Causal Inference II

Mixtape Session

Roadmap

Diff-in-diff Checklist

Craigslist Example

Basic suggestions going forward

Checklist

- We will be using this to go through steps of a checklist that has been presented in Roth, et. al. (2023) ”What’s Trending . . . ”
- I’ll be explaining the checklist in the context of my project with Christine Durrance and Melanie Guldi looking at the effect of online dating on families
- Empirical context will be Craigslist’s Personals

History of Online Dating

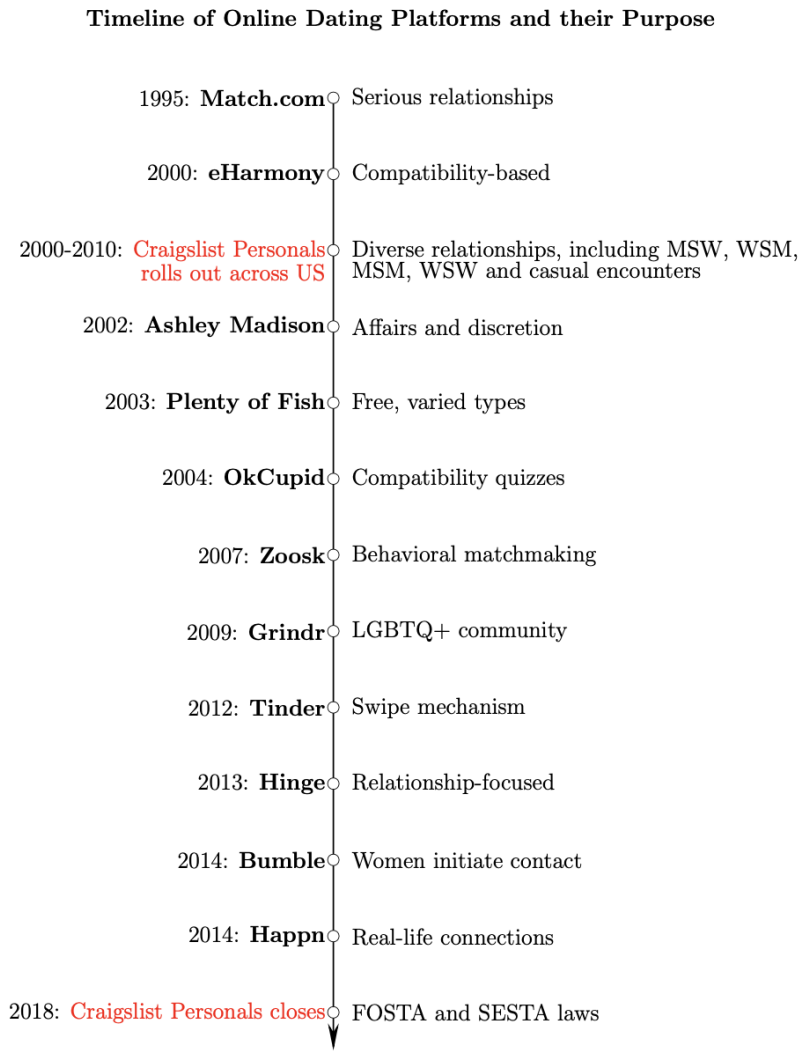


Figure 1: History of Major Online Dating Platforms

Data Selection

- We focused originally on birth rates for 15-44yo females and our unit of observation is the county
- We limited our sample period from 1995 (when Craigslist first opens) to 2007
- Why 2007? Three reasons:
  1. Birth rates plummeted following the Great Recession (demographic transition)
  2. Online dating apps like Tinder appear simultaneously so no geographic variation to exploit
  3. Wanted to avoid social media environment
- We keep in our dataset counties treated in 2008 to 2010, though, as a sample of counties that are going to be ”not-yet-treated”
- Used Wayback Machine to find exact month in which Craigslist Personals appeared in every county then assigned it to adjacent counties too, then treated a year as treated if it was ever treated

Conceptual Framework

- Our main focus has been birth rates because online dating has often been thought to reduce search frictions, thicken romance markets and lead to more efficient matches
  - → Therefore it could lead to longterm partnerships where families are an intentional choice
  - → Could therefore increase birth rates for the marginal couple
- But complaints have been rising in recent years that online dating leads to ”permanent dating” and that may reduce births if it substitutes people out of relationships intending for families
- We also focus on abortions, marriage, divorce, and STIs but for now I will only focus on birth rates

