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

By Kurt Schmidheiny and Sebastian Siegloch*

February 2, 2023

with corrections from November 29, 2023†

*Schmidheiny: University of Basel, Faculty of Business and Economics, CEPR and CESifo (e-mail: kurt.schmidheiny@unibas.ch); Siegloch: University of Cologne, CEPR, ZEW, IZA, and CESifo (e-mail: siegloch@wiso.uni-koeln.de).

† The lag/lead index ℓ was erroneously written as j in several instances.

Appendix

A Numerical Examples

In A, we illustrate various event study design setups 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 treatments of identical intensities (Subsection A.2), a single treatment of different signs and varying treatment intensities (Subsection A.3), and multiple treatments of different signs and varying intensities (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 Binary Treatment

Example 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 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	$D^{-3}_{i,t}$	$D^{-2}_{i,t}$	$D^{0}_{i,t}$	$D^{1}_{i,t}$	$D^{2}_{i,t}$	$D^{3}_{i,t}$	$D^{4}_{i,t}$	$\Delta D^{-3}_{i,t}$	$\Delta D^{-2}_{i,t}$	$\Delta D^{0}_{i,t}$	$\Delta D^{1}_{i,t}$	$\Delta D^{2}_{i,t}$	$\Delta D^{3}_{i,t}$	$\Delta D^{4}_{i,t}$
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	$T_{i,t+2}$	$T_{i,t+1}$	$T_{i,t}$	$T_{i,t\text{-}1}$	$T_{i,t\text{-}2}$	$T_{i,t\text{-}3}$	$T_{i,t\text{-}4}$	$\Delta T_{i,t+2}$	$\Delta T_{i,t+1}$	$\Delta T_{i,t}$	$\Delta T_{i,t\text{-}1}$	$\Delta T_{i,t\text{-}2}$	$\Delta T_{i,t\text{-}3}$	$\Delta T_{i,t\text{-}4}$
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 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 Treatments of Identical Intensities

Example 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*, the first event takes place in* 2004 *and the second in* 2006*.* $\ell = -1$ *is taken as the reference period.*

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

t	$D^{-3}_{i,t}$	$D^{-2}_{i,t}$	$D^{0}_{i,t}$	$D^{1}_{i,t}$	$D^{2}_{i,t}$	$D^{3}_{i,t}$	$D^{4}_{i,t}$	$\Delta D^{-3}_{i,t}$	$\Delta D^{-2}_{i,t}$	$\Delta D^{0}_{i,t}$	$\Delta D^{1}_{i,t}$	$\Delta D^{2}_{i,t}$	$\Delta D^{3}_{i,t}$	$\Delta D^{4}_{i,t}$
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

t	$T_{i,t+2}$	$T_{i,t+1}$	$T_{i,t}$	$T_{i,t\text{-}1}$	$T_{i,t\text{-}2}$	$T_{i,t\text{-}3}$	$T_{i,t\text{-}4}$	$\Delta T_{i,t+2}$	$\Delta T_{i,t+1}$	$\Delta T_{i,t}$	$\Delta T_{i,t\text{-}1}$	$\Delta T_{i,t\text{-}2}$	$\Delta T_{i,t\text{-}3}$	$\Delta T_{i,t\text{-}4}$
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 Treatments of Different Signs and Varying Intensities

Example 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 following four matrices show the explanatory variables for the event study and the distributed-lag model in levels and in first differences:

t	$D^{-3}_{i,t}$	$D^{-2}_{i,t}$	$D^{0}_{i,t}$	$D^{1}_{i,t}$	$D^{2}_{i,t}$	$D^{3}_{i,t}$	$D^{4}_{i,t}$	$\Delta D^{-3}_{i,t}$	$\Delta D^{-2}_{i,t}$	$\Delta D^{0}_{i,t}$	$\Delta D^{1}_{i,t}$	$\Delta D^{2}_{i,t}$	$\Delta D^{3}_{i,t}$	$\Delta D^{4}_{i,t}$
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

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 Treatments of Different Signs and Varying Intensities

Example 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 and the distributed-lag model in levels and in first differences:

t	$D^{-3}_{i,t}$	$D^{-2}_{i,t}$	$D^{0}_{i,t}$	$D^{1}_{i,t}$	$D^{2}_{i,t}$	$D^{3}_{i,t}$	$D^{4}_{i,t}$	$\Delta D^{-3}_{i,t}$	$\Delta D^{-2}_{i,t}$	$\Delta D^{0}_{i,t}$	$\Delta D^{1}_{i,t}$	$\Delta D^{2}_{i,t}$	$\Delta D^{3}_{i,t}$	$\Delta D^{4}_{i,t}$
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

B Application to Baker and Fradkin, 2017

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) henceforth BF2017.[1] 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 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 in 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 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 PBD leads to a 2.07% drop in aggregate job search.

