

# Problem Set 2 — Solution Write Up

Matias Palmunen

January 22, 2026

This write-up summarizes the main results of the second problem set in Financial Economics II. The analysis replicates and critically reassesses the Difference-in-Differences (DiD) event-study results in Gallagher [2014], which studies how households update flood risk beliefs following flood events, as proxied by insurance take-up.

Rest of the text is structured as follows: First, I provide a brief overview of the paper and its main empirical specification, which relies on a two-way fixed effects event-study design. I then replicate the baseline DiD event-study results and examine the sensitivity of inference to alternative standard error clustering choices. Next, I apply the Goodman–Bacon decomposition to diagnose how the two-way fixed effects estimator aggregates comparisons across treated and untreated cohorts under staggered treatment timing. Finally, I implement the imputation estimator proposed by Borusyak et al. [2024], which constructs counterfactual outcomes using only untreated observations, mitigating the imputation problem in standard TWFE methodology.

## 1 Summary of Gallagher [2014]

Gallagher [2014] studies whether experiencing a flood increases demand for flood insurance, interpreting insurance take-up as a proxy for belief updating about low-probability, high-impact risks. The analysis uses community-level panel data from the National Flood Insurance Program (NFIP) merged with flood occurrence data, with treatment defined as the first flood experienced by a community. Treatment timing varies across communities, resulting in staggered adoption.

The main empirical specification is a two-way fixed effects (TWFE) event study model:

$$\ln(\text{policies per capita})_{ct} = \sum_{\ell=-T}^T \beta_{\ell} D_{ct}^{\ell} + \lambda_c + \gamma_{st} + \varepsilon_{ct} \quad (1)$$

where  $D_{ct}^{\ell} = \mathbb{1}[t - T_c = \ell]$  indicates that community  $c$  is  $\ell$  years from its first flood (occurring at year  $T_c$ ),  $\lambda_c$  are community fixed effects, and  $\gamma_{st}$  are state-by-year fixed effects. The reference period is  $\ell = -1$ . The coefficients  $\beta_{\ell}$  trace out the dynamic treatment effect under identifying assumptions.

Identification relies on a parallel trends assumption: in the absence of treatment, outcomes for treated and control communities would have evolved similarly over time. In settings with staggered treatment timing and potentially dynamic effects, however, the TWFE estimator aggregates comparisons between treated and untreated communities as well as between earlier- and later-treated cohorts, which may complicate the causal interpretation of the dynamic coefficients.

Analyzing the robustness of the Gallagher [2014] results to the comparison problem is the main goal of this report.

## 2 DiD Event Study Replication

I estimate the DiD event-study specification in Gallagher [2014] using community and state-by-year fixed effects, with standard errors clustered at the state level to account for spatial correlation in shocks. The sample is restricted in two ways. First, communities treated in the first panel year are excluded, since no pre-treatment observations are available for them. Second, never-treated communities are retained throughout and serve as untreated controls.

Following Gallagher [2014] and the discussion in Baker et al. [2022], I bin relative-time indicators at  $\pm 11$  years around treatment. Binning pools sparsely populated event times in the tails of the distribution, improving precision without materially affecting estimates in the central event window. For comparison, I also report results without binning.

Figure 1 displays the estimated event-study coefficients with and without binning. Pre-treatment estimates ( $\ell < 0$ ) are close to zero and statistically insignificant in both specifications, providing no evidence of differential pre-trends. Post-treatment coefficients ( $\ell \geq 0$ ) exhibit a sharp increase in insurance take-up at the time of the flood, followed by a gradual decline over subsequent years. In the binned and unbinned specification, the estimated effect becomes statistically indistinguishable from zero at longer horizons, while, naturally, the unbinned specification shows substantially wider confidence intervals in the tails due to limited support. The qualitative pattern is similar across specifications, consistent with a temporary increase in insurance demand that fades over time.

Despite these patterns, two concerns arise when interpreting TWFE estimates in staggered DiD settings. First, inference may be sensitive to the choice of clustering level, motivating the comparison of alternative standard error specifications in the next section. Second, when treatment effects are dynamic, the TWFE estimator combines comparisons between treated and untreated units with comparisons that use already-treated units as controls, which may complicate the causal interpretation of the event-study coefficients. I assess the importance of these issues using a Goodman–Bacon decomposition in Section 4 and by implementing an imputation-based DiD estimator in Section 5.