Diff-in-diff checklist step 1

1. Is Everyone Treated at the Same Time?
   - → **If Yes:**
     - For balanced panels, TWFE specifications yield interpretable estimates.
   - → **If No:**
     - Document the number of units treated in each cohort in a Table

Count the units by cohort

Table: Number of counties by year Craigslist personals appeared

Treatment Cohort	Number of Counties Treated
Never treated	1,779
2000 cohort	9
2001 cohort	5
2002 cohort	12
2003 cohort	36
2004 cohort	58
2005 cohort	69
2006 cohort	341
2007 cohort	65
2008 cohort	66
2009 cohort	215
2010 cohort	1
Total counties	2,656

Diff-in-diff checklist: Step 2

2. Plot Treatment Rollout
   - → If not all units are treated at the same time, plot the treatment rollout.
     - Use tools like `Panelview` to visualize.
   - → Use color and labels to help readers and viewers decipher

Plotting the Treatment Rollout

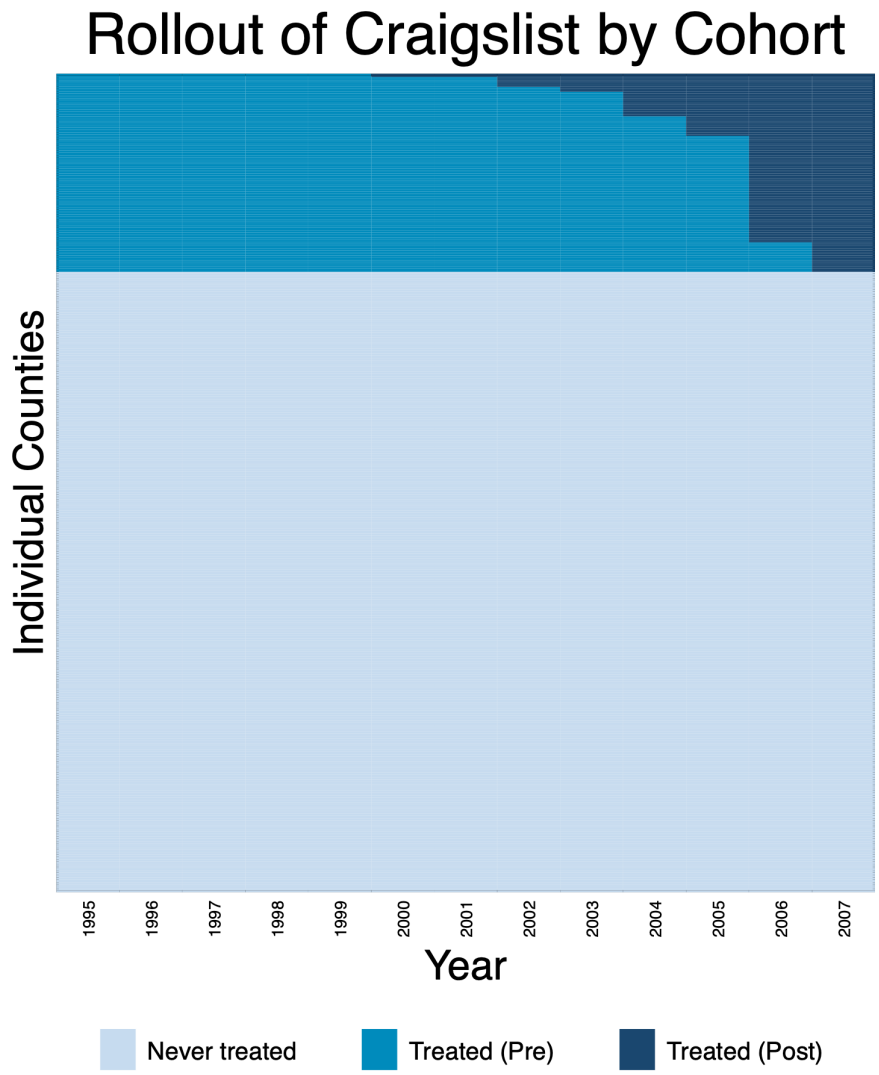


Figure: Rollout of Craigslist across US counties.