In the 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

$$\ln GJSI_{i,t} = \sum_{\ell=-3}^{4} \beta_\ell \Delta T_{i,t-\ell} + w'_{i,t}\xi + \mu_i + \theta_t + \varepsilon_{i,t}, \tag{1}$$

where $GJSI_{i,t}$ is the natural logarithm of the Google Job Search Index in state i and period t (year-month), $\Delta T_{i,t-\ell}$ is an indicator variable that indicates whether PBD in state i was changed

[1] Replication code in Stata and R is available in the Harvard Dataverse at https://doi.org/10.7910/DVN/LXMYV6 and in the JAE data archive at https://dx.doi.org/10.15456/jae.2023031.1844400605. We thank Jakob Miethe (LMU Munich) for collaborating on programming the R code.

$\ell \in [-3, 4]$ 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 varying intensities across US states over time. The authors analyze PBD increases and decreases in separate regressions and for different time windows. Increases in 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, $\Delta T_{i,t}$ 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 $\Delta T_{i,t}$ is switched on for decreases of 7 weeks or more. In the respective models, the event indicator $\Delta T_{i,t}$ is zero if (i) no change happened, (ii) a change of the same sign but with a 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.[2]

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 B.1, we replicate the main event study results for large increases and large decreases in the two respective samples by estimating equation (1). 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 B.1 are based on strong implicit assumptions and parameter restrictions embodied in equation (1), which speak directly to our main points raised in the previous sections. While the empirical model looks like a classic event study design and therefore innocuous at first sight, event indicators $\Delta T_{i,t}$ are not binned at the endpoints (cf. Remark 2) and no coefficient is normalized to zero (cf. Remark 1). This implies that dynamic

[2] 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.

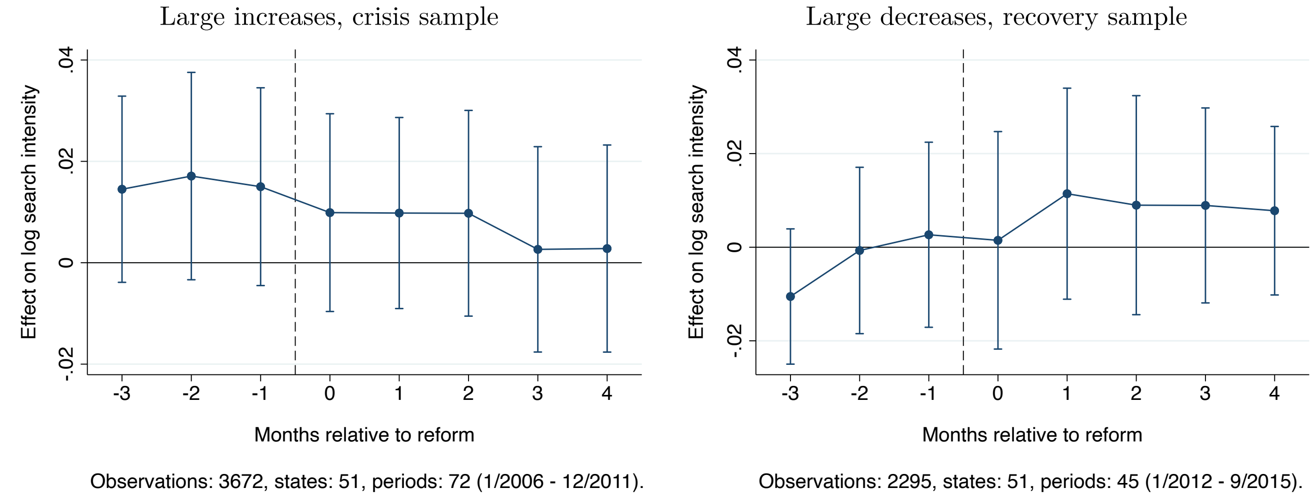
treatment effects are implicitly normalized to be zero four and more periods before the event as well as five and more periods after the event, i.e. $\beta_\ell = 0$ for all $\ell \leq -4$ and for all $\ell \geq 5$. In particular, the assumption $\beta_5 = 0$ is very strong since it assumes that the 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.

