Two-Period Difference-in-Differences with Panel Data

Duke Kunshan University
Economics 101, Professor Luyao Zhang
The Economic Principle in Econometrics
February 15, 2022

Jeffrey M. Wooldridge
Department of Economics
Michigan State University

1. Introduction and Overview
2. The $T = 2$ Panel Data DiD Case
3. Adding Covariates: Linear Regression Adjustment

1. Introduction and Overview

- Difference-in-Differences methods are increasingly popular for analyzing interventions.
  - Currie, Kleven, and Zwiers (2020, AER P&P).
- Observing units over at least two time periods allows controlling for unobserved heterogeneity.
  - Can also control for observed heterogeneity (covariates).
- Here we study panel data settings.

- Standard DiD methods use common or parallel trends assumptions.
  - Adding covariates can help.
- Also need “no anticipation” assumption.
- Common intervention date is relatively easy to handle.
  - $T = 2$ is the traditional special case.

2. The $T = 2$ Panel Data DiD Case

- Two periods, $t \in \{1,2\}$.
  - $t = 1$ is the control period.
  - Some units are "treated" just prior to $t = 2$.
- Treatment indicator $D_i \in \{0,1\}$ means treated in the second time period.
- For each i, we have four potential outcomes:

$$[Y_{i1}(0), Y_{i1}(1), Y_{i2}(0), Y_{i2}(1)]$$

- $Y_{i1}(0),\ Y_{i1}(1)$ are potential outcomes before the treatment is assigned.

- Common assumption:

$$Y_{i1}(1) = Y_{i1}(0)$$

  - Heckman, Ichimura, Todd (1997); Abadie (2005).
  - A strong form of "no anticipation."
  - Can relax it somewhat.
- Assume random sampling across i.

- Parameter of interest is the average treatment effect on the treated (ATT) in second period:

$$\tau_{2,att} \equiv E[Y_2(1) - Y_2(0)|D = 1]$$

- Stronger assumptions are needed to estimate

$$\tau_{2,ate} \equiv E[Y_2(1) - Y_2(0)]$$

- Henceforth, $\tau_2 \equiv \tau_{2,att}$.
- Missing data problem: We do not observe both $Y_{i2}(1)$ and $Y_{i2}(0)$ for any i.

- The data we observe are $\{(D_i, Y_{i1}, Y_{i2}) : i = 1, \ldots, N\}$, where

$$\begin{aligned} Y_{i1} &= (1 - D_i)Y_{i1}(0) + D_i Y_{i1}(1) \\ Y_{i2} &= (1 - D_i)Y_{i2}(0) + D_i Y_{i2}(1) \end{aligned}$$

- Suppose we only observe Y in $t = 2$.
  - ▸ Without additional information, we are stuck with the simple difference in means:

$$\hat{\tau}_{2,SDM} = N_1^{-1} \sum_{i=1}^{N} D_i Y_{i2} - N_0^{-1} \sum_{i=1}^{N} (1 - D_i) Y_{i2}$$

$$N_1 = \sum_{i=1}^{N} D_i, \;\; N_0 = N - N_1 = \sum_{i=1}^{N} (1 - D_i)$$

- Generally,

$$\begin{aligned} \text{plim}(\hat{\tau}_{2,SDM}) &= E[Y_2(1)|D=1] - E[Y_2(0)|D=0] \\ &= \tau_2 + \{E[Y_2(0)|D=1] - E[Y_2(0)|D=0]\} \\ &= \tau_2 + \text{selection bias} \end{aligned}$$

- Consistency of $\hat{\tau}_{2,SDM}$ requires no selection bias:

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

▸ Sufficient is assignment is independent of the outcome in the control state.