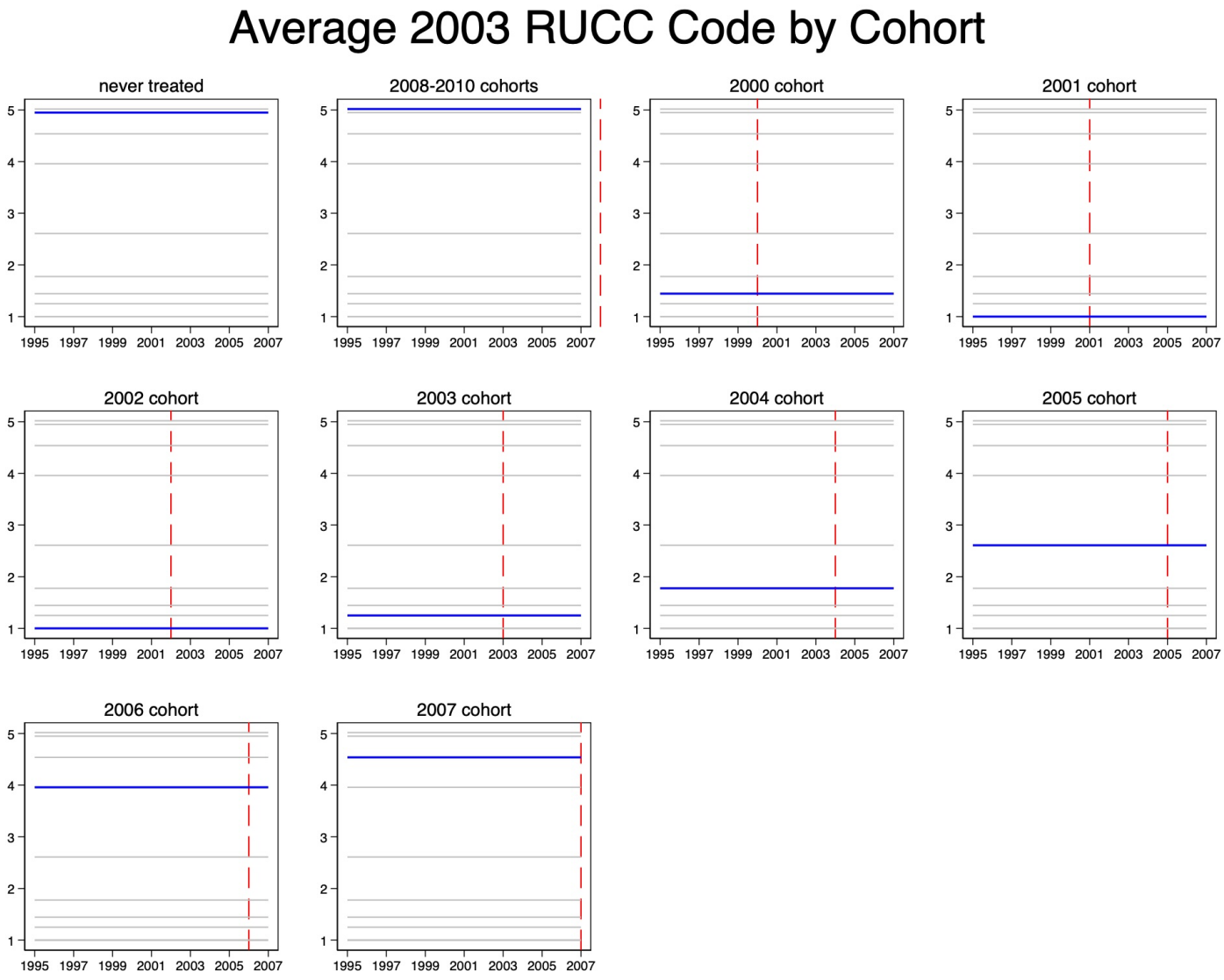
Diff-in-diff checklist: Step 3

3. Validity of Unconditional Parallel Trends Assumption
   - → **Covariate Selection:** Choose covariates predictive of missing potential outcome based on common sense, literature, and expertise.
   - → **Split-Sample Validation:** Use untreated outcomes with methods like Lasso to identify covariates if unsure.
   - → **Balance Check:** Check covariate balance using normalized differences.
   - → **Adjust Imbalanced Covariates:** Adjust if covariates are predictive of potential outcomes $[Y^0|D=1, Post]$.

Rural vs Urban counties

- We have a 9-digit ordinal score measuring how urbanized the county is called RUCC
- Counties that never got a Craigslist were more rural
- We therefore had originally wanted to use the 2008-2010 and other not-yet-treated counties as comparisons
- I'll plot rather than show the normalized difference for illustrative purposes

Plotting covariate imbalance



Comment

- Urban counties have different compositions by race, age, sex ratios, education shares, income all of which are *highly* predictive of untreated birth rate trends
- And even the 2008-2010 counties, despite our original hope, are very rural – just as rural even
- So we will run it both ways – we will run it without controls and then with a single covariate (RUCC)