Nonetheless an inherent normalization in the model. However, the normalized coefficients are outside of the effect window. The implicit assumption is that effects for $\ell < -3$ and $\ell > 4$ are zero. We added these coefficients to Panels A and B to make the underlying assumption explicit for both models. Moreover, none of the estimated event study coefficients β_ℓ, $\ell = -3, ...4$ is normalized to zero (cf. Remark 1). Nonetheless, there is an inherent normalization in the model. However, the normalized coefficients are outside of the effect window. The implicit assumption is that effects for $\ell < -3$ and $\ell > 4$ are zero. We added these coefficients to Panels A and B to make the underlying assumption explicit for both models.

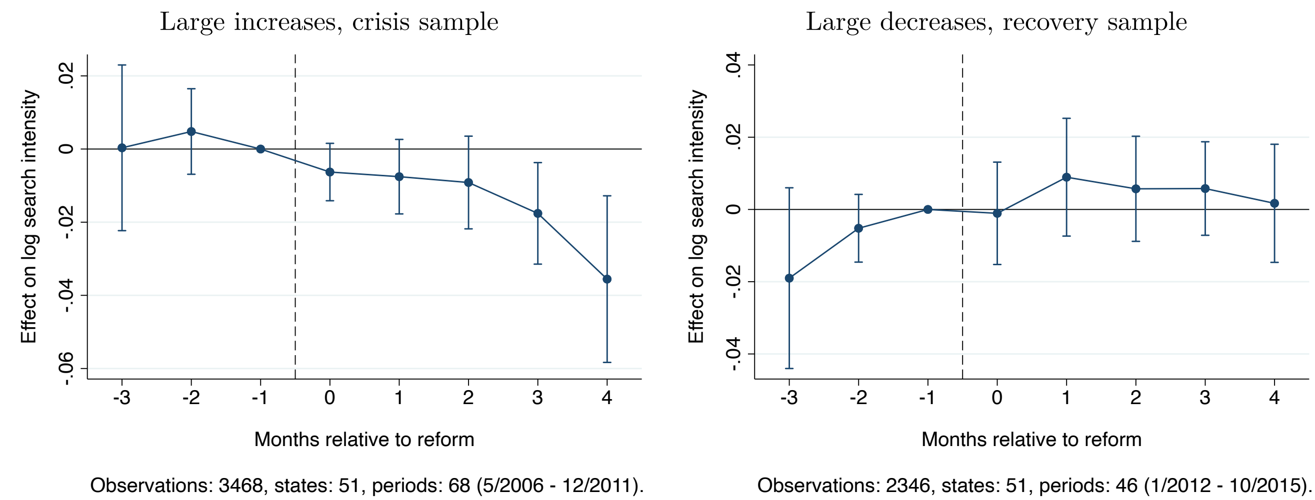
Next, we estimate equation (1) as an event study model with restrictions suggested in Sections 2.2 and 4. We bin endpoints according to Remark 2 and we normalized the pre-event coefficient $\beta_{-1} = 0$ according to Remark 1. As events can occur several times per state in our application, this leads to Case 2 "multiple treatments of identical intensities" in Section 4.2. The β-coefficients can be estimated by creating binned treatment indicators at the endpoints $\ell = -3$ and $\ell = 4$ 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 B.1 shows results with binned endpoints and normalized pre-event period. Other than the original results in Panel A, large increases in 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 before large increases in PBD, so the parallel trends assumption cannot be rejected before treatment. Hence, the estimated dynamic treatment effects are fully consistent with the simple difference-in-differences estimation. In contrast,

Figure B.1: 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 a 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 experience no changes in a certain month 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 specification 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.

