Event-Studies with a Continuous Treatment

By Brantly Callaway, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna*

A researcher can analyze a two-group/two-period binary difference-in-differences (DiD) design without having to make many choices. Assuming that treatment does not affect outcomes prior to its initiation, the average treatment effect on the treated is identified under a parallel trends assumption by the difference between average outcome changes across groups: one parameter, one parallel trends assumption, and one estimator.

A more complicated DiD design, however, requires researchers to make more decisions. This paper builds on Callaway, Goodman-Bacon and Sant'Anna (2024) and considers choices facing a researcher using DiD methods to study a treatment that begins at different times (staggered) *and* affects units to different degrees (treatment intensity). We discuss transparent ways to aggregate the large set of causal parameters that arise in this setting that convey heterogeneity by event-time and across doses, and we show how these choices can simplify estimation and inference. In contrast to parameters from common two-way fixed effects linear specifications, our summary parameters do not suffer from negative/non-transparent weighting issues.

I. Some causal parameters of interest

Consider a panel dataset with N units indexed by i, and T time periods indexed by t. Denote the time unit i is first treated by $G_i \in \mathcal{G} = \{2, \cdots, T, \infty\}$, where $G_i = \infty$ means that a unit is not treated by T ("never treated"). As we focus on staggered setups, G_i can be interpreted as the "treatment timing group". Let $D_i \in \mathcal{D} \subseteq [0, d_H]$, $d_H < \infty$, denote the treatment "dose" (or intensity) unit i received when it was first treated. We interpret D_i as the "dose group".

* Callaway: University of Georgia (email: brantly.callaway@uga.edu). Goodman-Bacon: Opportunity and Inclusive Growth Institute, Federal Reserve Bank of Minneapolis (email: andrew@goodman-bacon.com). Sant'Anna: Emory University (email: pedro.santanna@emory.edu). We thank Alex Bartik for guidance about the data used in this paper.

We adopt the potential outcome framework and write $Y_{i,t}(g, d)$ as the potential outcome of unit i at time t if such a unit is first treated in period g, with dose d; we write $Y_{i,t}(0) = Y_{i,t}(\infty, 0)$ for units that remain untreated by the last time period of available data. This notation defines group-time-dose-specific average treatment effects:

$$ATT(g, t, d) = E[Y_t(g, d) - Y_t(0)|G = g, D = d].$$

$ATT(g, t, d)$ is the average treatment effect in period t of (i) becoming treated in period g and (ii) experiencing dose d versus zero dose, among those in timing group g that received treatment dose d. $ATT(g, t, d)$ parameters describe what some treatment actually achieved. They embody many types of heterogeneity that researchers often want to report and interpret. For a given g and t, average effects at different doses describe group g's "dose-response" function. For a given d and g, differences in average causal effects across t represent treatment effect dynamics. Finally, for a given d and t, different average effects by g capture a combination of dynamics and differences in the effect of being treated at a given time.

With continuous treatments, another class of causal parameters that may be of interest are average causal response parameters, defined as

$$ACR(g, t, d) = \frac{\partial E[Y_t(g, \tilde{d})|G = g]}{\partial \tilde{d}}\bigg|_{\tilde{d}=d}.$$

$ACR(g, t, d)$ is the average causal response to a marginal change in the dose at d for all units in timing group g. ACR parameters answer causal questions about what level of treatment matters more or less. This slope parameter is also a function of g, t, and d, and can vary in these dimensions in meaningful ways.

In staggered DiD setups with continuous treatments, it may not be practical to estimate one dose-response type of function for each t and g. Thus, researchers may want to aggregate these "building blocks" by time-since-treatment (event-time; $e = t - g$) and/or across doses. Here,

we focus on summarizing $ATT(g,t,d)s$ across both margins, and then discuss how these aggregation choices shape estimation. See Callaway, Goodman-Bacon and Sant'Anna (2024) for aggregations of $ACR(g,t,d)$s.

II. Event-study-type parameters

We start discussing event-study aggregations that average over treatment dosages. For a given group g and time t, let $ATT^o(g,t) = E[ATT(g,t,D)|G=g, D>0]$ be the average ATT for that group in a given point in time, and let

$$ATT^{es}(e) = E[ATT^o(G, G+e)|G+e \in [2,T], D>0]$$

denote the average treatment effect among those that have been exposed to any treatment for exactly e periods, conditional on being observed having participated in the treatment for that number of periods ($G+e \in [2,T]$), and being ever-treated ($D>0$). When D is binary, $ATT^{es}(e)$ reduces to the event-study coefficient considered by Callaway and Sant'Anna (2021).