## 3 Robustness to standard errors

In panel settings with clustered errors, inappropriate clustering can lead to severely understated standard errors when regression residuals are correlated within groups over time or space. A common guideline is to cluster at least at the level at which treatment varies. However, when shocks are correlated at higher levels, clustering at a finer level may still yield invalid inference. In this application, treatment varies at the community level, but flood risk, insurance markets, and regulatory environments are plausibly correlated within states.

Following Gallagher [2014], I cluster standard errors at the state level. This choice is appropriate for three reasons. First, floods and disaster responses are spatially correlated within states due to shared weather and regulatory frameworks. Second, state-level clustering allows for arbitrary serial correlation within communities as well as cross-community correlation within states. Third, clustering at too fine a level (e.g. the community level) risks understating uncertainty when within-state dependence is present.

Figure 2 compares event-study estimates under three clustering approaches: community-level, state-level, and two-way state-by-year clustering. Community-level clustering, naturally, yields the smallest confidence intervals, while state-level and two-way clustering produce substantially wider intervals, reflecting stronger assumptions about residual dependence. Overall two-way clustering yields the largest average standard errors, providing a conservative upper bound.

Importantly, the qualitative pattern of results is stable across all clustering choices: pre-

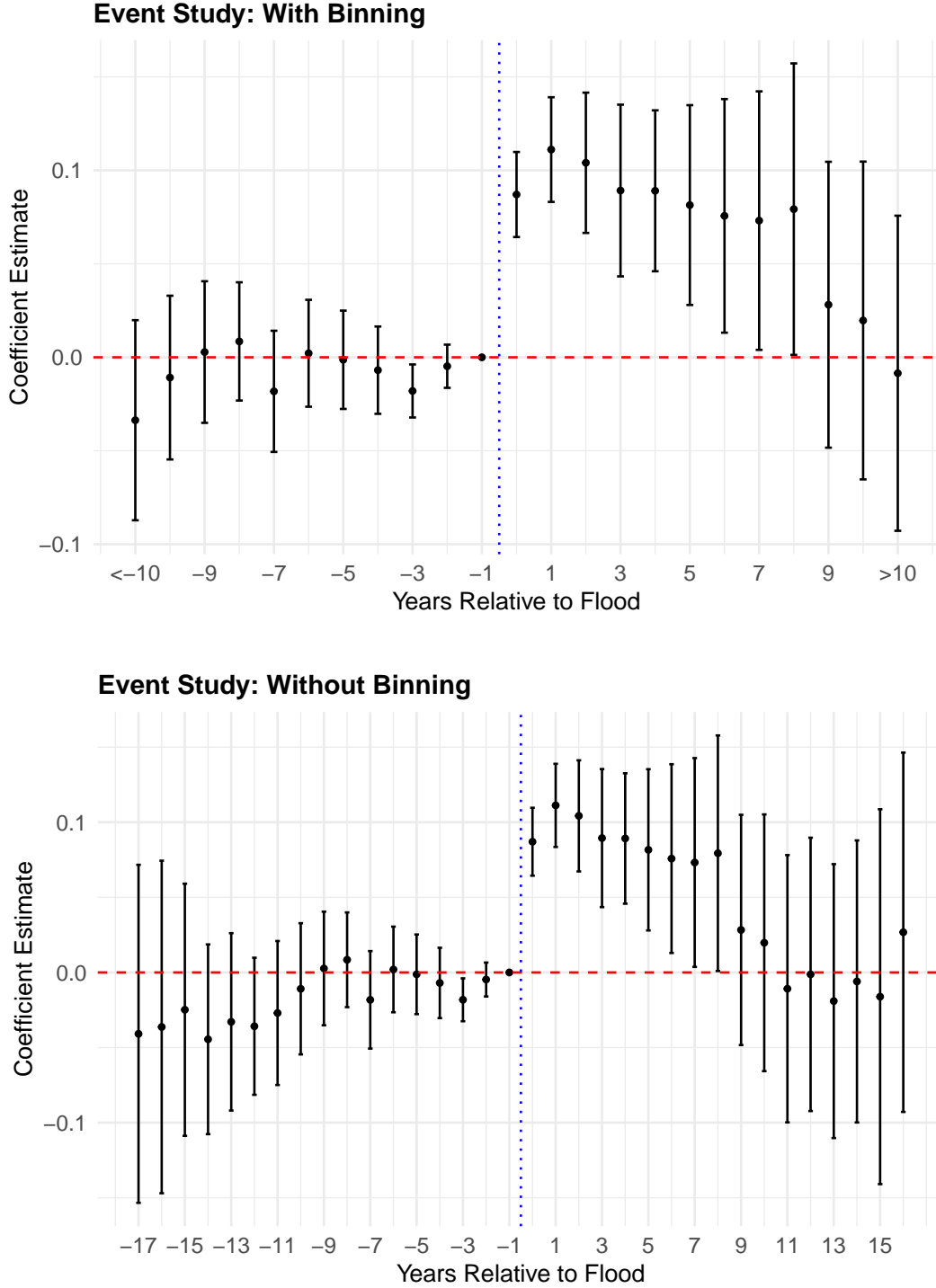


Figure 1: Event Study Analysis: Effect of Floods on Log Insurance Policies. Each panel displays point estimates (dots) with 95% confidence intervals (vertical bars) for event-time coefficients from a two-way fixed effects regression. The dashed horizontal line marks zero effect; the dotted vertical line marks the treatment period. The top panel bins endpoints at  $\pm 11$  years ( $\ell < -10$  and  $\ell > 10$ ) to pool sparse observations in the tails. The bottom panel includes all available event times without binning. Both specifications estimate model 1 using `feols` from the `fixest` package with community fixed effects ( $\lambda_c$ ), state-by-year fixed effects ( $\gamma_{st}$ ), and standard errors clustered at the state level. The reference period is  $\ell = -1$  (normalized to zero). X-axis labels show every other event time for readability.