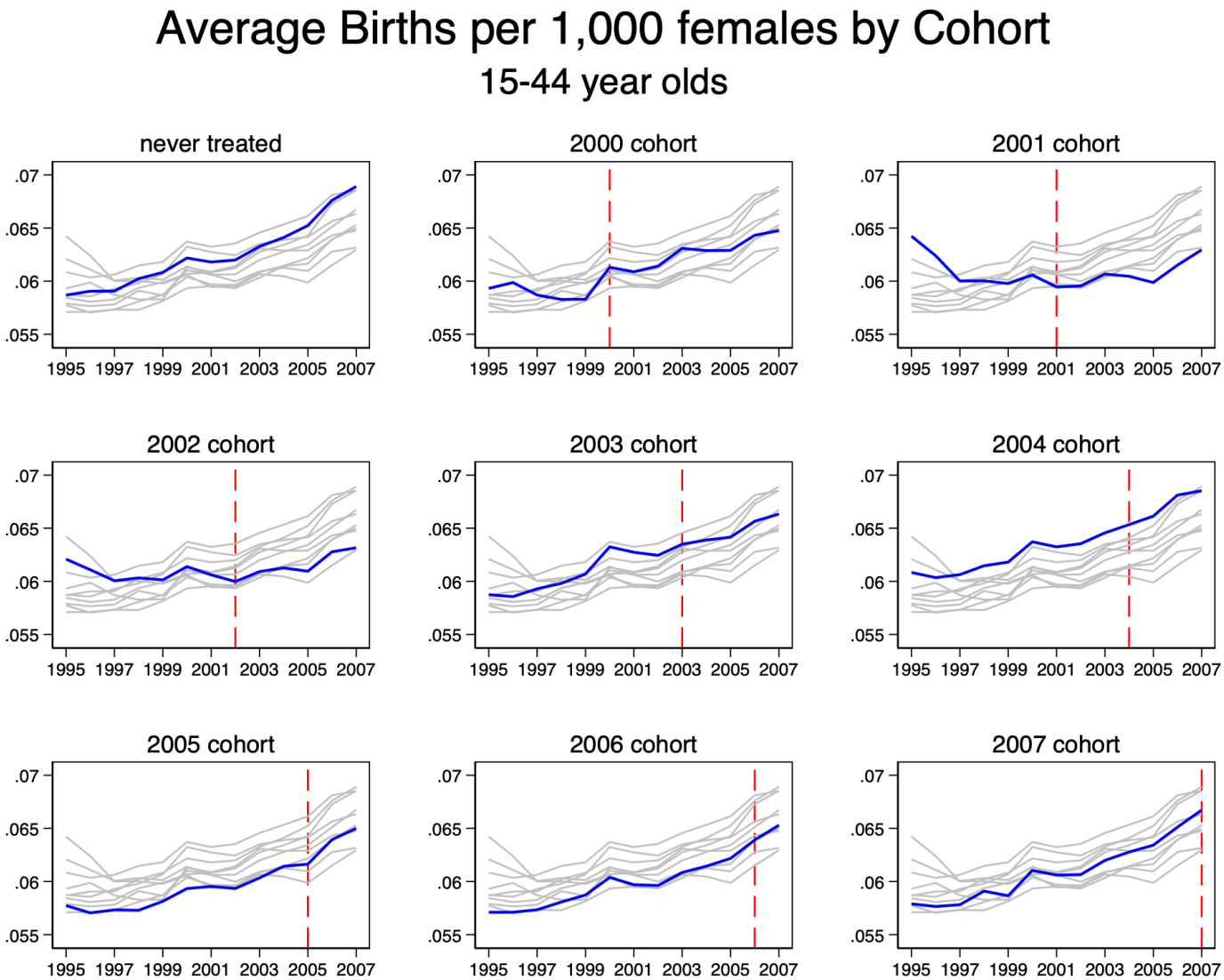
Diff-in-diff Checklist: Step 4

4. Check Overlap
   - → **Overlap Check:** Ensure sufficient overlap in propensity scores.
   - → **Adjustment for Lack of Overlap:** If lacking, consider a regression adjustment like BJS.
   - → We have overlap since the RUCC code has only 9 values so I'll skip this, but you'll want to check using a propensity score if you have continuous or several covariates

Diff-in-diff Checklist: Step 5

5. Plot Average Outcomes Across Cohorts
   → Visualize the evolution of average outcomes across cohorts.

Plotting Outcomes Across Cohorts



Diff-in-diff checklist: Step 5

5. Estimator Selection and Assumptions
   - → **Estimator Choice:** Choose an estimator based on comfort and assumptions.
   - → **Robust Method:** Use robust methods like CS, SA, dCDH, or BJS if agnostic to heterogeneous effects.
   - → **Event study:** Estimate simple averages as well as event studies
     - ■ Use long-differences with `long2` or universal base
   - → **Cluster standard errors:** Report with clustering at the aggregate level of treatment.

