The FE model, Difference-in-differences model, Linear probability model

Empirical Methods, Lecture 8

Tunga Kantarcı, FEB, Groningen University, Spring 2025

FE model

Consider the linear model

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \iota\alpha_i + \varepsilon_{it}.$$

y_{it}: an observation for individual i at time t.

$\boldsymbol{x}'_{it}$: K observations for K regressors for individual i at time t. $1 \times K$.

$\boldsymbol{\beta}$: vector of true coefficients. $K \times 1$.

ι: scalar with a value of 1. Greek letter 'iota'.

α_i: time invariant constant term specific to individual i in the panel. Potentially correlated with $\boldsymbol{x}'_{it}$. The model allows this. It captures individual heterogeneity.

ε_{it}: error term. It meets the OLS assumptions.

FE model

There are T observations available for each i. If we stack the T observations, we obtain

$$\mathbf{y}_i = \mathbf{X}_i'\boldsymbol{\beta} + \boldsymbol{\iota}\alpha_i + \boldsymbol{\varepsilon}_i.$$

$\mathbf{y}_i$: $T \times 1$.

$\mathbf{X}_i'$: T observations for i for K independent variables. $T \times K$.

$\mathbf{x}_{it}'$: row vector in row t of $\mathbf{X}_i'$. $1 \times K$. It contains k observations for k regressors for individual i at time t.

$\boldsymbol{\iota}$: column vector containing 1 in every row. $T \times 1$.

$\boldsymbol{\varepsilon}_i$: $T \times 1$.

FE model

There are N individuals. If we stack the N individuals, we obtain

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}.$$

$\mathbf{y}$: $NT \times 1$.

$\mathbf{X}$: $NT \times K$.

$\mathbf{D}$: has N diagonal elements. Each element of the diagonal is a vector, is the same, and is given by the column vector $\boldsymbol{\iota}$. All of the off-diagonal elements are $\mathbf{0}$ column vectors of size $T \times 1$. Hence, $\mathbf{D}$ is $NT \times N$.

$\boldsymbol{\alpha}$: $N \times 1$ since there are N different α_is.

This is the Least Squares Dummy Variable (LSDV) model.

FE model

For individual i, $T = 3$, $K = 3$,

$$\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix} = \begin{bmatrix} k_{i1} & l_{i1} & m_{i1} \\ k_{i2} & l_{i2} & m_{i2} \\ k_{i3} & l_{i3} & m_{i3} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \alpha_i + \begin{bmatrix} \varepsilon_i \\ \varepsilon_i \\ \varepsilon_i \end{bmatrix}$$

where k, l, m represent three different regressors. Putting them into row vector $\mathbf{x}'_{it}$,

$$\underbrace{\begin{bmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \end{bmatrix}}_{\mathbf{y}_i} = \underbrace{\begin{bmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ \mathbf{x}'_{i3} \end{bmatrix}}_{\mathbf{X}'_i} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\boldsymbol{\beta}} + \underbrace{\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\boldsymbol{\iota}} \alpha_i + \underbrace{\begin{bmatrix} \varepsilon_i \\ \varepsilon_i \\ \varepsilon_i \end{bmatrix}}_{\boldsymbol{\varepsilon}_i}$$

FE model

Assume $N = 3$. Stack N individuals to obtain

$$\underbrace{\begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{bmatrix}}_{\substack{\mathbf{y} \\ NT \times 1}} = \underbrace{\begin{bmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \\ \mathbf{X}_3' \end{bmatrix}}_{\substack{\mathbf{X} \\ NT \times K}} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}}_{\substack{\boldsymbol{\beta} \\ K \times 1}} + \underbrace{\begin{bmatrix} \boldsymbol{\iota} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\iota} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{\iota} \end{bmatrix}}_{\substack{\boldsymbol{D} \\ NT \times N}} \underbrace{\begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix}}_{\substack{\boldsymbol{\alpha} \\ N \times 1}} + \underbrace{\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}}_{\substack{\boldsymbol{\varepsilon} \\ N \times 1}}$$

where

$$\boldsymbol{D} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

LSDV model

As we control for time individual intercepts across all units of the data, the variation that is left to exploit for estimation is within individuals. Therefore, we interpret the estimation as exploiting within individual variation over time. This is also the same in the FE model we study below.

