Causal Inference using Difference-in-Differences

Lecture 10: Pre-tests

Pedro H. C. Sant'Anna
Emory University

January 2025

Pre-testing our assumptions

Testing our assumptions

- So far, we have taken our assumptions for granted.
- That is usually not very good practice.
- Now, we will start discussing how we can assess the plausibility of our assumptions.
- We will first talk about anticipation.
  - To relax it, we need to strengthen the parallel trends to hold for some pre-treatment periods, too!
- We then will talk about how to asses parallel trend assumptions.
  - Here, we will keep the no-anticipation assumption.
  - If both assumptions are violated, we may not be able to detect these!

Let's try something with empirical content this time

- Let's build our discussion on Fadlon and Nielsen (2021)
  - They study households' labor supply responses to fatal and severe nonfatal health shocks in the short run and medium run.
  - They leverage administrative data on Danish families and construct counterfactuals using households that experience the same event a few years apart (DiD)
  - Fatal events lead to considerable increases in surviving spouses' labor supply, which the evidence suggests is driven by families who experience significant income losses.
  - Nonfatal shocks have no meaningful effects on spousal labor supply, consistent with their adequate insurance coverage.

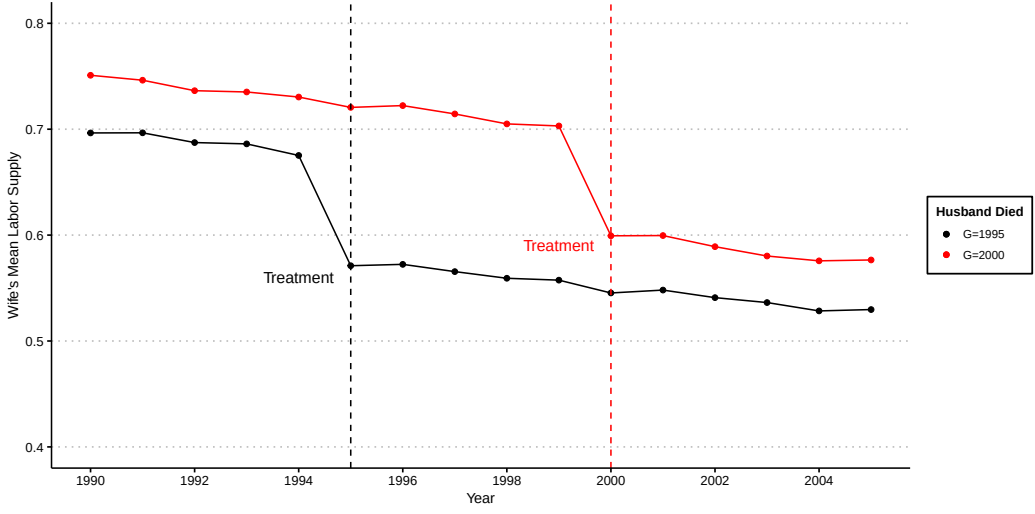
This was also inspired by the "Practical Guide to Event Studies" by David Novgorodsky and Bradley Setzler.

Let's try something with empirical content this time

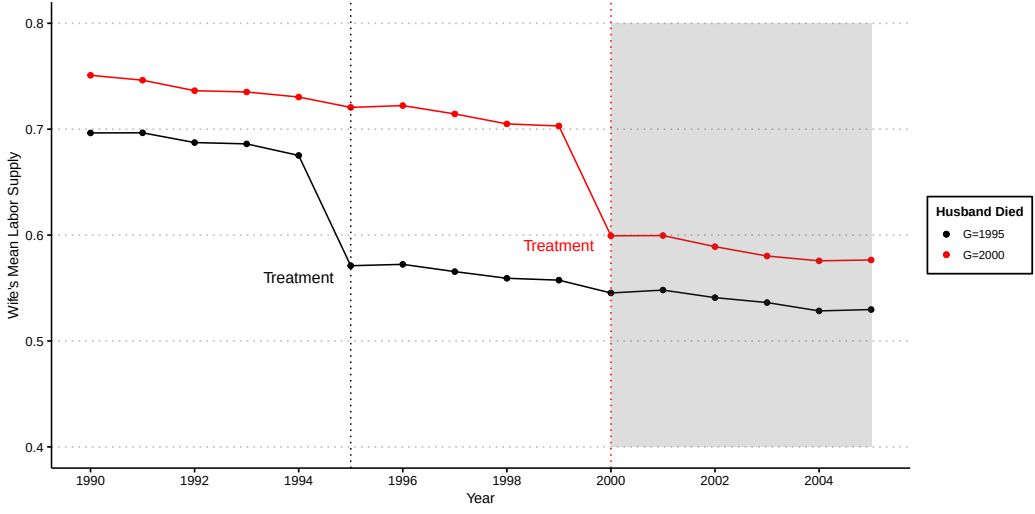
- Outcome of interest: wife's labor supply
- Treatment: time of spousal death
  - "Treated" group: Husband died in 1995.
  - "Comparison" group: Husband died in 2000.
- Data are available from 1990 to 2005.
- We will not replicate their result but rather simulate data and use their context to illustrate our main points.

Can we use all the data available?

