Econometric Policy Analysis under Uncertainty

Sargan Lectures of the Econometric Society - Sydney 2023

Raffaella Giacomini (UCL and Chicago Fed)[1]

[1] The lectures do not reflect the views of the Federal Reserve Bank of Chicago or the Federal Reserve System

Recall: Uncertainties in causal effects estimation

- We will consider different kinds of scientific uncertainties
  - Impulse-response analysis
    - Uncertainty about identifying assumptions (Lecture 1)
    - Uncertainty about the measurement of shocks (Lecture 2)
  - Average treatment effects (Lecture 3)
    - Uncertainty about the validity of the control group
    - Uncertainty about model specification

Lecture 3. Uncertainty about validity of control group

Based on Botosaru, Giacomini and Weidner (2023), "Forecasted Treatment Effects"

Set up

- We move from impulse-response analysis with time series data to treatment effect estimation with panel data
- Can we estimate causal effects of policies without **a valid control group?**. E.g.,
  - a policy implemented at the population level
  - potential control group but we are uncertain about validity of assumptions behind existing estimators (e.g., parallel paths for Diffs-in-Diffs)
- Can we do so without making strong assumptions about **model specification?**
- We look once again for a robust approach, in this case that does not require 1) a valid control group and 2) correct model specification

Idea

- Let's assume we have a panel of individual pre-treatment outcomes (large n, fixed, small T)
  - balanced or unbalanced panels
  - treatment time deterministic (staggered adoption ok)
- Idea: use individual pre-treatment outcomes to forecast individual-specific counterfactuals
- The parameter of interest is the ATT = cross-sectional average of differences between post-treatment outcomes and counterfactuals

Remarks:

- The forecast is not the goal, but an ingredient in the ATT
- The interest is not forecast accuracy, but consistent+ asy normal estimator for the ATT under weak assumptions on the DGP

Contribution

1. Point out that **average unbiasedness of forecasts** is sufficient for consistency and asymptotic normality of ATT estimator
2. Propose simple way to forecast individual counterfactuals using **polynomial regressions**
3. Show that this gives unbiased forecast of individual treatment effects (and thus average unbiasedness) under minimal assumptions on the DGP

Parameter of interest: ATT

- Observe outcomes y_{it} for $i = 1, ..., n$ and $t = 1, ..., \tau + h, \ h > 0$
- Program implemented at τ (deterministic). All i's treated at $t > \tau$
- Observed outcome: $y_{it} = 1(t \leq \tau) y_{it}(0) + 1(t > \tau) y_{it}(1)$, where $y_{it}(0)$ potential outcome if i untreated and $y_{it}(1)$ potential outcome if i treated
- The parameter of interest is

$$ATT_{\tau+h} = \sum_i \mathbb{E}\left[y_{i\tau+h}(1) - y_{i\tau+h}(0)\right]/n = \sum_i \mathbb{E}\left[y_{i\tau+h} - y_{i\tau+h}(0)\right]/n,$$

where $y_{i\tau+h}(0)$ is the counterfactual

- Note that the individual treatment effects $\mathbb{E}\left[y_{i\tau+h} - y_{i\tau+h}(0)\right]$ could be heterogeneous

Our proposal: FAT

- Parameter of interest

$$ATT_{\tau+h} = \sum_i \mathbb{E}\left[y_{i\tau+h} - y_{i\tau+h}(0)\right]/n$$

- Conventional approach: "Learn" $\sum_i \mathbb{E}\left[y_{i\tau+h}(0)\right]/n$ from a control group
- This paper:
  - **Forecast** $y_{i\tau+h}(0)$ from pre-treatment data $\{y_{it}\}_{t=1}^{\tau} \rightarrow \widehat{y}_{i\tau+h}(0)$
  - Estimate $ATT_{\tau+h}$ by the **Forecasted Average treatment effect**:

$$FAT_{\tau+h} = \frac{1}{n}\sum_{i=1}^{n}\left[y_{i\tau+h} - \widehat{y}_{i\tau+h}(0)\right]$$

  - Goal: find $\widehat{y}_{i\tau+h}(0)$ that makes FAT consistent for ATT under minimal assumptions on the DGP

High-level assumptions

Assumption (Average unbiasedness)

The forecast for time $\tau + h,\ h \geq 1$, *is unbiased on average:*

$$\frac{1}{n}\sum_{i}\mathbb{E}\left(\widehat{y}_{i\tau+h}\left(0\right) - y_{i\tau+h}\left(0\right)\right) = 0 \tag{1}$$

Assumption (CLT)