LSDV model

The LSDV model has two problems. First, it requires the inversion of a very large matrix due to $\boldsymbol{D}$. Second, it requires estimation of a large number of intercept terms contained in $\boldsymbol{\alpha}$. Could we avoid these problems?

FE model

Consider again the panel model for individual i at time t

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \iota\alpha_i + \varepsilon_{it}.$$

Take the average over all t for individual i to obtain

$$\bar{y}_i = \bar{\boldsymbol{x}}'_i\boldsymbol{\beta} + \iota\alpha_i + \bar{\varepsilon}_i,$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}.$$

Subtract the second equation from the first to obtain

$$y_{it} - \bar{y}_i = (\boldsymbol{x}'_{it} - \bar{\boldsymbol{x}}'_i)\boldsymbol{\beta} + \varepsilon_{it} - \bar{\varepsilon}_i.$$

The is the fixed effects transformation. The time invariant individual specific constant term α_i drops!

FE model

We have carried out the fixed effects transformation for individual i using his T observations. We need to consider the fact that we have n individuals in the panel data.

FE model

The panel model for N individuals described above is

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{D}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}.$$

Consider the model

$$\mathbf{y} = \mathbf{M}_D\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\upsilon}.$$

where

$$\mathbf{M}_D = \mathbf{I} - \mathbf{P}_D.$$

and

$$\mathbf{P}_D = \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'.$$

Apply the FWL theorem. The theorem states that the OLS estimator of $\boldsymbol{\beta}$ in the stated two models is the same and given by

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{OLS} &= ((\mathbf{M}_D\mathbf{X})'(\mathbf{M}_D\mathbf{X}))^{-1}(\mathbf{M}_D\mathbf{X})'\mathbf{y} = (\mathbf{X}'\mathbf{M}_D\mathbf{X})^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y} \\ &= \hat{\boldsymbol{\beta}}_{FE}.\end{aligned}$$

This is the fixed effects (FE) estimator.

FE model

Why does $\mathbf{M}_D$ demeans the data?

$$\mathbf{M}_D = \mathbf{I} - \mathbf{D}(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'$$

where $\mathbf{D} = \mathbf{I}_n \otimes \boldsymbol{\iota}_T$.

FE model

Assuming that $T = 3$ and $N = 2$,

$$\boldsymbol{M}_D = \begin{bmatrix} 1-\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} & 0 & 0 & 0 \\ -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} & 0 & 0 & 0 \\ -\frac{1}{T} & -\frac{1}{T} & 1-\frac{1}{T} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} \\ 0 & 0 & 0 & -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} \\ 0 & 0 & 0 & -\frac{1}{T} & -\frac{1}{T} & 1-\frac{1}{T} \end{bmatrix}$$

FE model

Then, assuming values for $\boldsymbol{X}$, $\boldsymbol{X}$ in deviation form is

$$\boldsymbol{M}_D\boldsymbol{X} = \begin{bmatrix} 1-\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} & 0 & 0 & 0 \\ -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} & 0 & 0 & 0 \\ -\frac{1}{T} & -\frac{1}{T} & 1-\frac{1}{T} & 0 & 0 & 0 \\ 0 & 0 & 0 & 1-\frac{1}{T} & -\frac{1}{T} & -\frac{1}{T} \\ 0 & 0 & 0 & -\frac{1}{T} & 1-\frac{1}{T} & -\frac{1}{T} \\ 0 & 0 & 0 & -\frac{1}{T} & -\frac{1}{T} & 1-\frac{1}{T} \end{bmatrix} \begin{bmatrix} 1 & 12 \\ 1 & 13 \\ 1 & 11 \\ 3 & 43 \\ 4 & 46 \\ 4 & 41 \end{bmatrix}$$

$$= \begin{bmatrix} 1-\frac{1+1+1}{3} & 12-\frac{12+13+11}{3} \\ 1-\frac{1+1+1}{3} & 13-\frac{12+13+11}{3} \\ 1-\frac{1+1+1}{3} & 11-\frac{12+13+11}{3} \\ 3-\frac{3+4+4}{3} & 43-\frac{43+46+41}{3} \\ 4-\frac{3+4+4}{3} & 46-\frac{43+46+41}{3} \\ 4-\frac{3+4+4}{3} & 41-\frac{43+46+41}{3} \end{bmatrix}$$

This shows how the $\boldsymbol{M}_D$ matrix demeans the data. $\boldsymbol{M}_D$ is indeed called the centering matrix.