The above parameters fully aggregate across doses and thus do not describe treatment effect dynamics for "higher-dose" or "lower-dose" groups, for instance. To tackle these questions, we can "partially" aggregate the doses in each group and time period to form the "dose-aware" event-study parameters

$$ATT^{es}_{d_1,d_2}(e)$$
$$= E[ATT^o_{d_1,d_2}(G, G+e)|d_1 \le D \le d_2, G+e \in [2,T]],$$

where $ATT^o_{d_1,d_2}(g, g+e) = E[ATT(g,t,D)|G = g, d_1 \le D \le d_2]$, and $0 < d_1 \le d_2$ are thresholds within the support of D. By picking different intervals, one can assess how average treatment effect dynamics vary across dosage groups. For instance, one could set d_1 and d_2 to split treated units into those with above- or below-median doses. Of course, one can entertain finer dose partitions. We recommend paying attention to the effective sample size in each chosen partition.

A feature of $ATT^{es}_{d_1,d_2}(e)$ is that it aggregates across several treatment dosages. In some applications, however, researchers may want to report detailed heterogeneity with respect to d. Aggregating over some event-times can facilitate reporting estimated dose-response functions. To formalize this idea, let e_1 and e_2 be two post-treatment event-times such that $0 \le e_1 \le e_2$ and let

$$ATT^{es}_{e1,e2}(d) = \frac{\sum_{e=e_1}^{e_2} E\left[ATT(G, G+e, d)\middle| G+e_2 \in [2,T], D=d\right]}{e_2 - e_1 + 1}$$

be the average treatment effect of receiving dose d, among all units that have been treated for at least e_2 periods, averaged over event-times e_1 to e_2. When $e_2 = e_1 = e$, $ATT^{es}_{e1,e2}(d)$ provides a dose-response function for all units that have been treated for exactly e periods. When $e_2 > e_1$, however, $ATT^{es}_{e_1,e_2}(d)$ averages these curves over event-time and can be used to summarize dose-responses in terms of "short-run" (e.g., $e_1 = 0, e_2 = 2$) and "long-run" (e.g., $e_1 = 3, e_2 = 4$) effects. When constructing these dose-response curves averaged over event-times, we impose that the data is balanced in event-time within the e_1 to e_2 window ($G+e_2 \in [2,T]$), so compositional changes are not a concern, but it is possible to relax this restriction.

Next, we show that these event-study functionals are identified under traditional DiD assumptions.

III. Identification of event-study parameters

The following assumptions are sufficient to identify the event-study parameters. Let $\mathcal{D}_+ = \mathcal{D} \setminus \{0\}$ denote the positive part of the support of D. Let $\Delta Y_t = Y_t - Y_{t-1}$, and write $W_t = D1\{t \ge G\}$.

ASSUMPTION 1: *The observed data consists of $\{Y_{i,1}, \ldots, Y_{i,T}, D_i, G_i\}_{i=1}^n$ which is independent and identically distributed.*

ASSUMPTION 2: *(i) $\mathcal{D}_+ = [d_L, d_U]$ with $0 < d_L < d_U < \infty$, (ii) $P(D=0) > 0$ and $dF_{D|G}(d|g) > 0$ for all $(g,d) \in (\mathcal{G} \setminus \{\infty\}) \times \mathcal{D}_+$, (iii) For all $g \in (\mathcal{G} \setminus \{\infty\})$ and $t = 2, \ldots, T$, $E[\Delta Y_t | G=g, D=d]$ is continuously differentiable in d on $\mathcal{D}_+$.*

ASSUMPTION 3: *(i) For all $g \in \mathcal{G}$ and $t = 1, \ldots, T$ with $t < g$, $Y_{i,t}(g,d) = Y_{i,t}(0)$ a.s. (ii) $W_{i,1} = 0$ a.s. and for $t = 2, \ldots, T$, $W_{i,t-1} = d$ implies that $W_{i,t} = d$ a.s.*

ASSUMPTION 4: *For all $(g,g') \in \mathcal{G} \times \mathcal{G}$, $t = 2, \ldots, T$, and $(d,d') \in \mathcal{D} \times \mathcal{D}$, $E[\Delta Y_t(0)|G=g, D=d] = E[\Delta Y_t(0)|G=g', D=d']$.*

Assumption 1 states that we have panel data. Assumption 2 states that we have a set of units that are never-treated and that treatment is continuous; if there are no never-treated units, we can restrict attention to periods $t = 1, \ldots, \bar{G} - 1$, where $\bar{G} = \max\{G_i : G_i < \infty\}$ is the time of the last-treated group. Assumption 3 imposes no-anticipation and that treatment is staggered. Finally, Assumption 4 is a parallel trends condition that states that, in the absence of treatment, the average evolution of the untreated potential outcomes is the same across time-dosage groups. These assumptions are similar to those in Callaway, Goodman-Bacon and Sant'Anna (2024).

