

# Nonlinear Difference-in-Differences with Repeated Cross Sections

**ASSA**

January 4, 2026  
Philadelphia, PA

Jeff Wooldridge  
Department of Economics  
Michigan State University

1. Introduction
2. Notation, Assumptions, and Identification
3. Estimation by Imputation and Pooled QMLE
4. Application

# 1. Introduction

- Much of the work on difference-in-differences for staggered interventions has focused on
  - ▶ Linear outcome models.
  - ▶ Panel data.
- Some papers allow repeated cross sections with linear outcome models:
  - ▶ Callaway and Sant'Anna (2021, Journal of Econometrics) [CS (2021)]
  - ▶ Borusyak, Jaravel, Spiess (2024, REStud) [BJS (2024)]
  - ▶ Deb, Norton, Wooldridge, Zabel (2025, WP) [DNWZ (2025)].

- What if we have repeated cross-sections and a binary response?
  - ▶ Almost certainly start with a linear probability model.
  - ▶ The (conditional) parallel trends assumption is sensitive to functional form; Roth and Sant'Anna (2023, Econometrica).
- As a robustness check, try a simple nonlinear model.
  - ▶ Extend Wooldridge (2023, Econometrics Journal) to repeated cross sections.
  - ▶ In some cases, a nonlinear model will pass pre-trends tests when a linear model does not.

- With  $Y_t$  the outcome, common situations that might call for a nonlinear model:
  - ▶  $Y_t$  is binary.
  - ▶  $Y_t$  is fractional.
  - ▶  $Y_t$  is a count.
  - ▶  $Y_t \geq 0$  is a corner solution.
- Want estimation methods that, like OLS, impose very few assumptions for consistency.
  - ▶ Pool quasi-maximum likelihood estimation (QMLE) in the linear exponential family (LEF).
  - ▶ Additional benefits from using the canonical link version of the mean function.

## 2. Notation, Assumptions, and Identification

- Identification assumes a stable population over time periods  $t = 1, \dots, T$ .
- First intervention is at time  $q$ ; more units treated through  $T$ .
- Potential outcomes for each treated state:

$$Y_t(g), g \in \{q, \dots, T, \infty\}, t \in \{1, 2, \dots, T\},$$

- ▶  $Y_t(\infty)$  is the outcome in the never treated state.
- Cohort dummies (exhaustive and mutually exclusive):

$$D_g, g \in \{q, \dots, T, \infty\}$$

- ▶  $D_\infty = 1$  means a never treated unit.

- Following the literature, we want to (at least) estimate the average treatment effects on the treated (ATTs):

$$\begin{aligned}\tau_{gt} &= E[Y_t(g) - Y_t(\infty) | D_g = 1] \\ &= E(Y_t | D_g = 1) - E[Y_t(\infty) | D_g = 1], t = g, \dots, T; g = q, \dots, T,\end{aligned}$$

- ▶  $E(Y_t | D_g = 1)$  is easy to estimate given a random sample at time  $t$ :  $\bar{Y}_{gt} = N_{gt}^{-1} \sum_{i=1}^{N_t} D_{itg} \cdot Y_{it}$ .

- ▶ Hard part is the counterfactual mean:

$$E[Y_t(\infty) | D_g = 1]$$

- Allow covariates  $\mathbf{X}_t$  (assume no bad controls); rule out spillovers (SUTVA).

**Assumption CNA (Conditional No Anticipation):** For  $g \in \{q, \dots, T\}$  and  $t \in \{1, \dots, g - 1\}$ ,

$$E[Y_t(g)|\mathbf{D}, \mathbf{X}_t] = E[Y_t(\infty)|\mathbf{D}, \mathbf{X}_t]. \quad \square$$

- ▶ Rules out anticipatory changes in the potential outcomes.
- ▶ Difficult to relax, can skip over pre-treatment periods just before the intervention.

**Assumption CPT (Conditional Parallel Trends):** For  $t = 1, 2, \dots, T$  and  $G(\cdot)$  a twice-continuously differentiable, strictly increasing function,

$$E[Y_t(\infty)|\mathbf{D}, \mathbf{X}_t] = G\left(\alpha + \sum_{g=q}^T \beta_g D_g + \mathbf{X}_t \boldsymbol{\kappa} + \sum_{g=q}^T (D_g \cdot \dot{\mathbf{X}}_{tg}) \boldsymbol{\eta}_g + \gamma_t + \mathbf{X}_t \boldsymbol{\pi}_t\right),$$

- ▶  $\dot{\mathbf{X}}_{tg} \equiv \mathbf{X}_t - E(\mathbf{X}_t | D_g = 1)$
- ▶ As normalizations,  $\gamma_t = 0$ ,  $\boldsymbol{\pi}_t = \mathbf{0}$ .  $\square$
- Except when  $G(z) = z$ , Assumption CPT imposes parallel trends on the linear index, not  $E[Y_t(\infty)|\mathbf{D}, \mathbf{X}_t]$ .