- How can we exploit the $T = 2$ panel structure?
- The difference-in-differences estimator is

$$\begin{aligned}\hat{\tau}_{2,DD} &= N_1^{-1} \sum_{i=1}^{N} D_i \Delta Y_i - N_0^{-1} \sum_{i=1}^{N} (1 - D_i) \Delta Y_i \\ &= \overline{\Delta Y}_{treat} - \overline{\Delta Y}_{control}\end{aligned}$$

$$\Delta Y_i = Y_{i2} - Y_{i1}$$

- **Question**: When does $\hat{\tau}_{2,DD}$ consistently estimate τ_2?
- Write TE_2 in a more complicated way:

$$
\begin{aligned}
TE_2 &= Y_2(1) - Y_2(0) = [Y_2(1) - Y_1(1)] - [Y_2(0) - Y_1(0)] \\
&\quad + [Y_1(1) - Y_1(0)] \\
&\equiv \Delta Y(1) - \Delta Y(0) + TE_1
\end{aligned}
$$

$$\Delta Y(d) \equiv Y_2(d) - Y_1(d),\ d \in \{0, 1\}$$

$$TE_2 = \text{[change in treated state]} - \text{[change in untreated state]}$$
$$+ \text{ [TE in pre-treatment period]}$$

- Take expectation conditional on $D = 1$:

$$\tau_2 = E(TE_2|D = 1) = E[\Delta Y(1)|D = 1] - E[\Delta Y(0)|D = 1] + \tau_1$$

- More precisely,

$$\tau_{2,att} = E[\Delta Y(1)|D = 1] - E[\Delta Y(0)|D = 1] + \tau_{1,att}$$

Assumption NA (No Anticipation): With D the treatment indicator,

$$E[Y_1(1) - Y_1(0)|D = 1] = 0. \square$$

- $E[Y_1(1) - Y_1(0)|D = 1]$ is the ATT before the intervention.
- Units cannot change behavior in anticipation in ways that would affect pre-treatment potential outcomes.
- Sufficient is the strongest form of no anticipation:

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

- Under Assumption NA,

$$\tau_{2,att} = E[\Delta Y(1)|D = 1] - E[\Delta Y(0)|D = 1]$$

- We can always estimate $E[\Delta Y(1)|D = 1]$ because, for each i,

$$\Delta Y_i = \Delta Y_i(1) \text{ when } D_i = 1$$

- Use the average change over the treated units:

$$N_1^{-1} \sum_{i=1}^{N} D_i \Delta Y_i = N_1^{-1} \sum_{i=1}^{N} D_i \Delta Y_i(1) \overset{p}{\to} E[\Delta Y(1) | D = 1]$$

