Difference-in-Differences with Panel Data

Slides 6: **Nonlinear Difference-in-Differences**

ESTIMATE: The Reduced Form
December 14-15, 2021

Jeff Wooldridge
Department of Economics
Michigan State University

1. Why Nonlinear Difference-in-Differences?
2. Nonlinear Models with $T = 2$
3. General Common Timing Case
4. Adding Covariates
5. Staggered Interventions
6. Some Simulations

1. Why Nonlinear Difference-in-Differences?

- Economic outcomes often are limited in some important way.
- At disaggregated levels, the response variable Y_{it} is often binary.
- At all levels of aggregation, Y_{it} might be a fractional response:

$0 \leq Y_{it} \leq 1$.

- Nonnegative outcomes are also important: $Y_{it} \geq 0$.
  - Could be a count variable, continuous, or mixed.
  - Zero is often an important value.

- Standard argument for using nonlinear models for limited dependent variables:
  - They provide a better “fit” than linear models.
  - Marginal effects are sometimes more plausible.
- In DID settings, using nonlinear mean functions can make the parallel (common) trends assumption more plausible.
- The PT assumption is not generally invariant to transformations.
  - If PT holds in logs, say, it is unlikely it holds in levels.

- Roth and Sant'Anna (2020, WP) on PT invariance to transformations.
- Athey and Imbens (2006, Econometrica): "Changes-in-Changes" formulations.
  - With discrete outcomes, only get bounds on the average treatment effect on the treated (ATT).
- With functional form assumptions, can identified ATTs.
- Current work: Extension of linear case in Wooldridge (2021, WP).

2. Nonlinear Models with $T = 2$

- Binary treatment indicator, D.
  - Treatment in second period.
- Potential outcomes are $Y_t(0)$, $Y_t(1)$, $t \in \{1,2\}$.
- We want the ATT in $t = 2$:

$$\tau_2 = E[Y_2(1) - Y_2(0)|D = 1]$$

- Impose a no anticipation assumption:

$$Y_1 = Y_1(1) = Y_1(0)$$

- Linear parallel trends (in untreated state):

$$E[Y_1(0)|D] = \alpha + \beta D$$

$$E[Y_2(0)|D] = \alpha + \beta D + \gamma_2$$

$$E[Y_2(0)|D] - E[Y_1(0)|D] = (\alpha + \beta D + \gamma_2) - (\alpha + \beta D) = \gamma_2$$

  - Trend in mean does not depend on D.
- Especially for discrete outcomes, linear parallel trends might be unrealistic.

- For a strictly increasing, continuously differentiable function $G(\bullet)$,

$$E[Y_1(0)|D] = G(\alpha + \beta D)$$

  - Treatment can be systematically related to $Y_1(0)$.

- In the second period,

$$E[Y_2(0)|D] = G(\alpha + \beta D + \gamma_2)$$

- A nonlinear transformation of the means satisfies PT:

$$G^{-1}(E[Y_2(0)|D]) - G^{-1}(E[Y_1(0)|D]) = \gamma_2$$

- Suppose $Y_t(0)$ is binary:

$$Y_t(0) = 1[Y_t^*(0) > 0]$$

$$Y_t^*(0) = \alpha + \beta D + \gamma f2_t + U_{it}$$

$$U_{it} \text{ continuous, independent of } D,\ t = 1,2$$

$$U_{i1},\ U_{i2} \text{ identically distributed (may be correlated) with CDF } F(\bullet)$$

- Then

$$E[Y_t(0)|D] = P[Y_t(0) = 1|D] = P[\alpha + \beta D + \gamma_t + U_t > 0|D]$$
$$= 1 - F[-(\alpha + \beta D + \gamma_t)] \equiv G(\alpha + \beta D + \gamma_t)$$

- PT assumption for a linear model holds for the latent variable $Y_t^*(0)$:

$$E[Y_t^*(0)|D] = \alpha + \beta D + \gamma_t,\ t = 1, 2$$

  - PT fails for $E[Y_t(0)|D]$.

- Exponential example:

$$E[Y_t(0)|D] = \exp(\alpha + \beta D + \gamma_t),\ t = 1,2 \ \ (\gamma_1 \equiv 0)$$

$$\frac{E[Y_2(0)|D]}{E[Y_1(0)|D]} = \exp(\gamma_2)$$

  - Does not depend on D.

- Equivalently, the growth in the mean does not depend on D:

$$\log\{E[Y_2(0)|D]\} - \log\{E[Y_1(0)|D]\} = \gamma_2$$

- General case with $G(\bullet)$ continuously differentialy, strictly increasing:

$$E[Y_1(0)|D] = G(\alpha + \beta D)$$

- By no anticipation,

$$E(Y_1|D) = E[Y_1(0)|D] = G(\alpha + \beta D)$$

  - So α and β are identified using $t = 1$ data.

- What about identifying the parameter of interest,

$$\tau_2 = E[Y_2(1)|D = 1] - E[Y_2(0)|D = 1]?$$

- We observe D, $Y_1 = Y_1(0) = Y_1(1)$, and

$$Y_2 = (1 - D)Y_2(0) + DY_2(1)$$

- First term in τ_2 is easy:

$$E(Y_2|D = 1) = E[Y_2(1)|D = 1]$$

- Given a random sample of size N, number of treated units is

$$N_1 = \sum_{i=1}^{N} D_i$$

- As usual, a consistent estimator of $E[Y_2(1)|D = 1]$ is

$$\bar{Y}_{12} = N_1^{-1} \sum_{i=1}^{N} D_i Y_{i2} = \left(\frac{N_1}{N}\right)^{-1} \left(N^{-1} \sum_{i=1}^{N} D_i Y_{i2}\right)$$

  - Average of treated units in $t = 2$.