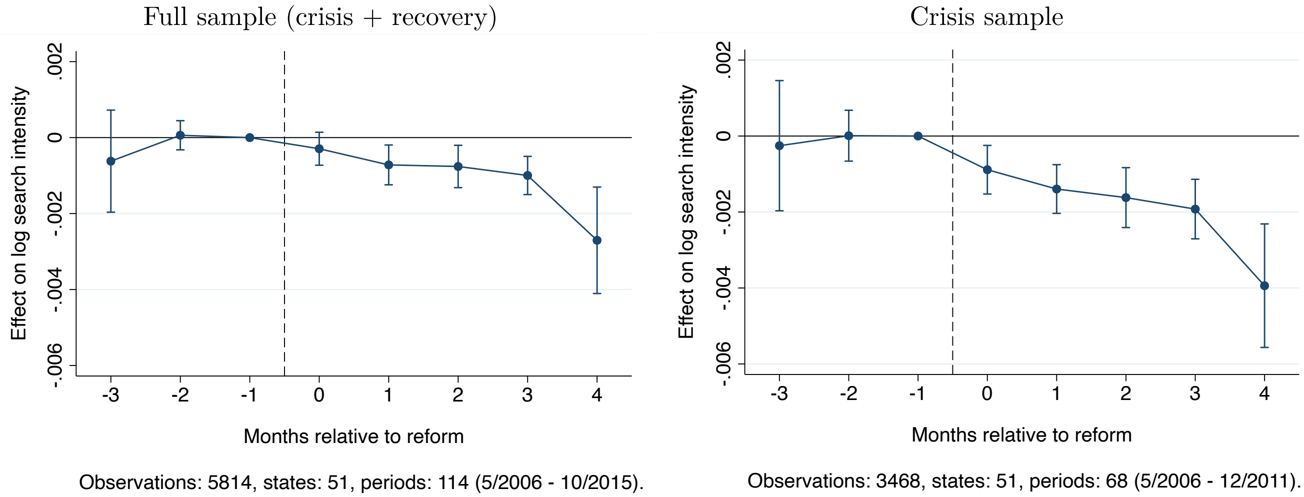
the large decreases occurring during the recovery period after the Great Recession do not seem to have a systematic effect on search intensity as shown in the right graph of Panel B of Figure B.1.

Note that the number of observations differs between Panels A and B of Figure B.1. This is 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 can generate all leads up until $\underline{\ell} = -3$ as we observe treatment status until 12/2015. However, we can calculate the first binned endpoint for a specification with four lags $D^4_{i,t}$ only in 5/2006. Consequently, our sample is $\bar{\ell} = 4$ periods shorter and $4 \cdot 51 = 204$ observations smaller. An analogous argument applies to 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 the time 1/2012. However, we cannot generate all leads on 12/2015. We have to shorten our estimation sample by $|\underline{\ell}| - 1 = 2$ 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 PBD before 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 B.1 shows that binning endpoints leads to convincing event study coefficients, which match the difference-in-differences estimates, the zero-one model 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 B.2 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

Figure B.2: Generalized Event Study Design



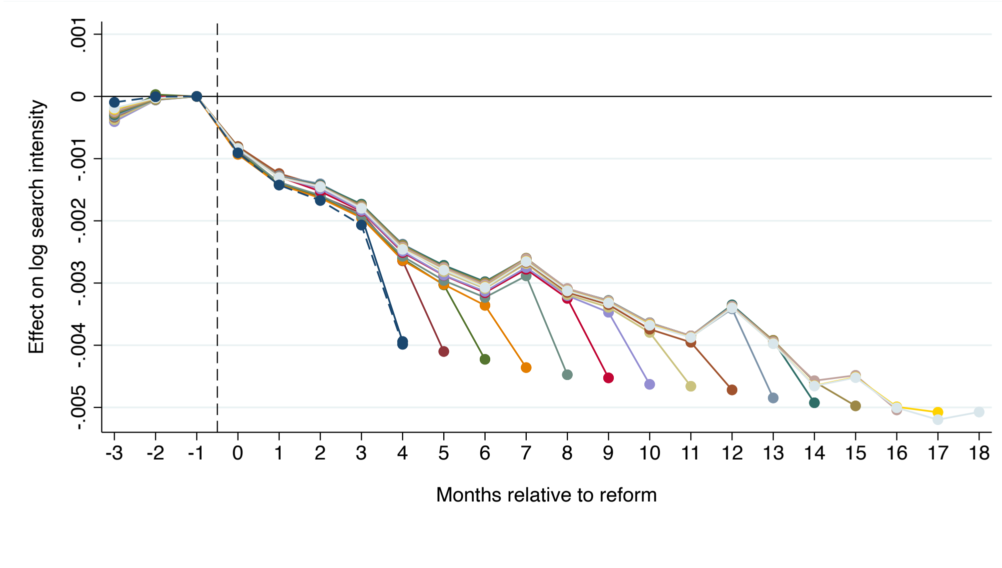
Notes: The figure plots the results when applying the generalized event study with binned endpoints and normalization at -1 to the setting in (Baker and Fradkin, 2017). The graphs show the dynamic effect of an increase in the potential benefit duration by one week on log search intensity as measured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