- The difficult term to estimate is

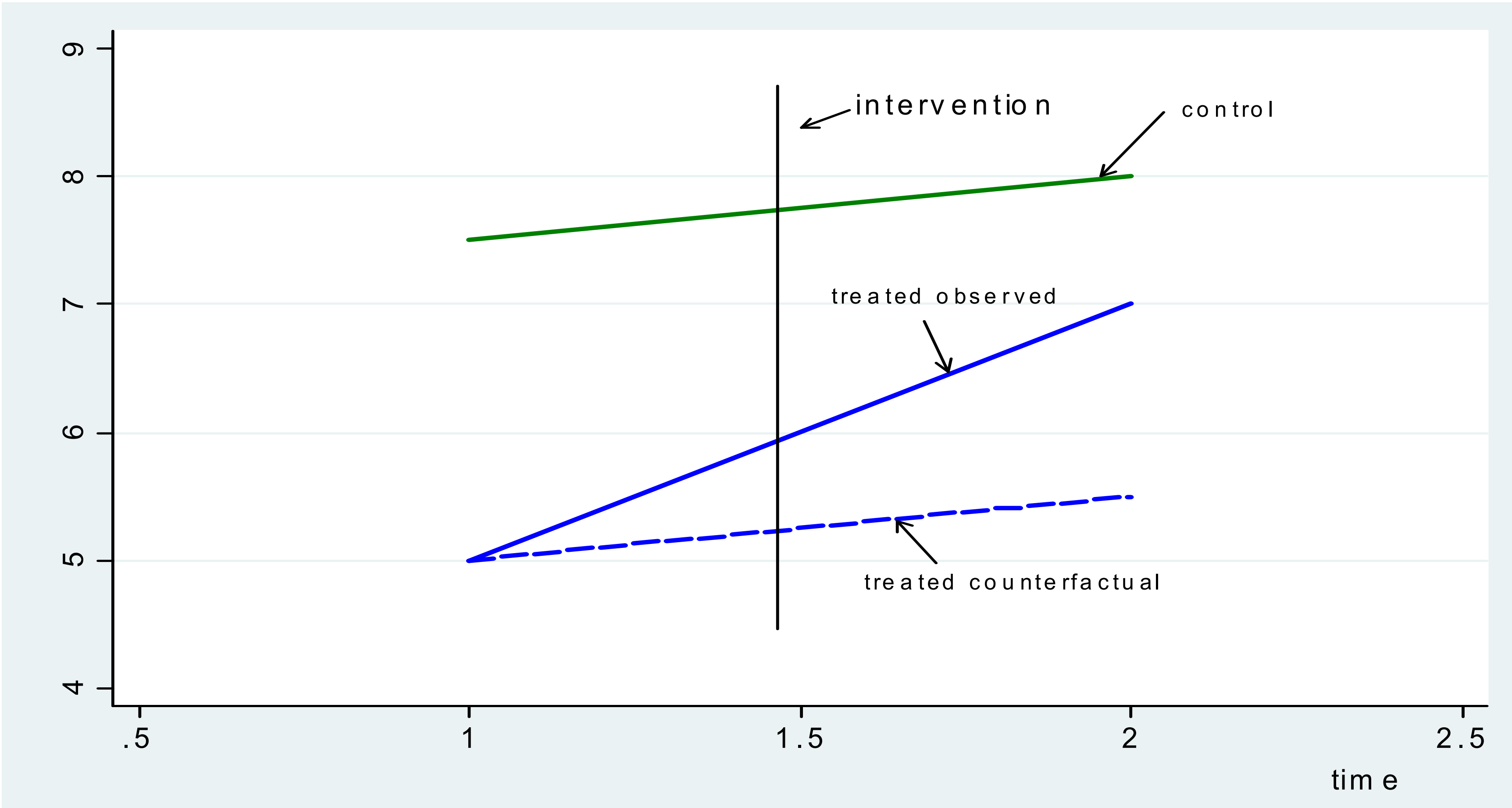
$$E[\Delta Y(0) | D = 1]$$

  - We do not observe $\Delta Y_i(0)$ when $D_i = 1$.

Assumption CT (Common or Parallel Trends): With D the treatment indicator,

$$E[\Delta Y(0)|D = 1] = E[\Delta Y(0)|D = 0]. \ \square$$

- In the untreated state, the average change is the same for treated and control.
  - ▸ No selection bias in the *trend* in the control state.
- Importantly, D can be correlated with $Y_1(0)$.
- D can be correlated with the change in the treated state, $\Delta Y(1) = Y_2(1) - Y_1(1)$.



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

- Under Assumption CT,

$$N_0^{-1}\sum_{i=1}^{N}(1-D_i)\Delta Y_i = N_0^{-1}\sum_{i=1}^{N}(1-D_i)\Delta Y_i(0) \xrightarrow{p} E[\Delta Y(0)|D=0]$$
$$= E[\Delta Y(0)|D=1]$$

- Conclusion: Under Assumptions NA and CT,

$$\text{plim}(\hat{\tau}_{2,DD}) = \text{plim}(\overline{\Delta Y}_{treat} - \overline{\Delta Y}_{control}) = \tau_{2,att}$$

Equivalent Methods for Obtaining $\hat{\tau}_{2,DD}$

1. Cross-Sectional OLS on Differences

- Run the regression

$$\Delta Y_i \text{ on } 1, D_i, \; i = 1, \ldots, N$$

and then $\hat{\tau}_{2,DD}$ is the coefficient on D_i.

  - ▸ Heteroskedasticity-robust inference.

- Contrast the difference-in-means regression for $t = 2$:

$$Y_{i2} \text{ on } 1, D_i, \; i = 1, \ldots, N$$

  - ▸ This produces $\hat{\tau}_{2,SDM}$.

2. Pooled OLS on Levels

- Define a second period dummy:

$$f2_t = 0 \text{ if } t = 1$$

$$f2_t = 1 \text{ if } t = 2$$

- The time-varying treatment indicator is

$$W_{it} = D_i \cdot f2_t$$

  - ▸ In both simple and complicated settings, useful to have W_{it}.

```
. * w = train

. list id year lwage train d d92 in 19/34, sep(2)

     +----------------------------------------+
     |  id   year      lwage   train   d   d92 |
     |----------------------------------------|
 19. | 166   1991   .9376874       0   1     0 |
 20. | 166   1992   1.432148       1   1     1 |
     |----------------------------------------|
 21. | 189   1991    .836248       0   0     0 |
 22. | 189   1992    1.46736       0   0     1 |
     |----------------------------------------|
 23. | 193   1991   2.493875       0   0     0 |
 24. | 193   1992   1.822662       0   0     1 |
     |----------------------------------------|
 25. | 209   1991   .7654077       0   0     0 |
 26. | 209   1992   1.978185       0   0     1 |
     |----------------------------------------|
 27. | 212   1991   2.179983       0   1     0 |
 28. | 212   1992   2.048526       1   1     1 |
     |----------------------------------------|
 29. | 218   1991   1.962152       0   0     0 |
 30. | 218   1992   2.275641       0   0     1 |
     |----------------------------------------|
 31. | 243   1991   1.683546       0   0     0 |
 32. | 243   1992   1.690503       0   0     1 |
     |----------------------------------------|
 33. | 259   1991   2.089011       0   1     0 |
 34. | 259   1992   2.147384       1   1     1 |
     |----------------------------------------|
```

- Can show $\hat{\tau}_{2,DD}$ is the coefficient on W_{it} in the regression

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

▸ Cluster standard errors to account for serial correlation and heteroskedasticity.

- More traditional is to use:

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

▸ In more complicated settings (adding covariates, nonlinear models), better to use W_{it}.

▸ Sometimes have to compute an average partial (marginal) effect with respect to W_{it}, not D_i.

3. Two-Way Fixed Effects

- Think of an equation for random draw i:

$$Y_{it} = \tau_2 W_{it} + \theta_2 f2_t + C_i + U_{it}, \; t = 1,2$$

- Estimate by fixed effects. Inclusion of $f2_t$ means TWFE.
- $\hat{\tau}_{2,DD}$ is the coefficient on W_{it}.
  - Cluster standard errors to account for serial correlation.
- Recall the TWFE dummy variable regression where $ch_i = 1[h = i]$:

$$Y_{it} \text{ on } W_{it}, c1_i, c2_i, ..., cN_i, f2_t, \; t = 1,2; \; i = 1,\ldots,N$$

- Equivalence of POLS and TWFE: It is enough to control for D_i rather than N unit dummies.

Stata Commands

```
xtset cid tid

reg D.y w, vce(robust)

reg y w d f2, vce(cluster cid)

xtreg y w f2, fe vce(cluster cid)
```

- Standard errors allow for heteroskedasticity and serial correlation.

3. Adding Covariates: Linear Regression Adjustment

- Common to include controls (covariates) in DiD settings.
  - Can help with failure of the CT Assumption.
- Current thinking: Only use pre-treatment controls.
  - Can ruin the CT assumption by conditioning on controls affected by intervention.
- Let $\mathbf{X}_i$ be a $1 \times K$ vector of pre-intervention covariates.

Adding Covariates in Levels has No Effect

- The POLS regressions

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

and

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

give *identical* estimates on W_{it}, $\hat{\tau}_2$.

- Follows from Wooldridge (2021, Working Paper).

```
reg y w d f2, vce(cluster id)
reg y w d f2 x1 ... xK, vce(cluster id)
```

- Equivalence makes sense.
- If we write

$$Y_{it} = \tau_2 W_{it} + \theta_2 f2_t + \mathbf{X}_i \boldsymbol{\beta} + C_i + U_{it},\ t = 1,2$$

then using FE eliminates $\mathbf{X}_i$.

- Moreover, if we use FD (equivalent to FE with $T = 2$),

$$\Delta Y_i = \tau_2 D_i + \theta_2 + \Delta U_i$$

Adding Covariates to Relax Common Trends

- How can we use time-constant covariates to adjust the basic DiD estimator?
- Recall under no anticipation:

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

- $E[\Delta Y(1)|D = 1]$ is estimable as before.
- $E[\Delta Y(0)|D = 1]$ is still the problem.
- Want to relax Assumption CT:

$$E[\Delta Y(0)|D = 1] = E[\Delta Y(0)|D = 0]$$

Assumption CCT (Conditional Common Trends): For D the treatment indicator and covariates $\mathbf{X}$,

$$E[Y_2(0) - Y_1(0)|D, \mathbf{X}] = E[Y_2(0) - Y_1(0)|\mathbf{X}]. \; \square$$

- CT holds within each subpopulation determined by values of $\mathbf{X}$.
  - Think of partitioning population by values $\mathbf{x} \in \mathcal{X}$.
- Assumption CCT simply says: D is unconfounded for $\Delta Y(0)$ conditional on $\mathbf{X}$.
  - Cross sectional case: D is unconfounded for $Y(0)$ conditional on $\mathbf{X}$.

