ, respectively, can therefore be estimated with standard fixed effects estimation including time period dummies.[6] These estimates are consistent and asymptotically normal as $N \to \infty$ under the standard assumptions for panel data models with large N and small T. Alternatively, both models can be estimated with standard ordinary least squares in first differences and including time period dummies as in equations (6) and (7). The first difference estimator is consistent and asymptotically normal under the same assumption as the fixed effects estimator in levels. Depending on the autocorrelation of the error term, either fixed effects estimation or estimation in first differences is more efficient.

It is important to assure that the model is econometrically identified such that the dynamic effects, β_j, are distinguished from secular time trends, θ_t. Borusyak and Jaravel (2017) show that with an infinite effect window, $[\underline{j}, \bar{j}] = [-\infty, \infty]$, the dynamic effects are only identified up to a linear trend.[7] However, limiting the effect window as in Remark 1 (by binning of the event dummies in the event study model) introduces additional restrictions that allow separately identifying dynamic effects, β_j, and secular time trends, θ_t. Intuitively, observations outside of the effect window are similar to never-treated units: they serve as a control group and help to bin down secular time trends.[8] Hence, binning of endpoints – a practical necessity in many application – allows for identification even in the absence of never-treated units and is an alternative to dropping an additional pre-treatment indicator or resorting to unit random fixed effects. At the same time, choosing the length of the effect window determines whether a unit-year observation is assigned to the treatment or the control group and hence is not innocuous.

The model is econometrically identified if (i) each lag and lead j is observed in the outcome

[6] Throughout the paper, we understand identification as the purely mechanical recovery of the coefficients of interest, β, and not their causal interpretation.

[7] In practice, an infinite effect window implies including the maximal number of leads and lags that can be observed in the data.

[8] Formally, let us adjust equation (5) in Borusyak and Jaravel (2017) to our notation for the case of an infinite effect window: $y_{it} = \sum_{j=-\infty}^{\infty} \beta_j \mathbb{1}[t = e_i + j] + \mu_i + \theta_t + \varepsilon_{it} = \sum_{j=-\infty}^{\infty} (\beta_j + \lambda \cdot j)\mathbb{1}[t = e_i + j] + (\mu_i + \lambda \cdot e_i) + (\theta_t - \lambda \cdot t) + \varepsilon_{it}$. Dynamic treatment effects are therefore only identified up to the linear trend $\lambda \cdot j$. Equivalently, in the first difference distributed-lag representation, we can write $\Delta y_{it} = \sum_{j=-\infty}^{\infty} \gamma_j \mathbb{1}[t = e_i + j] + \theta_t + \Delta\varepsilon_{it} = \sum_{j=-\infty}^{\infty} (\gamma_j + \lambda)\mathbb{1}[t = e_i + j] - \lambda + \theta_t$. The underidentification problem exists because $\sum_{j=-\infty}^{\infty} \lambda \mathbb{1}[t = e_i + j] = \lambda$. However, with a limited effect window, the equation does not hold anymore because $\sum_{j=\underline{j}+1}^{\bar{j}} \lambda \mathbb{1}[t = e_i + j] = 0 \neq \lambda$ if $t < e_i - |\underline{j}| + 1$ or $t > e_i + \bar{j}$.

window $[\underline{t}, \bar{t}]$ for at least one unit but not necessarily for the same unit, (ii) for at least one observed endpoint ($\underline{j}$ or $\bar{j}$) in t there is at least one other unit which is treated after $(t + \underline{j})$ or before $(t - |\bar{j}|)$. Condition (ii) guaranties that at least one endpoint is directly identified from a comparison with a control group for which period t is outside of its effect window. Condition (ii) is automatically satisfied in the presence of at least one never-treated unit. Condition (i) identifies all other effects either from a direct comparison with a control group or from an iterative comparison of effects. The identified endpoint allows backing out all other treatment effects and all time fixed effects iteratively. This is equivalent to the econometric identification in staggered treatment difference-in-differences designs. In Appendix B, we present intuitive examples demonstrating how identification is achieved.