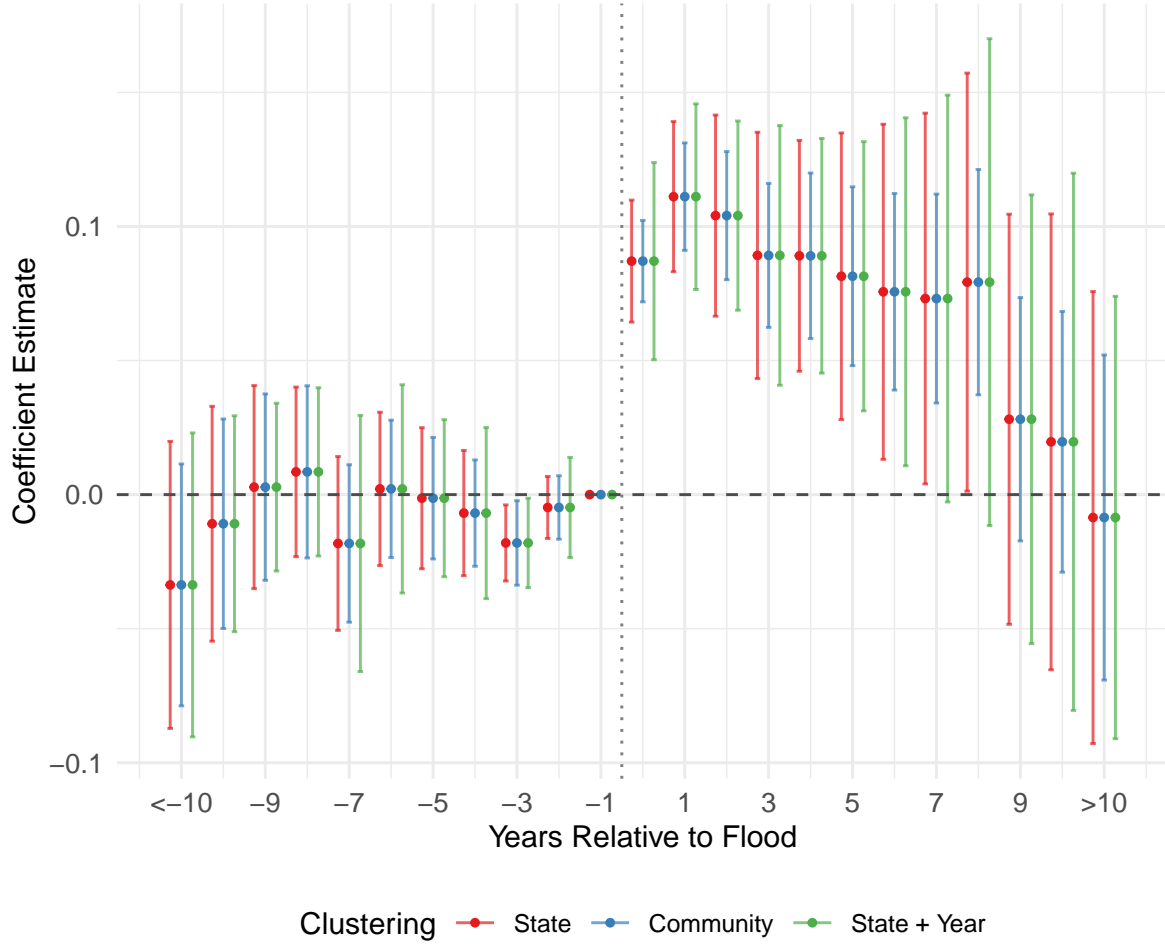


Figure 2: Comparison of Confidence Intervals by Clustering Method. Event-time coefficients (points) with 95% confidence intervals (error bars) from three clustering specifications applied to the same TWFE regression with binning at  $\pm 11$  years: community-level (blue), state-level (red), and two-way state-year (green). Points are horizontally offset for visibility. Wider confidence intervals reflect larger standard errors. Two-way clustering yields the most conservative inference, accounting for both within-state spatial correlation and common time shocks.

treatment coefficients remain statistically insignificant, while post-treatment effects are positive and decay over time.

## 4 Bacon-Goodman decomposition

I next implement the Bacon–Goodman decomposition to examine how the TWFE estimator aggregates different  $2 \times 2$  difference-in-differences comparisons in this staggered treatment setting. The decomposition partitions the TWFE estimate into contributions from treated–untreated, earlier–vs–later treated, and later–vs–earlier treated comparisons. Table 1 and Figure 3 summarize the results.

Most of the identifying variation in the TWFE estimator comes from treated–vs–untreated comparisons, which receive approximately 75% of the total weight. These comparisons use never-treated or not-yet-treated communities as controls and, under the parallel trends assumption, identify the average treatment effect on the treated. A smaller share of the weight is placed on earlier–vs–later treated comparisons, which are valid during periods when the later-treated cohort has not yet been exposed.