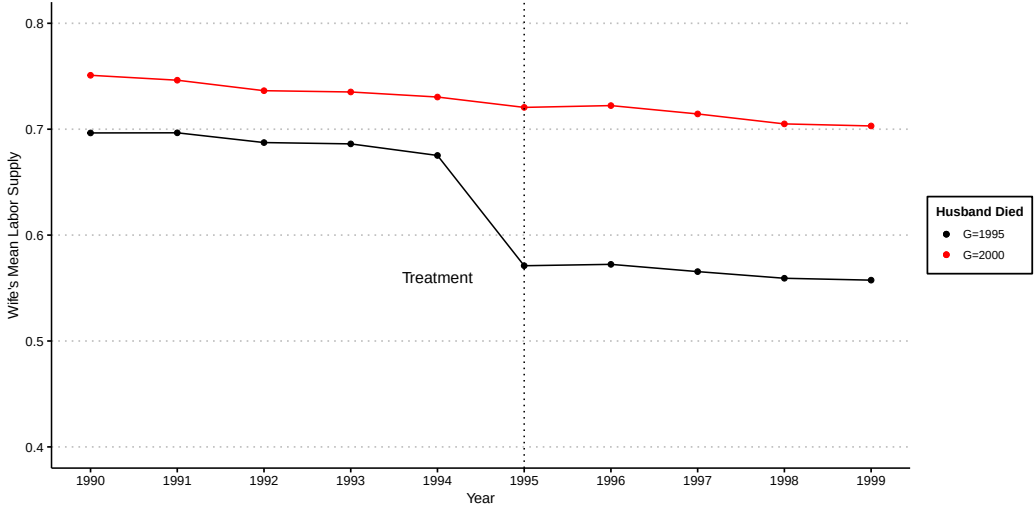
Let's take a look at the data



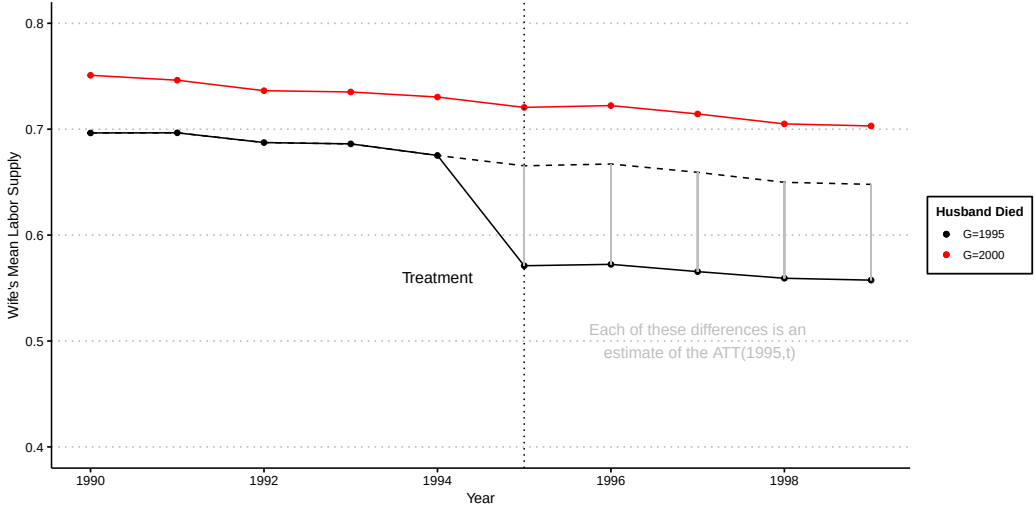
Let's take a look at the data: everybody is treated in the grey area



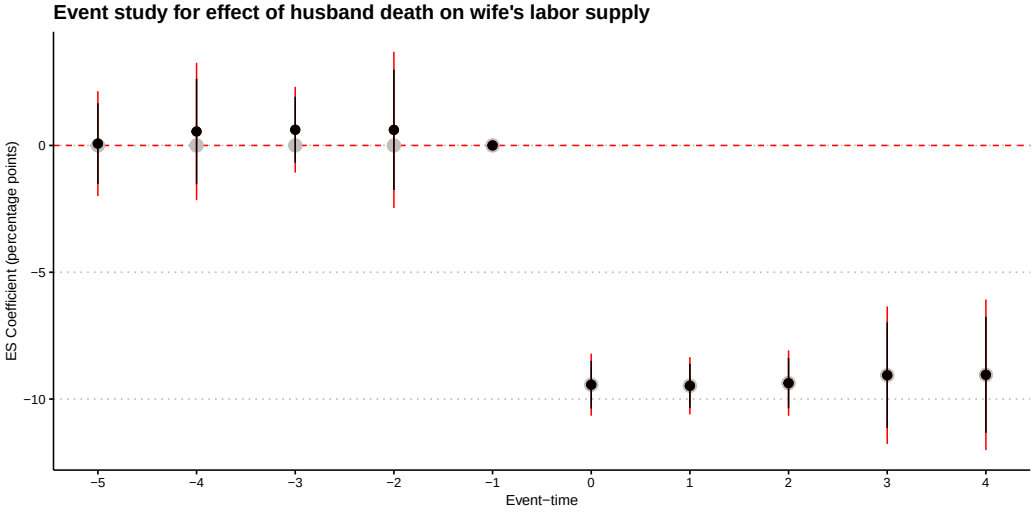
Let’s take a look at the data: subset where we can learn



Imposing PT + No-anticipation to get estimates of ATT(g,t)



Event-study plot: fixing the estimand's formula



More about ATT(g,t) estimates in pre-treatment periods

- The above event-study plot uses the following estimand **for all time periods**:

$$\begin{aligned}\textbf{ATT}(\mathbf{g},\mathbf{t}) = &\left(\mathbb{E}\left[Y_{i,t}|G_i = g\right] - \mathbb{E}\left[Y_{i,g-1}|G_i = g\right]\right) \\ &- \left(\mathbb{E}\left[Y_{i,t}|G_i = \infty\right] - \mathbb{E}\left[Y_{i,g-1}|G_i = \infty\right]\right)\end{aligned}$$

- We have shown that this is "required" for all post-treatment periods.