Let U_t be a generic binary variable that takes value one if a unit is part of a user-chosen "clean" comparison group at time t; e.g., the not-yet-treated indicator $U_t = 1\{G > t \cup D = 0\}$. Define $\theta^o_{d_1,d_2}(g,t) = E\big[Y_t - Y_{g-1}|G = g, d_1 \leq D \leq d_2\big] - E[Y_t - Y_{g-1}|U_{\max\{t,g-1\}} = 1, d_1 \leq D \leq d_2 \cup D = 0]$, and $\theta^o(g,t) = \theta^o_{d_L,d_H}(g,t)$. In addition, for every unit i such that $G_i + e_2 \in [2,T]$, let $\widetilde{Y}_i^{e_1,e_2}(g) = 1\{G_i = g\}\big[\sum_{e=e_1}^{e_2}\big(Y_{i,g+e} - Y_{i,g-1}\big)\big/(e_2 - e_1 + 1) - \sum_{e=e_1}^{e_2} E[Y_{g+e} - Y_{g-1}|U_{g+e_2} = 1]\big/(e_2 - e_1 + 1)\big]$.

THEOREM 1: *Under Assumptions 1 to 4, (i)* $ATT^{es}(e) = E[\theta^o(G, G+e)|G+e \in [2,T], D > 0]$*, (ii)* $ATT^{es}_{d_1,d_2}(e) = E[\theta^o_{d_1,d_2}(G, G+e)|G+e \in [2,T], d_1 \leq D \leq d_2]$*, and (iii)* $ATT^{es}_{e_1,e_2}(d) = E[\widetilde{Y}^{e_1,e_2}(G)|G + e_2 \in [2,T], D = d]$.

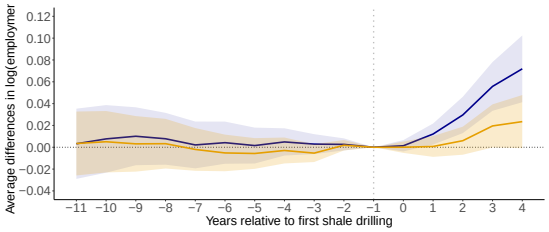
Theorem 1(i) shows that we can essentially ignore treatment intensity when focusing on the event-study-type parameters $ATT^{es}(e)$, and therefore, use the estimators from staggered DiD setups with a binary treatment to estimate these parameters. For instance, with $U_t = 1\{G > t \cup D = 0\}$, one can use Callaway and Sant'Anna (2021) event-study estimators with the not-yet-treated units as a comparison group; this involves "labeling" units with $D = 0$ as $G = \infty$. Theorem 1(ii) shows that this is still the case when one wants to present event-studies that only partially aggregate across dosages. More specifically, to estimate $ATT^{es}_{d_1,d_2}(e)$, one needs to first subset the observations only to contain units that received treatment dose between d_1 and d_2, or those that received zero dose, $D = 0$. Within this subset of units, one can then "ignore" the treatment intensity and proceed as estimating $ATT^{es}(e)$.

Theorem 1(iii) has a slightly different flavor than the previous parts, as $ATT^{es}_{e_1,e_2}(d)$ is a dose-response parameter. However, it also has important, practical implications. More specifically, it highlights that one can rely on the dose-response-curve estimators proposed by Callaway, Goodman-Bacon and Sant'Anna (2024) for two-periods setups, as $\widetilde{Y}_i^{e_1,e_2}(g)$ essentially involves only one pre- and post-treatment period. In setups with a large number of cross-sectional units, one can leverage Chen, Christensen and Kankanala (2023)'s data-driven nonparametric estimator discussed in Callaway, Goodman-Bacon and Sant'Anna (2024). When the sample size is limited, one can choose a flexible parametric model (e.g., splines with a fixed number of knots) to approximate the dose-response curve.

