# Causal Analysis with Unit-Level Event Studies: Application to the Child Penalty

Dmitry Arkhangelsky, Kazuharu Yanagimoto, & Tom Zohar

CEMFI

15-th Research Workshop, BdE – CEMFI
November 13, 2024

# Starting point

- Event-study regressions

$$Y_{i,t} = \text{fixed effects} + \sum_{h\geq 0} \tau_h \mathbf{1}\{t - E_i = h\} + \text{error}$$

  are extremely common in economics
  - Most popular tool in applied micro (Currie et al., '20, Goldsmith-Pinkham, '24)
  - Can view $\hat{\tau}_h$ as an impulse response function of horizon $h$
    - Which sometimes can be interpreted causally

- In many applications $\tau_h$ is not the final goal
  - Instead, want to understand how it **reacts to policies**, i.e., conduct causal analysis

# Child penalties

- Use **child penalties** (CPs) introduced in Kleven et al. '19 as a running example
  - Outcomes $Y_{i,t}$ are **labor earnings**, event $E_i$ is the **birth of the first child**
    - We will use Dutch administrative data for this
  - Can debate about the causal interpretation of $\tau_h$
    - **Not** the point of this paper
- Want to learn how CPs react to changes in childcare policies
  - Use the reform in the Netherlands in 2005, which expanded access to childcare
  - But first, we need to define what we mean by CP - **measurement**
- Current empirical literature approaches this question in an ad hoc way
  - Typically interacting event times with policy indicators
    - Raises econometric problems (Goldsmith-Pinkham et al., '24)
    - But more importantly, neither transparent nor flexible

# This paper

- Methodology to study how policy changes affect individual reactions to events
  - Two separate steps: **measurement** and **policy analysis**
- Measurement: construct **unit-level** event studies, i.e., $\hat{\tau}_{i,h}$
  - Borrowing methods from linear panel data literature
- Policy analysis: use $\hat{\tau}_{i,h}$ as a LHS outcome in policy regressions
  - Focusing on situations where policy variation is not experimental
    - And varies at a group level, e.g., geographic

# What do we achieve

- Introduce a coherent framework that improves over current ad hoc approaches
- Our methodology is transparent and modular
  - Users can change measurement and policy analysis steps separately
    - Unlike the current practice
- Policy evaluation step can use any innovation from the methodological literature
  - E.g., new diff-in-diff or synthetic control methods, binscatter, double ML, etc.
- Demonstrate advantages of the new strategy using the CP application

# Measurement

# Model

- The observed outcomes follow a strictly exogenous linear panel data model:

$$
Y_{i,t} = \alpha_i + \lambda_{g(i),t}(X_i) + \sum_{h \geq -h_0} \tau_{i,e,h} \mathbf{1}\{E_i = e\}\mathbf{1}\{t = e + h\} + \varepsilon_{i,t},
$$

$$
\mathbb{E}_{g(i)}[\varepsilon_{i,t} | E_i, X_i, \alpha_i, \boldsymbol{\tau}_i] = 0
$$

  where $X_i$ are unit-level covariates, and $g(i)$ is the group unit $i$ belongs to

- Assume $\varepsilon_{i,t}$ are conditionally independent across units within the group
- In our application:
  - $X_i$ is gender, birth cohort, and education level
  - $g(i)$ is the municipality individual $i$ resides in

# Discussion

- The object of interest $\tau_{i,e,h}$ varies over three dimensions:
  - $i$ – heterogeneity, $e$ – state dependence, and $h$ - dynamics
  - Only have a chance to study $\tau_{i,h} := \tau_{i,E_i,h}$, other effects cannot be recovered
    - Complicated object, more on this later
- Special role of $g(i)$ - will become apparent in policy analysis
  - Policy variation will be at this level
- Model is over-identified and can be tested
  - Using standard pretrends tests as in the event study regressions

