Two-Way Fixed Effects, the Two-Way Mundlak Regression, and Difference-in-Differences Estimators

Jeff Wooldridge
Department of Economics
Michigan State University

Brown University
Department of Economics
April 19, 2022

1. Introduction

- For panel data, two-way fixed effects (TWFE) is a staple in empirical research.
  - Applied to "structural" models – say, production functions.
  - Applied to policy analysis – difference-in-differences.
- Used for all configurations of N and T.
- With small T, large N, time effects often absorbed into covariates.
  - Analyze as a one-way FE estimator.

- One-way Mundlak regression has proven useful for many purposes.

  ▸ Leads to simple, robust, regression-based comparisons between FE and RE estimation: Arellano (1993).

  ▸ Produces insight into the pre-testing problem with Hausman tests.

  ▸ Suggests how to allow heterogeneity to correlate with covariates in nonlinear models: Mundlak-Chamberlain device.

- Wooldridge (2019, J of E): The one-way Mundlak regression applies to unbalanced panels.
  - In the linear case, Mundlak still produces the complete-cases FE estimator.
  - Suggests correlated random effects approaches for heterogeneous slopes and nonlinear models.

- Current paper: Equivalence between the TWFE estimator and the obvious two-way Mundlak regression.
- Equivalence is simple but useful.
  - Further reveals the workings of TWFE.
  - Applications to staggered interventions and DiD.

- Advantages of TWFE for intervention analysis:

1. We know properties of TWFE when the panel is unbalanced.
2. It is easy to test the null that treatment effects are homogeneous in a robust way.
3. Immediate extensions of the TWFE etimator to removing unit-specific trends can be applied with heterogenous treatment effects.

- Advantages of POLS: Can be extended to nonlinear models.

2. Equivalence of TWFE and Two-Way Mundlak

- Motivation for TWFE estimation:

$$y_{it} = \mathbf{x}_{it}\boldsymbol{\beta} + c_i + f_t + u_{it}, \; t = 1, \ldots, T; i = 1, \ldots, N$$

  - $\mathbf{x}_{it}$ is $1 \times K$.
  - c_i are the unit-specific effects.
  - f_t are the time-specific effects.

- Equivalence results are algebraic.
- The two-way dummy variable regression (TWFE):

$$y_{it} \text{ on } \mathbf{x}_{it}, 1, c2_i, ..., cN_i, f2_t, ..., fT_t,\ \ t = 1, \ldots, T;\ i = 1, \ldots, N.$$

  - Coefficients on $\mathbf{x}_{it}$ are $\hat{\boldsymbol{\beta}}_{FE}$ $(K \times 1)$.
- $\mathbf{x}_{it}$ only includes variables that have some variation across i and t.

- Alternatively, consider the *two-way Mundlak regression*.

$$\bar{\mathbf{x}}_{i\cdot} \equiv T^{-1} \sum_{t=1}^{T} \mathbf{x}_{it}$$

$$\bar{\mathbf{x}}_{\cdot t} \equiv N^{-1} \sum_{i=1}^{N} \mathbf{x}_{it}$$

- Pooled OLS of

$$y_{it} \text{ on } 1, \mathbf{x}_{it}, \bar{\mathbf{x}}_{i\cdot}, \bar{\mathbf{x}}_{\cdot t},\ t = 1,\ldots,T;\ i = 1,\ldots,N.$$

  - Let $\hat{\boldsymbol{\beta}}_M$ be the coefficients $\mathbf{x}_{it}$.

THEOREM: Let $\ddot{\mathbf{x}}_{it} \equiv \mathbf{x}_{it} - \bar{\mathbf{x}}_{i\cdot} - \bar{\mathbf{x}}_{\cdot t} + \bar{\mathbf{x}}$. Provided the $K \times K$ matrix

$$\sum_{i=1}^{N} \sum_{t=1}^{T} \ddot{\mathbf{x}}_{it}' \ddot{\mathbf{x}}_{it}$$

is nonsingular,

$$\hat{\boldsymbol{\beta}}_M = \hat{\boldsymbol{\beta}}_{FE}$$

Moreover, in the extended regression

$$y_{it} \text{ on } 1, \mathbf{x}_{it}, \bar{\mathbf{x}}_{i\cdot}, \bar{\mathbf{x}}_{\cdot t}, \mathbf{z}_i, \mathbf{m}_t, t = 1, \ldots, T; i = 1, \ldots, N$$

for time-constant variables $\mathbf{z}_i$ and unit-constant variables $\mathbf{m}_t$, the coefficients on $\mathbf{x}_{it}$ are still $\hat{\boldsymbol{\beta}}_{FE}$. $\square$