At the same time, roughly 10% of the total weight comes from later–vs–earlier treated comparisons. In these comparisons, already-treated units serve as controls, so differences in outcomes reflect post-treatment dynamics rather than untreated counterfactual trends. When treatment effects are dynamic, as suggested by the learning-and-decay mechanism in Gallagher [2014], these comparisons do not have a clean interpretation and may distort the estimated dynamic effects.

Table 1: Bacon-Goodman Decomposition: Weights and Estimates by Comparison Type

type	Total Weight	Avg Estimate	Contribution
Earlier vs Later Treated	0.1674	0.1035	0.0173
Later vs Earlier Treated	0.1017	0.2102	0.0214
Treated vs Untreated	0.7309	0.1147	0.0839

*Note:* Decomposition of the TWFE estimator from a specification with community and year fixed effects.

Each row shows a comparison type’s total weight in the overall estimate, the weighted average  $2 \times 2$  DiD estimate for that type, and its contribution to the overall TWFE coefficient (weight  $\times$  estimate). Treated vs. Untreated comparisons use never-treated or not-yet-treated units as controls. Earlier vs. Later comparisons use later-treated units as controls before their treatment. Later vs. Earlier comparisons (potentially problematic) use already-treated units as controls.

Overall the Bacon–Goodman decomposition, and the large contribution from good comparisons and low contribution of bad comparisons, suggests that the TWFE estimates are primarily driven by comparisons with a clear causal interpretation, but that a non-negligible share of the weight arises from potentially problematic comparisons. This motivates the use of alternative estimators that explicitly exclude already-treated units as controls, which I implement in the next section.