# Estimation

- Estimation is done using OLS, treating $\tau_{i,h}$ as parameters
  - Analogous to the proposal in Borusyak et al. '24 and Arellano & Bonhomme '12
  - The model is over-identified, so other methods are possible
    - E.g., can restrict comparisons to individuals with similar event times
    - Or consider different ways of adjusting for $\alpha_i$
- The resulting estimator satisfies:

$$\hat{\tau}_{i,h} = \tau_{i,h} + \text{measurement error} + \text{estimation error},$$

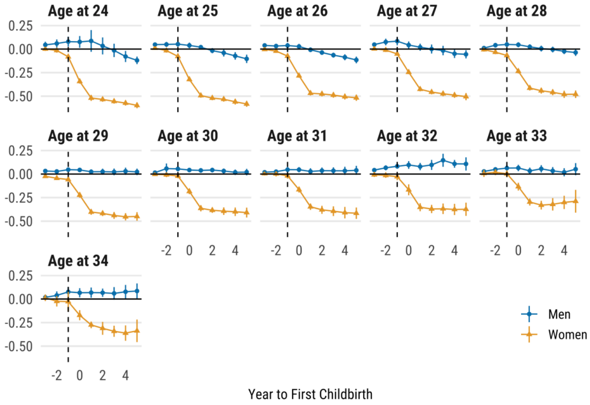
where $\mathbb{E}_{g(i)}[\text{errors}|E_i, X_i, \alpha_i, \boldsymbol{\tau}_i] = 0$

  - Only the estimation error vanishes with the sample size
  - Estimation errors are correlated within the group (will be important later)

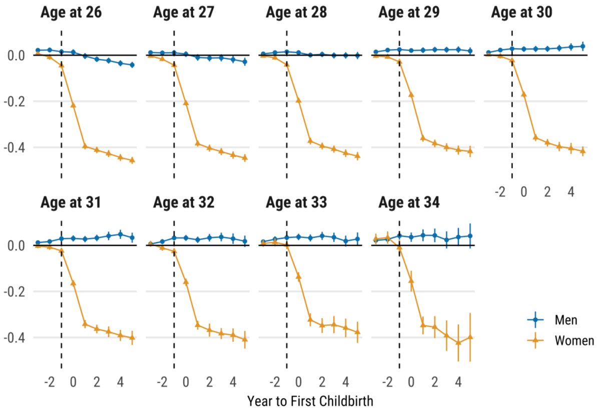
# Back to CPs

- Report average $\hat{\tau}_{i,h}$, conditioning on gender, age at first birth and education
  - Normalizing by average earnings of non-parents, i.e. $\mathbb{E}_{g(i)}[\alpha_i + \lambda_{g(i),t}(X_i)|E_i, X_i]$
- Three education categories
  - High school, vocational education, and bachelor's degree
- Look at averages to understand if the measurement model is reasonable
  - Correct model $\Rightarrow$ no effects prior to the childbirth
    - At least several years before

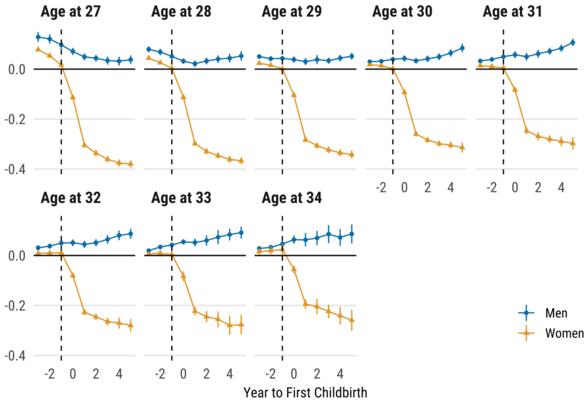
# Average results: High school



# Average results: Vocational



# Average results: Bachelor



# Discussion

- The model is more reasonable for women than men
  - Though it works quite well for men as well
- The model is unreasonable for younger individuals with a bachelor's degree
  - The effects three years before the event are substantial
- Overall - works surprisingly well on average
  - Perhaps yearly frequency helps; perhaps there is not that much selection
- In what follows, restrict the sample to subgroups where the model works well