- Proof uses Frisch-Waugh partialling out.
- Coefficients on $\bar{\mathbf{x}}_{i\cdot}$ and $\bar{\mathbf{x}}_{\cdot t}$ *do* change with the inclusion of $\mathbf{z}_i$, $\mathbf{m}_t$.
  - Basis of robust, regression-based Hausman tests.
- Suppose a regressor is an interaction of the form

$$x_{it} = z_i \cdot m_t$$

  - Then

$$\bar{x}_{i\cdot} = z_i \cdot \bar{m}, \;\; \bar{x}_{\cdot t} = \bar{z} m_t$$

  - Mundlak regression: Sufficient to include z_i, m_t as controls.

3. Application to Difference-in-Differences

Common Timing

- T time periods.
  - $t = 1,\ldots,q-1$ are control periods.
  - Intervention happens at $t = q$, remains in place.
- Treatment indicator:

$$w_{it} = d_i \cdot p_t = d_i \cdot (fq_t + \cdots + fT_t)$$

$$d_i = 1 \text{ if (eventually) treated}$$

$$p_t = fq_t + \cdots + fT_t = 1 \text{ if a post treatment period}$$

- Estimate a single treatment effect.
- Equation that motivates TWFE:

$$y_{it} = \tau w_{it} + c_i + g_t + u_{it}, t = 1, \ldots, T; i = 1, 2, \ldots, N$$

$$\bar{w}_{i\cdot} = d_i \bar{p}$$
$$\bar{w}_{\cdot t} = \bar{d} p_t$$

- TWM regression is equivalent to the DID (Mundlak) regression

$$y_{it} \text{ on } 1, w_{it}, d_i, p_t, t = 1, \ldots, T; i = 1, \ldots, N$$

- The TWFE estimator has the familiar form

$$\hat{\tau}_{FE} = \hat{\tau}_{DD} = \left(\bar{y}_1^{post} - \bar{y}_0^{post}\right) - \left(\bar{y}_1^{pre} - \bar{y}_0^{pre}\right)$$

- Enough to control for d_i; unit-specific fixed effects not needed.
- Using separate time dummies $f2_t$, ..., fT_t in place of p_t has no effect on $\hat{\tau}_{DD}$.
- Adding time-constant controls, $\mathbf{x}_i$ or $d_i \cdot \mathbf{x}_i$, has no effect on $\hat{\tau}_{DD}$.

- Allow TEs to change over treatment period, add covariates.
- Motivating equation:

$$
\begin{aligned}
y_{it} = {} & \tau_q(d_i \cdot fq_t) + \cdots + \tau_T(d_i \cdot fT_t) + [d_i \cdot fq_t \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_1)]\boldsymbol{\gamma}_q \\
& + \cdots + [d_i \cdot fT_t \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_1)]\boldsymbol{\gamma}_T \\
& + (fq_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_q + \cdots + (fT_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_T + c_i + g_t + u_{it}
\end{aligned}
$$

$$
\mathbf{x}_1 \equiv N_1^{-1} \sum_{t=1}^{N} d_i \mathbf{x}_i
$$

- Estimate by TWFE.

- Same estimates using TWM while including time-constant variables d_i, $\mathbf{x}_i$, $d_i \cdot \mathbf{x}_i$.
- Need time dummies for fq_t, ..., fT_t.
  - Harmless to include fs_t for $s = 2, \ldots, q-1$.

$$
\begin{aligned}
y_{it} = {} & \tau_q(w_{it} \cdot fq_t) + \cdots + \tau_T(w_{it} \cdot fT_t) \\
& + [w_{it} \cdot fq_t \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_1)]\boldsymbol{\gamma}_q + \cdots + [w_{it} \cdot fT_t \cdot (\mathbf{x}_i - \bar{\mathbf{x}}_1)]\boldsymbol{\gamma}_T \\
& + \alpha + \zeta d_i + \mathbf{x}_i\boldsymbol{\xi} + (d_i \cdot \mathbf{x}_i)\boldsymbol{\lambda} \\
& + \theta_2 f2_t + \cdots + \theta_T fT_t \\
& + (f2_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_2 + \cdots + (fT_t \cdot \mathbf{x}_i)\boldsymbol{\delta}_T + e_{it}
\end{aligned}
$$