- As usual, the second part is harder.

$$E[Y_2(0)|D] = G(\alpha + \beta D + \gamma_2)$$

$$E[Y_2(0)|D = 1] = G(\alpha + \beta + \gamma_2)$$

- Need to estimate α, β, and γ_2.
- α and β are identified using $t = 1$ and $D \in \{0, 1\}$:

$$E(Y_1|D) = E[Y_1(0)|D] = G(\alpha + \beta D)$$

- γ_2 is then identified using $t = 2$ and $D = 0$.

$$E(Y_2|D = 0) = E[Y_2(0)|D = 0] = G(\alpha + \gamma_2)$$

- Can estimate all parameters at once.
- Define $f2_t = 1$ if $t = 2$, zero otherwise.
- Time-varying treatment indicator:

$$W_t = D \cdot f2_t, \; t = 1, 2$$

  - $W_t = 1$ for treated units in period two.
- Use the $W_t = 0$ ("untreated") observations to estimate α, β, γ_2.
  - $D = 0$ in both periods, $t = 1$ for $D = 1$.

- Linear case: Obtain $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}_2$ from the OLS regression

$$Y_{it} \text{ on } 1, D_i, f2_t,\ \ t = 1,2;\ i = 1,\ldots,N \text{ if } W_{it} = 0$$

- Then ATT for $t = 2$ is estimated as

$$\hat{\tau}_2 = \bar{Y}_{12} - \left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_2\right)$$

- This is the previous imputation estimator in the simplest case.
- It produces usual DID estimator:

$$\hat{\tau}_2 = (\bar{Y}_{12} - \bar{Y}_{11}) - (\bar{Y}_{02} - \bar{Y}_{01})$$

- With nonlinear $G(\cdot)$, for robust use quasi-MLE in the linear exponential family (LEF).
  - But can use other methods to estimate

$$E(Y_{it}|D_i, W_{it} = 0) = G(\alpha + \beta D_i + \gamma_2 f2_t)$$

- Benefits if we use the canonical link function in the chosen LEF.
- Then, can show that $\hat{\tau}_2$ is equivalent to the average partial effect of W_t evaluated at $D = 1, f2_t = 1$.

- Specifically, let $\hat{\alpha}$, $\hat{\beta}$, $\hat{\gamma}_2$, and $\hat{\delta}_2$ be the QMLEs from using all of the data and estimating the mean function

$$E(Y_{it}|D_i, W_{it}) = G(\alpha + \beta D_i + \gamma_2 f2_t + \delta_2 W_{it}), t = 1,2$$

  - Only needs to be the mean function for $W_t = 0$.

- Then

$$\hat{\tau}_2 = G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_2 + \hat{\delta}_2\right) - G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_2\right) = \bar{Y}_{12} - G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_2\right)$$

  - Convenient for obtaining standard error of $\hat{\tau}_2$.

- Canonical link pairs (mean + LLF):
  - Linear + Normal (leads to OLS)
  - Logistic + Bernoulli (binary and fractional)
  - Logistic + Binomial (nonnegative, known upper bound)
  - Exponential + Poisson (nonnegative, no natural upper bound)

3. General Common Timing Case

- Allow any $T > 2$. Intervention occurs at $1 < q \leq T$.
- Treatment indicator is D. Potential outcomes are $Y_t(0)$, $Y_t(1)$.
- No anticipation:

$$Y_t = Y_t(0) = Y_t(1),\ 1 \leq t < q$$

- ATTs for each treated period:

$$\tau_r = E[Y_r(1) - Y_r(0)|D = 1], r = q, q+1, \ldots, T$$

- Common or Parallel Trends Assumption: For a known, strictly increasing, continuously differentiable function $G(\cdot)$ and parameters α, β, and $\gamma_2, \ldots, \gamma_T$,

$$E[Y_t(0)|D] = G(\alpha + \beta D + \gamma_t), t = 1, 2, \ldots, T \ (\gamma_1 \equiv 0)$$

- Equivalently,

$$G^{-1}(E[Y_t(0)|D]) - G^{-1}(E[Y_{t-1}(0)|D]) = \gamma_t - \gamma_{t-1}, t = 2, \ldots, T$$

  - Transformation of mean does not depend on D.

- ATTs still have the form

$$\tau_r = E(Y_r|D = 1) - G(\alpha + \beta + \gamma_r), r = q, \ldots, T$$

- Time-varying treatment indicator:

$$W_{it} = D_i(fq_t + \cdots + fT_t) \equiv D_i p_t$$

  - $W_{it} = 1$ for a treated unit in an intervention period.

- α, β, γ_2, ..., γ_T estimated by pooling $W_{it} = 0$ observations.

$$\hat{\tau}_r = \bar{Y}_{1r} - G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r\right), r = q, \ldots, T$$

$$\bar{Y}_{1r} = N_1^{-1} \sum_{i=1}^{N} D_i Y_{ir}$$

- When $G(\cdot)$ is the canonical link, $\hat{\tau}_r$ can be obtained as estimated APEs with respect to W_{it} in the mean function

$$\begin{aligned} E(Y_{it}|D_i) = G[&\alpha + \beta D_i + \gamma_2 f2_t + \cdots + \gamma_T fT_t \\ &+ \delta_q(W_{it} \cdot fq_t) + \cdots + \delta_T(W_{it} \cdot fT_t)] \end{aligned}$$

- Pooled estimation using all data.

$$\hat{\tau}_r = G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r + \hat{\delta}_r\right) - G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r\right), r = q, \ldots, T$$

  - Because of algebraic equivalence, $G(\alpha + \beta + \gamma_t)$ only has to be the mean function when $W_t = 0$.
  - $W_{it} \cdot fr_t = W_{it} \cdot D_i \cdot fr_t$ and APE calculations are the same.

4. Adding Covariates

- Still assume no anticipation.
- Parallel trends conditional on $\mathbf{X}$:

$$\begin{aligned} E[Y_t(0)|D,\mathbf{X}] &= G[\alpha + \beta D + \mathbf{X}\boldsymbol{\kappa} + (D \bullet \mathbf{X})\boldsymbol{\xi} \\ &\quad + \gamma_2 f2_t + \cdots + \gamma_T fT_t + (f2_t \bullet \mathbf{X})\boldsymbol{\pi}_2 + \cdots + (fT_t \bullet \mathbf{X})\boldsymbol{\pi}_T] \end{aligned}$$

  - No interactions between D and the fs_t.
  - No triple interactions $D \bullet fs_t \bullet \mathbf{X}$.