Short-gaps no controls

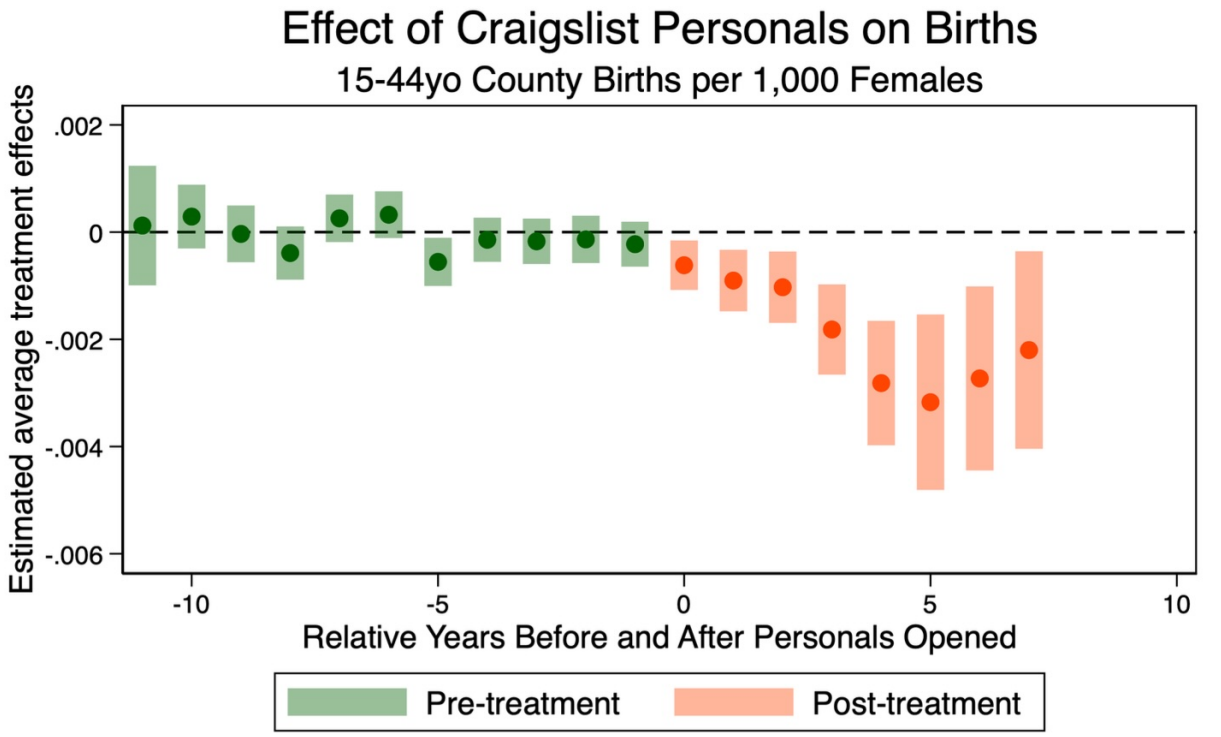
