SFI Léman PhD Financial Econometrics
Cross Section and Panel Data – Problem Set 2 – Solution

Andreas Fuster, SFI@EPFL

January 24, 2026

(a) We read (skim through) the paper.

(b) First, we prepare the event study plot. We bin the relative time indicators beyond -10 and 10. We also retain communities that have never experienced a flood. Lastly, we omit the indicator for the relative time -1. Depending on your design choices, the plot you obtained might be slightly different but should follow the same pattern as in Figure 1. Here, we used the plotting command "event_plot" - in the solution do-file, you can find another approach that prepares the plot "by hand"; this might be useful in case you cannot easily pass your estimates to the plotting command (e.g. if your results are generated as a matrix from a custom-made code).

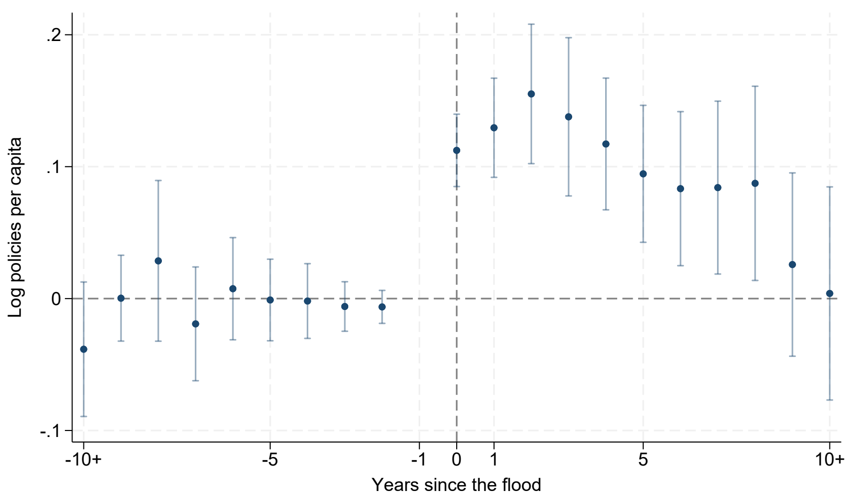

Figure 1: This figure presents the dynamic treatment effects of a flood on (log) number of flood insurance policies per capita.

Since we omitted the relative time -1, we can interpret the coefficients as the percentage increase/decrease in the number of flood insurance policies per capita for a treated (flooded) community relative to the year before a flood. The author clusters the standard errors at the state level, which is (perhaps excessively) conservative. The choice of clustering here depends on what level of aggregation we believe the flood "treatment" is applied at. One could also argue for simply clustering at the community level, in which case we observe the standard errors to be, as expected, smaller. In this particular case, the choice of clustering does not have the potential to alter our conclusions as the treatment effect is large and strongly significant.

(c) The approach in the paper might be problematic in the presence of heterogeneous treatment effects. As we learned in the lecture, staggered difference-in-differences estimates might be biased in that case. For example, if treatment effects vary across time, the staggered TWFE DiD might be biased because it implicitly uses the earlier-treated units as controls for the later-treated units. In this exercise, we try to quantify the extent of the problem by carrying out the Goodman-Bacon decomposition (Figure 2). We note that comparisons with the never-treated units account for most of the weight (around 70%); the estimated average treatment effect implied by observations in that group is only slightly smaller than the overall effect (which is 0.121). On the other hand, the potentially problematic late-to-early comparisons have a total weight of around 10%. For this group, the implied effect is nearly two times higher than the estimated ATT (around 0.21). Hence, these comparisons pull our treatment effect estimate upwards. However, their overall weight is not large, and they do not drive our result.