The forecast errors $\{u_{i\tau+h}\} := y_{i\tau+h} - \widehat{y}_{i\tau+h}(0)$ *satisfy CLT:*

$$\frac{\frac{1}{\sqrt{n}}\sum_{i} u_{i\tau+h}}{\bar{\sigma}_n} \Rightarrow \mathcal{N}\left(0,1\right), \tag{2}$$

where $\bar{\sigma}_n^2 := \mathrm{Var}(\frac{1}{\sqrt{n}}\sum_i u_{i\tau+h}) < \infty$

Consistency and asymptotic normality of FAT

Theorem (Consistency and asymptotic normality)

Let Assumptions 1 and 2 hold. Then $\widehat{\mathrm{FAT}}_h$ *satisfies:*

$$\frac{\sqrt{n}\left(\widehat{\mathrm{FAT}}_h - \mathrm{ATT}_h\right)}{\bar{\sigma}_n} \Rightarrow \mathcal{N}\left(0, 1\right).$$

- We just need to find forecasts of counterfactuals that are unbiased on average
- Then, as long as there are no common unpredictable shocks at time $\tau + h$ we have consistency and asymptotic normality of $\widehat{\mathrm{FAT}}_h$

Solution: polynomial regressions (PR)

- $q_i \in \{0, 1, 2, ..., \tau - 1\}$ max order of polynomial time trend
- $R_i \in \{q_i + 1, ..., \tau\}$ number of pre-treatment time periods used for estimation
- $\mathcal{T}_i = \{\tau - R_i + 1, ..., \tau\}$ estimation window

1. Regress $y_i \equiv (y_{i\tau - R_i + 1}, ..., y_{i\tau})$ on a polynomial time trend:

$$\widehat{\alpha}_i^{(q_i, R_i)} = \arg \min_{\alpha_i \in \mathbb{R}^{q_i+1}} \sum_{t \in \mathcal{T}_i} \left(y_{it} - \sum_{k=0}^{q_i} \alpha_{ik} t^k \right)^2$$

2. The PR forecast of counterfactuals at $\tau + h$ is

$$\widehat{y}_{i\tau+h}^{PR}(0) = \sum_{k=0}^{q_i} \widehat{\alpha}_{ik}^{(q_i, R_i)} (\tau + h)^k$$

Simple to compute

Example

If $h = 1$ and we choose $R_i = q_i + 1$, there is not need to run a regression as PR simplifies to

- $q_i = 0$ (use last period)

$$\widehat{y}^{PR}_{i\tau+1}(0) = y_{i\tau},$$

- $q_i = 1$ (use last two periods)

$$\widehat{y}^{PR}_{i\tau+1}(0) = 2y_{i\tau} - y_{i\tau-1},$$

- $q_i = 2$ (use last three periods)

$$\widehat{y}^{PR}_{i\tau+1}(0) = 3y_{i\tau} - 3y_{i\tau-1} + y_{i\tau-2},$$

- ... (use last $q_i + 1$ periods)

$$\widehat{y}^{PR}_{i\tau+1}(0) = \sum_{t=\tau-q_i}^{\tau} w_{it} y_{it}, \; w_{it} = (-1)^{\tau-t} \begin{pmatrix} q_i + 1 \\ \tau - t + 1 \end{pmatrix}$$

Result: unbiasedness for large class of DGPs

Theorem

Suppose $\{y_{it}(0)\}_{t=1}^{\tau+h}$ *can be written as the* **sum of (up to) two unobserved components***:*

$$y_{it}(0) = y_{it}^{(s)}(0) + y_{it}^{(u)}(0), \ \ t = 1, ..., \tau + h,$$

$$\begin{aligned} y_{it}^{(s)}(0) &= \textit{mean stationary process,} \\ y_{it}^{(u)}(0) &= \textit{unit root process,} \end{aligned}$$

Then PR gives unbiased estimators of the **individual treatment effects:**

$$\mathbb{E}\left(y_{i\tau+h}(0) - \widehat{y}_{i\tau+h}^{PR}(0)\right) = 0$$

Result: unbiasedness for large class of DGPs

Theorem

Suppose $\{y_{it}(0)\}_{t=1}^{\tau+h}$ *can be written as the sum of (up to) three unobserved components where (for sure) one is a* **deterministic polynomial time trend**

$$y_{it}(0) = y_{it}^{(s)}(0) + y_{it}^{(u)}(0) + y_{it}^{(p)}(0), \;\; t = 1, ..., \tau + h,$$

$$\begin{aligned} y_{it}^{(s)}(0) &= \textit{mean stationary process,} \\ y_{it}^{(u)}(0) &= \textit{unit root process,} \\ y_{it}^{(p)}(0) &= \sum_{k=0}^{q_{i0}} \alpha_{ik}^{(p)} t^k, \;\; q_{i0} \in \{0, 1, 2, ...\}. \end{aligned}$$