- Recall how unconfoundedness implies identification:

$$
\begin{aligned}
E[\Delta Y(0)|D=1] &= E\{E[\Delta Y(0)|D=1,\mathbf{X}]|D=1\} \text{ (iterated expectations)} \\
&= E\{E[\Delta Y(0)|D=0,\mathbf{X}]|D=1\} \text{ (CCT)}
\end{aligned}
$$

- The function

$$
m_0(\mathbf{x}) \equiv E[\Delta Y(0)|D=0,\mathbf{X}=\mathbf{x}] = E[\Delta Y(0)|\mathbf{X}=\mathbf{x}]
$$

is estimable using data on $D_i = 0$ (control group) because

$$
\Delta Y_i = \Delta Y_i(0) \text{ when } D_i = 0
$$

  - Generally, can only estimate $m_0(\mathbf{x})$ for $\mathbf{x} \in \text{Support}(\mathbf{X}|D=0)$.

Assumption OV (Overlap): For all $\mathbf{x} \in \text{Support}(\mathbf{X})$,

$$p(\mathbf{x}) \equiv P(D = 1|\mathbf{X} = \mathbf{x}) > 0. \ \square$$

- Overlap means that

$$\text{Support}(\mathbf{X}|D = 1) \subset \text{Support}(\mathbf{X}|D = 0)$$

- CCT + Overlap means we nonparametrically identify

$$E[\Delta Y(0)|D = 1] = E[m_0(\mathbf{X})|D = 1]$$

- Natural starting point: Apply linear regression adjustment to the cross-sectional data set

$$\{(\Delta Y_i, D_i, \mathbf{X}_i), i = 1, \ldots, N\}$$

- Let $\hat{\alpha}_0, \hat{\boldsymbol{\beta}}_0$ be from

$$\Delta Y_i \text{ on } 1, \mathbf{X}_i \text{ using } D_i = 0$$

- The RA estimate of τ_2 is

$$\begin{aligned}\hat{\tau}_{2,RA} &= N_1^{-1} \sum_{i=1}^{N} D_i \Delta Y_i - N_1^{-1} \sum_{i=1}^{N} D_i \left(\hat{\alpha}_0 + \mathbf{X}_i \hat{\boldsymbol{\beta}}_0\right) \\ &= \overline{\Delta Y}_{treat} - \left(\hat{\alpha}_0 + \bar{\mathbf{X}}_1 \hat{\boldsymbol{\beta}}_0\right)\end{aligned}$$

$$\bar{\mathbf{X}}_1 \equiv N_1^{-1} \sum_{i=1}^{N} \mathbf{X}_i$$

▸ Interpretation as imputation estimator.

- Same as the coefficient on D_i in the pooled regression

$$\Delta Y_i \text{ on } 1, D_i, \mathbf{X}_i, D_i \cdot (\mathbf{X}_i - \bar{\mathbf{X}}_1), i = 1, \ldots, N$$

  - Need to center around $\mathbf{X}_i$ around $\bar{\mathbf{X}}_1$.
- Some interest in coefficients on the interactions, $D_i \cdot (\mathbf{X}_i - \bar{\mathbf{X}}_1)$.
  - Moderating effects of the X_j.
- Słoczyński (2021, REStat): Bias in estimating τ_{ate} when dropping $D_i \cdot (\mathbf{X}_i - \bar{\mathbf{X}})$ when $P(D_i = 1)$ is far from 1/2.