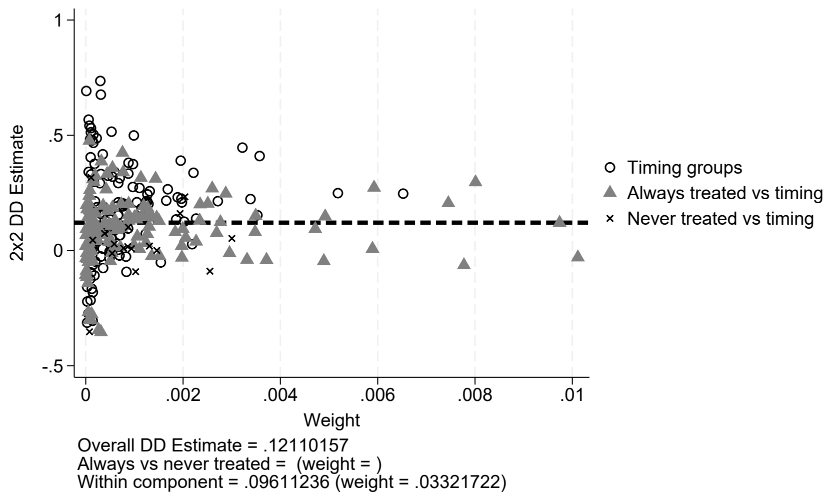

Figure 2: This figure presents the Goodman-Bacon decomposition of our treatment effect.

(d) Next, we use the imputation estimator as explained in the exercise (see the do-file for the implementation). Figure 3 presents the resulting event-study plot. First, we note that the plot is very similar to the one using TWFE although the estimated effects are slightly smaller. Note that if you also generated this plot from the package, your estimates in the pre-period will likely look different. The reason for this is that the package implements a separate procedure to test for pre-trends so their pre-treatment coefficients come from another "parallel" model estimated only on the sample of pre-treatment observations.

(e) Lastly, we estimate the ATT using both our custom solution and did_imputation package. Depending on whether you used the time fixed-effects or the state-time fixed effects you should have obtained the ATT of 0.0891 or 0.0797. If everything was done correctly, the estimate obtained "by hand" should be the same as the one from the package. In either case, we note that the ATT obtained with the imputation estimator is lower than that obtained using the traditional TWFE difference-in-differences approach. This corroborates our findings from the Goodman-Bacon decomposition in point (c) which suggested that our estimate might be biased upwards due to the implicit late-to-early comparisons. We also note that our approach leads to much lower standard errors than those returned by the package. The intuition here is that when running the second-stage regression which retrieves the ATT, we ignore the additional uncertainty implied by the first-stage imputation step. One solution to this problem is presented by Boryusak et al. and implemented in the package. Another option could be jointly bootstrapping the two steps of the estimator.

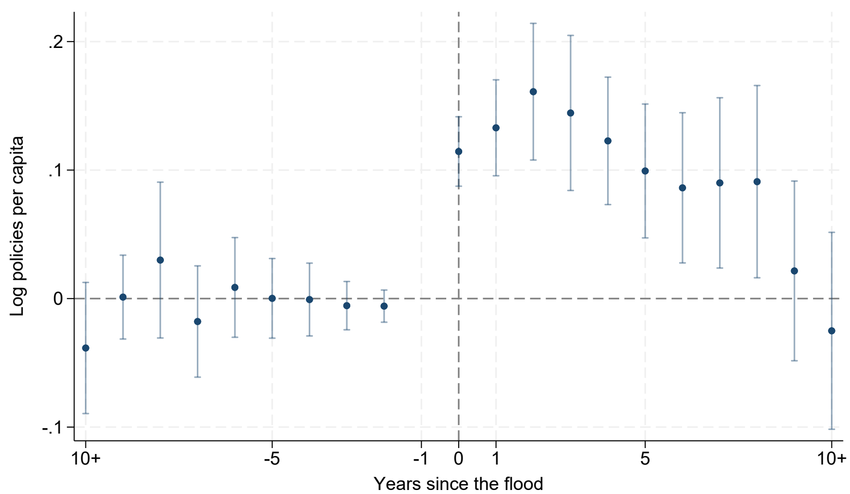

Figure 3: This figure presents the dynamic treatment effects of a flood on (log) number of flood insurance policies per capita using the imputation estimator.