Then PR gives unbiased estimators of the **individual treatment effects:**

$$\mathbb{E}\left(y_{i\tau+h}(0) - \widehat{y}_{i\tau+h}^{PR}(0)\right) = 0 \textbf{ if } q_{i0} \geq q_0$$

In words

- It is not necessary to have a correctly specified model for counterfactuals to obtain unbiased estimators of (heterogeneous) treatment effects $\rightarrow$ robustness
- If DGPs are stationary or have a stochastic trend, any PR satisfies unbiasedness
  - This is actually true of any forecast that can be written as a weighted average of pre-treatment data with weights summing to 1
- If we are sure that individual DGPs have a deterministic trend, it has to be a polynomial time trend + PR needs a large enough order to satisfy unbiasedness

Excluded DGPs

- Our results cover DGPs such as

$$y_{it}(0) = \mu_i + \rho_i y_{it-1}(0) + \varepsilon_{it},\ \mathbb{E}(\varepsilon_{it}) = 0,\ var(\varepsilon_{it}) < \infty,$$

  - with either $|\rho_i| < 1$ and $\mathbb{E}(y_{i0}(0)|\mu_i, \rho_i) = \frac{\mu_i}{1-\rho_i}$ ("stationary initial condition"),
  - or $\rho_i = 1$ (unit root)

- Results **do not** apply if $|\rho_i| < 1$ and $\mathbb{E}(y_{i0}(0)|\mu_i, \rho_i) \neq \frac{\mu_i}{1-\rho_i}$
  - Forecast may still perform well if $\mathbb{E}(y_{i0}(0)|\mu_i, \rho_i)$ is well-approximated by a polynomial time trend

- CLT assumption rules out strong dependence, e.g., **macro shocks that cannot be approximated by a polynomial** and that affect all individuals between τ and $\tau + h$

Parameters to choose

To implement our procedure, must choose R_i and q_i

Consider q_i :

- For unbiased forecasts in stationary or stochastic trends DGPs any q_i is ok
- For unbiased forecasts in DGPs with deterministic trends, we need $q_i \geq q_{i0}$
- In both cases forecast weights increase with q_i, so variance goes up
- Without assuming DGP for $y_{it}(0)$, cannot choose optimally
- Larger polynomial order q_i can mitigate bias due to nonstationary initial condition in short time series (simulations)

Parameters to choose

Consider R_i :

- Under mean stationarity, large R_i gives more precise estimates
- On the other hand, short R_i guards against violation of stationarity due to parameter change
- So short R_i may be preferable

Practical recommendation:
Set $R_i = q_i + 1$, and report FAT for range of values for q_i e.g., $q_i = 0, 1, ..., 3$.

Alternative: Model-Based (MB) forecast

Results so far covered many DGPs, but what if you have a **correctly specified model**, e.g. including **covariates** or lags that may help forecast? E.g.

$$y_{it}(0) = \rho y_{it-1}(0) + \delta_i t + \varepsilon_{it}$$

- Consistently estimate common parameter, $\widehat{\rho}$, and then do PR on residuals

$$\widehat{\alpha}^{(i,q_i,R_i)} = \arg\min_{\alpha \in \mathbb{R}^{q_i+1}} \sum_{t \in \mathcal{T}_i} \left(y_{it} - \widehat{\rho} y_{i\tau} - \sum_{k=0}^{q_i} \alpha_k t^k \right)^2$$

- The MB forecast is

$$\widehat{y}_{i\tau+1}^{MB}(0) = \widehat{\rho} y_{i\tau} + \sum_{k=0}^{q_i} \widehat{\alpha}_k^{(i,q_i,R_i)} (\tau + 1)^k$$

- Result: forecast is biased, but $FAT_{\tau+h}^{MB}$ consistent and asy normal

What if we have a control group?