## 5 Imputation estimator

I next implement the imputation estimator of Borusyak et al. [2024], which delivers a clean estimate of the average treatment effect on the treated (ATT) by constructing counterfactual outcomes using only never-treated and not-yet-treated observations. By excluding already-treated units from the control group, this approach avoids the contamination that can arise in TWFE estimators under staggered treatment timing and dynamic effects.

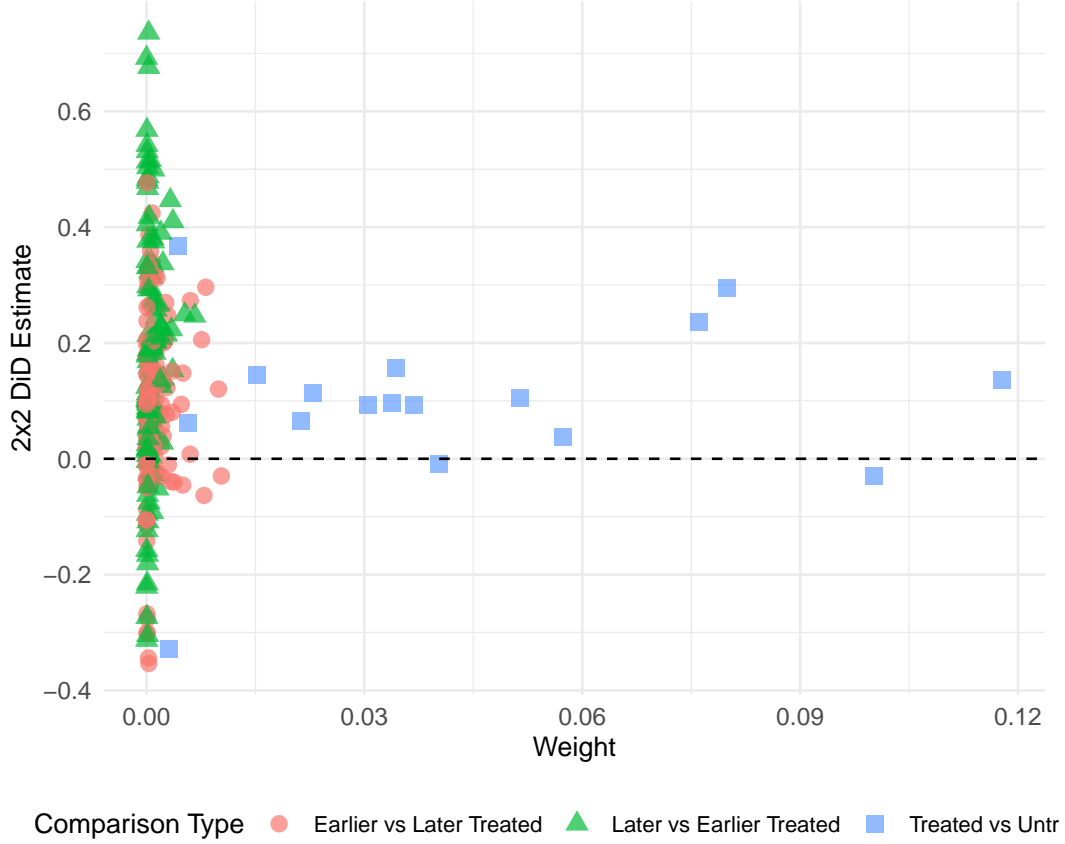


Figure 3: Bacon-Goodman Decomposition of TWFE Estimator. Each point represents a  $2 \times 2$  difference-in-differences comparison between two timing groups, with the x-axis showing the weight assigned to that comparison in the overall TWFE estimate and the y-axis showing the estimated treatment effect from that comparison. Point shapes and colors distinguish three comparison types: Treated vs. Never Treated (comparisons using never-treated units as controls), Earlier vs. Later Treated (comparisons where the later-treated group serves as control before their treatment), and Later vs. Earlier Treated (potentially problematic comparisons using already-treated units as controls). The decomposition uses a TWFE specification with community and year fixed effects, as the `bacondecomp` package does not support state-by-year fixed effects.