IV. An application to fracking

Bartik et al. (2019*a*) use a staggered and non-binary treatment variable to study the local economic effects of hydraulic fracking. Fracking is only possible in areas with underground shale formations that can be fractured ("fracked") to release hydrocarbons. The costs and yield of fracking—"prospectivity"—depend on geologic factors that vary continuously across areas over a given shale formation. We slightly modify the DiD research design in Bartik et al. (2019*a*) by exploiting variation in the timing of fracking activity across shale formations from 2001-2014 (G_i, hand-collected by the authors) and continuous variation in prospectivity score across counties (D_i, purchased from Rystad Energy); see Bartik et al. (2019*b*). We denote counties with zero prospectivity score as "never-treated" and set $G_i = \infty$ for them. We focus on the log of total county employment as our main outcome, and use not-yet-treated units as the comparison group in all estimates below.

Figure 1 display results for $ATT^{es}_{d_1,d_2}(e)$ using two sets of (d_1, d_2): the orange curve sets $d_1 = 0.20$ and $d_2 = 3.95$, where 0.20 and 3.95 are the minimum and the median fracking exposure among counties with positive exposure. We refer to this group as "low dose". The blue curve sets d_1 slightly above 3.95 and $d_2 = 9.35$, where the latter is the maximum fracking exposure. We refer to this group as "high dose". Pre-trends seem to be parallel for both groups for 11 years prior to fracking, supporting the paral-

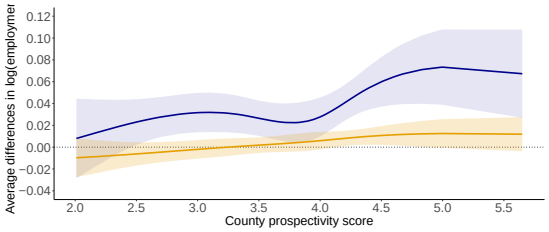


Notes: Solid lines denotes estimates of $ATT^{es}_{d_1,d_2}(e)$ using Callaway and Sant'Anna (2021). Shaded areas are 95% pointwise confidence intervals. The orange (blue) curve sets d_1 and d_2 to 0.2 and 3.95 (3.96 and 9.35).

FIGURE 1. EVENT STUDY ESTIMATES FOR HIGH AND LOW-DOSE GROUPS

lel trends assumption. "Low-dose" counties have treatment effects that are not statistically different from zero until 3 years after fracking activity begins, when they are about 2%. "High-dose" counties, on the other hand, have larger average effects that grow from 2% in the year after fracking to 7% higher employment four years after fracking.

Figure 2 displays results for time-averaged dose-response curves, $ATT^{es}_{e_1,e_2}(d)$: the orange curve sets e_1 and e_2 to 0 and 2 ("short-run") and the blue curve uses 3 and 4 ("long-run"). We use cubic splines with internal knots at the $25th$, $50th$, and $75th$ percentile of the dose. This figure echoes the conclusions from Figure 1 that short-run effects are smaller than long-run effects and counties with higher prospectivity scores have larger employment effects from fracking. This aggregation, however, shows where in the distribution of prospectivity these effects are largest. For example, in the longer-run, average employment effects are similarly large for all counties with scores above about 4.



Notes: Solid lines denotes estimates of $ATT^{es}_{e_1,e_2}(d)$ using Callaway, Goodman-Bacon and Sant'Anna (2024). Shaded areas are 95% pointwise confidence intervals. The orange (blue) curve sets e_1 and e_2 to 0 and 2 (3 and 4).

FIGURE 2. ESTIMATED DOSE-RESPONSE CURVES FOR SHORT AND LONG-RUN EFFECTS

V. Conclusion: choosing an aggregation

Because of the sheer number of "building block" parameters in a staggered and continuous DiD design (e.g., $ATT(g,t,d)$), researchers first need to *choose* a way to aggregate them. This choice should be driven by the research question. In our example, interest in the dynamic effects of fracking suggests event-study aggregations for bins of prospectivity. Interest in the relationship between natural resource endowments and local economies suggests dose-response aggregations in different post-fracking time windows.

Once aggregate summary parameters have been chosen, however, many simple DiD estimation tools are available. Thus, in practice, careful choices about aggregation help with both the interpretation and application of "new" DiD methods.