- Without a control group we need to rule out common shocks between τ and $\tau + h$
- Could account for common shocks if have a control group as long as they lead to the same average forecast bias in the treated and control groups:

Assumption

For each $i = 1, \ldots, n$, let the forecast $\widehat{y}_{i\tau+h}(0)$ be a function of pre-treatment data such that

$$\frac{1}{n} \sum_{i \in Treated} \mathbb{E} \left(y_{i\tau+h}(0) - \widehat{y}_{i\tau+h}(0) \right) = \frac{1}{n} \sum_{i \in Control} \mathbb{E} \left(y_{i\tau+h}(0) - \widehat{y}_{i\tau+h}(0) \right)$$

DFAT estimator

- One can then estimate the ATT by taking the difference in FAT estimators, "DFAT"

$$DFAT = \frac{1}{n} \sum_{i \in \text{Treated}} (y_{i\tau+h} - \widehat{y}_{i\tau+h}(0)) - \frac{1}{n} \sum_{i \in \text{Control}} (y_{i\tau+h} - \widehat{y}_{i\tau+h}(0))$$

- Reminiscent of Diffs in Diffs estimator:

$$DiD = \frac{1}{n} \sum_{i \in \text{Treated}} (y_{i\tau+h} - y_{i\tau}) - \frac{1}{n} \sum_{i \in \text{Control}} (y_{i\tau+h} - y_{i\tau})$$

- What's different?
  - DFAT does not require restricting heterogeneity (no parallel paths assumption)
  - So valid under more general assumptions than DiD (as long as DGP has no deterministic trend or the trend is a polynomial)

Simulations: worst-case scenario for FAT

- DGP (small n, high persistence, non-stationary initial condition):

$$
\begin{aligned}
y_{it} &= \mu_i + .9 y_{it-1} + t + u_{it}, t \in \{1, \ldots, 6\}, n = 50 \\
\mu_i &\sim U[-1, 1] \\
u_{it} &\sim \mathcal{N}(0, 1) \\
y_{i0} &\sim \mathcal{N}(1, 2)
\end{aligned}
$$

- Balanced panel with $\tau = 5$, and $h = 1$

- Questions:
  - Choice of polynomial order q?
  - PR vs. MB forecast?

Simulations: Forecasts

- PR forecasts:
  1. for each i, regress $\{y_{it}\}_{t=1}^{5}$ on polynomial in t of order q,
  2. forecast is $\widehat{y}_{i6}(0) = \sum_{k=0}^{q} \widehat{\alpha}_{ik}^{(q,q+1)} 6^k$
- MB forecasts:
  1. Anderson-Hsiao estimator of ρ with
     - ★ IV y_{it-3} with time trend accounted for (correctly specified)
     - ★ IV y_{it-2} with time trend not accounted for (misspecified)
  2. for each i, regress $\{y_{it} - \widehat{\rho} y_{it-1}\}_{t=1}^{5}$ on polynomial in t of order q
  3. forecast is $\widehat{y}_{i6}(0) = \widehat{\rho} y_{i5} + \sum_{k=0}^{q} \widehat{\alpha}_{ik}^{(q,q+1)} 6^k$
- *FAT*

$$FAT_6 = \frac{1}{n} \sum_{i=1}^{n} \left[y_{i6} - \widehat{y}_{i6}(0) \right]$$

Bias

		$q = 0$	$q = 1$	$q = 2$	$q = 3$
MB correctly specified	PR	4.69	0.61	−0.06	0.01
		(0.16)	(0.21)	(0.36)	(0.65)
	MB	1.03	−0.04	−0.17	−0.25
		(3.38)	(0.69)	(0.75)	(1.42)
MB misspecified	PR	4.69	0.61	−0.06	0.01
		(0.16)	(0.21)	(0.36)	(0.65)
	MB	186.7	2.36	0.2	−0.42
		(5905.57)	(86.1)	(10.98)	(15.9)

- PR forecast works well
- Choosing larger q can reduce bias due to nonstationary initial condition
- MB forecast sensitive to misspecification

Relationship with literature

1. Treatment effects without a comparison group
   - Bayesian methods, using Kalman-filter, e.g., Varian (2014)
     - ★ **Strong parametric assumptions**
   - "Regression discontinuity in time" (popular in applied environmental economics)
     - ★ **Needs high frequency data around treatment + local estimation before and after treatment mixes short and long-run effects**

Relationship with literature

2. Forecasting with panel data, e.g., Empirical Bayes (Liu, Moon, Schorfheide, 2020)
   - **Strong parametric assumptions + focus on forecast accuracy yields biased forecasts**
3. Unbiased forecasts
   - One time series with large T: Dufour (1984)
   - Short panels: Mavroeidis at al. (2015)
     - **More restrictive DGPs, e.g., stationarity and symmetric errors**

Related literature

4. Synthetic controls, matrix completion
    - **Requires one treated and many potential controls + black box in terms of DGPs**
5. Heterogeneous treatment effects
    - OLS or TWFE generally inconsistent
    - Solutions assume the existence of a valid control group in every period, e.g., Callaway and Sant'Anna (2020); Sun and Abraham (2020); Goodman-Bacon (2021)
    - No standard DID estimator for the case of staggered, heterogeneous treatment, e.g., Baker, Larcker, Wang (2022)
        - ⋆ **We don't need a control group**