I implement the estimator in four steps. First, I estimate the community and state-by-year fixed effects using only untreated observations. Second, I impute counterfactual untreated outcomes for treated observations using the estimated fixed effects.<sup>1</sup> Third, I compute unit-time treatment effects as the difference between observed and imputed outcomes. Finally, I aggregate these unit-time effects by averaging over treated observations to obtain the ATT.

Figure 4 presents the corresponding event-study estimates. Pre-treatment coefficients are close to zero and statistically insignificant, providing support for parallel trends in untreated potential outcomes. Post-treatment effects are positive and statistically significant, with a sharp increase at the time of the flood followed by a gradual decline over subsequent years. The dynamic pattern closely mirrors that obtained from the TWFE event study, indicating that the main qualitative findings are not driven by problematic treated-versus-treated comparisons.

Table 2 compares ATT estimates and standard errors across three implementations: a manual imputation estimator with naive standard errors, a manual estimator with state-level clustering, and the `did_imputation` package implementation. Point estimates are very similar across methods, with ATT estimates of approximately 0.08. The naive manual implementation yields implausibly small standard errors, reflecting a failure to account for serial and state level correlation in residuals. Once state-level clustering is applied, the manual standard errors increase substantially and become comparable to those produced by the package implementation.

Table 2: Comparison of ATT Estimates and Standard Errors Across Imputation Implementations

Method	ATT Estimate	Standard Error	SE Note
Manual Implementation	0.0830	0.0045	Incorrect (naive)
Manual Implementation (clustered)	0.0830	0.0254	Correct (clustered)
Package ( <code>did_imputation</code> )	0.0797	0.0195	Correct (clustered)

*Note:* Average treatment effect on the treated (ATT) estimates from three implementations of the imputation estimator. All three use the same two-stage procedure: (1) estimate fixed effects from untreated observations, (2) compute treatment effects as  $\hat{\tau}_{it} = Y_{it} - \hat{Y}_{it}(0)$  where  $\hat{Y}_{it}(0) = \hat{\lambda}_i + \hat{\gamma}_{st}$ . The first row shows a manual implementation with naive standard errors that ignore clustering and first-stage estimation uncertainty. The second row applies state-level clustering. The third row uses the `did_imputation` package implementation. Point estimates are nearly identical across methods, but naive standard errors severely understate uncertainty by failing to account for within-state correlation and two-stage estimation.

Overall, the imputation estimator confirms the main findings from the TWFE analysis while providing treatment effect estimates with a clear causal interpretation in a staggered DiD setting.

## 6 Conclusion

This assignment replicates and extends the main findings of Gallagher [2014] on the effect of floods on flood insurance demand using modern difference-in-differences methods. I reproduce the original TWFE event-study results, diagnose the role of staggered treatment timing using the Goodman–Bacon decomposition, and implement the imputation estimator of Borusyak et al. [2024], which yields treatment effect estimates with a clear causal interpretation under dynamic effects.

Across all approaches, the estimated effects are qualitatively similar: floods generate a sharp increase in insurance take-up that gradually fades over time. While the Bacon decomposition

<sup>1</sup>For visualization purposes, counterfactuals are imputed for all observations (treated and not yet treated) in order to plot pre-treatment event-time coefficients. When computing the ATT, only treated observations are used.