- Also same as running separate regressions:

$$\Delta Y_i \text{ on } 1, \mathbf{X}_i \text{ with } D_i = 0 \ \ (\hat{\alpha}_0, \hat{\boldsymbol{\beta}}_0)$$
$$\Delta Y_i \text{ on } 1, \mathbf{X}_i \text{ with } D_i = 1 \ \ (\hat{\alpha}_1, \hat{\boldsymbol{\beta}}_1)$$

$$\hat{\tau}_{2,RA} = (\hat{\alpha}_1 - \hat{\alpha}_0) + \bar{\mathbf{X}}_1\left(\hat{\boldsymbol{\beta}}_1 - \hat{\boldsymbol{\beta}}_0\right)$$

because, by the algebra of OLS,

$$\overline{\Delta Y}_{treat} = \hat{\alpha}_1 + \bar{\mathbf{X}}_1\hat{\boldsymbol{\beta}}_1$$

- $\hat{\tau}_{2,RA}$ is also the coefficient on $W_{it} = D_i \cdot f2_t$ in the POLS regression

$$Y_{it} \text{ on } 1, W_{it}, W_{it} \cdot (\mathbf{X}_i - \bar{\mathbf{X}}_1), D_i, f2_t, \mathbf{X}_i, D_i \cdot \mathbf{X}_i, f2_t \cdot \mathbf{X}_i$$

across $t = 1,2; i = 1,\ldots,N$.

- Fixed effects applied to

$$Y_{it} = \tau_2 W_{it} + W_{it} \cdot (\mathbf{X}_i - \bar{\mathbf{X}}_1)\boldsymbol{\delta}_2 + \theta_2 f2_t + (f2_t \cdot \mathbf{X}_i)\boldsymbol{\xi}_2 + C_i + U_{it},\ t = 1,2$$

also gives $\hat{\tau}_{2,RA}$ as coefficient on W_{it}.

  - Extended TWFE.

- The basic TWFE analysis drops

$$W_{it} \cdot (\mathbf{X}_i - \bar{\mathbf{X}}_1),\ f2_t \cdot \mathbf{X}_i$$

- Equivalence of POLS and ETWFE: Special case of Wooldridge (2021, working paper).

• Robust or cluster-robsust standard errors do not account for sampling error in $\bar{\mathbf{X}}_1$.

• In Stata, at least two possibilities.

1. Use `teffects ra` with ΔY_i as the outcome variable.

2. Use POLS, do *not* center $\mathbf{X}_i$ about $\bar{\mathbf{X}}_1$, and use `vce(uncond)` with `margins`.

  ▸ Useful in more complicated settings.

• Adjusting for sampling error in $\bar{\mathbf{X}}_1$ often has practically small effects.

- Stata commands:

```
sum x1 if d
gen x1_dm = x1 - r(mean)
sum x2 if d
gen x2_dm = x2 - r(mean)
...
sum xK if d
gen xK_dm = xK - r(mean)
```

- Get the ATT directly (not quite the correct standard errors):

```
reg d_y i.d x1 ... xK i.d#c.x1_dm ...
   i.d#c.xK_dm, vce(robust)
reg y i.w i.w#x1_dm ... i.w#xK_dm
   d f2 x1 ... xK i.d#x1 ... i.d#xK
   i.f2#x1 ... i.f2#xK, vce(cluster cid)
xtreg y i.w i.w#c.x1_dm ... i.w#c.xK_dm
   i.f2 i.f2#c.x1 ... i.f2#c.xK,
   fe vce(cluster cid)
```