REFERENCES

Bartik, Alexander W., Janet Currie, Michael Greenstone, and Christopher R. Knittel. 2019*a*. "The Local Economic and Welfare Consequences of Hydraulic Fracturing." *American Economic Journal: Applied Economics*, 11(4): 105–55.

Bartik, Alexander W., Janet Currie, Michael Greenstone, and Christopher R. Knittel. 2019*b*. "Replication package for: The Local Economic and Welfare Consequences of Hydraulic Fracturing." American Economic Association. `https://www.aeaweb.org/journals/dataset?id=10.1257/app.20170487`.

Callaway, Brantly, and Pedro H. C. Sant'Anna. 2021. "Difference-in-differences with multiple time periods." *Journal of Econometrics*, 225(2): 200–230.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. C. Sant'Anna. 2024. "Difference-in-Differences with a Continuous Treatment." *arXiv:2107.02637*.

Chen, Xiaohong, Timothy M. Christensen, and Sid Kankanala. 2023. "Adaptive Estimation and Uniform Confidence Bands for Nonparametric Structural Functions and Elasticities." *Review of Economic Studies*, Forthcoming.

Proof of Theorem 1

First, notice that part (i) is a special case of part (ii) with $d_1 = d_L$ and $d_2 = d_H$. Thus, the proof of part (i) follows directly from part (ii).

For part (ii), it suffices to show that, for every $g \in \mathcal{G} \setminus \{\infty\}$, and every time period $t = 2, \ldots, T$, $\theta^o_{d_1,d_2}(g,t) = E[ATT(g,t,D)\big|G = g, d_1 \leq D \leq d_2]$, where $0 < d_1 \leq d_2$. Towards that end, note that under Assumptions 1 to 4, we have that, for any $d \in [d_1, d_2]$, $g \in \mathcal{G} \setminus \{\infty\}$,

$$\begin{aligned} ATT(g,t,d) =& E[Y_t - Y_{g-1}\big|G = g, D = d] \\ & - E[Y_t - Y_{g-1}\big|U_{\max\{t,g-1\}} = 1, d_1 \leq D \leq d_2 \cup D = 0], \end{aligned} \tag{A1}$$

which follows from Theorem 3.1 and Appendix SA of Callaway, Goodman-Bacon and Sant'Anna (2024), and the fact that all units with $U_{\max\{t,g-1\}} = 1$ are untreated by time t. From (A1) and the law of iterated expectations, it follows that

$$\begin{aligned} E[ATT(g,t,D)\big|G = g, d_1 \leq D \leq d_2] =& E[Y_t - Y_{g-1}\big|G = g, d_1 \leq D \leq d_2] \\ & - E[Y_t - Y_{g-1}\big|U_{\max\{t,g-1\}} = 1, d_1 \leq D \leq d_2 \cup D = 0], \end{aligned} \tag{A2}$$

and the right-hand side of (A2) is the definition of $\theta^o_{d_1,d_2}(g,t)$. This establishes that $\theta^o_{d_1,d_2}(g,t) = E[ATT(g,t,D)\big|G = g, d_1 \leq D \leq d_2]$ and concludes the proof of part (ii) of Theorem 1. As mentioned above, part (i) of Theorem 1 follows by taking $d_1 = d_L$ and $d_2 = d_H$.

Next, we prove part (iii) that $ATT^{es}_{e_1,e_2}(d) = E[\widetilde{Y}^{e_1,e_2}(G)\big|G + e_2 \in [2,T], D = d]$. Towards that end, first notice that, for every $d \in \mathcal{D}_+$, every group $g \in \mathcal{G} \setminus \{\infty\}$ and time period $t \in [2,T]$ such that $t \leq g + e_2, g + e_2 \in [2,T]$, $e_2 \geq 0$, we have from Appendix SA of Callaway, Goodman-Bacon and Sant'Anna (2024) that under Assumptions 1 to 4,

$$ATT(g,t,d) = E[Y_t - Y_{g-1}\big|G = g, D = d] - E[Y_t - Y_{g-1}\big|U_{g+e_2} = 1].$$

Thus, it follows that, for a given group g that satisfies the above restrictions, and for any $0 \leq e_1 \leq e_2$,

$$\frac{\sum_{e=e_1}^{e_2} ATT(g,g+e,d)}{e_2 - e_1 + 1} = \sum_{e=e_1}^{e_2} \frac{E[Y_{g+e} - Y_{g-1}\big|G = g, D = d] - E[Y_{g+e} - Y_{g-1}\big|U_{g+e_2} = 1]}{e_2 - e_1 + 1}. \tag{A3}$$