- Pooled OLS regression

$$
\begin{aligned}
y_{it} \text{ on } & w_{it} \cdot d_i \cdot fq_t, \ldots, w_{it} \cdot d_i \cdot fT_t \\
& w_{it} \cdot d_i \cdot fq_t \cdot \dot{\mathbf{x}}_i, \ldots, w_{it} \cdot d_i \cdot fT_t \cdot \dot{\mathbf{x}}_i \\
& 1, d_i, \mathbf{x}_i, d_i \cdot \mathbf{x}_i, f2_t, \ldots, fT_t, f2_t \cdot \mathbf{x}_i, \ldots, fT_t \cdot \mathbf{x}_i,
\end{aligned}
$$

$$\dot{\mathbf{x}}_i = \mathbf{x}_i - \bar{\bar{\mathbf{x}}}_1$$

▸ Estimation is same without w_{it}.

▸ Introducing w_{it} is convenient for obtaining standard errors that account for sampling error in $\bar{\bar{\mathbf{x}}}_1$.

Staggered Interventions

- Assume:
  - First intervention at $1 < q \leq T$.
  - No reversibility.
  - Existence of a never treated group.
- Cohort dummies (exhaustive, mutually exclusive):

$$d_{iq}, d_{i,q+1}, \ldots, d_{iT}, d_{i\infty}$$

- TWFE estimation includes treatment indicators

$$d_{ig} \cdot fs_t,\ s \in \{g, \ldots, T\};\ g \in \{q, \ldots, T\}$$

- Estimated effects:

$$\hat{\tau}_{gs},\ s \in \{g, \ldots, T\};\ g \in \{q, \ldots, T\}$$

  - Time average of $d_{ig} \cdot fs_t$ is proportional to d_{ig}.
  - Cross-sectional average of $d_{ig} \cdot fs_t$ is proportional to fs_t,

$s \in \{g, \ldots, T\}$.

- With covariate interactions in TWFE

$$fs_t \cdot \mathbf{x}_i, d_{ig} \cdot fs_t \cdot \mathbf{x}_i$$

  - In Mundlak, need to include

$$\mathbf{x}_i, d_{iq} \cdot \mathbf{x}_i, ..., d_{iT} \cdot \mathbf{x}_i$$

- For TWM, it is enough to control for

$$d_{iq}, ..., d_{iT}, \mathbf{x}_i, d_{iq} \cdot \mathbf{x}_i, ..., d_{iT} \cdot \mathbf{x}_i$$

- The POLS and RE versions of Mundlak are also identical.

4. Identification of ATTs and Imputation Estimation

- Wooldridge (2005, REStat): TWFE with heterogeneous slopes and using intensity of treatment for heterogeneous effects.
- TWFE under recent scrutiny for staggered interventions.
- de Chaisemartin and D’Haultfœuille (2020, AER), Goodman-Bacon (2021, J of E), Callaway and Sant’Anna (2021, J of E), Sun and Abraham (2021, J of E).

- First intervention period is $t = q$.
- Subsequent treatment in each period after q, up to T.
  - Might have gaps.
- No reversibility.
- Initially, a never treated group.

- Define potential outcomes:

$$y_t(\infty)\text{: never treated state}$$

$$y_t(g),\, g \in \{q, q+1, \ldots, T\}\text{: first exposure in } g$$

- Define treatment cohorts by dummies: d_q, ..., d_T, d_∞.
  - $d_g = 1$ if unit first enters treatment in period g.
  - $d_\infty = 1$ indicates never treated.

- Define ATTs relative to the never treated state:

$$\tau_{gs} \equiv E[y_s(g) - y_s(\infty)|d_g = 1],\ g = q, \ldots, T;\ s = g, \ldots, T$$

  - For cohort g, can estimate ATTs for $s = g, g+1, \ldots, T$.

- Case without covariates is a special case.

Assumption CNAS (Conditional No Anticipation, Staggered):

For treatment cohorts $g = q,\ q+1,\ \ldots, T$,

$$E[y_t(g) - y_t(\infty)|d_g = 1, \mathbf{x}] = 0,\ t < g.\ \square$$

▸ No anticipatory treatment effects for each subpopulation defined by $\mathbf{x}$.