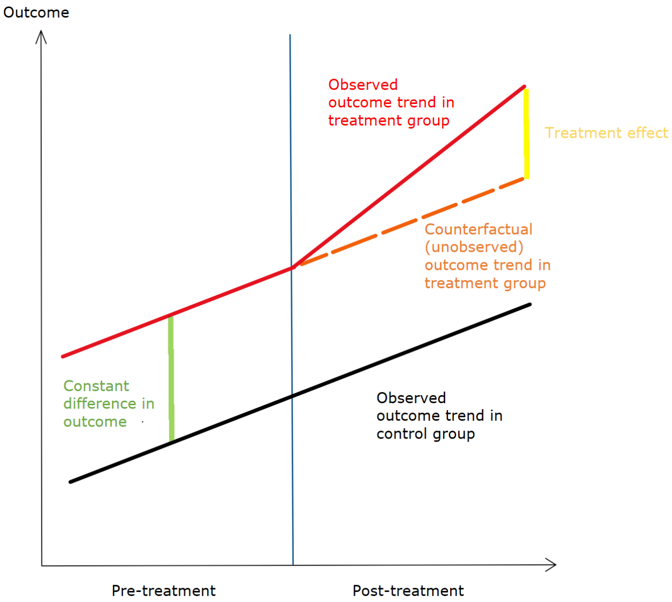
DiD model

Apart from the IV model, there are other models that allow causal effect claims. Here we study the difference-in-difference (DiD) model.

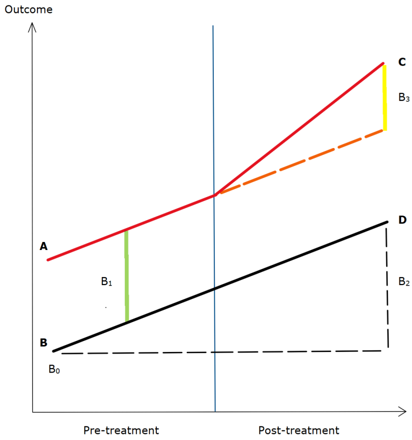
DiD is a compelling and widely used model.

However, it requires unique data. This is the price you pay, for the compelling model you buy.

DiD model


Outcome
Observed
outcome trend in
treatment group
Treatment effect
Counterfactual
(unobserved)
outcome trend in
treatment group
Constant
difference in
outcome
Observed
outcome trend in
control group
Pre-treatment
Post-treatment

DiD model

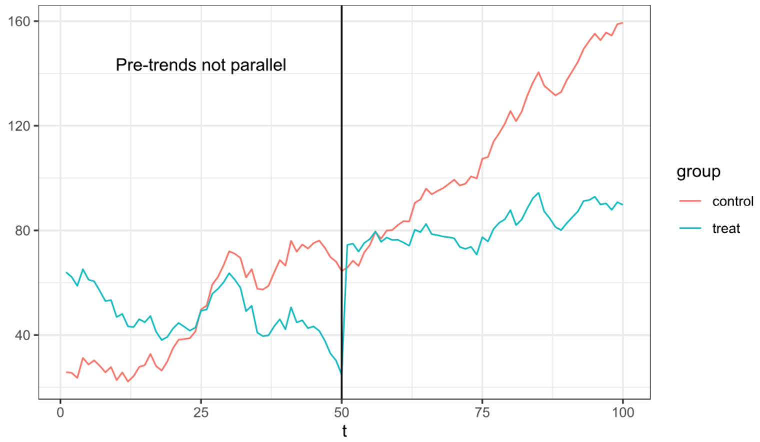


$Y_{it} = B_0 + B_1 \text{ Treatment} + B_2 \text{ Post}_t + B_3 \text{ Treatment} * \text{Post}_t + u_{it}$

Coef.	Calculation	Interpretation
B_0	B	Baseline average
B_1	A-B	Difference between two groups pre-intervention
B_2	D-B	Time trend in control group
B_3	(C-A)-(D-B)	Difference in changes over time