Next, notice that from the linearity property of conditional expectations and from the fact that the data is balanced in event-time (so there are no compositional changes across event-times), we have that, for every g such that $g + e_2 \in [2,T]$, and every $d \in \mathcal{D}_+$,

$$\begin{aligned} E[\widetilde{Y}^{e_1,e_2}(G)\big|G = g, D = d] &= E\left[\frac{\sum_{e=e_1}^{e_2}\left(Y_{g+e} - Y_{g-1}\right) - E[Y_{g+e} - Y_{g-1}\big|U_{g+e_2} = 1]}{e_2 - e_1 + 1}\middle| G = g, D = d\right] \\ &= \sum_{e=e_1}^{e_2} \frac{E[Y_{g+e} - Y_{g-1}\big|G = g, D = d] - E[Y_{g+e} - Y_{g-1}\big|U_{g+e_2} = 1]}{e_2 - e_1 + 1} \\ &= \frac{\sum_{e=e_1}^{e_2} ATT(g,g+e,d)}{e_2 - e_1 + 1}, \end{aligned} \tag{A4}$$

where the last equality follows from (A3).

From the definition of conditional expectations and its linearity property, it follows from (A4) that

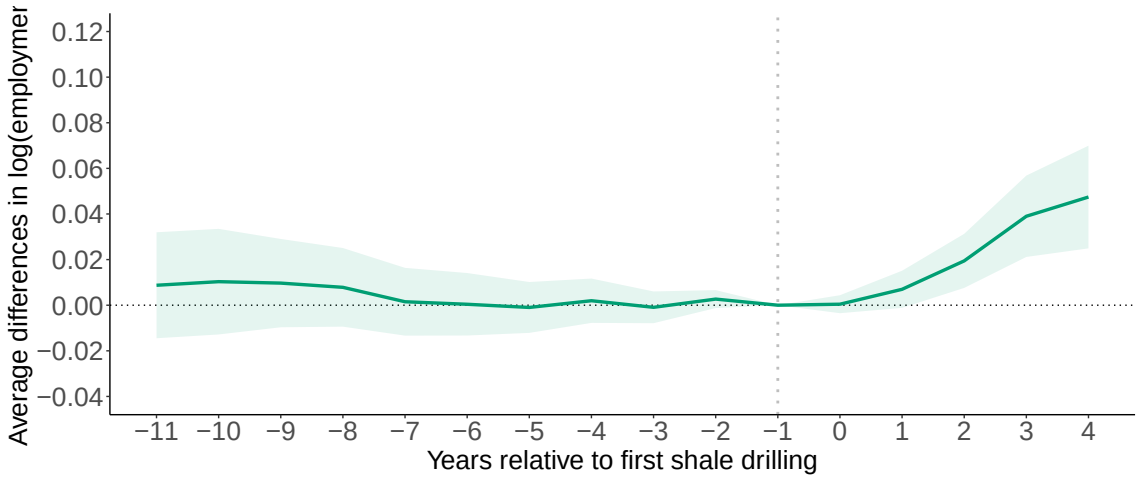
$$E[\widetilde{Y}^{e_1,e_2}(G)\big|G + e_2 \in [2,T], D = d] = \frac{\sum_{e=e_1}^{e_2} E\left[ATT(G,G+e,d)\middle|G + e_2 \in [2,T], D = d\right]}{e_2 - e_1 + 1},$$

which is what we wanted to show. This concludes the proof of Theorem 1. ■

Additional plots for empirical application

We now complement the empirical analysis of our main text related to Bartik et al. (2019*a*). As discussed in Section IV, Bartik et al. (2019*a*) use a staggered and non-binary treatment variable to study the local economic effects of hydraulic fracking, and we slightly modify the DiD research design in their paper by exploiting variation in the timing of fracking activity across shale formations from 2001-2014 (G_i, hand-collected by the authors) and continuous variation in prospectivity score across counties (D_i, purchased from Rystad Energy); see Bartik et al. (2019*b*). We denote counties with zero prospectivity score as "never-treated" and set $G_i = \infty$ for them. We use the log of total county employment as the outcome of interest and use not-yet-treated units as the comparison group in all estimates below.