- Still want ATTs for $r = q, q+1, \ldots, T$:

$$\begin{aligned}\tau_r &= E[Y_r(1) - Y_r(0)|D = 1] \\ &= E(Y_r|D = 1) - E[G(\alpha + \beta + \gamma_r + \mathbf{X}(\boldsymbol{\kappa} + \boldsymbol{\xi} + \boldsymbol{\pi}_r))|D = 1]\end{aligned}$$

- α, β, $\boldsymbol{\kappa}$, $\boldsymbol{\xi}$, γ_2, ..., γ_T, $\boldsymbol{\pi}_2$, ..., $\boldsymbol{\pi}_T$ can be estimated using a quasi-MLE in the LEF restricted to $W_{it} = 0$.

- Imputation estimator:

$$\hat{\tau}_r = \bar{Y}_{1r} - N_1^{-1} \sum_{i=1}^{N} D_i G\left(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r + \mathbf{X}_i\left(\hat{\boldsymbol{\kappa}} + \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\pi}}_r\right)\right)$$

$$\bar{Y}_{1r} = N_1^{-1} \sum_{i=1}^{N} D_i Y_{ir}, N_1 = \sum_{i=1}^{N} D_i, r = q, \ldots, T$$

• In the canonical link case, equivalent to pooling all data to estimate

$$\begin{aligned}
E(Y_{it}|D_i,\mathbf{X}_i) = G[&\alpha + \beta D_i + \mathbf{X}_i\boldsymbol{\kappa} + (D_i \cdot \mathbf{X}_i)\boldsymbol{\xi} \\
&+ \gamma_2 f2_t + \cdots + \gamma_T fT_t + (f2_t \cdot \mathbf{X}_i)\boldsymbol{\pi}_2 + \cdots + (fT_t \cdot \mathbf{X}_i)\boldsymbol{\pi}_T \\
&+ \delta_q(W_{it} \cdot fq_t) + \cdots + \delta_T(W_{it} \cdot fT_t) \\
&+ (W_{it} \cdot fq_t \cdot \mathbf{X}_i)\boldsymbol{\xi}_q + (W_{it} \cdot fT_t \cdot \mathbf{X}_i)\boldsymbol{\xi}_T\Big]
\end{aligned}$$

▸ For $\hat{\tau}_r$, obtain the APE of W_t at $D = 1, fr_t = 1, fs_t = 0$ all $s \neq r$, and average across the subsample $D_i = 1$.

▸ Same as using $W_{it} \cdot D_i$ in place of W_{it}.

- For $r = q, \ldots, T$,

$$\hat{\tau}_r = N_1^{-1} \sum_{i=1}^{N} D_i \Big[G\Big(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r + \hat{\delta}_t + \mathbf{X}_i\big(\hat{\boldsymbol{\kappa}} + \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\pi}}_r + \hat{\boldsymbol{\xi}}_r\big)\Big) - G\Big(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r + \mathbf{X}_i\big(\hat{\boldsymbol{\kappa}} + \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\pi}}_r\big)\Big) \Big]$$

$$= \bar{Y}_{1r} - N_1^{-1} \sum_{i=1}^{N} D_i G\Big(\hat{\alpha} + \hat{\beta} + \hat{\gamma}_r + \mathbf{X}_i\big(\hat{\boldsymbol{\kappa}} + \hat{\boldsymbol{\xi}} + \hat{\boldsymbol{\pi}}_r\big)\Big)$$

- In Stata, the `margins` command is convenient.
  - Use `vce(uncond)` to account for sampling error in

$\{\mathbf{X}_i : i = 1, \ldots, N\}$.

```
logit y i.w#c.fq ... i.w#c.fT
 i.w#c.fq#c.x1 ... i.w#c.fq#c.xK
 ... i.w#c.fT#c.x1 ... i.w#c.fT#c.xK
 d x1 ...xK c.d#c.x1 ... c.d#c.xK
 f2 ... fT c.f2#c.x1 ... c.f2#c.xK
 ... c.fT#c.x1 ... c.fT#c.xK,
 noomitted vce(cluster id)
```

```
margins, dydx(w) at(d = 1 f2 = 0 ... fq_min1 = 0
 fq = 1 fq_plus1 = 0 ... fT = 0)
 subpop(if d == 1) vce(uncond)
margins, dydx(w) at(d = 1 f2 = 0 ... fq_min1 = 0
 fq = 0 fq_plus1 = 1 ... fT = 0)
 subpop(if d == 1) vce(uncond)
⋮
margins, dydx(w) at(d = 1 f2 = 0 ... fq_min1 = 0
 fq = 0 fq_plus1 = 0 ... fT = 1)
 subpop(if d == 1) vce(uncond)
```

```
. use did_common_6_binary, clear

. tab year

       year |      Freq.     Percent        Cum.
------------+-----------------------------------
       2001 |      1,000       16.67       16.67
       2002 |      1,000       16.67       33.33
       2003 |      1,000       16.67       50.00
       2004 |      1,000       16.67       66.67
       2005 |      1,000       16.67       83.33
       2006 |      1,000       16.67      100.00
------------+-----------------------------------
      Total |      6,000      100.00

. tab d if f06

          d |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |        618       61.80       61.80
          1 |        382       38.20      100.00
------------+-----------------------------------
      Total |      1,000      100.00

. tab y

          y |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      3,468       57.80       57.80
          1 |      2,532       42.20      100.00
------------+-----------------------------------
      Total |      6,000      100.00
```

```
. * Sample ATTs:

. sum te_i if d & f04

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        te_i |        382     .078534    .6054086         -1          1

. sum te_i if d & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        te_i |        382     .117801    .6055278         -1          1

. sum te_i if d & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        te_i |        382    .1099476    .6177221         -1          1
```

```
. * Linear model (standard errors not adjusted for xbar):
.
. reg y c.d#c.f04 c.d#c.f05 c.d#c.f06 ///
>         c.d#c.f04#c.x_dm c.d#c.f05#c.x_dm c.d#c.f06#c.x_dm ///
>         f02 f03 f04 f05 f06 ///
>         c.f02#c.x c.f03#c.x c.f04#c.x c.f05#c.x c.f06#c.x ///
>         d x c.d#c.x, vce(cluster id)

Linear regression                                 Number of obs     =      6,000
                                                  F(19, 999)        =      73.84
                                                  Prob > F          =     0.0000
                                                  R-squared         =     0.1458
                                                  Root MSE          =     .45722

                                     (Std. err. adjusted for 1,000 clusters in id)
----------------------------------------------------------------------------------
                 |               Robust
               y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------+----------------------------------------------------------------
       c.d#c.f04 |   .0462206   .0367915     1.26   0.209    -.0259768    .1184181
                 |
       c.d#c.f05 |   .0755069   .0375981     2.01   0.045     .0017265    .1492873
                 |
       c.d#c.f06 |   .0445457   .0378681     1.18   0.240    -.0297645     .118856
                 |
c.d#c.f04#c.x_dm |  -.0185417   .0885739    -0.21   0.834    -.1923539    .1552704
                 |
c.d#c.f05#c.x_dm |   .0453048    .085807     0.53   0.598    -.1230779    .2136875
                 |
c.d#c.f06#c.x_dm |    .130662   .0841938     1.55   0.121    -.0345549    .2958789
```