- However, in pre-treatment periods, we can use "short-differences" instead of "long-differences":

$$\begin{aligned}\textbf{ATT}(\mathbf{g},\mathbf{t}) = &\left(\mathbb{E}\left[Y_{i,t}|G_i = g\right] - \mathbb{E}\left[Y_{i,t-1}|G_i = g\right]\right) \\ &- \left(\mathbb{E}\left[Y_{i,t}|G_i = \infty\right] - \mathbb{E}\left[Y_{i,t-1}|G_i = \infty\right]\right) \quad \text{for } t < g.\end{aligned}$$

More about ATT(g,t) estimates in pre-treatment periods

- This is because, if we were to believe in PT in all pre-treatment periods, we would have, for all $t < g$,

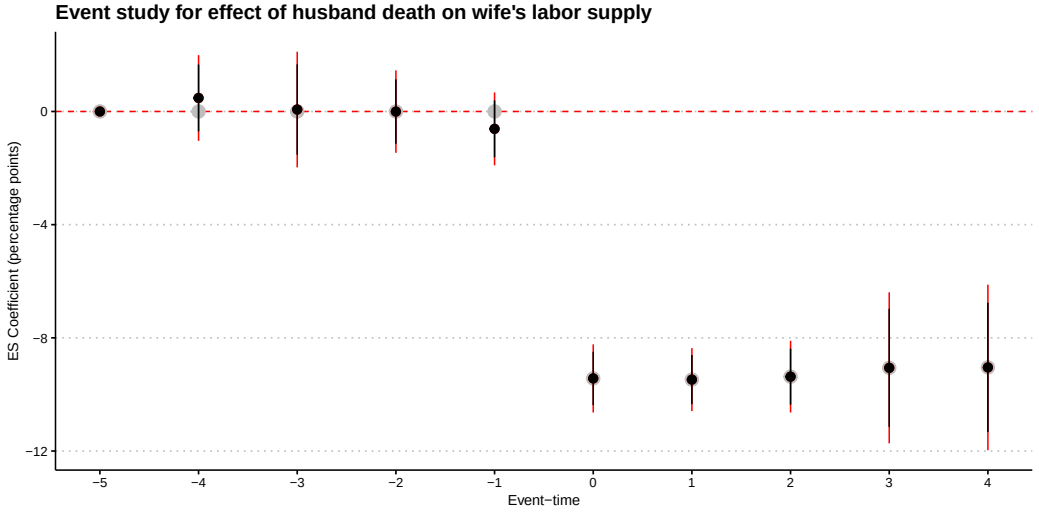
$$\mathbb{E}\left[Y_{i,t}(\infty)|G_i = g\right] - \mathbb{E}\left[Y_{i,t-1}(\infty)|G_i = g\right] = \mathbb{E}\left[Y_{i,t}(\infty)|G_i = \infty\right] - \mathbb{E}\left[Y_{i,t-1}(\infty)|G_i = \infty\right].$$

- Now, by No-Anticipation (and SUTVA), and simple manipulations, this is equal to

$$\left(\mathbb{E}\left[Y_{i,t}|G_i = g\right] - \mathbb{E}\left[Y_{i,t-1}|G_i = g\right]\right) - \left(\mathbb{E}\left[Y_{i,t}|G_i = \infty\right] - \mathbb{E}\left[Y_{i,t-1}|G_i = \infty\right]\right) = 0.$$

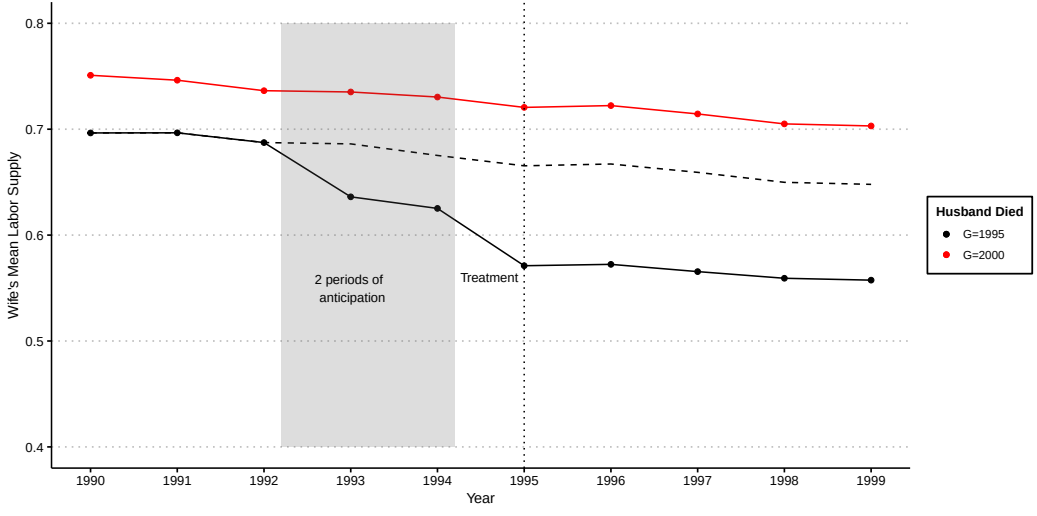
- No need to do "long-differences" with pre-treatment periods.

Event-study plot: using short-differences in pre-treatment

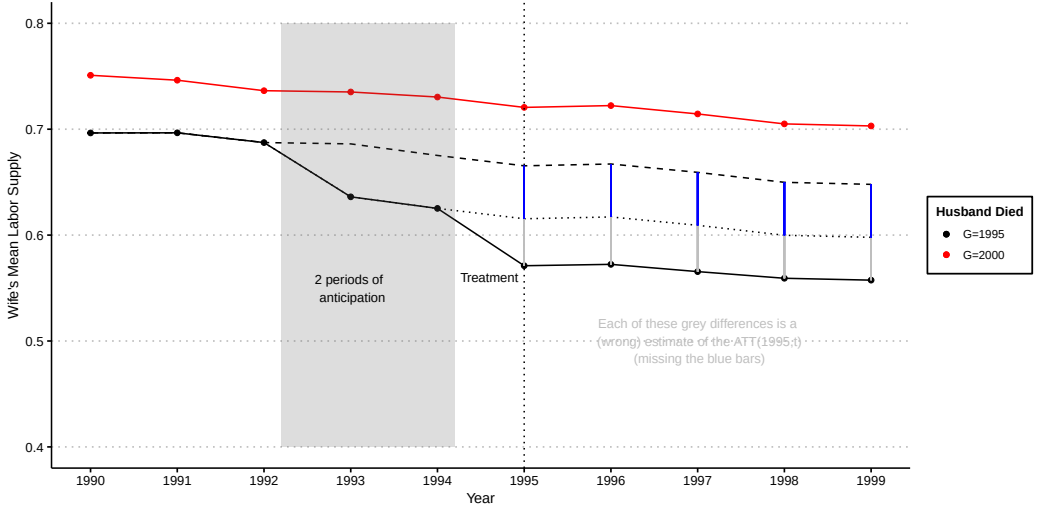


What if we have treatment anticipation for G=1995?