2.4 A Numerical Example

We illustrate the equivalence defined in Result 1 using the following

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

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

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

The left matrix shows how the event indicator b_{it}^{j} is binned at endpoints $\underline{j}$ and $\bar{j}$ as described in Remark 1. Moreover, the need for standardization (cf. Remark 3) becomes visible as all row sums in the left matrix are equal to one. Binning also implies that if the reform had happened on or before $t = \underline{t} - \bar{j} = 2000 - 4 = 1996$ (rather than in $t = 2005$

as assumed in the example), $b_{it}^{\bar{j}} = b_{it}^{4} = 1$ for all t from 2000 to 2010. In this case of an always-treated unit, $b_{it}^{\bar{j}} = b_{it}^{4}$ is a constant and its effect is absorbed in the unit fixed effect μ_i. By the analogous argument, events on or after $t = \bar{t} + |\underline{j}| = 2010 + 3 = 2013$ imply $b_{it}^{\underline{j}} = b_{it}^{-3} = 1$ for all t from 2000 to 2010 whose effect is absorbed in μ_i. It therefore suffices to know all events from time period $\underline{t} - \bar{j} + 1 = 1997$ to $t = \bar{t} + |\underline{j}| - 1 = 2012$ to estimate the model, see Remark 2.

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

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

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

In example 1, the event study effects are calculated according to equation (5) from the distributed-lag/lead coefficients as $\beta_{-3} = -(\gamma_{-1} + \gamma_{-2})$, $\beta_{-2} = -\gamma_{-1}$, $\beta_{-1} = 0$, $\beta_0 = \gamma_0$, $\beta_1 = \gamma_0 + \gamma_1$, $\beta_2 = \gamma_0 + \gamma_1 + \gamma_2$, $\beta_3 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3$, $\beta_4 = \gamma_0 + \gamma_1 + \gamma_2 + \gamma_3 + \gamma_4$.

3 Generalization

In many applications, treatment may occur repeatedly and be of different intensities across units and/or time. A common solution is to dichotomize treatment variables and use a dummy variable that is switched on for large policy changes (see, e.g., Simon, 2016; Fuest et al., 2018). However, the parameter estimates of such an artificial dichotomization are hard to interpret both in magnitude and direction.[9] Moreover, the dichotomization of the treatment variable eliminates valuable information which could otherwise be used to identify the magnitude of the effect.

In this section, we therefore propose a generalized event study approach which accommodates institutional set-ups where multiple events with known treatment intensity take place. In particular, we show how effects of single events with varying treatment intensity, multiple events of identical intensity and multiple events of different intensities can be estimated using the full information of the treatment intensity.

3.1 Equivalence

Let us consider the general case where each unit i may be treated in any period t and treatment intensity is known but may be different in each period. In this case, d_{it} is no longer a dummy variable, but indicates the direction and intensity of events for each i and t. Typically, those kind of models are estimated in a distributed-lag model in first differences as given in (7), where treatment status, x_{it}, is defined recursively as $x_{it} = x_{i,t-1} + \Delta x_{it} = x_{i,t-1} + d_{it}$ with initial value $x_{i,\underline{t}-\bar{j}} = 0$.[10]

However, the event study approach can also accommodate this general case because the equivalence Result 1 and Remarks 1 to 4 also hold in the general case. The only difference is that the binned treatment variable b_{it}^j needs to be generated as in our alternative definition (3) and not as in the more common definition (2).

By Remark 5, the event study coefficients in the standard case measure the average event effect. These average reform effects might differ across units because of (i) heterogeneous, unit-specific event effect for the same event, (ii) homogeneous treatment effect but different

[9] To see this, consider the following case: each unit is treated once, there are two types of treatment: a small reform $d_{it}^s = 1$ or a large reform $d_{it}^\ell = 2$; treatments are distributed randomly in time and treatment effects are linear in the intensities of the reform. Ignoring small events and applying the standard event dummy set-up yields $d_{it}^s = 0$, $d_{it}^\ell = 1$. In this case, units with small reforms become part of the the control group although they respond to the reform. This induces a bias in the time fixed effects and thereby also in the treatment coefficients. Depending on elasticity of the treatment effect with respect to the reform intensity, the share of large vs. small reforms and the relative size of effect window to observation window, it is possible that estimates in the model only using the large reforms can be larger, smaller or identical to the model using all reforms. A possible fix for this ambiguity is to exclude units with small events from the sample, in which case, the model is, however, estimated on a different and possibly selective sample.