f02	.0009375	.0509461	0.02	0.985	-.0990362	.1009113
f03	-.0363479	.0531606	-0.68	0.494	-.1406672	.0679714
f04	.0227675	.0718298	0.32	0.751	-.118187	.1637221
f05	.0047325	.068471	0.07	0.945	-.129631	.1390961
f06	.132355	.0694451	1.91	0.057	-.0039201	.26863
c.f02#c.x	-.0118991	.0455013	-0.26	0.794	-.1011881	.0773899
c.f03#c.x	.0392213	.0484515	0.81	0.418	-.0558571	.1342997
c.f04#c.x	.0613781	.0705436	0.87	0.384	-.0770525	.1998087
c.f05#c.x	.0941201	.0663857	1.42	0.157	-.0361513	.2243914
c.f06#c.x	.0185686	.0705219	0.26	0.792	-.1198194	.1569567
d	-.3353013	.0434568	-7.72	0.000	-.4205783	-.2500243
x	.1262778	.0447227	2.82	0.005	.0385166	.214039
c.d#c.x	-.0628457	.0421785	-1.49	0.137	-.1456143	.019923
_cons	.3866191	.0456444	8.47	0.000	.2970492	.4761891

```
. * Pooled logit:

. logit y i.w#c.d#c.f04 i.w#c.d#c.f05 i.w#c.d#c.f06 ///
>         i.w#c.d#c.f04#c.x i.w#c.d#c.f05#c.x i.w#c.d#c.f06#c.x ///
>         f02 f03 f04 f05 f06 c.f02#c.x c.f03#c.x c.f04#c.x c.f05#c.x c.f06#c.x ///
>         d x c.d#c.x, noomitted vce(cluster id)

--------------------------------------------------------------------------------
                |               Robust
              y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
----------------+---------------------------------------------------------------
    w#c.d#c.f04 |
             1  |   .9412159   .5137085     1.83   0.067    -.0656342    1.948066
                |
    w#c.d#c.f05 |
             1  |   .7852733   .4969686     1.58   0.114    -.1887672    1.759314
                |
    w#c.d#c.f06 |
             1  |   .2581332   .4880446     0.53   0.597    -.6984167    1.214683
                |
w#c.d#c.f04#c.x |
             1  |  -.3314925   .4347067    -0.76   0.446    -1.183502     .520517
                |
w#c.d#c.f05#c.x |
             1  |  -.0874416    .423375    -0.21   0.836    -.9172414    .7423582
                |
w#c.d#c.f06#c.x |
             1  |   .2483893   .4193698     0.59   0.554    -.5735604    1.070339
```

f02	.0159557	.254879	0.06	0.950	-.4835981	.5155094
f03	-.1853553	.2622365	-0.71	0.480	-.6993293	.3286188
f04	.0539743	.3137998	0.17	0.863	-.5610619	.6690106
f05	-.0386572	.3041224	-0.13	0.899	-.6347261	.5574118
f06	.4795352	.3138648	1.53	0.127	-.1356284	1.094699
c.f02#c.x	-.0749433	.2420313	-0.31	0.757	-.549316	.3994294
c.f03#c.x	.2054979	.2502045	0.82	0.411	-.284894	.6958897
c.f04#c.x	.2938375	.3180849	0.92	0.356	-.3295974	.9172724
c.f05#c.x	.4540765	.30813	1.47	0.141	-.1498472	1.058
c.f06#c.x	.1561729	.3322175	0.47	0.638	-.4949615	.8073073
d	-2.185559	.2805608	-7.79	0.000	-2.735448	-1.63567
x	.5048528	.2036075	2.48	0.013	.1057896	.9039161
c.d#c.x	.0588332	.2287561	0.26	0.797	-.3895205	.5071869
_cons	-.4500012	.2046174	-2.20	0.028	-.8510439	-.0489586

```
. margins, dydx(w) at(d = 1 f02 = 0 f03 = 0 f04 = 1 f05 = 0 f06 = 0) ///
>         subpop(if d == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   .0886639   .0326848     2.71   0.007     .0246029    .1527249
------------------------------------------------------------------------------

. margins, dydx(w) at(d = 1 f02 = 0 f03 = 0 f04 = 0 f05 = 1 f06 = 0) ///
>         subpop(if d == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   .1217999   .0355845     3.42   0.001     .0520556    .1915441
------------------------------------------------------------------------------

. margins, dydx(w) at(d = 1 f02 = 0 f03 = 0 f04 = 0 f05 = 0 f06 = 1) ///
>         subpop(if d == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   .1073639   .0371242     2.89   0.004     .0346018    .1801261
------------------------------------------------------------------------------
```

```
. * Callaway and Sant'Anna (2021, Journal of Econometrics)
.
. gen first_treat = 0

. replace first_treat = 2004 if d
(2,292 real changes made)

. csdid y x, ivar(id) time(year) gvar(first_treat)
.....
Difference-in-difference with Multiple Time Periods
Outcome model  :
Treatment model:
------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
g2004        |
 t_2001_2002 |   .0395675   .0423073     0.94   0.350    -.0433533    .1224884
 t_2002_2003 |   -.009403   .0413075    -0.23   0.820    -.0903642    .0715583
 t_2003_2004 |   .0278723   .0461201     0.60   0.546    -.0625215    .1182661
 t_2003_2005 |   .0610949   .0457254     1.34   0.182    -.0285252     .150715
 t_2003_2006 |   .0367104   .0448462     0.82   0.413    -.0511866    .1246074
------------------------------------------------------------------------------
Control: Never Treated

See Callaway and Sant'Anna (2020) for details
```