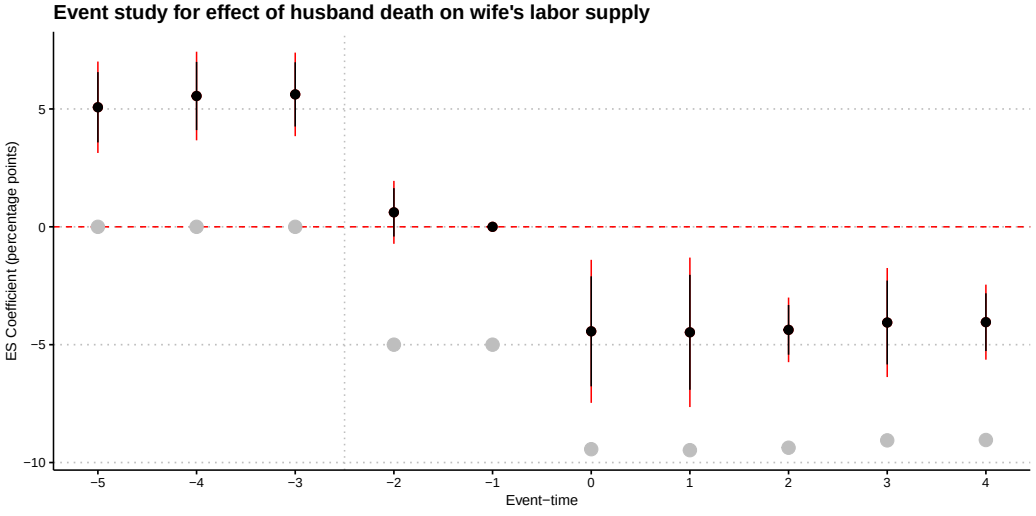
Two periods of treatment anticipation for G = 1995 (only)



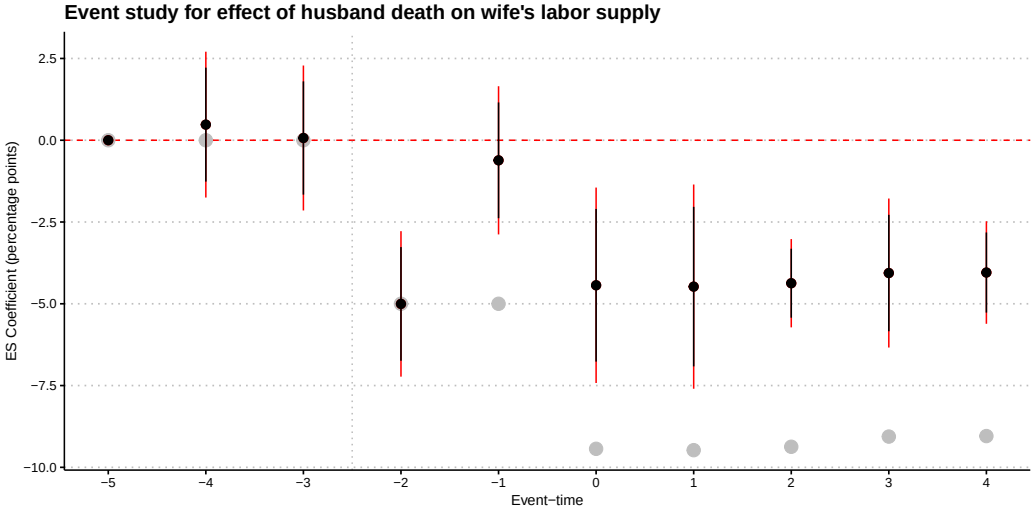
Two periods of treatment anticipation for G = 1995 (only)



Event-study with universal baseline



Event-study with short-differences



Pre-testing our assumptions

Relaxing the No-anticipation assumption

Relaxing the No-anticipation assumption

- When we know the number of periods units anticipate the treatment, we can actually relax the no-anticipation assumption.

Assumption (Limited-Anticipation)

For all units i, $Y_{i,t}(g) = Y_{i,t}(\infty)$ for all groups in pre-treatment periods $t < g - k$, where $k \geq 0$ is known.

- This assumption states that units anticipate their treatment by at most k periods, and k is known.

Strengthening the PT

- Interestingly, because we are relaxing the no-anticipation assumption, we need to strengthen our PT assumptions

Assumption (Stronger Parallel Trends Assumption)

For all $t \geq g - k$,

$$\mathbb{E}\left[Y_{i,t}(\infty)|G_i = g\right] - \mathbb{E}\left[Y_{i,t-1}(\infty)|G_i = g\right] = \mathbb{E}\left[Y_{i,t}(\infty)|G_i = \infty\right] - \mathbb{E}\left[Y_{i,t-1}(\infty)|G_i = \infty\right].$$