Assumption CCTS (Conditional Common Trends, Staggered):

For cohort indicators d_g and covariates $\mathbf{x}$,

$$E[y_t(\infty) - y_1(\infty)|d_q, \ldots, d_T, \mathbf{x}] = E[y_t(\infty) - y_1(\infty)|\mathbf{x}],\ t = 2, \ldots, T.\ \square$$

- Also assume conditional expectations are linear in **x**:

$$E[y_1(\infty)|d_q,\ldots,d_T,\mathbf{x}] = \alpha + \sum_{g=q}^{T}\beta_g d_g + \mathbf{x}\boldsymbol{\kappa} + \sum_{g=q}^{T}(d_g\cdot\mathbf{x})\boldsymbol{\eta}_g$$

  - ▸ Without covariates, above always holds.

- For $t \geq 2$:

$$E[y_t(\infty)|d_q,\ldots,d_T,\mathbf{x}] = \alpha + \sum_{g=q}^{T}\beta_g d_g + \mathbf{x}\boldsymbol{\kappa} + \sum_{g=q}^{T}(d_g\cdot\mathbf{x})\boldsymbol{\eta}_g + \gamma_t + \mathbf{x}\boldsymbol{\pi}_t$$

$$E[y_t(\infty)|d_q,\ldots,d_T,\mathbf{x}] - E[y_1(\infty)|d_q,\ldots,d_T,\mathbf{x}] = \gamma_t + \mathbf{x}\boldsymbol{\pi}_t$$

- Define time-varying treatment indicator:

$$\begin{aligned} w_{it} &= d_{iq} \bullet (fq_t + \cdots + fT_t) + \cdots + d_{iT} \bullet pT_t \\ &\equiv d_{iq} \bullet pq_t + \cdots + d_{iT} \bullet pT_t \end{aligned}$$

- For observable variables:

$$\begin{aligned} E(y_{it}|d_{iq},\ldots,d_{iT},\mathbf{x}_i,w_{it}=0) &= \alpha + \sum_{g=q}^{T} \beta_g d_{ig} + \mathbf{x}_i\boldsymbol{\kappa} + \sum_{g=q}^{T} (d_{ig} \bullet \mathbf{x}_i)\boldsymbol{\eta}_g \\ &+ \sum_{s=2}^{T} \gamma_s fs_t + \sum_{s=2}^{T} (fs_t \bullet \mathbf{x}_i)\boldsymbol{\pi}_s \end{aligned}$$

Imputation Procedure:

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

$$\left(\alpha, \beta_q, \ldots, \beta_T, \boldsymbol{\kappa}, \boldsymbol{\eta}_q, \ldots, \boldsymbol{\eta}_T, \gamma_2, \ldots, \gamma_T, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_T\right)$$

by pooled OLS. The explanatory variables are

$$\begin{gathered}1,\ d_{iq},\ \ldots,\ d_{iT},\ \mathbf{x}_i,\ d_{iq} \cdot \mathbf{x}_i,\ \ldots,\ d_{iT} \cdot \mathbf{x}_i, \\ f2_t,\ \ldots,\ fT_t,\ f2_t \cdot \mathbf{x}_i,\ \ldots,\ fT_t \cdot \mathbf{x}_i\end{gathered}$$

2. For cohort $g \in \{q, \ldots, T\}$, impute $y_{is}(\infty)$ for $w_{is} = 1$:

$$\hat{y}_{igs}(\infty) \equiv \hat{\alpha} + \hat{\beta}_g + \mathbf{x}_i \hat{\boldsymbol{\kappa}} + \mathbf{x}_i \hat{\boldsymbol{\eta}}_g + \hat{\gamma}_s + \mathbf{x}_i \hat{\boldsymbol{\pi}}_s, \; s = g, \ldots, T$$