Empirical replication

Shover et al. (2019): Effect of legal cannabis laws (staggered) adoption in US states on opioid overdose mortality rates

- Outcome variable = annual opioid overdose mortality rate

The issue:

- Bachhuber et al. (2014) find decrease. Shover et al. (2019) find increase. Both use two-way fixed effects (biased)

Existing staggered adoption approaches

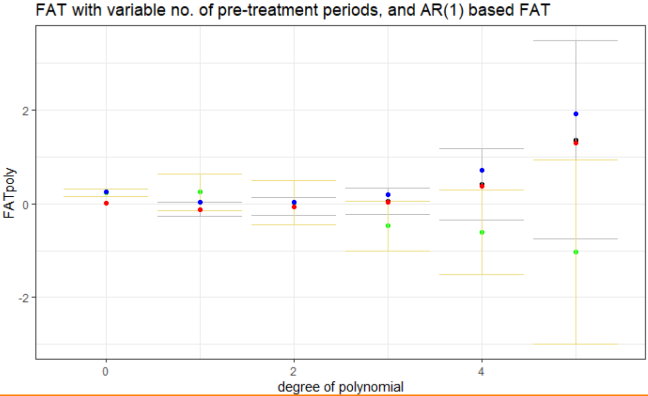
First we redo the analysis, but removing bias of original studies:

- Various staggered DID approaches $\rightarrow$ no significant effect
- Ceneralized SC approach of Xu (2017) and matrix completion approach of Athey et al. (2021) $\rightarrow$ no significant effect
- Can FAT replicate the no significant effect result without using the control group (= robustness check)?

FAT

$R = q + 1$, different q's: mostly positive effect, but statistically insignificant

- black dot = all states (grey CI),
- red dot = states adopting before 2010, blue dot = states adopting after 2010
- green dot = MB FAT (orange CI)



Lecture 3 conclusion

- Propose a forecast-based estimator (FAT) for the average treatment effects when there are only treated units and no control units
- The estimator is based on forecasted counterfactuals via local polynomial regressions, which we show are unbiased for a large class of DGPs
- We show consistency and asymptotic normality of FAT given unbiasedness of the forecasts
- Consistency and asymptotic normality also with (biased) model-based forecasts of counterfactuals, but less robust and sensitive to misspecification

Overall conclusion

- We discussed how to approach econometric policy analysis when we are uncertain about typical assumptions on identification, measurement and existence of a control group
- A running theme has been looking for robust methods, i.e., drop the problematic assumption and see what you can still say
- Robustness often means "giving up" something...
  - informative confidence bands in impulse response analysis
  - ability to control for common shocks in treatment effect estimation
- ...but, paraphrasing Manski, better to know what you don't know than make policy decisions driven by "incredible certitude"

References

- Athey, S., M. Bayati, N. Doudchenko, G. Imbens, and K. Khosravi (2021). Ma- trix completion methods for causal panel data models. Journal of the American Statistical Association
- Bachhuber, M., B. Saloner, C. Cunningham, and C. Barry (2014). Medical cannabis laws and opioid analgesic overdose mortality in the united states, 1999-2010. JAMA Intern Med.
- Baker, A., D. F. Larcker, and C. C. Y. Wang (2022). How much should we trust stag- gered difference-in-differences estimates? Journal of Financial Economics
- Callaway, B. and P. H. C. SantâAnna (2021). Difference-in-differences with multiple time periods. Journal of Econometrics
- Dufour, J.-M. (1984). Unbiasedness of predictions from estimated autoregressions when the true order is unknown. Econometrica
- Liu, L., H. Moon, and F. Schorfheide (2020). Forecasting with dynamic panel data models. Econometrica

References

- Goodman-Bacon, A. (2021). Difference-in-differences with variation in treatment timing. Journal of Econometrics
- Mavroeidis, S., Y. Sasaki, and I. Welch (2015). Estimation of heterogeneous au- toregressive parameters with short panel data. Journal of Econometrics
- Shover, C., C. Davis, S. Gordon, and K. Humphreys (2019). Association between medical cannabis laws and opioid overdose mortality has reversed over time. PNAS
- Sun, L. and S. Abraham (2020). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics
- Varian, H. (2014). Big data: New tricks for econometrics. Journal of Economic
- Xu, Y. (2017). Generalized synthetic control method: Causal inference with inter- active fixed effects models. Political Analysis Perspectives