5. Staggered Interventions

- Now we have different treatment cohorts.
- First group (cohort) is $g = q$, last is $g = T$.
- No reversibility.
  - Potential outcome in the never treated state: $Y_t(\infty)$.
  - Potential outcome if first treatment period is g: $Y_t(g)$, $g = q, \ldots, T$.
- Assume a never treated group.
  - Easy to relax.

- For $g \in \{q, q+1, \ldots, T\}$ we are interested in the following ATTs:

$$\begin{aligned}\tau_{gr} &= E[Y_r(g) - Y_r(\infty)|D_g = 1] \\ &= E[Y_r(g)|D_g = 1] - E[Y_r(\infty)|D_g = 1],\ r = g, g+1, \ldots, T\end{aligned}$$

- Because $Y_r = Y_r(g)$ when $D_g = 1$, $E[Y_r(g)|D_g = 1]$ is always estimable:

$$\bar{Y}_{gr} = N_g^{-1} \sum_{i=1}^{N} D_{ig} \cdot Y_{ir} = N_g^{-1} \sum_{i=1}^{N} D_{ig} \cdot Y_{ir}(g) \xrightarrow{p} E[Y_r(g)|D_g = 1]$$

$$N_g = \sum_{i=1}^{N} D_{ig}$$

- To estimate $E[Y_r(\infty)|D_g = 1]$, we assume the following.

1. No anticipation: For cohorts $g = q, ..., T$,

$$Y_t(g) = Y_t(\infty), t = 1, \ldots, g - 1$$

    - Can be relaxed.

2. Conditional common or parallel trends: For a known, strictly increasing function $G(\cdot)$,

$$E[Y_{it}(\infty)|D_{iq},\ldots,D_{iT},\mathbf{X}_i] = G\left[\alpha + \sum_{g=q}^{T}\beta_g D_{ig} + \mathbf{X}_i\boldsymbol{\kappa} + \sum_{g=q}^{T}(D_{ig}\cdot\mathbf{X}_i)\boldsymbol{\xi}_g + \sum_{s=2}^{T}\gamma_s fs_t + \sum_{s=2}^{T}(fs_t\cdot\mathbf{X}_i)\boldsymbol{\pi}_s\right]$$

▸ No anticipation means we observe $Y_{it}(\infty)$ for eventually treated cohorts prior to intervention.

▸ Conditional PT means no interactions allowed among D_{ig} and fs_t.

▸ Linear case from before: $G(z) \equiv z$.

- To estimate τ_{gr}, we need to estimate

$$E[Y_r(\infty)|D_g = 1] = E\Big[G\Big(\alpha + \beta_g + \mathbf{X}_i\boldsymbol{\kappa} + \mathbf{X}_i\boldsymbol{\xi}_g + \gamma_r + \mathbf{X}_i\boldsymbol{\pi}_r\Big)\Big|D_g = 1\Big]$$

- Use an imputation approach.

1. Using $W_{it} = 0$ observations, estimate the parameters

$$\Big(\alpha, \beta_q, \ldots, \beta_T, \boldsymbol{\kappa}, \boldsymbol{\xi}_q, \ldots, \boldsymbol{\xi}_T, \gamma_2, \ldots, \gamma_T, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_T\Big)$$

using pooled quasi-MLE in the linear exponential family (LEF).

  - Bernoulli ($0 \leq Y_{it} \leq 1$), Poisson ($Y_{it} \geq 0$) attractive.

2. Impute $Y_{ir}(\infty)$ for $W_{ir} = 1$:

$$\hat{Y}_{igr}(\infty) \equiv G\Big(\hat{\alpha} + \hat{\beta}_g + \mathbf{X}_i\hat{\boldsymbol{\kappa}} + \mathbf{X}_i\hat{\boldsymbol{\xi}}_g + \hat{\gamma}_r + \mathbf{X}_i\hat{\boldsymbol{\pi}}_r\Big), r = g, \ldots, T$$

- Estimate of τ_{gr}: For $r = g, \ldots, T$,

$$
\begin{aligned}
\hat{\tau}_{gr} &= N_g^{-1} \sum_{i=1}^{N} D_{ig}\left[Y_{ir} - \hat{Y}_{igr}(\infty) \right] \\
&= \bar{Y}_{gr} - N_g^{-1} \sum_{i=1}^{N} D_{ig} G\left(\hat{\alpha} + \hat{\beta}_g + \mathbf{X}_i \hat{\boldsymbol{\kappa}} + \mathbf{X}_i \hat{\boldsymbol{\xi}}_g + \hat{\gamma}_r + \mathbf{X}_i \hat{\boldsymbol{\pi}}_r \right)
\end{aligned}
$$

  - Extension of linear imputation estimators.
  - Can obtain an analytical standard error or use bootstrap.
- Can apply to Tobit models, two-part models, and so on.
  - Replace the index $G(\cdot)$ with the conditional mean from the chosen model.