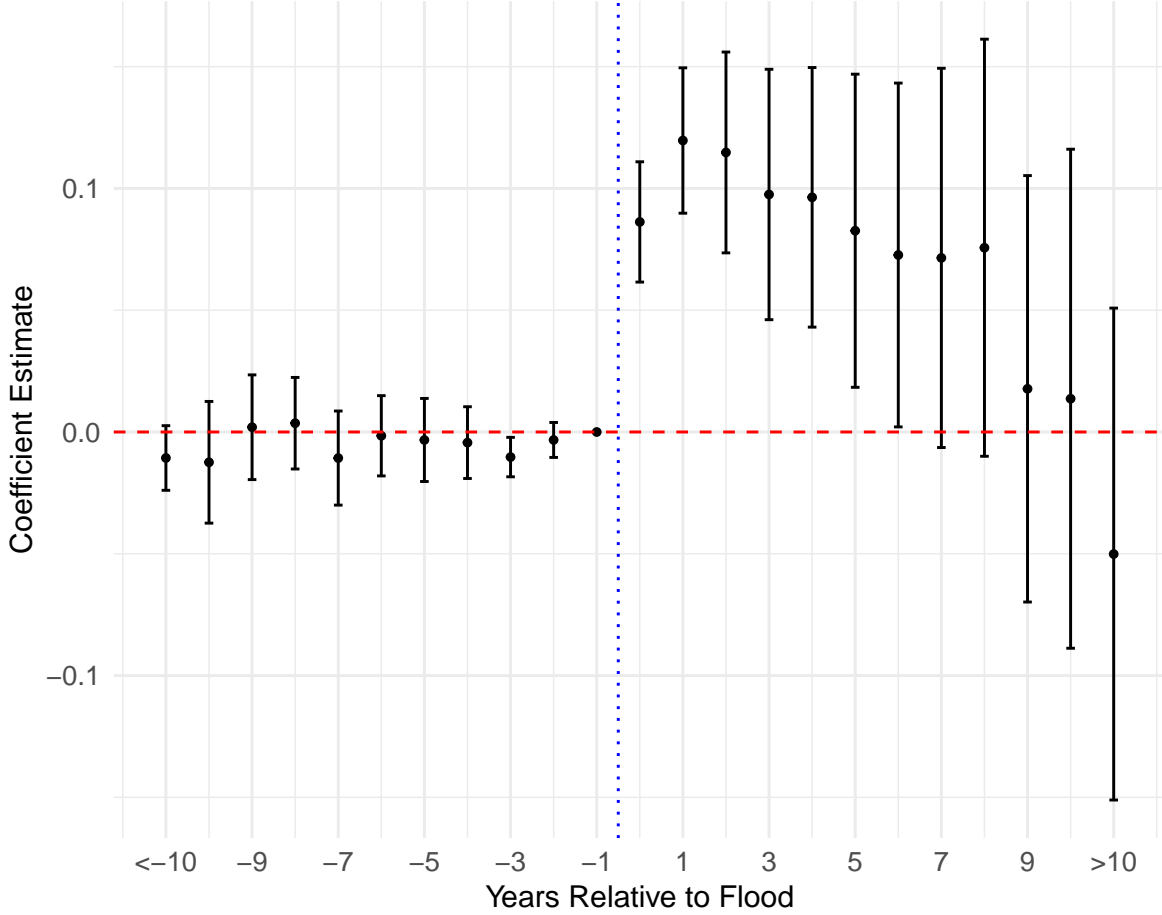


Figure 4: Event Study: Imputation Estimator (Manual Implementation). Point estimates (dots) with 95% confidence intervals (vertical bars) for event-time coefficients from a two-stage imputation procedure following Borusyak et al. [2024]. First stage: community and state-by-year fixed effects ( $\hat{\lambda}_i$ ,  $\hat{\gamma}_{st}$ ) are estimated using only untreated observations (never-treated and not-yet-treated). Second stage: unit-time treatment effects  $\hat{\tau}_{it} = Y_{it} - (\hat{\lambda}_i + \hat{\gamma}_{st})$  are regressed on binned event-time indicators with binning at  $\pm 11$  years ( $\ell < -10$  and  $\ell > 10$ ). The dashed horizontal line marks zero; the dotted vertical line marks treatment. Reference period is  $\ell = -1$  (normalized to zero). Standard errors are clustered at the state level.



reveals that a non-negligible share of the TWFE estimate is driven by treated-versus-treated comparisons, the imputation estimator confirms that the main conclusions are not an artifact of these comparisons.

## References

- Andrew C Baker, David F Larcker, and Charles CY Wang. How much should we trust staggered difference-in-differences estimates? *Journal of Financial Economics*, 144(2):370–395, 2022.
- Kirill Borusyak, Xavier Jaravel, and Jann Spiess. Revisiting event study designs: Robust and efficient estimation. *Review of Economic Studies*, 91(6):3253–3285, 2024.
- Justin Gallagher. Learning about an infrequent event: Evidence from flood insurance take-up in the united states. *American Economic Journal: Applied Economics*, pages 206–233, 2014.