generalized event study design are measured on the same scale as the simple difference-in-differences model and can be readily compared (see below for more details).

Merging the 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 B.1 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 only on the crisis sample, which mainly covers increases. The right graph in Figure B.2 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 periods 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

Figure B.3: Varying the Effect Window (Crisis Sample)



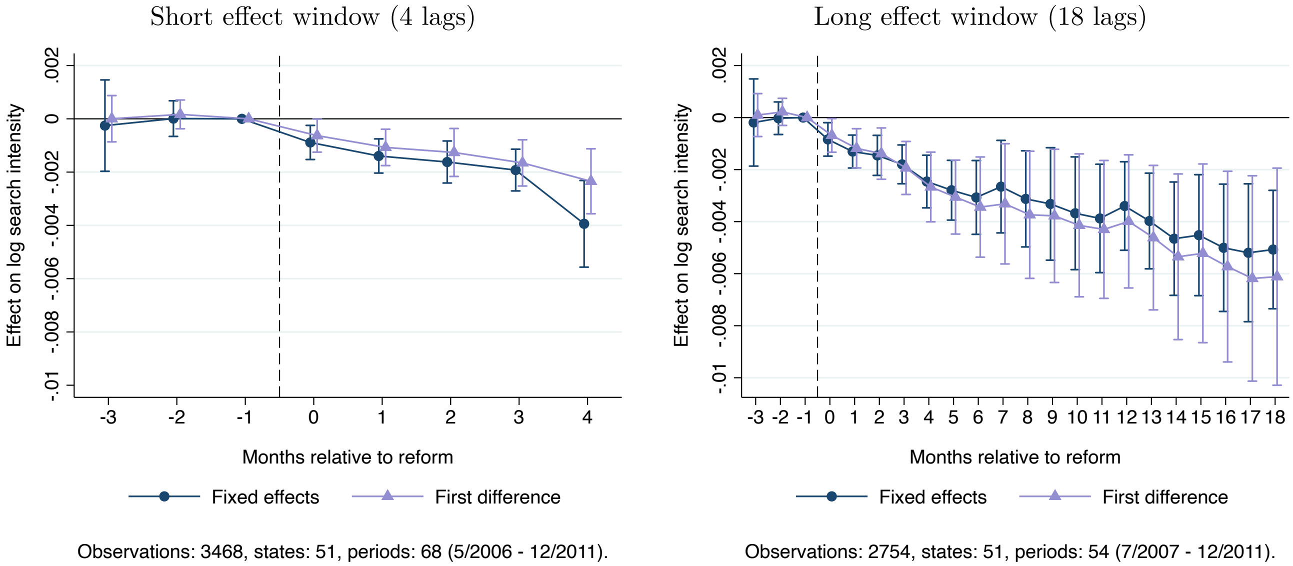
Notes: The figure plots the results when applying the generalized event study with binned endpoints and normalization at -1 to the setting in (Baker and Fradkin, 2017). The graph shows the dynamic treatment effects of an increase in the potential benefit duration by one week on log search intensity as measured by the 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 shortened sample (dashed blue line). Confidence intervals are omitted.

endpoints comes along with the assumption that dynamic treatment effects have fully materialized after $\bar{\ell}$ periods. In Figure B.2, we see that effects are still on the decline four months after the reform. Moreover, the slope of the event study graph becomes steeper between lag 3 and lag 4. As argued in Remark 3, this is an indication that 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 B.3. 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

Figure B.4: 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 graph shows the dynamic treatment effects of an increase in the potential benefit duration by one week on log search intensity as measured by the 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.

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 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 B.4, 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.[3]

[3] 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 the sample size grows both models should eventually yield identical estimates.

References

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

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

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.