• With a canonical link in the LEF, same as estimating a conditional mean on the full sample with many interactions.

$$
\begin{aligned}
E(Y_{it}|D_{iq},\ldots,D_{iT},\mathbf{X}_i) = G\Bigg[& \alpha + \sum_{g=q}^{T} \beta_g D_{ig} + \mathbf{X}_i\boldsymbol{\kappa} + \sum_{g=q}^{T} (D_{ig} \cdot \mathbf{X}_i)\boldsymbol{\eta}_g \\
& + \sum_{s=2}^{T} \gamma_s fs_t + \sum_{s=2}^{T} (fs_t \cdot \mathbf{X}_i)\boldsymbol{\pi}_s \\
& + \sum_{g=q}^{T} \sum_{s=g}^{T} \delta_{gs} (W_{it} \cdot D_{ig} \cdot fs_t) \\
& + \sum_{g=q}^{T} \sum_{s=g}^{T} (W_{it} \cdot D_{ig} \cdot fs_t \cdot \mathbf{X}_i)\boldsymbol{\lambda}_{gs} \Bigg]
\end{aligned}
$$

- Partial effect still taken with respect to W_t.
- Interactions must include the D_{ig} to get the $\hat{\tau}_{gr}$ as average partial effects.
- Average over the different subsamples determined by the D_g and fr_t.
- Centering the covariates can make the parameters more interpretable, but has no effect on properly computed APEs.
  - In exponential case, useful for estimating percentage effects.
- Correctly done, `margins` in Stata provides proper standard errors.

```
. * Generated data, T = 6, three treated periods.
. * y a corner solution outcome.

. use did_staggered_6_corner, clear

. xtset id year

Panel variable: id (strongly balanced)
 Time variable: year, 2001 to 2006
         Delta: 1 unit

. sum dinf d4 d5 d6 if year == 2001

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        dinf |      1,000        .503    .5002412          0          1
          d4 |      1,000        .277    .4477404          0          1
          d5 |      1,000        .163    .3695505          0          1
          d6 |      1,000        .057    .2319586          0          1

. sum y

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
           y |      6,000    6.356815    30.53533          0   977.2437

. count if y == 0
  2,194
```

```
. * The sample ATTs:
.
. gen te_4i = y4 - yinf

. sum te_4i if d4 & f04

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       te_4i |        277    2.079871    11.41619  -31.82291   74.87006

. sum te_4i if d4 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       te_4i |        277    5.265336     47.0484  -474.6237   402.3652

. sum te_4i if d4 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       te_4i |        277    4.622355    19.60196    -37.414   168.5657
```

```
. gen te_5i = y5 - yinf

. sum te_5i if d5 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       te_5i |        163    4.392324    29.06417  -41.48495   273.8863

. sum te_5i if d5 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       te_5i |        163    5.179047    32.92998  -200.6624   179.0408

.
. gen te_6i = y6 - yinf

. sum te_6i if d6 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       te_6i |         57    11.63881    71.78052  -19.07541   529.1949

. * did_staggered_6_poisson.do shows the poisson regression with
. * covariates centered.
```

```
. poisson y i.w#c.d4#c.f04 i.w#c.d4#c.f05 i.w#c.d4#c.f06 ///
>         i.w#c.d5#c.f05 i.w#c.d5#c.f06 ///
>         i.w#c.d6#c.f06 ///
>         i.w#c.d4#c.f04#c.x i.w#c.d4#c.f05#c.x i.w#c.d4#c.f06#c.x ///
>         i.w#c.d5#c.f05#c.x i.w#c.d5#c.f06#c.x ///
>         i.w#c.d6#c.f06#c.x ///
>         f02 f03 f04 f05 f06 ///
>         c.f02#c.x c.f03#c.x c.f04#c.x c.f05#c.x c.f06#c.x ///
>         d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x, noomitted vce(cluster id)
note: you are responsible for interpretation of noncount dep. variable.

------------------------------------------------------------------------------------
                 |               Robust
               y | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-----------------+------------------------------------------------------------------
    w#c.d4#c.f04 |
               1 |   .4798855   .5537199     0.87   0.386    -.6053856    1.565157
                 |
    w#c.d4#c.f05 |
               1 |    1.26722   .5110232     2.48   0.013     .2656328    2.268807
                 |
    w#c.d4#c.f06 |
               1 |   .6365721   .5979153     1.06   0.287    -.5353202    1.808465
                 |
    w#c.d5#c.f05 |
               1 |   .3369498    1.14253     0.29   0.768    -1.902368    2.576267
                 |
    w#c.d5#c.f06 |
               1 |  -.0489807   .9182506    -0.05   0.957    -1.848719    1.750757
                 |
    w#c.d6#c.f06 |
               1 |   2.035604   1.182908     1.72   0.085     -.282853    4.354062
                 |
```

```
w#c.d4#c.f04#c.x |
              1  |  -.2052884   .4326644    -0.47   0.635    -1.053295    .6427182
                 |
w#c.d4#c.f05#c.x |
              1  |  -.1320868   .4592519    -0.29   0.774    -1.032204    .7680304
                 |
w#c.d4#c.f06#c.x |
              1  |   .2263718   .5356826     0.42   0.673    -.8235468     1.27629
                 |
w#c.d5#c.f05#c.x |
              1  |   .9717867   .9239081     1.05   0.293      -.83904    2.782613
                 |
w#c.d5#c.f06#c.x |
              1  |   1.201889   .7255394     1.66   0.098    -.2201419     2.62392
                 |
w#c.d6#c.f06#c.x |
              1  |  -.3027545   .6868041    -0.44   0.659    -1.648866    1.043357
                 |
```