- Job training data set; use only 1991 (pre) and 1992 (post).
- Outcome variable is *lwage* $= \log(wage)$.

```
. use wagepan_did, replace

. xtset id year

Panel variable: id (strongly balanced)
 Time variable: year, 1990 to 1992
         Delta: 1 unit

. sort id year

. * Three years, 1990, 1991, 1992. 1992 is the single post treatment period.
. * Here, two-period DID.
.
. drop if year == 1990
(545 observations deleted)

. egen d = sum(train), by(id)
```

```
. list id year lwage train d d92 in 19/34, sep(2)

     +----------------------------------------------+
     |  id   year       lwage   train   d   d92 |
     |----------------------------------------------|
 19. | 166   1991    .9376874       0   1     0 |
 20. | 166   1992    1.432148       1   1     1 |
     |----------------------------------------------|
 21. | 189   1991     .836248       0   0     0 |
 22. | 189   1992     1.46736       0   0     1 |
     |----------------------------------------------|
 23. | 193   1991    2.493875       0   0     0 |
 24. | 193   1992    1.822662       0   0     1 |
     |----------------------------------------------|
 25. | 209   1991    .7654077       0   0     0 |
 26. | 209   1992    1.978185       0   0     1 |
     |----------------------------------------------|
 27. | 212   1991    2.179983       0   1     0 |
 28. | 212   1992    2.048526       1   1     1 |
     |----------------------------------------------|
 29. | 218   1991    1.962152       0   0     0 |
 30. | 218   1992    2.275641       0   0     1 |
     |----------------------------------------------|
 31. | 243   1991    1.683546       0   0     0 |
 32. | 243   1992    1.690503       0   0     1 |
     |----------------------------------------------|
 33. | 259   1991    2.089011       0   1     0 |
 34. | 259   1992    2.147384       1   1     1 |
     |----------------------------------------------|
```

```
. * Basic DiD. Estimates are all the same:
.
. reg D.lwage train, vce(robust)

Linear regression                               Number of obs     =        545
                                                F(1, 543)         =       4.09
                                                Prob > F          =     0.0435
                                                R-squared         =     0.0085
                                                Root MSE          =     .45454

------------------------------------------------------------------------------
             |               Robust
     D.lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       train |   .0960654     .04748     2.02   0.044     .0027985    .1893324
       _cons |   .0599069   .0217649     2.75   0.006     .0171532    .1026605
------------------------------------------------------------------------------

. reg D.(lwage train d92), nocons vce(robust)

------------------------------------------------------------------------------
             |               Robust
     D.lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       train |
         D1. |   .0960654     .04748     2.02   0.044     .0027985    .1893324
             |
         d92 |
         D1. |   .0599069   .0217649     2.75   0.006     .0171532    .1026605
------------------------------------------------------------------------------
```

```
. * Pooled OLS:

. reg lwage train d d92, vce(cluster id)

Linear regression                               Number of obs     =      1,090
                                                F(3, 544)         =      20.23
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0547
                                                Root MSE          =     .50746

                                  (Std. err. adjusted for 545 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       train |   .0960654   .0475018     2.02   0.044      .002756    .1893748
           d |   .2091146     .05122     4.08   0.000     .1085014    .3097278
         d92 |   .0599069   .0217749     2.75   0.006     .0171338      .10268
       _cons |    1.45915   .0260405    56.03   0.000     1.407997    1.510302
------------------------------------------------------------------------------
```

```
. * Fixed Effects:

. xtreg lwage train d92, fe vce(cluster id)

Fixed-effects (within) regression               Number of obs     =      1,090
Group variable: id                              Number of groups  =        545

R-squared:                                      Obs per group:
     Within  = 0.0415                                         min =          2
     Between = 0.0580                                         avg =        2.0
     Overall = 0.0290                                         max =          2

                                                F(2,544)          =      10.62
corr(u_i, Xb) = 0.0803                          Prob > F          =     0.0000

                                   (Std. err. adjusted for 545 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       train |   .0960654   .0474799     2.02   0.044      .002799    .1893319
         d92 |   .0599069   .0217648     2.75   0.006     .0171535    .1026603
       _cons |   1.512867   .0097352   155.40   0.000     1.493744     1.53199
-------------+----------------------------------------------------------------
     sigma_u |  .46243458
     sigma_e |   .3214084
         rho |   .6742751   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```

```
. * Adding time constant controls, educ and exper (in 1990) to the pooled
. * regression does nothing except for changing the degrees-of-freedom used
. * to construct standard errors. And the R-squared increases a lot:
.
. reg lwage train d d92 educ exper, vce(cluster id)

Linear regression                               Number of obs     =      1,090
                                                F(5, 544)         =      28.48
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1361
                                                Root MSE          =     .48557

                                   (Std. err. adjusted for 545 clusters in id)
------------------------------------------------------------------------------
             |               Robust
       lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
       train |   .0960654   .0475456     2.02   0.044       .00267    .1894609
           d |   .2048787   .0489374     4.19   0.000     .1087493    .3010081
         d92 |   .0599069   .0217949     2.75   0.006     .0170943    .1027194
        educ |    .104157   .0127503     8.17   0.000     .0791112    .1292028
       exper |   .0568412   .0142865     3.98   0.000     .0287778    .0849046
       _cons |   .0632674   .1845499     0.34   0.732    -.2992502    .4257851
------------------------------------------------------------------------------
```

```
. * Full regression adjustment. Demean educ and exper using treated subsample:
.
. qui sum educ if d
. gen educ_dm = educ - r(mean)
. qui sum exper if d
. gen exper_dm = exper - r(mean)