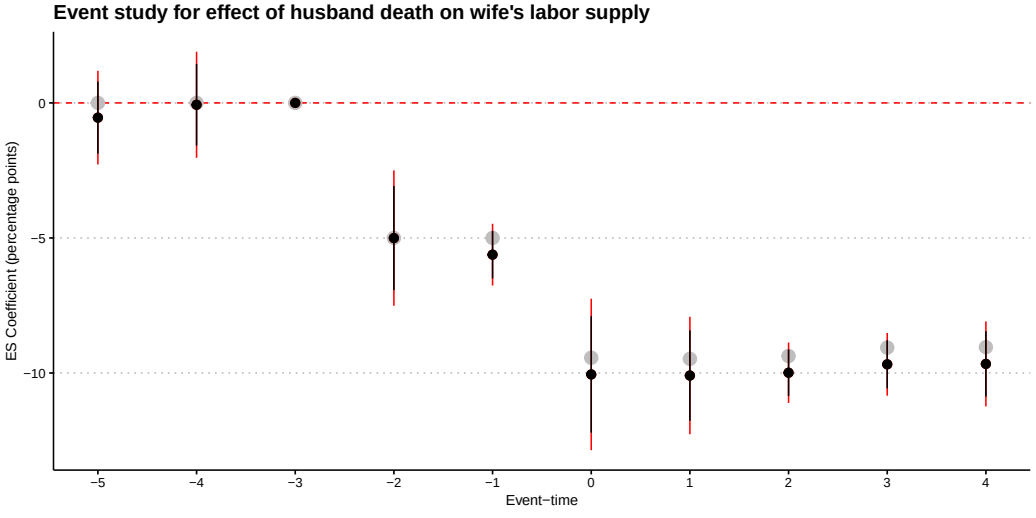
The parallel trends (PT) assumption states that, in the absence of treatment, the evolution of the outcomes among the treated units is, on average, the same as the evolution of the outcomes among the untreated units, **in some pre-treatment and all post-treatment periods**.

Identification result for ATT(g,t)'s with anticipation

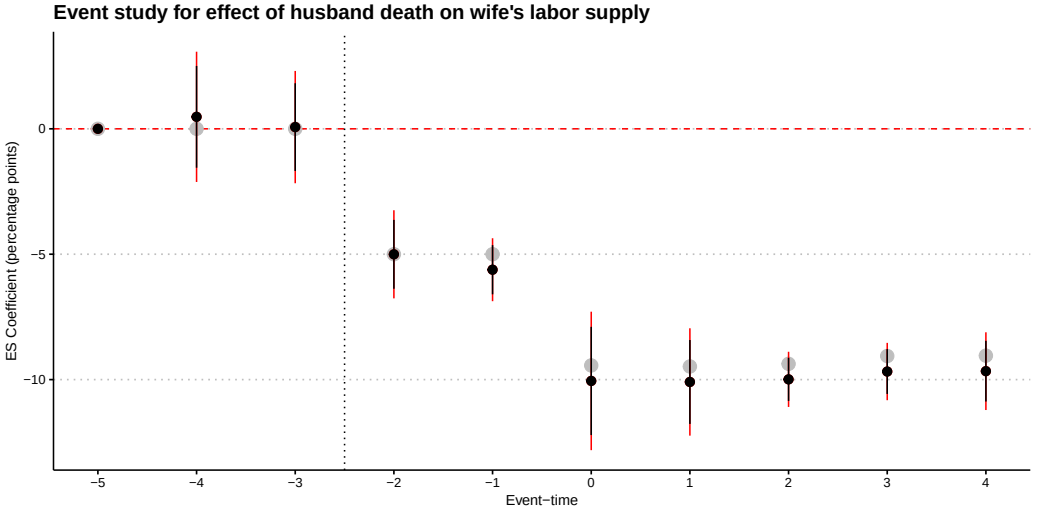
- Following the same steps as we did in Lecture 6, we can show that, for all $t \geq g - k$,

$$\begin{aligned} \mathbf{ATT(g,t)} = &\left(\mathbb{E}\left[Y_{i,t} | G_i = g\right] - \mathbb{E}\left[Y_{i,g-1-k} | G_i = g\right]\right) \\ &- \left(\mathbb{E}\left[Y_{i,t} | G_i = \infty\right] - \mathbb{E}\left[Y_{i,g-1-k} | G_i = \infty\right]\right) \end{aligned}$$