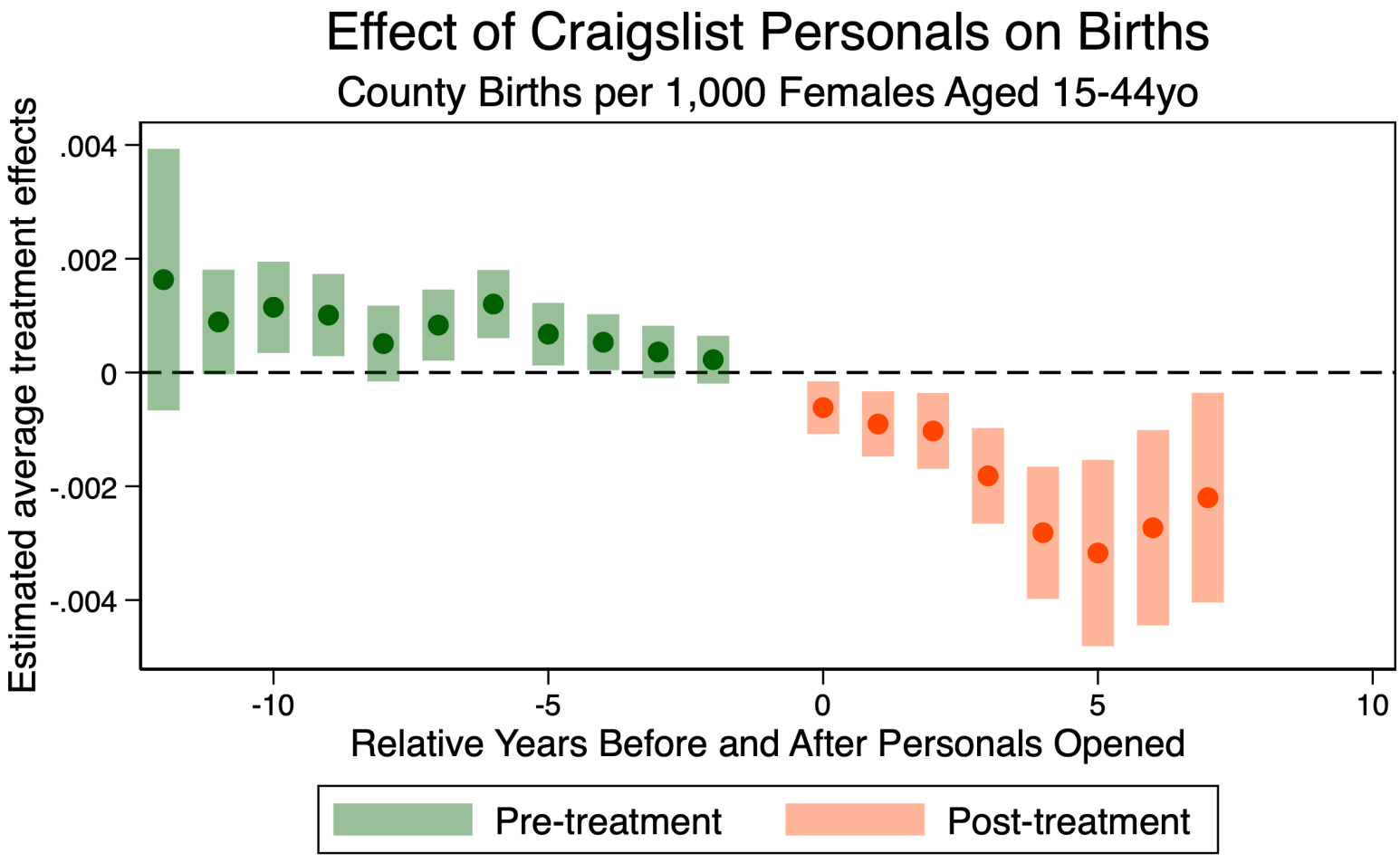