[10] Of course, distributed lag models in levels are suitable, too.

events across units.[11] The generalization removes (parts of) the latter sources of heterogeneity at the cost of an additional assumption, sacrificing parts of the non-parametric appeal of event study designs. At the same time, estimates of the treatment effect become more precise if the assumption is correct.

3.2 A Numerical Example

In the following, we provide a brief generic numerical example for the general case.

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

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

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

[11] On an abstract level, the distinction between the two is somewhat arbitrary because you could interpret an event of the same size happening in a different unit (or time) as a different event.

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

3.3 Typical Cases and Applications

In this subsection we briefly discuss typical cases of the generalized event study design and provide selected examples from recent applications in top five journals if there are any. Numerical examples for the special cases in 3.3.1 to 3.3.3 can be found in Appendix C.

3.3.1 Single Events of Varying Treatment Intensity

Consider the case where each unit receives one treatment, but treatment intensity d_i differs across units (see Appendix C.1 for a numerical example). Note that the event study specification in levels looks like a regression of levels (y_{it}) on changes ($d_{i,t-j}$). However, the identical distributed-lag model in levels shows that the model is derived from an intuitive regression of levels (y_{it}) on levels (x_{it}), i.e. on the cumulated changes d_i. Besides the classic event study, this case is quite frequently applied as it fits an institutional setting where a shock at some aggregate level hits units at a disaggregate level with different intensities (see, e.g., Alsan and Wanamaker, 2018; Charles et al., 2018; Clemens et al., 2018; Goodman-Bacon, 2018).

3.3.2 Multiple Events of Identical Intensity

Consider the case where events of identical intensity take place repeatedly for a unit. In this case, d_{it} is an event dummy that takes the value 1 in any period where an event took place and 0 in other periods. The binned treatment variable b_{it}^j is again generated by definition (3). At the endpoints $\underline{j}$ and $\bar{j}$, the binned treatment variable b_{it}^j is the backward cumulated event dummy $d_{i,t+|\underline{j}|}$ and the forward cumulated event dummy $d_{i,t-\bar{j}}$, respectively (see Appendix

C.2 for a numerical example.) Note how the more common definition of the binned treatment variable in definition (2) fails in this case.

In the case of multiple events, the parameter β_j estimates the average effect of all single and multiple events j periods after the event. Interestingly, the application of this model is relative rare and mostly used in papers applying the original model used in the finance literature (see, e.g. Dube et al., 2011, for an exception). However, many institutional set-ups would fit the model well, such as hospital admissions or firm switches. Some studies only focus on the first of potentially many events. Two things are important to note here. First, depending on the proximity of multiple events pre-treatment and post-treatment effect might overlap, which could lead to a bias. Second, following Remark 2, it is important to ensure that no event happened outside the observation window for the dependent variable but inside the window for the treatment indicator.

3.3.3 Multiple Events of Different Intensities and Direction

Last, we consider the most general case, described in Section 3.1. There are many settings that fit this model, such as multiple tax changes or minimum wage hikes, and correspondingly many applications. Traditionally, the corresponding models were framed as distributed lag models rather than event study designs (Suárez Serrato and Zidar, 2016; Drechsler et al., 2017; Fuest et al., 2018). A special case is when events have a different direction. Assume that d_{it} is a variable that takes the value 1 in periods with a "positive" treatment, -1 in periods with a "negative" treatment and 0 in periods without a treatment. The parameter β_j estimates the average effect j periods after the event of all "positive" treatments and – with reversed sign – all "negative" treatments. In other words, the effects of "positive" and "negative" treatments are assumed symmetric with opposing signs. A typical example would be the introduction of a new law in some period and the abolition of the law in some later period, or the opening and closing of plants across regions. We are not aware of any recent application of this case, while the model could be suitable for studying settings, such as introduction of a policy that will eventually sunset, or the opening and closing of plants.