Event-study with universal baseline



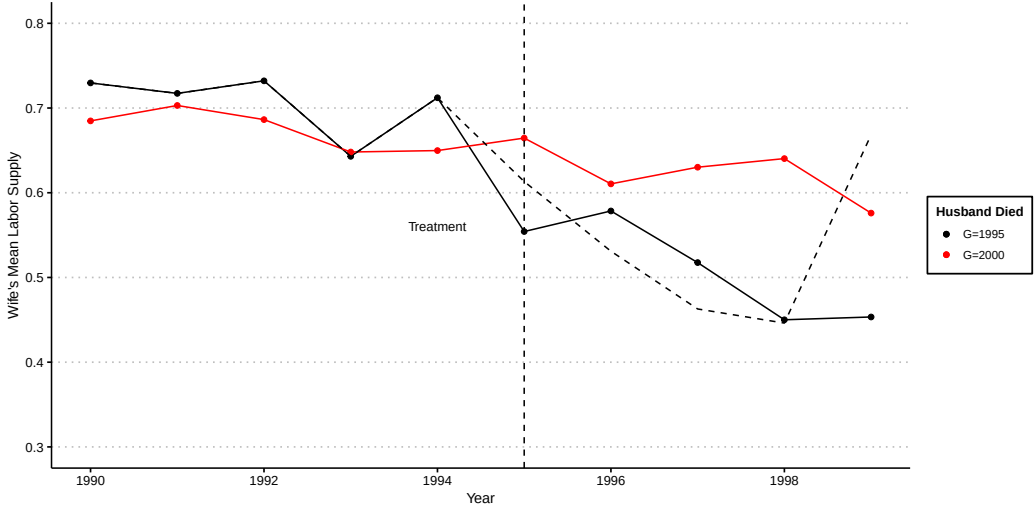
Event-study with short-differences



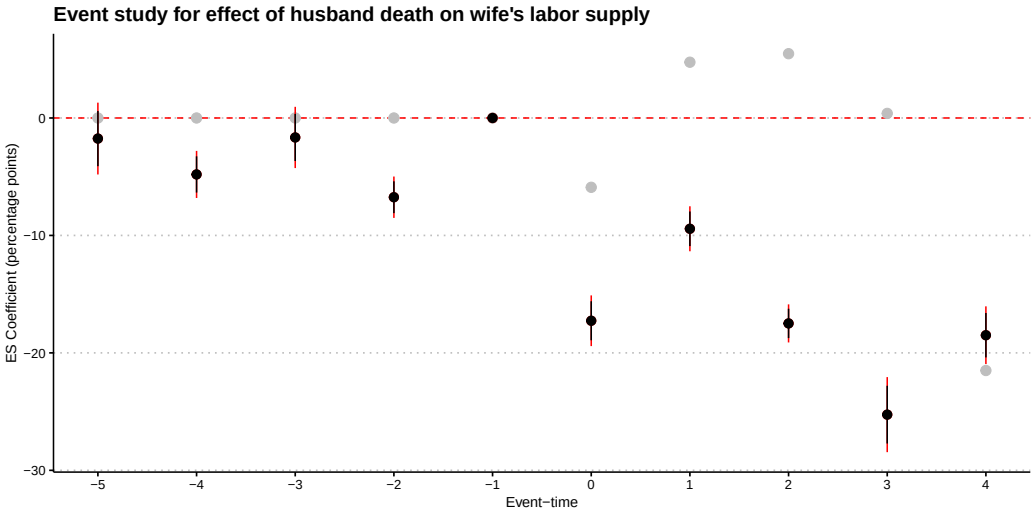
Pre-testing our assumptions

Violating the PT assumption (but not the no-anticipation assumption)

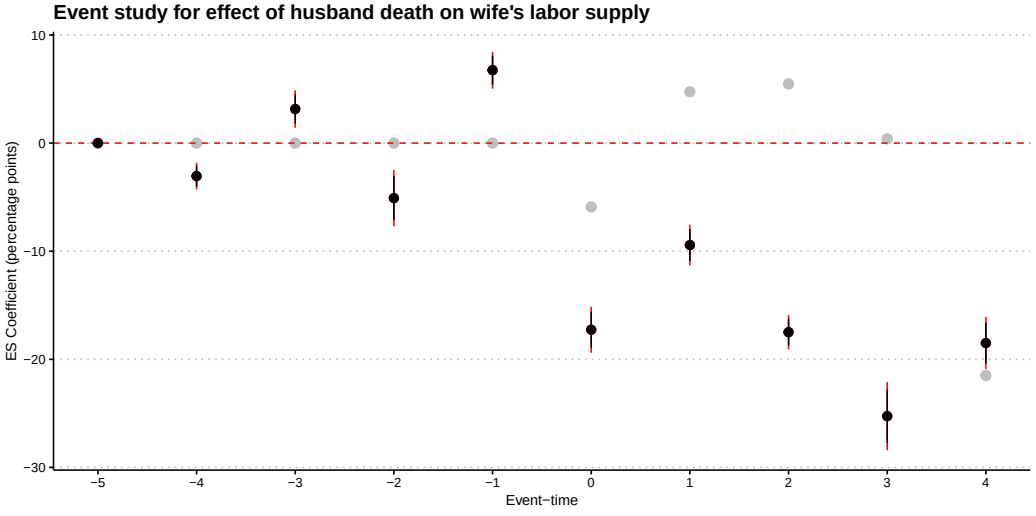
What if PT is violated?



What if PT is violated?



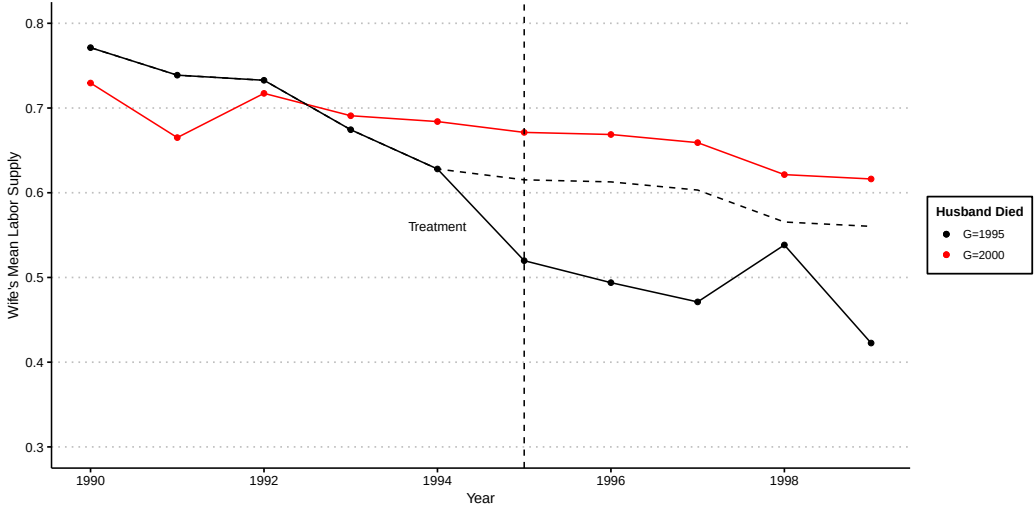
What if PT is violated?



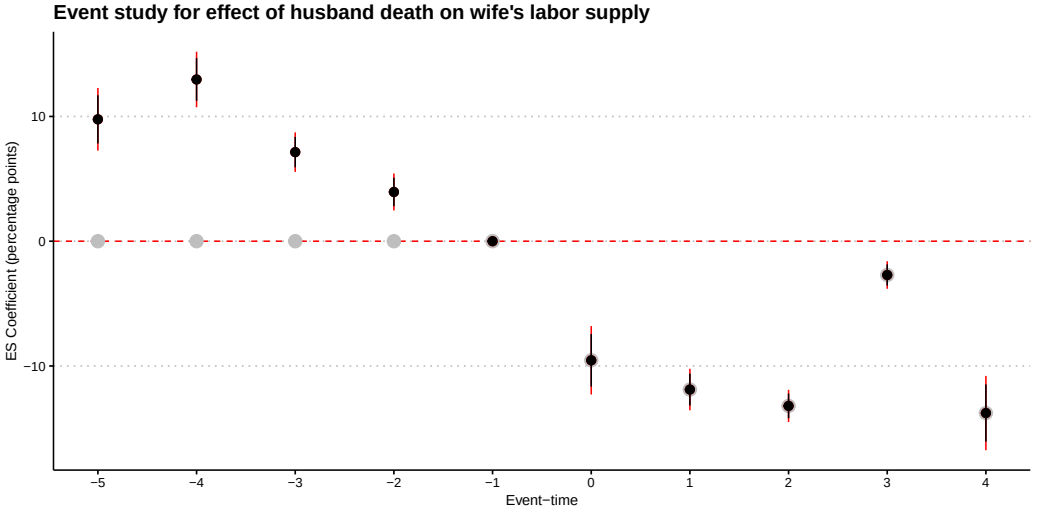
We can’t “estimate” ourselves out of this type of PT violations!