In the main text, we report in Figure 1 estimates of $ATT^{es}_{d_1,d_2}(e)$ using two sets of (d_1, d_2): the orange curve sets $d_1 = 0.20$ and $d_2 = 3.95$, where 0.20 and 3.95 are the minimum and the median fracking exposure among counties with positive exposure ("low dose"), whereas the blue curve sets d_1 slightly above 3.95 and $d_2 = 9.35$, where 9.35 is the maximum fracking exposure ("high dose"). In some applications, we expect researchers also to want to report an "overall" event-study aggregation, $ATT^{es}(e)$, as discussed in our main text. Figure B1 presents estimates of such event-study coefficients using the event-study estimators proposed by Callaway and Sant'Anna (2021).



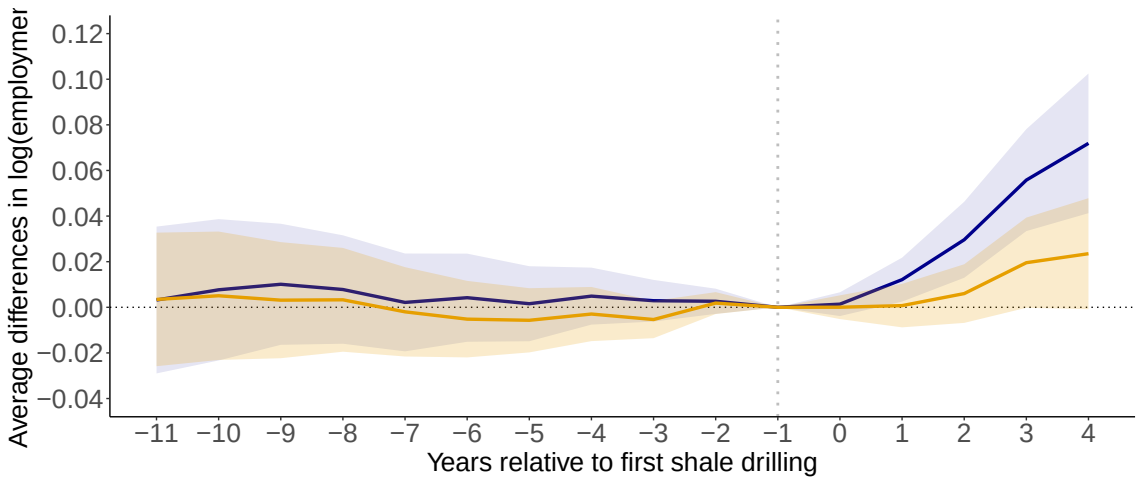
Notes: Solid lines denotes estimates of $ATT^{es}(e)$ using Callaway and Sant'Anna (2021). Shaded areas are 95% pointwise confidence intervals.

Figure B1. Overall event-study estimates

As one should expect, the event-study estimates in Figure B1 are an average of the "high dose" and "low dose" event-study estimates in Figure 1 from the main text (which we reproduce as Figure B2 to facilitate comparisons). From Figure B1, one can see that non-parallel pre-trends are not a major concern, and that longer-run effects are stronger than shorter-run ones.

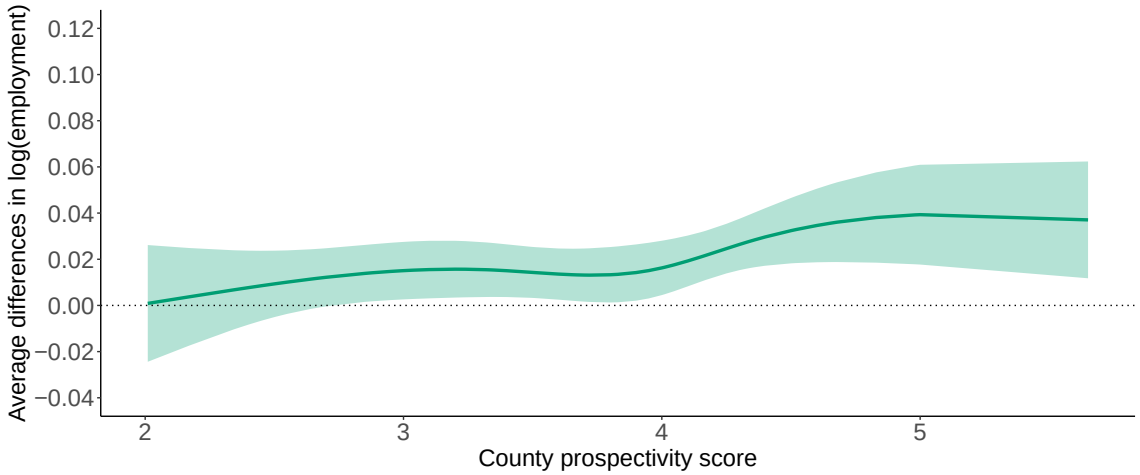
We next move to estimates of time-averaged dose-response curves, $ATT^{es}_{e_1,e_2}(d)$. Figure 2 in the main text displays results for time-averaged dose-response curves, $ATT^{es}_{e_1,e_2}(d)$ using Callaway, Goodman-Bacon and Sant'Anna (2024)'s estimators with cubic splines: the orange curve sets e_1 and e_2 to 0 and 2 ("short-run"), and the blue curve uses 3 and 4 ("long-run"). We reproduce Figure 2 as Figure B4 below to facilitate comparisons. Similar to the above, we expect that some researchers may be interested in reporting an "overall" dose-response curve, e.g., by setting $e_1 = 0$ and $e_2 = 4$. We report estimates of this in Figure B3.