4 Conclusion

In this paper, we have made remarks about the standard event study design and its generalization. First, event study designs and distributed-lag models are equivalent and lead to numerically identical parameter estimates after correct transformation of the parameters. Second, binning of the endpoints of the effect window allows identification of both secular time fixed effects and dynamic treatment effects even when no never-treated units are present. Third, classic event study designs using dummy variables as event indicators can be naturally generalized to models that account for multiple events of different intensities and signs.

References

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

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

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

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

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

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

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

de Chaisemartin, C. and X. D'Haultfoeuille (2018). Two-way fixed effects estimators with heterogeneous treatment effects. *UCSB Working Paper*.

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

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

Freyaldenhoven, S., C. Hansen, and J. M. Shapiro (2018). Pre-event Trends in the Panel Event-study Design. *Working Paper*.

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

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

Roth, J. (2018). Pre-test with Caution Event-study Estimates After Testing for Parallel Trends. *Working Paper*.

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

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

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

Appendix A Proof of Equivalence

In the following we show that the standard event study is numerically equivalent to the distributed-lag model. We start by rewriting the equation (1) in first differences. The first-difference model, which omits the first time period $\underline{t}$ and is defined for all $t = \underline{t}+1, ..., \bar{t}$, is given by

$$\Delta y_{it} = y_{it} - y_{i,t-1} = \sum_{j=\underline{j}}^{\bar{j}} \beta_j \Delta b_{it}^j + \theta_t + \Delta\varepsilon_{it}$$

where $\Delta b_{it}^j = b_{it}^j - b_{i,t-1}^j$ and where individual fixed effects μ_i cancel.

From (3), it follows that $b_{it}^j = d_{i,t-j}$ and $b_{i,t-1}^j = d_{i,t-j-1}$ for $\underline{j} < j < \bar{j}$. Moreover, verify that $\Delta b_{it}^{\underline{j}} = -d_{i,t-\underline{j}-1}$ and $\Delta b_{it}^{\bar{j}} = d_{i,t-\bar{j}}$. We can therefore rewrite for all $t = \underline{t}+1, ..., \bar{t}$

$$\begin{aligned}
\Delta y_{it} &= \sum_{j=\underline{j}}^{\bar{j}} \beta_j \Delta b_{it}^j + \theta_t + \Delta\varepsilon_{it} \\
&= \beta_{\underline{j}} \Delta b_{it}^{\underline{j}} + \sum_{j=\underline{j}+1}^{\bar{j}-1} \beta_j (b_{it}^j - b_{i,t-1}^j) + \beta_{\bar{j}} \Delta b_{it}^{\bar{j}} + \theta_t + \Delta\varepsilon_{it} \\
&= -\beta_{\underline{j}} d_{i,t-\underline{j}-1} + \sum_{j=\underline{j}+1}^{\bar{j}-1} \beta_j (d_{i,t-j} - d_{i,t-j-1}) + \beta_{\bar{j}} d_{i,t-\bar{j}} + \theta_t + \Delta\varepsilon_{it} \\
&= -\beta_{\underline{j}} d_{i,t-\underline{j}-1} + \sum_{j=\underline{j}+1}^{\bar{j}-1} \beta_j d_{i,t-j} - \sum_{j=\underline{j}+1}^{\bar{j}-1} \beta_j d_{i,t-j-1} + \beta_{\bar{j}} d_{i,t-\bar{j}} + \theta_t + \Delta\varepsilon_{it} \\
&= -\beta_{\underline{j}} d_{i,t-\underline{j}-1} + \beta_{\underline{j}+1} d_{i,t-\underline{j}-1} + \sum_{j=\underline{j}+2}^{\bar{j}-1} \beta_j d_{i,t-j} - \sum_{j=\underline{j}+2}^{\bar{j}-1} \beta_{j-1} d_{i,t-j} - \beta_{\bar{j}-1} d_{i,t-\bar{j}} + \beta_{\bar{j}} d_{i,t-\bar{j}} + \theta_t + \Delta\varepsilon_{it} \\
&= (\beta_{\underline{j}+1} - \beta_{\underline{j}}) d_{i,t-\underline{j}-1} + \sum_{j=\underline{j}+2}^{\bar{j}-1} (\beta_j - \beta_{j-1}) d_{i,t-j} + (\beta_{\bar{j}} - \beta_{\bar{j}-1}) d_{i,t-\bar{j}} + \theta_t + \Delta\varepsilon_{it} \\
&= \gamma_{\underline{j}+1} d_{i,t-\underline{j}-1} + \sum_{j=\underline{j}+2}^{\bar{j}-1} \gamma_j d_{i,t-j} + \gamma_{\bar{j}} d_{i,t-\bar{j}} + \theta_t + \Delta\varepsilon_{it} \\
&= \sum_{j=\underline{j}+1}^{\bar{j}} \gamma_j d_{i,t-j} + \theta_t + \Delta\varepsilon_{it} = \sum_{j=\underline{j}+1}^{\bar{j}} \gamma_j \Delta x_{i,t-j} + \theta_t + \Delta\varepsilon_{it}
\end{aligned}$$