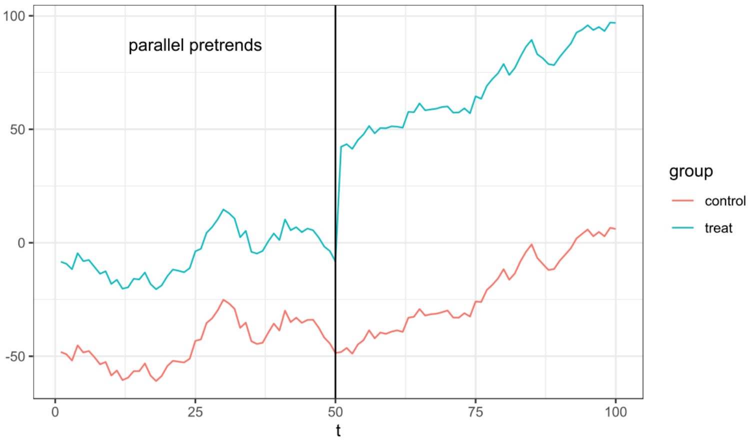
The key assumption is the parallels trends assumption.

DiD model



Source: `https://www.r-bloggers.com/2021/10/`

DiD model



parallel pretrends
100
50
0
-50
0
25
50
75
100
t
group
control
treat

DiD model

$$Y_{it} = B_0 + B_1 \text{ Treatment}_i + B_2 \text{ Post}_t + B_3 \text{ Treatment}_i * \text{Post}_t + u_{it}$$

DiD model

- $E [Y_{it} \mid Treatment_i = 1, Post_t = 1] = B_0 + B_1 + B_2 + B_3$
- $E [Y_{it} \mid Treatment_i = 1, Post_t = 0] = B_0 + B_1$
- $E [Y_{it} \mid Treatment_i = 0, Post_t = 1] = B_0 + B_2$
- $E [Y_{it} \mid Treatment_i = 0, Post_t = 0] = B_0$

DiD model

- $E[\, Y_{it} \mid Treatment_i = 1, Post_t = 0] = B_0 + B_1$
  $-$
  $E[\, Y_{it} \mid Treatment_i = 0, Post_t = 0] = B_0$
  $=$
  B_1

DiD model

- $E[\, Y_{it} \mid Treatment_i = 0, Post_t = 1] = B_0 + B_2$
  -
  $E[\, Y_{it} \mid Treatment_i = 0, Post_t = 0] = B_0$
  =
  B_2

DiD model

- [
  $E[Y_{it} \mid Treatment_i = 1, Post_t = 1] = B_0 + B_1 + B_2 + B_3$
  -
  $E[Y_{it} \mid Treatment_i = 1, Post_t = 0] = B_0 + B_1$
  ]
  -
  [
  $E[Y_{it} \mid Treatment_i = 0, Post_t = 1] = B_0 + B_2$
  -
  $E[Y_{it} \mid Treatment_i = 0, Post_t = 0] = B_0$
  ]
  =
  B_3

This is the treatment effect.

LPM

Consider the linear model

$$y_{it} = \boldsymbol{x}'_{it}\boldsymbol{\beta} + \varepsilon_{it}$$

where y_{it} is a dummy variable.

LPM

$$\begin{aligned} y_i &= \boldsymbol{x}_i'\boldsymbol{\beta} + \varepsilon_i \\ \mathrm{E}\left[y_i \mid \boldsymbol{x}_i\right] &= \mathrm{E}\left[\boldsymbol{x}_i'\boldsymbol{\beta} \mid \boldsymbol{x}_i\right] + \mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] \\ 0 * \mathrm{P}\left[y_i = 0 \mid \boldsymbol{x}_i\right] + 1 * \mathrm{P}\left[y_i = 1 \mid \boldsymbol{x}_i\right] &= \boldsymbol{x}_i'\boldsymbol{\beta} \\ \mathrm{P}\left[y_i = 1 \mid \boldsymbol{x}_i\right] &= \boldsymbol{x}_i'\boldsymbol{\beta} \end{aligned}$$

if $\mathrm{E}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \mathbf{0}$.

This is the linear probability model.

The coefficients have a direct probability increase interpretation.

LPM

One problem of the LPM is that the error is heteroskedastic. That is, it can be shown that

$$\operatorname{Var}\left[\varepsilon_i \mid \boldsymbol{x}_i\right] = \boldsymbol{x}_i'\boldsymbol{\beta}\left(1 - \boldsymbol{x}_i'\boldsymbol{\beta}\right).$$

This is easy to tackle. Use the heteroskedastic robust S.E. estimator!

LPM

Another problem of the LPM is that the predictions can lie out of the unit interval, $(0, 1)$.

This does not make sense. Why?

This is not easy to tackle. We need a different model. The probit and logit models are developed to address this. We do not cover these models in this course.