3. For $s = g, \ldots, T$, obtain the imputation estimator of τ_{gs}:

$$
\begin{aligned}
\hat{\boldsymbol{\tau}}_{gs} &= N_g^{-1}\sum_{i=1}^{N} d_{ig}[y_{is} - \hat{y}_{igs}(\infty)] \\
&= \bar{y}_{gs} - N_g^{-1}\sum_{i=1}^{N} d_{ig}\Big(\hat{\alpha} + \hat{\beta}_g + \mathbf{x}_i\hat{\boldsymbol{\kappa}} + \mathbf{x}_i\hat{\boldsymbol{\eta}}_g + \hat{\gamma}_s + \mathbf{x}_i\hat{\boldsymbol{\pi}}_s\Big) \\
&= \bar{y}_{gs} - \Big(\hat{\alpha} + \hat{\beta}_g + \bar{\mathbf{x}}_g\hat{\boldsymbol{\kappa}} + \bar{\mathbf{x}}_g\hat{\boldsymbol{\eta}}_g + \hat{\gamma}_s + \bar{\mathbf{x}}_g\hat{\boldsymbol{\pi}}_s\Big) \\
\bar{\mathbf{x}}_g &\equiv N_1^{-1}\sum_{i=1}^{N} d_{ig}\mathbf{x}_i
\end{aligned}
$$

5. Equivalence Between POLS and Imputation

- Starting point: Act *as if*

$$
\begin{aligned}
E(y_{it}|d_{iq},\ldots,d_{iT},\mathbf{x}_i) = {} & \alpha + \sum_{g=q}^{T}\beta_g d_{ig} + \mathbf{x}_i\boldsymbol{\kappa} + \sum_{g=q}^{T}(d_{ig}\cdot\mathbf{x}_i)\boldsymbol{\eta}_g \\
& + \sum_{s=2}^{T}\gamma_s fs_t + \sum_{s=2}^{T}(fs_t\cdot\mathbf{x}_i)\boldsymbol{\pi}_s \\
& + \sum_{g=q}^{T}\sum_{s=g}^{T}\tau_{gs}(w_{it}\cdot d_{ig}\cdot fs_t) \\
& + \sum_{g=q}^{T}\sum_{s=g}^{T}(w_{it}\cdot d_{ig}\cdot fs_t\cdot\dot{\mathbf{x}}_{ig})\boldsymbol{\xi}_{gs}
\end{aligned}
$$

• Imputation:

(i) Set $w_{it} = 0$, runs pooled OLS to estimate parameters.

(ii) For the $w_{it} = 1$ subsample, obtain

$$\widehat{te}_{it} = y_{it} - \left[\hat{\alpha} + \sum_{g=q}^{T} \hat{\beta}_g d_{ig} + \mathbf{x}_i \hat{\boldsymbol{\kappa}} + \sum_{g=q}^{T} (d_{ig} \cdot \mathbf{x}_i) \hat{\boldsymbol{\eta}}_g \right. \\ \left. + \sum_{s=2}^{T} \hat{\gamma}_s fs_t + \sum_{s=2}^{T} (fs_t \cdot \mathbf{x}_i) \hat{\boldsymbol{\pi}}_s \right]$$

$$\hat{\tau}_{gs} = N_g^{-1} \sum_{i=1}^{N} d_{ig} \cdot \widehat{te}_{is}$$

• POLS obtains the $\tilde{\tau}_{gs}$ as the coefficients on $w_{it} \cdot d_{ig} \cdot fs_t$ in the regression using all data:

$$
\begin{aligned}
y_{it} \text{ on } & 1, d_{iq}, \ldots, d_{iT}, \mathbf{x}_i, d_{iq} \cdot \mathbf{x}_i, \ldots, d_{iT} \cdot \mathbf{x}_i, \\
& f2_t, \ldots, fT_t, f2_t \cdot \mathbf{x}_i, \ldots, fT_t \cdot \mathbf{x}_i, \\
& w_{it} \cdot d_{iq} \cdot fq_t, \ldots, w_{it} \cdot d_{iq} \cdot fT_t, \ldots, \\
& w_{it} \cdot d_{i,q+1} \cdot f(q+1)_t, \ldots, w_{it} \cdot d_{i,q+1} \cdot fT_t, \ldots, w_{it} \cdot d_{iT} \cdot fT_t, \\
& w_{it} \cdot d_{iq} \cdot fq_t \cdot \dot{\mathbf{x}}_{iq}, \ldots, w_{it} \cdot d_{iq} \cdot fT_t \cdot \dot{\mathbf{x}}_{iT}, \\
& w_{it} \cdot d_{i,q+1} \cdot f(q+1)_t \cdot \dot{\mathbf{x}}_{iq}, \ldots, w_{it} \cdot d_{i,q+1} \cdot fT_t \cdot \dot{\mathbf{x}}_{iT}, \ldots, \\
& w_{it} \cdot d_{iT} \cdot fT_t \cdot \dot{\mathbf{x}}_{iT}
\end{aligned}
$$