f02	-.1661607	.5602147	-0.30	0.767	-1.264161	.93184
f03	.9632906	.4634932	2.08	0.038	.0548607	1.871721
f04	.543864	.4113719	1.32	0.186	-.26241	1.350138
f05	.4526241	.3643914	1.24	0.214	-.26157	1.166818
f06	.4978902	.3614676	1.38	0.168	-.2105733	1.206354
c.f02#c.x	.7084648	.5577776	1.27	0.204	-.3847592	1.801689
c.f03#c.x	-.361231	.4055552	-0.89	0.373	-1.156105	.4336426
c.f04#c.x	.2070722	.3604028	0.57	0.566	-.4993044	.9134488
c.f05#c.x	.1624938	.3205892	0.51	0.612	-.4658495	.7908371
c.f06#c.x	.2472584	.313374	0.79	0.430	-.3669434	.8614602
d4	-.4357018	.4858549	-0.90	0.370	-1.38796	.5165563
d5	-.3063094	.6184987	-0.50	0.620	-1.518545	.9059257
d6	-.9770025	.5265438	-1.86	0.064	-2.009009	.0550044
x	.3293786	.1887492	1.75	0.081	-.0405631	.6993202
c.d4#c.x	-.6559586	.4395996	-1.49	0.136	-1.517558	.2056407
c.d5#c.x	-.9547599	.5199871	-1.84	0.066	-1.973916	.0643961
c.d6#c.x	-.2140135	.4812769	-0.44	0.657	-1.157299	.7292719
_cons	1.202868	.229064	5.25	0.000	.7539108	1.651825

```
. margins, dydx(w) at(d4 = 1 d5 = 0 d6 = 0 f02 = 0 f03 = 0 f04 = 1 f05 = 0 f06 = 0) ///
>         subpop(if d4 == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   1.017501   1.033521     0.98   0.325    -1.008164    3.043166
------------------------------------------------------------------------------

. margins, dydx(w) at(d4 = 1 d5 = 0 d6 = 0 f02 = 0 f03 = 0 f04 = 0 f05 = 1 f06 = 0) ///
>         subpop(if d4 == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |    6.00713   2.162626     2.78   0.005      1.76846     10.2458
------------------------------------------------------------------------------

. margins, dydx(w) at(d4 = 1 d5 = 0 d6 = 0 f02 = 0 f03 = 0 f04 = 0 f05 = 0 f06 = 1) ///
>         subpop(if d4 == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   4.569667   1.369919     3.34   0.001     1.884675    7.254658
------------------------------------------------------------------------------
```

```
. margins, dydx(w) at(d4 = 0 d5 = 1 d6 = 0 f02 = 0 f03 = 0 f04 = 0 f05 = 1 f06 = 0) ///
>         subpop(if d5 == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   7.170127   3.355386     2.14   0.033     .5936913    13.74656
------------------------------------------------------------------------------

. margins, dydx(w) at(d4 = 0 d5 = 1 d6 = 0 f02 = 0 f03 = 0 f04 = 0 f05 = 0 f06 = 1) ///
>         subpop(if d5 == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   7.185492   2.781751     2.58   0.010      1.73336    12.63762
------------------------------------------------------------------------------

. margins, dydx(w) at(d4 = 0 d5 = 0 d6 = 1 f02 = 0 f03 = 0 f04 = 0 f05 = 0 f06 = 1) ///
>         subpop(if d6 == 1) noestimcheck vce(uncond)

------------------------------------------------------------------------------
             |            Unconditional
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         1.w |   13.73294   10.32555     1.33   0.184    -6.504777    33.97065
------------------------------------------------------------------------------
```

```
. * Imputation is the same (no standard errors):
.
. poisson y f02 f03 f04 f05 f06 ///
>         c.f02#c.x c.f03#c.x c.f04#c.x c.f05#c.x c.f06#c.x ///
>         d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x if ~w
note: you are responsible for interpretation of noncount dep. variable.

------------------------------------------------------------------------------
           y | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
         f02 |  -.1661604   .0602748    -2.76   0.006    -.2842969   -.0480239
         f03 |   .9632911   .0596965    16.14   0.000      .846288    1.080294
         f04 |   .5438644   .0602786     9.02   0.000     .4257205    .6620082
         f05 |   .4526244   .0628473     7.20   0.000      .329446    .5758028
         f06 |   .4978898   .0619736     8.03   0.000     .3764238    .6193557
             |
   c.f02#c.x |   .7084644   .0556699    12.73   0.000     .5993535    .8175753
             |
   c.f03#c.x |  -.3612316   .0586636    -6.16   0.000    -.4762102   -.2462529
             |
   c.f04#c.x |   .2070717   .0574569     3.60   0.000     .0944582    .3196852
             |
   c.f05#c.x |   .1624933   .0600223     2.71   0.007     .0448518    .2801349
             |
   c.f06#c.x |    .247259   .0592517     4.17   0.000     .1311277    .3633902
```

```
             |
          d4 |  -.4357019   .0682781    -6.38   0.000    -.5695245   -.3018793
          d5 |  -.3063131   .0773811    -3.96   0.000    -.4579772    -.154649
          d6 |  -.9770024   .1088114    -8.98   0.000    -1.190269   -.7637361
           x |    .329379   .0475337     6.93   0.000     .2362146    .4225435
             |
     c.d4#c.x |  -.6559584   .0624641   -10.50   0.000    -.7783858   -.5335309
             |
     c.d5#c.x |   -.954755   .0718444   -13.29   0.000    -1.095567   -.8139425
             |
     c.d6#c.x |  -.2140135   .0791491    -2.70   0.007    -.3691428   -.0588841
             |
       _cons |   1.202868   .0492831    24.41   0.000     1.106275    1.299461
------------------------------------------------------------------------------

. predict double yh
(option n assumed; predicted number of events)

. gen teyh = y - yh

. sum teyh if d4 & f04

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teyh |        277    1.017501    12.06767  -3.527664   81.15445
```

```
. sum teyh if d4 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teyh |        277     6.00713    37.30788  -3.129414   399.5018

. sum teyh if d4 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teyh |        277    4.569665    20.86058   -3.46407   175.1626

. sum teyh if d5 & f05

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teyh |        163    7.170124    43.97455  -3.200809   411.3746

. sum teyh if d5 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teyh |        163    7.185487    36.58664  -3.444643   305.0844

. sum teyh if d6 & f06

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teyh |         57    13.73293    78.90365  -4.725423   570.4094
```

```
. * Linear Model:

. reg y i.w#c.d4#c.f04 i.w#c.d4#c.f05 i.w#c.d4#c.f06 ///
>         i.w#c.d5#c.f05 i.w#c.d5#c.f06 ///
>         i.w#c.d6#c.f06 ///
>         i.w#c.d4#c.f04#c.x_dm4 i.w#c.d4#c.f05#c.x_dm4 i.w#c.d4#c.f06#c.x_dm4 ///
>         i.w#c.d5#c.f05#c.x_dm5 i.w#c.d5#c.f06#c.x_dm5 ///
>         i.w#c.d6#c.f06#c.x_dm6 ///
>         f02 f03 f04 f05 f06 ///
>         c.f02#c.x c.f03#c.x c.f04#c.x c.f05#c.x c.f06#c.x ///
>         d4 d5 d6 x c.d4#c.x c.d5#c.x c.d6#c.x, noomitted vce(cluster id)

------------------------------------------------------------------------------------------
                     |               Robust
                   y | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
---------------------+--------------------------------------------------------------------
        w#c.d4#c.f04 |
                  1  |  -.1455549   1.671318    -0.09   0.931    -3.425251    3.134142
                     |
        w#c.d4#c.f05 |
                  1  |   5.288825   2.478633     2.13   0.033     .4249014    10.15275
                     |
        w#c.d4#c.f06 |
                  1  |   2.816552   2.138228     1.32   0.188    -1.379381    7.012485
                     |
        w#c.d5#c.f05 |
                  1  |   6.697316   3.623153     1.85   0.065    -.4125477    13.80718
                     |
        w#c.d5#c.f06 |
                  1  |   5.562791   3.329713     1.67   0.095     -.971242    12.09682
                     |
        w#c.d6#c.f06 |
                  1  |    12.0696   10.58551     1.14   0.254    -8.702791      32.842
```

```
                          |
w#c.d4#c.f04#c.x_dm4 |
                       1  |  -1.534406   2.608612    -0.59   0.557    -6.653393    3.584581
                          |
w#c.d4#c.f05#c.x_dm4 |
                       1  |  -2.391267   3.602391    -0.66   0.507    -9.460388    4.677855
                          |
w#c.d4#c.f06#c.x_dm4 |
                       1  |  -.3870517   5.016994    -0.08   0.939    -10.23211    9.458004
                          |
w#c.d5#c.f05#c.x_dm5 |
                       1  |   6.001162   10.44713     0.57   0.566    -14.49967    26.50199
                          |
w#c.d5#c.f06#c.x_dm5 |
                       1  |   8.149612   7.926036     1.03   0.304    -7.403976     23.7032
                          |
w#c.d6#c.f06#c.x_dm6 |
                       1  |  -1.548522   9.214616    -0.17   0.867    -19.63074     16.5337
                          |
```