Figure B3 echoes the conclusions from the "long-term" dose-response results in Figure B4 that counties with higher prospectivity scores have larger employment effects from fracking. Figure B3 also highlights that average employment effects in the first 4 years after fracking are similarly large



Notes: Solid lines denotes estimates of $ATT^{es}_{d_1,d_2}(e)$ using Callaway and Sant'Anna (2021). Shaded areas are 95% pointwise confidence intervals. The orange (blue) curve sets d_1 and d_2 to 0.2 and 3.95 (3.96 and 9.35).

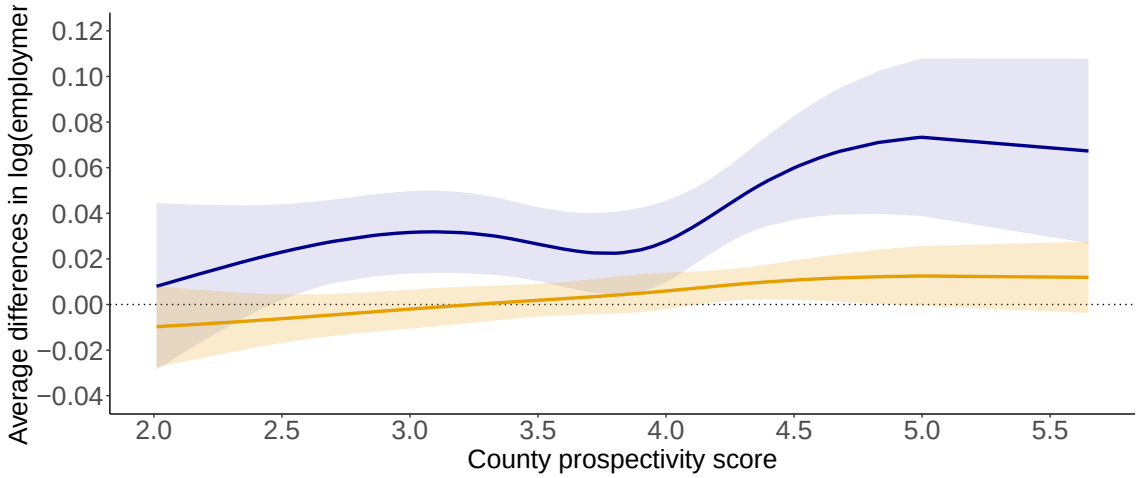
FIGURE B2. EVENT STUDY ESTIMATES FOR HIGH AND LOW-DOSE GROUPS



Notes: Solid lines denotes estimates of $ATT^{es}_{e_1,e_2}(d)$ using Callaway, Goodman-Bacon and Sant'Anna (2024), with $e_1 = 0$ and $e_2 = 4$. Shaded areas are 95% pointwise confidence intervals.

FIGURE B3. TIME-AVERAGED ESTIMATED DOSE-RESPONSE CURVES

for all counties with scores above about 2.5; it is not just the most fracking-amenable counties that drive its labor market effects.



Notes: Solid lines denotes estimates of $ATT^{es}_{e_1,e_2}(d)$ using Callaway, Goodman-Bacon and Sant'Anna (2024). Shaded areas are 95% pointwise confidence intervals. The orange (blue) curve sets e_1 and e_2 to 0 and 2 (3 and 4).

FIGURE B4. ESTIMATED DOSE-RESPONSE CURVES FOR SHORT AND LONG-RUN EFFECTS