When Should We Use Linear Fixed Effects Regression Models for Causal Inference with Longitudinal Data? (Imai & Kim 2016)

Jeremy Cohen

Princeton University

30 March 2017

Linear Unit-Fixed Effects Regression Model

Given a balanced, longitudinal data set of observations for N units over T time periods, then for each i=1, 2, . . . , N and t = 1, 2, . . . , T:

$$Y_{it} = \alpha_i + \beta X_{it} + \epsilon_{it} \tag{1}$$

where:

- Y_{it} is the outcome variable for unit i at time t
- α_i is a fixed but unknown intercept for unit i
- X_{it} is the binary treatment assignment for unit i at time t
- β is the causal effect to be estimated
- ϵ_{it} is a disturbance term for unit i at time t

Importantly, the unit-specific intercept, α_i, is designed to capture all unobserved, time-invariant confounders, such that $\alpha_i = h(\mathbf{U}_i)$, where $\mathbf{U}_i$ is a vector of unobserved time-invariant confounders and $h(\cdot)$ is an arbitrary and unknown function.

Linear Unit-Fixed Effects Regression Model

Assumptions for identifiability of β:

1. Linearity (see last slide)
2. Strict Exogeneity

For each $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$:

$$\epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$$

where $\mathbf{X}_i$ is a T x 1 vector of treatment variables for unit i. In the classic linear fixed effects model (**LM-FE**) mean independence

$$\mathbb{E}(\epsilon_{it}|\mathbf{X}_i, \mathbf{U}_i) = \mathbb{E}(\epsilon_{it}|\mathbf{X}_i, \alpha_i) = \mathbb{E}(\epsilon_{it})$$

is sufficient to identify β

Linear Unit-Fixed Effects Regression Model

Taken together, these assumptions allow us to obtain a least squares estimate of β by regressing the deviation of the outcome variable from its mean on the deviation of the treatment variable from its mean (hence why these are sometimes called "difference" models):

$$\hat{\beta}_{FE} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} \left\{ (Y_{it} - \bar{Y}_i) - \beta(X_{it} - \bar{X}_i) \right\}^2 \quad (2)$$

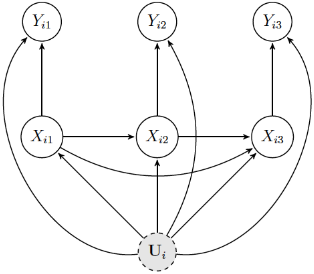
Note that because units that do not vary in the treatment variable across values of t contribute no information to the estimation of β, the causal estimand here is the average contemporaneous effect of X_{it} on Y_{it} *for units that vary in treatment status over time*:

$$\tau = \mathbb{E}(Y_{it}(1) - Y_{it}(0) | C_i = 1) \quad (3)$$

where $C_i = \mathbf{1}\left\{0 < \sum_{t=1}^{T} X_{it} < T\right\}$. Then under Assumptions 1 and 2, $\beta = \tau$.

Nonparametric Model

We can also look at this nonparametrically, with DAGs, which yields the **NPSEM-FE**



Formally,

$$Y_{it} = g_1(X_{it}, \mathbf{U}_i, \epsilon_{it}) \tag{4}$$

$$X_{it} = g_2(X_{i1}, \dots, X_{i,t-1}, \mathbf{U}_i, \eta_{it}) \tag{5}$$

Nonparametric Model

Equations (4) and (5) again

$$Y_{it} = g_1(X_{it}, \mathbf{U}_i, \epsilon_{it})$$

$$X_{it} = g_2(X_{i1}, \ldots, X_{i,t-1}, \mathbf{U}_i, \eta_{it})$$

This model is the nonparametric generalization of **LM-FE**:

1. Linearity is a special case of equation (4): $g_1(X_{it}, \mathbf{U}_i, \epsilon_{it}) = h(\mathbf{U}_i) + \beta X_{it} + \epsilon_{it}$
2. Strict exogeneity, $\epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{U}_i\}$, holds because, although we do not include the error terms in our DAG, if we did, Y_{it} would act as a collider on the path between any ϵ_{it} and the set of unit-level treatment variables and unobserved confounders, $\{\mathbf{X}_i, \mathbf{U}_i\}$

We cannot draw any more arrows on this DAG without making it inconsistent with **LM-FE**

Additional Causal Assumptions

Since we cannot draw any more arrows, we are left with additional causal assumptions for the identification of our causal estimand in the fixed effects case:

1. No unobserved, time-*varying* confounders
2. Past outcomes do not directly affect current outcomes
3. Past outcome do not directly affect current treatment
4. Past treatments do not directly affect current outcomes

Additional Causal Assumptions: Potential Outcomes

In the potential outcomes framework, we can add two more assumptions to our original assumptions (of linearity and strict exogeneity) to account for all causal assumptions needed to identify β in **LM-FE**

1. Linearity
2. Strict Exogeneity
3. **No Carryover Effect**
4. **Sequential Ignorability with $\mathbf{U}_i$**

No Carryover Effect

For each $i = 1, 2, \ldots, N$ and $t = 1, 2, \ldots, T$, the potential outcome is given by:

$$Y_{it}(X_{i1}, X_{i2}, \ldots, X_{i,t-1}, X_{it}) = Y_{it}(X_{it})$$

The right hand sides of the linear model and the nonparametirc model include only contemporaneous treatment effects, implying that past treatments do not affect current outcomes (the final assumption identified from the **NPSEM-FE**).