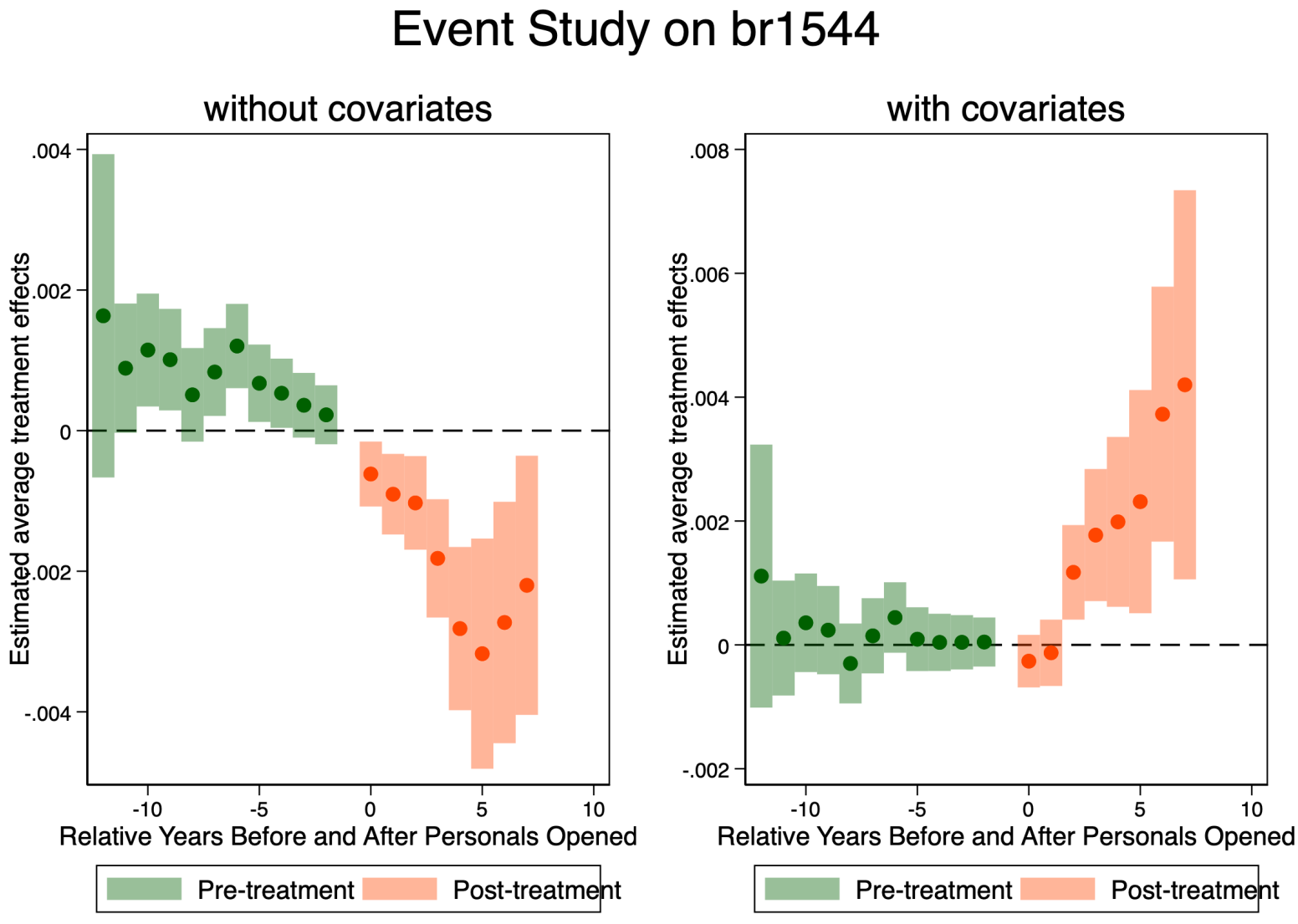
Figure 9: Event study plots estimated using Callaway and Sant'Anna. Comparison group is the not-yet-treated cohorts, including the 2008-2010 cohorts who were treated after the end of our panel period. Data is from 1995 to 2007 and based on county level data.