where $\gamma_j = \beta_j - \beta_{j-1}$ and where we define $\Delta x_{it} = d_{it}$. Hence the event study model in differences is equivalent to a distributed-lag model in differences with $\bar{j}$ lags and $|\underline{j}| - 1$ leads.

The re-parametrization holds accordingly in levels for all $t = \underline{t}, ..., \bar{t}$ as

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

where treatment status x_{it} is defined recursively as $x_{it} = x_{i,t-1} + \Delta x_{it} = x_{i,t-1} + d_{it}$ with initial value $x_{i,\underline{t}-\bar{j}} = 0$. The respective level equation is only defined up to a constant. We can therefore set the initial treatment status $x_{\underline{t}-\bar{j}}$ to an arbitrary value, e.g. $x_{\underline{t}-\bar{j}} = 0$. Adding a constant c to x_{it} and $x_{i,t-1}$ does not affect Δx_{it}.

Appendix B Identification

In the following, we present intuitive cases that demonstrate how identification is achieved. The empirical model is described by equation (7), hence

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

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

Consider the following seven cases with an effect window from $\underline{j} = -2$ to $\bar{j} = 1$.

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

The matrix of explanatory variables in Case 1 is given by

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

where $t1$, $t2$ and $t3$ are dummy variables for the three time periods 1, 2 and 3, respectively.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

All three dynamic effects are directly identified in direct comparison to a never-treated unit. The matrix of explanatory variables is full rank. This case shows that the observation window for the dependent variable can be shorter than the effect window.

Appendix C More Numerical Examples

C.1 Single Events of Varying Treatment Intensity

Example C.1. *We assume a panel that runs from* $\underline{t} = 2000$ *to* $\bar{t} = 2010$ *and an effect window from* $\underline{j} = -3$ *to* $\bar{j} = 4$*. For individual* i*, the single event of intensity* $d_i = 0.1$ *takes place at* $e_i = 2005$*.*

The explanatory variables for the event study in levels, b^j_{it}, and in first differences, Δb^j_{it}, are

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

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

t	$x_{i,t+2}$	$x_{i,t+1}$	x_{it}	$x_{i,t-1}$	$x_{i,t-2}$	$x_{i,t-3}$	$x_{i,t-4}$	$\Delta x_{i,t+2}$	$\Delta x_{i,t+1}$	Δx_{it}	$\Delta x_{i,t-1}$	$\Delta x_{i,t-2}$	$\Delta x_{i,t-3}$	$\Delta x_{i,t-4}$
2000	0	0	0	0	0	0	0							
2001	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2002	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

C.2 Multiple Events of Identical Intensity

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

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

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

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

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