We also need to be careful with pre-trends

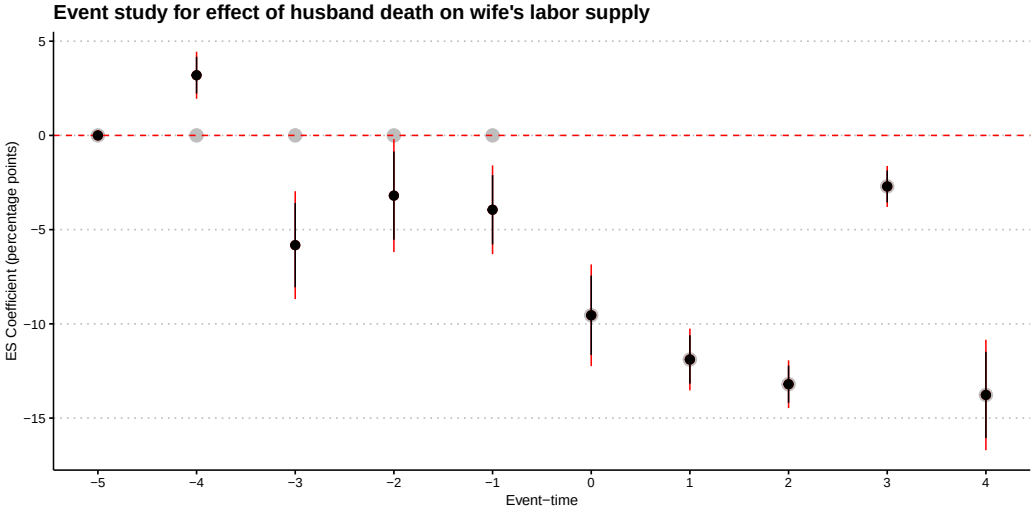
What if PT is violated only in pre-treatments?



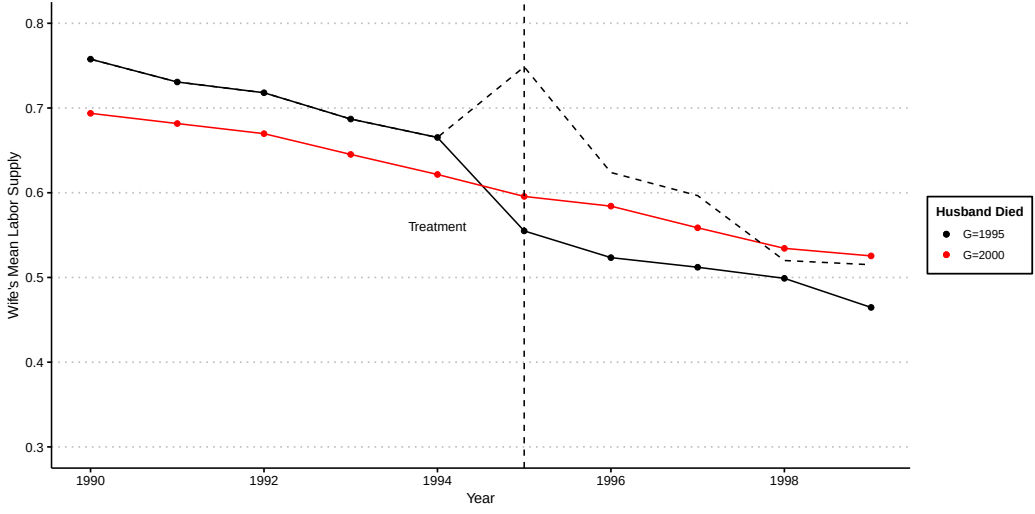
What if PT is violated only in pre-treatments?



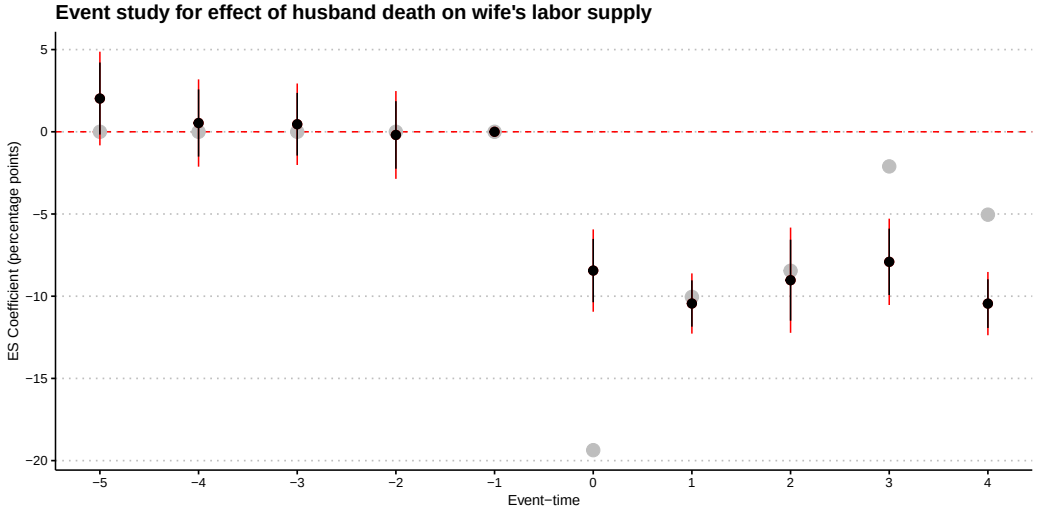
What if PT is violated only in pre-treatments?