$$\dot{\mathbf{x}}_{ig} \equiv \mathbf{x}_i - \bar{\mathbf{x}}_g$$

- The appendix shows that

$$\tilde{\tau}_{gs} = \hat{\tau}_{gs}, g = q, \ldots, T; s = g, \ldots, T$$

  - Extension from regression adjustment in the cross-sectional case.
- Also, the estimates $\hat{\alpha}$, $\hat{\beta}_g$, $\hat{\boldsymbol{\kappa}}$, $\hat{\boldsymbol{\eta}}_g$, $\hat{\gamma}_s$, $\hat{\boldsymbol{\pi}}_s$ from the first step of the imputation method are the same as the POLS estimates.
- Not the same imputation as BJS (2021), who use fixed effects in the first step.
  - But the estimates with time-constant covariates are still identical.
  - BJS (2021) allows time-varying covariates.

- RE = POLS.
  - ▸ Under classical error components assumptions, RE = GLS.
  - ▸ Improving over POLS requires allowing more general patterns of serial correlation and maybe time-varying variances.
- Equivalently, use TWFE (dropping all time-constant variables).
- Imputation, POLS, RE and TWFE are all the same.
  - ▸ Problem with the standard analysis is with the model, not the estimator.

- Advantages of POLS/TWFE over Imputation:

1. Standard errors easy to obtain without resampling.
   - Including aggregation and linear combinations of $\hat{\tau}_{gs}$.
2. Obtain estimates of the $\boldsymbol{\xi}_{gs}$ – which give the "moderating" effects.

6. All Units Eventually Treated

- Regression approach (POLS/ETWFE) extends immediately if all units are treated by period T.
- Generally, the TEs are

$$y_s(g) - y_s(T),\ g = q,\ldots,T-1;\ s = g,\ldots,T$$

  - ▸ The gain in period s from first being treated in the earlier period g rather than the last period.

- The identified parameters are

$$\tau_{(g:T),s} \equiv E[y_s(g) - y_s(T)|d_g = 1],\ g = q, \ldots, T-1;\ s = g, \ldots, T$$

- The NA and CT assumptions are stated for the potential outcome $y_t(T)$.
- If there *could* have been a never treated group, under (the strong form of) NA

$$y_t(T) = y_t(\infty),\ t < T$$

$$\tau_{(g:T),s} = \tau_{(g:\infty),s},\ s < T$$

- Cannot estimate a TE for the $d_T = 1$ cohort.

7. Testing and Relaxing Parallel Trends

- Need at least two pre-treatment periods.
- Suppose $T = 3$, intervention at $t = 3$.
- Without covariates, run the regression

$$\Delta y_{i2} \text{ on } 1, d_i, i = 1, \ldots, N$$

  - ▸ Heteroskedasticity-robust t statistic on d_i.

- Two pooled OLS approaches yield the same statistic.

1. Run the regression

$$y_{it} \text{ on } 1, d_i, f2_t, d_i \cdot f2_t, f3_t, d_i \cdot f3_t, t = 1,2,3; i = 1,\ldots,N$$

and use the cluster-robust t statistic on $d_i \cdot f2_t$.

2. Run the regression

$$y_{it} \text{ on } 1, d_i, f2_t, d_i \cdot t, f3_t, d_i \cdot f3_t, t = 1,2,3; i = 1,\ldots,N$$

and use the cluster-robust t statistic on $d_i \cdot t$.

- Statistics are identical.
  - ▸ Coefficients on $d_i \cdot f3_t$ can be very different.

- When including $d_i \cdot t$, the coefficient on $d_i \cdot f3_t$ is a DiDiD estimator:

$$\begin{aligned}
\hat{\tau}_3 &= N_1^{-1} \sum_{i=1}^{N} d_i \cdot \Delta^2 y_{i3} - N_0^{-1} \sum_{i=1}^{N} (1 - d_i) \cdot \Delta^2 y_{i3} \\
&= [(\bar{y}_{3,treat} - \bar{y}_{2,treat}) - (\bar{y}_{2,treat} - \bar{y}_{1,treat})] \\
&\quad - [(\bar{y}_{3,control} - \bar{y}_{2,control}) - (\bar{y}_{2,control} - \bar{y}_{1,control})] \\
&= (\overline{\Delta y}_{3,treat} - \overline{\Delta y}_{3,control}) - (\overline{\Delta y}_{2,treat} - \overline{\Delta y}_{2,control})
\end{aligned}$$

- Testing strategies in the general case:

1. In the full POLS regression, add interactions $d_{ig} \cdot fs_t$ for $s < g$, do joint test: the event study regression.
   - Not a sensible correction.