f02	-1.432423	3.368666	-0.43	0.671	-8.042895	5.178049
f03	4.808135	2.685276	1.79	0.074	-.461293	10.07756
f04	2.574639	2.6855	0.96	0.338	-2.695229	7.844507
f05	2.058054	2.334671	0.88	0.378	-2.523368	6.639476
f06	1.717605	3.16678	0.54	0.588	-4.496699	7.93191
c.f02#c.x	3.707984	3.898847	0.95	0.342	-3.942884	11.35885
c.f03#c.x	-2.24394	2.146963	-1.05	0.296	-6.457014	1.969134
c.f04#c.x	1.070393	2.568546	0.42	0.677	-3.969972	6.110759
c.f05#c.x	.7396578	2.404356	0.31	0.758	-3.97851	5.457825
c.f06#c.x	2.432101	3.621197	0.67	0.502	-4.673923	9.538126
d4	-.6732229	3.071511	-0.22	0.827	-6.700577	5.354131
d5	-.647893	3.189825	-0.20	0.839	-6.907419	5.611633
d6	-2.281338	2.931903	-0.78	0.437	-8.034733	3.472056
x	3.560941	2.190103	1.63	0.104	-.736789	7.858671
c.d4#c.x	-4.43147	3.546717	-1.25	0.212	-11.39134	2.528399
c.d5#c.x	-5.120498	3.631408	-1.41	0.159	-12.24656	2.005565
c.d6#c.x	-3.426569	3.452896	-0.99	0.321	-10.20233	3.349192
_cons	2.326338	1.7649	1.32	0.188	-1.136999	5.789674

6. Some Simulations

Common Intervention, Logit Mean

- $Y_{it}(0)$ a binary variable, generated to depend on heterogeneity, C_i.
- Logit mean is correctly specified.
- All serial correlation due to C_i.
- $P(D_i = 1) \approx 0.39$.
- $P[Y_{it}(0) = 1] \approx 0.38$; $P[Y_{it}(1) = 1] \approx 0.46$.
- $N = 1,000$, $T = 6$, 1,000 Monte Carlo Replications.

	Sample ATT	Logit (Pooled Bernoulli)		Linear (Pooled OLS)		CS (2021)	
$N = 1{,}000$	Mean	Mean	SD	Mean	SD	Mean	SD
τ_4	0.082	0.081	0.021	–0.043	0.029	–0.041	0.036
τ_5	0.120	0.119	0.025	–0.027	0.032	–0.025	0.039
τ_6	0.166	0.165	0.027	0.0005	0.032	0.0026	0.039

Staggered Intervention, Exponential Mean

- $Y_{it}(g)$ a count variable, generated to depend on heterogeneity, C_i.
- Exponential mean is correctly specified.
- Conditional distribution of $Y_{it}(g)$ is mixture of Poisson and lognormal.
- All serial correlation due to C_i.
- $N = 500$, $T = 6$, 1,000 Monte Carlo Replications.

	Sample ATT	Exponential (Pooled Poisson)		Linear (Pooled OLS)		CS (2021)	
$N = 500$	Mean	Mean	SD	Mean	SD	Mean	SD
τ_{44}	5.35	5.34	0.73	3.93	0.94	3.99	0.97
τ_{45}	13.20	13.21	1.61	10.31	1.94	10.37	1.87
τ_{46}	20.07	20.07	2.38	14.72	2.90	15.62	2.78
τ_{55}	12.02	12.03	1.66	9.67	2.01	10.50	1.93
τ_{56}	24.91	24.91	3.30	19.86	3.88	21.75	3.66
τ_{66}	4.72	4.76	1.15	–0.04	2.11	3.05	1.93

7. Extensions

- Without a never treated group, drop terms depending on D_{iT}, which is always unity.
- Under NA, all effects for $t < T$ are still ATTs with respect to $Y_t(\infty)$.
- For $t = T$, estimate with respect to $Y_t(T)$ under a modified PT assumption.
- Can test/correct for violations of PT by adding

$$D_{ig} \cdot t, \ldots, D_{iT} \cdot t$$

to any of the nonlinear models.

- Might even add

$$D_{ig} \cdot t \cdot \mathbf{X}_i, \ldots, D_{iT} \cdot t \cdot \mathbf{X}_i$$