. reg D.lwage train c.train#c.educ_dm c.train#c.exper_dm ///
>         d92 educ exper, nocons vce(robust)

Linear regression                               Number of obs     =        545
                                                F(6, 539)         =       4.36
                                                Prob > F          =     0.0003
                                                R-squared         =     0.0461
                                                Root MSE          =     .45514

------------------------------------------------------------------------------------
                   |               Robust
           D.lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------------+----------------------------------------------------------------
             train |    .095888   .0472991     2.03   0.043     .0029748    .1888012
                   |
  c.train#c.educ_dm |  -.0281286   .0364286    -0.77   0.440    -.0996881    .0434309
                   |
c.train#c.exper_dm |   -.044728   .0355761    -1.26   0.209    -.1146128    .0251568
                   |
               d92 |  -.0245192   .2273547    -0.11   0.914     -.471129    .4220906
              educ |    .007221   .0150001     0.48   0.630    -.0222448    .0366868
             exper |  -.0001651    .017045    -0.01   0.992     -.033648    .0333177
------------------------------------------------------------------------------------
```

```
. * POLS version:

. reg lwage train c.train#c.educ_dm c.train#c.exper_dm ///
>         d d92 educ exper c.d#c.educ c.d#c.exper c.d92#c.educ c.d92#c.exper, ///
>         vce(cluster id)

Linear regression                               Number of obs     =      1,090
                                                F(11, 544)        =      14.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1383
                                                Root MSE          =      .4863

                                   (Std. err. adjusted for 545 clusters in id)
------------------------------------------------------------------------------
                  |               Robust
            lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
------------------+-----------------------------------------------------------
            train |    .095888   .0473209     2.03   0.043      .002934    .188842
                  |
c.train#c.educ_dm |  -.0281286   .0364454    -0.77   0.441    -.0997195   .0434623
                  |
c.train#c.exper_dm |  -.044728   .0355924    -1.26   0.209    -.1146435   .0251875
                  |
                d |  -.4698976   .5744642    -0.82   0.414    -1.598337   .6585423
              d92 |  -.0245192   .2274591    -0.11   0.914     -.471325   .4222866
             educ |   .0954815   .0163737     5.83   0.000      .063318    .127645
            exper |   .0545992   .0186165     2.93   0.004     .0180302   .0911682
```

```
                |
    c.d#c.educ  |   .0504608   .0408213     1.24   0.217    -.0297259    .1306476
                |
   c.d#c.exper  |   .0264654   .0369413     0.72   0.474    -.0460997    .0990305
                |
  c.d92#c.educ  |    .007221    .015007     0.48   0.631    -.0222577    .0366998
                |
 c.d92#c.exper  |  -.0001651   .0170529    -0.01   0.992    -.0336627    .0333324
                |
          _cons |   .1720381   .2396264     0.72   0.473    -.2986682    .6427445
---------------------------------------------------------------------------------
```

```
. xtreg lwage train c.train#c.educ_dm c.train#c.exper_dm ///
>         d92 c.d92#c.educ c.d92#c.exper, fe vce(cluster id)

Fixed-effects (within) regression               Number of obs     =      1,090
Group variable: id                              Number of groups  =        545

                                     (Std. err. adjusted for 545 clusters in id)
------------------------------------------------------------------------------
                 |               Robust
           lwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-----------------+------------------------------------------------------------
           train |    .095888   .0472115     2.03   0.043     .0031488    .1886272
                 |
c.train#c.educ_dm |  -.0281286   .0363611    -0.77   0.440     -.099554    .0432968
                 |
c.train#c.exper_dm |  -.044728   .0355102    -1.26   0.208    -.1144819    .0250259
                 |
             d92 |  -.0245192   .2269334    -0.11   0.914    -.4702924     .421254
                 |
     c.d92#c.educ |    .007221   .0149723     0.48   0.630    -.0221896    .0366316
                 |
    c.d92#c.exper |  -.0001651   .0170135    -0.01   0.992    -.0335853     .033255
                 |
           _cons |   1.512867     .00973   155.49   0.000     1.493754     1.53198
-----------------+------------------------------------------------------------
         sigma_u |  .46266869
         sigma_e |   .3218322
             rho |   .6739185   (fraction of variance due to u_i)
------------------------------------------------------------------------------
```