2. In the full POLS regression, add heterogenous linear trends

$$d_{iq} \cdot t, \ldots, d_{iT} \cdot t$$

and use a joint test.
   - Works as a correction if the differences in trends are linear in t.

- The imputation equivalence result holds when adding heterogeneous trends.
  - So the test is identical to using only the $w_{it} = 0$ observations and doing a joint test on

$$d_{iq} \bullet t, ..., d_{iT} \bullet t$$

  - Implication: Provided fully heterogeneous TEs are allowed, the test for pre-trends is not contaminated by using the long regression.
  - Same property as the BJS (2021) test for pre-trends.

8. Small Simulation Study

- $N = 500$, $T = 6$, staggered entry at $q = 4$.
- One covariate. Common trends holds conditional on x.
- $R^2 = 0.241$.
- Approximate cohort shares:

$$\rho_\infty = 0.358,\ \rho_4 = 0.290,\ \rho_5 = 0.225,\ \rho_6 = 0.127$$

- 1,000 replications.
- CS = Callaway and Sant'Anna (2021, Journal of Econometrics).

	ATT	POLS		CS		Het. Trends	
$N = 500$	Mean	Mean	SD	Mean	SD	Mean	SD
τ_{44}	3.99	3.99	0.288	3.99	0.362	3.99	0.396
τ_{45}	4.19	4.19	0.289	4.20	0.367	4.20	0.513
τ_{46}	4.59	4.60	0.316	4.60	0.372	4.61	0.662
τ_{55}	3.03	3.03	0.326	3.03	0.446	3.02	0.423
τ_{56}	3.62	3.63	0.358	3.63	0.430	3.62	0.521
τ_{66}	2.05	2.04	0.474	2.04	0.644	2.05	0.546

- Rejection rate of event study test (9 df, 5% level): 0.061
- Rejection rate of common trends test (3 df, 5% level): 0.045

- $N = 500$, $T = 6$, staggered entry at $q = 4$.
- One covariate.
- Heterogeneous linear trends based on D_{ig}.
- $R^2 = 0.314$.
- Approximate cohort shares:

$$\rho_\infty = 0.358,\ \rho_4 = 0.290,\ \rho_5 = 0.225,\ \rho_6 = 0.127$$

- 1,000 replications.

	ATT	POLS		CS		Het. Trends		Event Study	
$N = 500$	Mean	Mean	SD	Mean	SD	Mean	SD	Mean	SD
τ_{44}	3.99	4.34	0.288	4.24	0.362	3.99	0.396	4.74	0.367
τ_{45}	4.19	4.82	0.289	4.70	0.367	4.20	0.513	5.19	0.366
τ_{46}	4.59	5.55	0.317	5.35	0.372	4.61	0.662	5.85	0.369
τ_{55}	3.03	3.34	0.326	3.20	0.446	3.02	0.423	3.70	0.409
τ_{56}	3.62	4.19	0.358	3.96	0.430	3.62	0.521	4.46	0.415
τ_{66}	2.05	2.43	0.474	2.16	0.644	2.05	0.546	2.69	0.551

- Rejection rate of event study test (9 df, 5% level): 0.266
- Rejection rate of common trends test (3 df, 5% level): 0.401

9. Concluding Remarks

- Equivalence between TWFE and TWM has applications to DiD estimators with staggered entry.
  - "Extended" TWFE allows for flexible treatment effects.
  - In leading cases, equivalent to imputation.
- TWFE has some resilience to unbalanced panels.
  - Allows sample selection to depend on unobserved heterogeneity.

- Can accomodate staggered exit.
  - Cohorts are now indexed by entry and exit date.
- The pooled OLS methods extend to nonlinear models and pooled QMLE.
  - Binary and Fractional: Logit mean, Bernoulli QMLE.
  - Nonnegative: Exponential mean, Poisson QMLE.