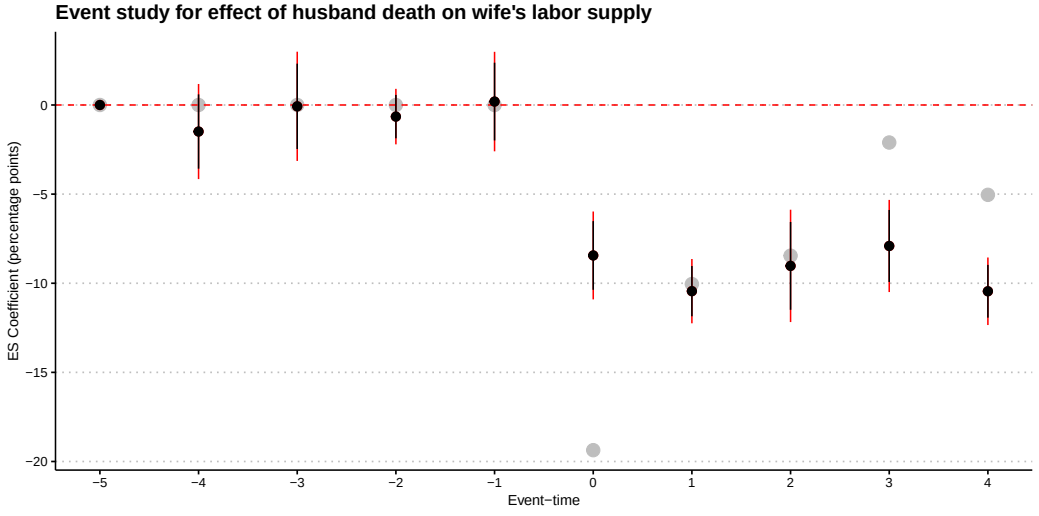
What if PT is violated only in post-treatments?



What if PT is violated only in post-treatments?



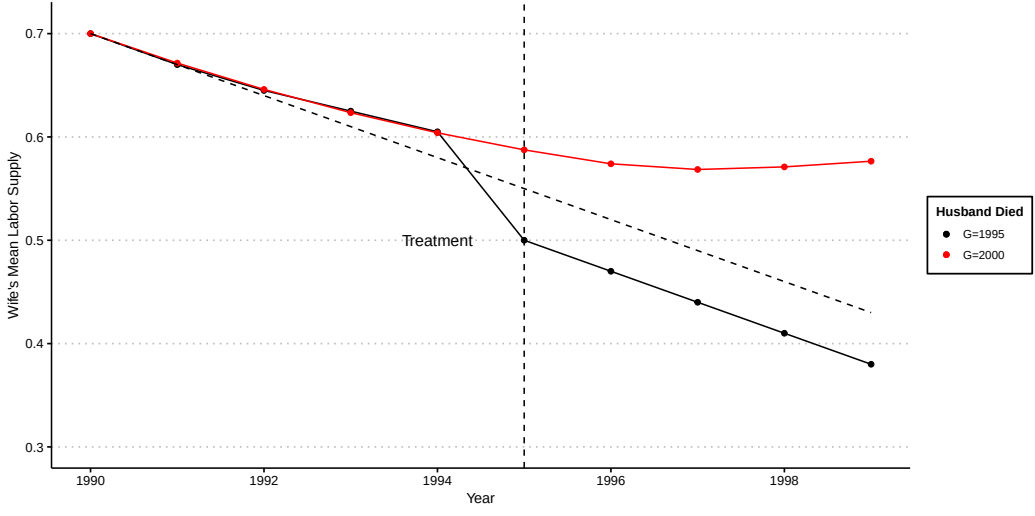
What if PT is violated only in post-treatments?



Pre-testing our assumptions

What if everything is violated?

What if both PT and No-anticipation assumptions are violated?



Main take-way messages

Main take-way messages

- With multiple time periods, we need to be careful with the implementation of our DiD estimator or own that we are restricting pre-trends.
- Also, there are many “easy” ways of assessing the plausibility of our assumptions.
- Overall, we can test for No-Anticipation (limited anticipation) by maintaining PT.
- We can also pre-test for PT by maintaining the No-Anticipation (or limited anticipation).
- If both are violated, things can look good but be very bad!
- We also need to be aware that non-parallel pre-treatment trends may not necessarily be bad (but most empirical researchers will fight me on this).
- There are more dangers of pre-test, like the issues described by Roth (2022).

What else should be in this lecture?

- This lecture only touches on the very basic points about pre-tests.
- So far, we have not been able to incorporate the pioneering work of Jon Roth on the topic (Roth, 2022; Rambachan and Roth, 2023).
- In future versions of this lecture, we will talk about the pitfalls of using standard pre-tests to assess the plausibility of PT based on Roth (2022).
- We will also talk about the important topic of how to do sensitivity analysis wrt violations of PT as discussed in Rambachan and Roth (2023).
- Meanwhile, check Jon's lecture at `https://raw.githack.com/Mixtape-Sessions/Advanced-DID/main/Slides/03-violations.pdf`.

References

Fadlon, Itzik and Torben Heien Nielsen, "Family Labor Supply Responses to Severe Health Shocks: Evidence from Danish Administrative Records†," *American Economic Journal: Applied Economics*, 2021, *13* (3), 1–30.

Rambachan, Ashesh and Jonathan Roth, "A More Credible Approach to Parallel Trends," *Review of Economic Studies*, 2023, *90* (5), 2555–2591.

Roth, Jonathan, "Pre-test with Caution: Event-study Estimates After Testing for Parallel Trends," *American Economic Review: Insights*, 2022, *4* (3), 305–322.