- Define a time-varying treatment indicator:

$$W_t = D_q \cdot pq_t + D_{q+1} \cdot p(q+1)_t + \cdots + D_T \cdot pT_t,$$

$$pg_t = fg_t + \cdots + fT_t$$

$fs_t$  are time period dummies

- Imposing Assumptions CNA and CPT gives

$$E(Y_t | \mathbf{D}, \mathbf{X}_t, W_t = 0) = G \left( \alpha + \sum_{g=q}^T \beta_j D_g + \mathbf{X}_t \boldsymbol{\kappa} + \sum_{g=q}^T (D_g \cdot \dot{\mathbf{X}}_{tg}) \boldsymbol{\eta}_g + \sum_{s=2}^T \gamma_s fs_t + \sum_{s=2}^T (fs_t \cdot \mathbf{X}_t) \boldsymbol{\pi}_s \right)$$

### 3. Estimation by Imputation and Pooled QMLE

- Identification suggests estimating  $E[Y_t(\infty)|D_g = 1]$  by imputation.
  - ▶ BJS (2024) in linear panel data case.
  - ▶ Wooldridge (2023) in nonlinear panel data case.
  - ▶ DNWZ (2025) in linear repeated cross sections case.
- Imputation makes inference more difficult, does not provide all information.
- Pooling across all observations makes inference easier.

- In sampling the data, let  $t(i)$  denote the observed time period for unit  $i$ .
- Observed data are

$$\{Y_{i,t(i)}, \mathbf{X}_{i,i(t)}, D_{i,t(i),g}, g = q, \dots, T, i = 1, \dots, N\}$$

- With independence across  $i$ , inference need only be made robust to distributional misspecification (heteroskedasticity in linear case).
- With cluster sampling or clustered assignment, need to cluster at a suitable level.
  - ▶ Common is cohort assignment is at, say, the county level.
- Estimates and valid inference can be based on a single conditional expectation.

$$\begin{aligned}
E[Y_{it(i)} | \mathbf{D}_{i,t(i)}, \mathbf{X}_{i,t(i)}] = & G \left[ \alpha + \sum_{g=q}^T \beta_g D_{i,t(i),g} \right. \\
& + \mathbf{X}_{i,t(i)} \boldsymbol{\kappa} + \sum_{g=q}^T (D_{i,t(i),g} \cdot \dot{\mathbf{X}}_{i,t(i),g}) \boldsymbol{\eta}_g \\
& + \sum_{s=2}^T \gamma_s f_s t + \sum_{s=2}^T (f_s t \cdot \mathbf{X}_{i,t(i)}) \boldsymbol{\pi}_s \\
& + \sum_{g=q}^T \sum_{s=g}^T \delta_{gs} (W_{i,t(i)} \cdot D_{i,t(i),g} \cdot f_s t) \\
& \left. + \sum_{g=q}^T \sum_{s=g}^T (W_{i,t(i)} \cdot D_{i,t(i),g} \cdot f_s t \cdot \dot{\mathbf{X}}_{i,t(i),g}) \boldsymbol{\xi}_{gs} \right]
\end{aligned}$$

- Obtain the average partial effects (APEs) with respect to  $W_{i,t(i)}$ .
  - ▶ Average across subsamples  $(g, t)$  to obtain  $\hat{\tau}_{gt}$ .
  - ▶ Average by exposure times.
  - ▶ Can obtain a single overall weighted effect.
  - ▶ In Stata, `margins with subpop()` and `vce(uncond)` does all the work.
- When  $G^{-1}(\cdot)$  is the canonical link function, imputation = pooled QMLE.
  - ▶ Handy for checking calculations.
- The  $\hat{\delta}_{gt}$ , and aggregated versions, are also of some interest.
  - ▶ Proportionate effects or log odds.

## Leads and Lags (Event Study) Estimation

- To the earlier conditional mean function, add the pre-treatment indicators and interactions:

$$NW_{i,t(i)} \cdot D_{i,t(i),g} \cdot f_{s_t}, NW_{i,t(i)} \cdot D_{i,t(i),g} \cdot f_{s_t} \cdot \dot{X}_{i,t(i),g}, s = 1, \dots, g - 2$$

- ▶ Coefficients for  $s = g - 1$  normalized to zero.
- ▶  $NW_{i,t(i)} = 1 - W_{i,t(i)}$ .
- ▶ Aggregate time until treatment using APEs with respect to  $NW_{i,t(i)}$ ; margins.
  - ▶ Produces, along with the ATTs, an aggregated event study graph.
- Imputation and pooled estimation are still the same with the canonical link.
  - ▶ Linear + OLS, logit, flogit, exponential + Poisson.

## 4. Application

- Effects of Secure Communities immigration enforcement on SNAP enrollment (binary).
  - ▶ Individual survey data from the ACS, 2005-2016.
  - ▶ Hispanic U.S. citizens who reported not moving states.
  - ▶ Treatment at the county level; need to aggregate to PUMA.
  - ▶ Four treated cohorts: 2009, 2010, 2011, 2012.
  - ▶ More than 500,000 observations; 682 PUMAs across 32 states.
  - ▶  $\mathbf{X}_t$  includes 31 state dummies; gender; age; education; poverty status.
  - ▶ Used in DNWZ (2025) in linear analysis.

## Weighted Aggregated Effects

	Lags Only	Leads and Lags
Linear	−0.0143 (0.0049)	−0.0135 (0.0067)
Logit	−0.0099 (0.0051)	−0.0133 (0.0071)

- Stata code available on request.
  - ▶ Standard errors clustered at the PUMA level.
  - ▶ Estimating several hundred parameters (12 time periods, four cohorts,  $\mathbf{X}_t$  includes 35 variables).
  - ▶ Lots of empty cells; need to be careful with state dummies.
  - ▶ Run time is several hours.