# Policy Analysis

# Causal framework

- Consider policies that vary at the group level, denote them $W_g$
- View $\tau_{i,h}$ as the realization of the underlying potential outcome

$$\tau_{i,h} = \tau_{i,E_i(W_{g(i)}),h}(W_{g(i)})$$

- Two different effects of the policy
  - Direct effect, i.e., $\tau_{i,e,h}(w)$ as a function of $w$
  - Indirect effect through the event times $E_i(w)$
- Separating between the two is impossible in general (mediation analysis)
  - Additional problem: construct $\hat{\tau}_{i,h}$ for units with certain value of $E_i$
    - Introduces a sample selection problem

# Additional details

- Assume away the indirect effect, i.e., $E_i(w)$ does not vary with $w$
  - Automatically addresses the sample selection problem
  - Empirically is reasonable in our application
- If view $W_g$ as random $\Rightarrow$ can use any cross-sectional method
  - E.g., IV, Double ML, or any other modern evaluation method
  - Simply using $\hat{\tau}_{i,h}$ as the LHS variable
    - The error in $\hat{\tau}_{i,h}$ is mean-zero at $g$ level
- But $W_g$ is not random in our application

# Policy intervention in the Netherlands

- Before 2005:
  - Childcare services were subsidized at different rates by municipalities
  - Heterogeneous accountabilities of childcare services across municipalities
- After 2005:
  - Services are unified and subsidized at the same rate by the central government
  - Accountabilities are improved

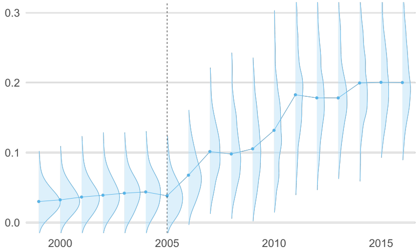



Figure: Distribution of the childcare index

# Dynamics

- Expansion is not random
  - But there is a variation over time, i.e., $W_g = (W_{g,1}, \ldots, W_{g,T})$
  - Will use this variation
- Assume limited dynamics

$$\tau_{i,h}(W_g) = \tau_{i,h}(W_{g,E_i-1}, W_{g,E_i+h}),$$

  where $h \geq 0$
  - Only a small part of the expansion path matters
  - No effect for $h < 0$ (test this)

- Seems to be reasonable in the application but can be more flexible

# Addressing confounding

- $W_g$ is not random
  - Childcare provision expansion can be systematically related to underlying CPs
- Use variation across event times + two-way model:

$$\tau_{i,h} = \alpha_{E_i}(X_i) + \beta_{g(i)}(X_i) + \tau_{-1}(X_i)W_{g,E_i-1} + \tau_{cont}(X_i)W_{g,E_i+h} + \nu_{i,h},$$
$$\mathbb{E}[\nu_{i,h}|X_i, E_i, W_{g(i)}] = 0,$$

  which relies on **cross-sectional** comparisons
  - To utilize the time-series dimension, can compare across horizons
- The error $\nu_{i,h}$ is **not** mean-zero at the cluster level conditional on $E_i, X_i$
  - But it is unrelated to the policy path $W_{g(i)}$

# Implementation

- Use constructed $\hat{\tau}_{i,h}$ (or rather its normalized version)
  - For a particular subsample of ages
- Final regression:

$$\hat{\tau}_{i,h} = \alpha_{E_i}(X_i) + \beta_{g(i)}(X_i) + \delta_{-1}(X_i)W_{g,E_i-1} + \delta_{cont}(X_i)W_{g,E_i+h} + \text{error}$$

  where the error has three components:
  1. Estimation error from the measurement equation (correlated at the group level)
  2. Measurement error from idiosyncratic errors $\varepsilon_{i,t}$ (independent across units)
  3. Policy evaluation error (correlated at the group level)
- To do inference use clustering at $g(i)$ level
  - Addressing all the errors we have above