Long differences no controls

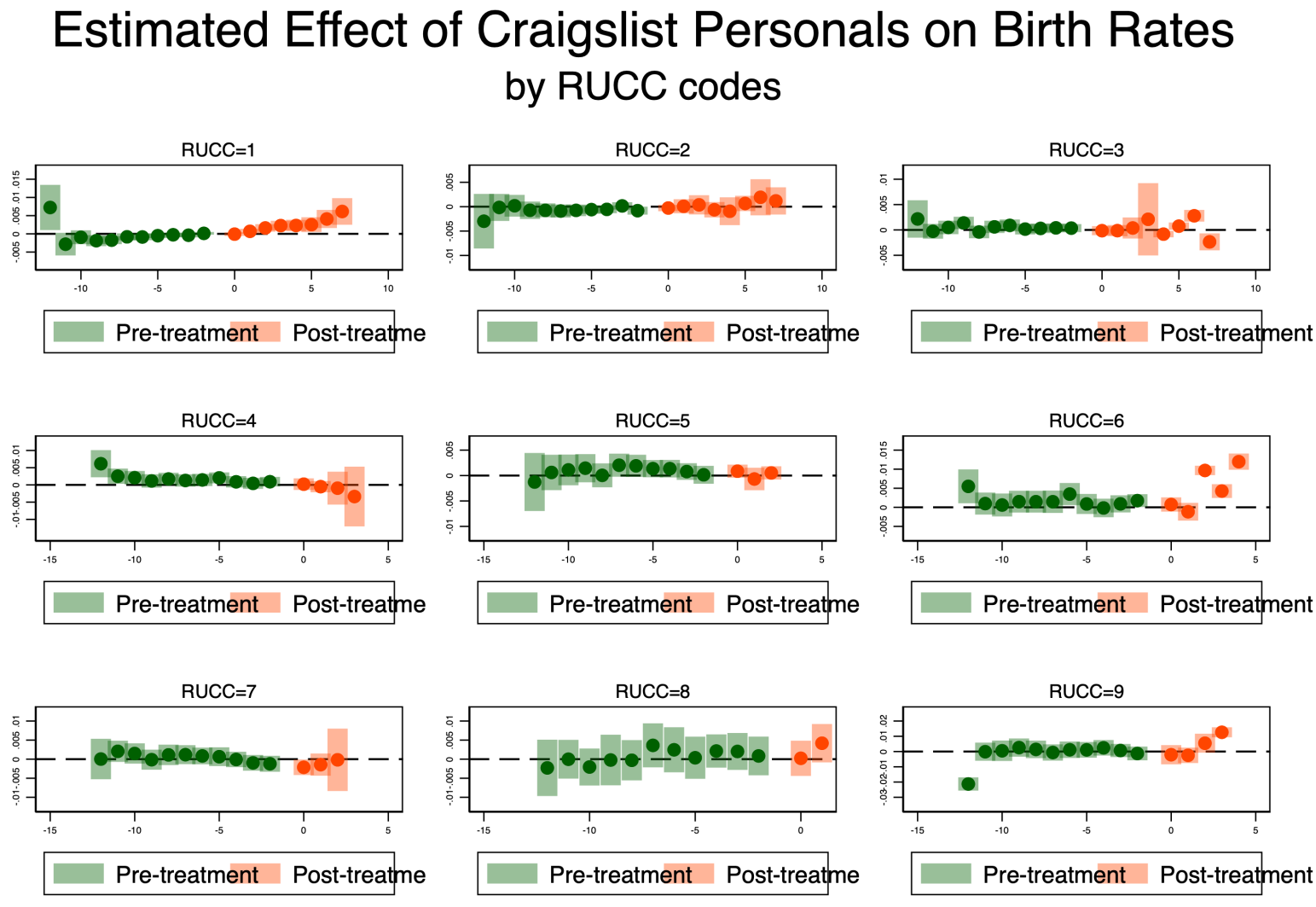


Note: Uses not-yet-treated including the 2008-2010 but not the never treated counties (i.e., model specification 2 from table estimates). Circles are average ATT(g,t) per relative time period across all groups and bands are 95% uniform confidence intervals. Mean birth rate variable was approximately 0.06 in 2000.

Long differences without and with RUCC control (!)



Running it within each RUCC code



Smaller RUCC codes are more urbanized counties. Models estimated separately with CS and not-yet-treat

Diff-in-diff checklist: Step 6

6. Sensitivity Analysis for Parallel Trends Violations
   → **Consider Sensitivity Analysis:** Use tools like `honestdid` for PT violations.

Diff-in-Diff Checklist: Step 7

7. Throw in the towel
   - → You may have to concede that for any number of reasons, parallel trends with or without covariates is implausible — it happens.
   - → Consider then some alternatives:
     - **Panel Method:** If imbalance or PT assumptions fail, consider matrix completion or nuclear norm regularization.
     - **Cross-Section:** Cross-sectional unconfoundedness may also be appropriate.
   - → Don't force it — diff-in-diff doesn't make estimates causal; parallel trends makes diff-in-diff causal.

Roadmap

Diff-in-diff Checklist

Craigslist Example

Basic suggestions going forward

Concluding remarks on DD

- You're probably going to write a paper using DiD at least once in your life, but probably more
- Even if you don't, you're going to read a lot of papers using DiD, referee them, or advise students using them
- It's in your best interest to make the fixed cost investment in the new econometrics of DiD because the old methods are mostly harmful
- Good news is we are at the conclusion of this wave of papers, software is now widely available, solutions tend to have common features, and overall presentations (static and dynamic) aren't all that different
- Lots of surveys have been written (Roth, et al. 2023; our new JEL) and you'll have the slides and recordings here

Concluding remarks

- Stronger assumptions needed to include covariates, and bias can be large
- But you need the correct covariates, and I suggested using sample splitting for selection and LASSO to identify relevant covariates that affect **untreated potential outcomes**
- Don’t control for covariates that could be affected by the outcome

Concluding remarks

- Main problem in differential timing is heterogeneity and the use of already-treated units as controls
- Differential timing and lacking a priori theory, you *should* avoid the canonical TWFE specification
- With heterogeneity, canonical TWFE is biased, signs can flip, weights can be negative, etc etc.
- Robust DiD methods do not place restrictions on treatment effect heterogeneity, but cannot solve everything
- But when you lose parallel trends, you're probably going in the direction of need a new design

Concluding remarks

- Devote time to designing your diff-in-diff – you'll thank yourself as it keeps you honest and helps identify problems before having peeked at the outcomes
- Much harder psychologically to pull back when you've become invested in a story, so do not go to the outcomes until you've pre-registered your study to some degree (even if not formally)

Conclusion

- Thank you for coming to the workshop
- I hope that this was valuable for you
- We will be doing this until demand drops off, so keep your eyes peeled