# Preliminary results

- First step: test dynamics + the two way model
  - Use $\hat{\tau}_{i,h}$ for negative $h$
- Find null effects $\Rightarrow$ in line with our assumptions

| | -3 | -2 | -1 |
|---|---|---|---|
| cci_3 | -0.025 | -0.022 | -0.008 |
| | (0.079) | (0.079) | (0.090) |
| cci_2 | 0.114 | 0.106 | 0.085 |
| | (0.072) | (0.085) | (0.100) |
| cci_1 | 0.018 | 0.036 | 0.068 |
| | (0.061) | (0.076) | (0.089) |

Figure: Women, Vocational training

| | -3 | -2 | -1 |
|---|---|---|---|
| cci_3 | 0.072 | -0.047 | 0.062 |
| | (0.094) | (0.134) | (0.153) |
| cci_2 | 0.024 | 0.000 | -0.053 |
| | (0.101) | (0.113) | (0.139) |
| cci_1 | -0.006 | 0.049 | -0.001 |
| | (0.103) | (0.112) | (0.137) |

Figure: Women, Bachelor degree

# Policy results, I

- Start with the effect of $W_{g(i),E_i-1}$
  - I.e., the policy level the year **before** the childbirth

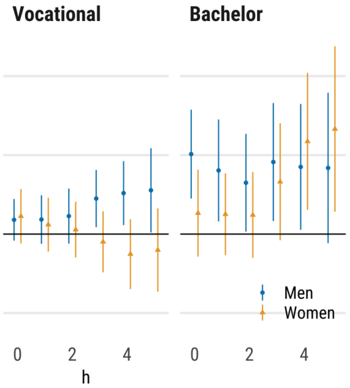



Figure: $\hat{\hat{\delta}}_{-1}(X_i)$, municipal-level s.e.

# Policy results, II

- Continue with the effect of $W_{g(i),E_i+h}$
  - I.e., the contemporaneous effect of the policy

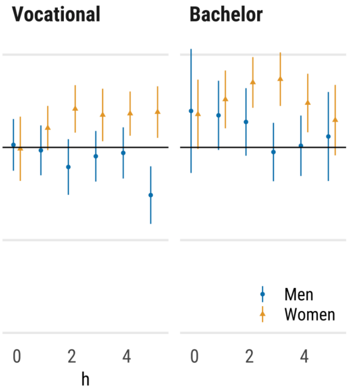



Figure: $\hat{\delta}_{cont}(X_i)$, municipal-level s.e.

# Discussion

- Results are preliminary; still thinking about the mechanisms
- Few initial thoughts:
  - Results for the vocational group are intriguing
    - Opposite movement for men and women
  - Large contemporaneous effects for $h = 3, 4$
    - When we expect the childcare to matter most
- Will explore more to think about the effects of second children, type of occupation
  - All feasible within our framework

# Concluding comments

# What have we done?

- New framework with two crucial steps:
  - Measurement and policy analysis
  - Can be applied very broadly
- Two steps rely on conceptually different assumptions
  - These assumptions are testable
- Final estimation procedure is linear $\Rightarrow$ can be done in one step
  - But this step is very specific and depends on the intermediate methods
  - It is much more transparent to report and validate each step separately
- Inference is not a problem $\Rightarrow$ simply use group-level clustered s.e.

# Going forward

- View the separation of measurement and policy analysis as a general principle
  - Applicable to many empirical problems in economics
  - Generating insights that are harder to get using conventional tools
- Our particular approach can be generalized
  - More complicated measurement models (perhaps nonlinear)
  - Other policy evaluation assumptions, e.g., synthetic control type methods
- Individual-level measurement can be used more broadly
  - To think about variance decompositions or empirical Bayes computations
    - Need additional assumptions on the measurement step