Sequential Ignorability with $\mathbf{U}_i$

One way to account for treatment assignment under the assumption of strict ignorability is to assume sequential ignorability with $\mathbf{U}_i$, i.e. at time t we randomize the current treatment, X_{it} conditional on past treatments. For each $i = 1, 2, \ldots, N$:

$$\{Y_{it}(1), Y_{it}(0)\}_{t=1}^{T} \perp\!\!\!\perp X_{i1} | \mathbf{U}_i$$

$$\vdots$$

$$\{Y_{it}(1), Y_{it}(0)\}_{t=1}^{T} \perp\!\!\!\perp X_{it'} | X_{i1}, \ldots, X_{i,t'-1}, \mathbf{U}_i$$

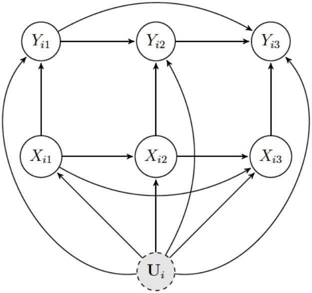
$$\vdots$$

$$\{Y_{it}(1), Y_{it}(0)\}_{t=1}^{T} \perp\!\!\!\perp X_{iT} | X_{i1}, \ldots, X_{i,T-1}, \mathbf{U}_i$$

Note that we must still assume no time-*varying* confounder and no effect of past outcomes $Y_{it'}$ on current treatment X_{it} where $t' < t$.
This corresponds to assumptions 1 and 3 under **NPSEM-FE** and the strict exogeneity assumption

Relaxing Assumptions: $Y_{it'} \rightarrow Y_{it}$

Easy. As long as we condition on past treatments and unobserved time-invariant confounders, this is not a threat to identification. Why?



(a) past outcome affects current outcome

Every non-causal path between the treatment X_{it} and any outcome $Y_{it'}$ is blocked where $t \neq t'$

Relaxing Assumptions: $X_{it'} \to Y_{it}$

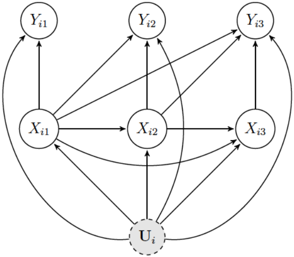
We could condition on lagged treatment variable(s). Which yields:

$$Y_{it} = \alpha_i + \beta_1 X_{it} + \beta_2 X_{i,t-1} + \epsilon it \tag{6}$$

in the linear case and

$$Y_{it} = g_1(X_{i1}, \ldots, X_{it}, \mathbf{U}_i, \epsilon_{it}) \tag{7}$$

And sequential ignorability still holds because any non-causal path between ϵ_{it} and $\{\mathbf{X}_i, \mathbf{U}_i\}$ contains a collider Y_{it}



(b) past treatments affect current outcome

Relaxing Assumptions: $X_{it'} \to Y_{it}$ (cont'd)

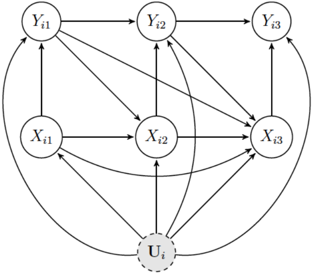
BUT we cannot actually adjust for $\mathbf{U}_i$ and all prior treatments at the same time. If we:

- Adjust for $\mathbf{U}_i$ then the comparison of treated and control units must be done across values of t within each i. We can no longer adjust for $X_{i1}, \ldots, X_{i,t-1}$ since no two observations for i have the same treatment history
- Adjust for $X_{i1}, \ldots, X_{i,t-1}$, then we need to compare observations across is within t and can no longer adjust for $\mathbf{U}_i$

Relaxing Assumptions $Y_{it'} \rightarrow X_{it}$

Clear violation of Sequential Ignorability. Common Approach: Instrumental Variables

$$Y_{it} = \alpha_i + \beta X_{it} + \rho Y_{i,t-1} + \epsilon_{it} \tag{8}$$



(d) instrumental variables

However, the validity of each instrument relies on the satisfaction of the exclusion restriction.

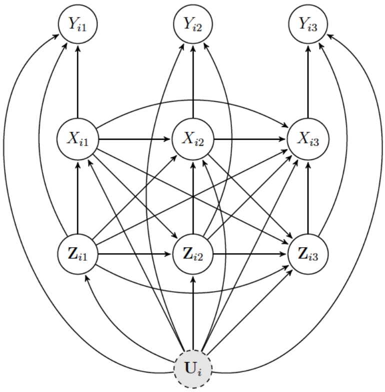
Adjusting for Observed Time-Varying Confounders

Let $\mathbf{Z}_{it}$ be a vector of (pre-treatment) observed, time-varying confounders. Then:

$$Y_{it} = \alpha_i + \beta X_{it} + \boldsymbol{\delta}^T \mathbf{Z}_{it} + \epsilon_{it} \tag{9}$$

Then with $\mathbf{Z_i}$ added to our strict exogeneity assumption, $\epsilon_{it} \perp\!\!\!\perp \{\mathbf{X}_i, \mathbf{Z}_i, \mathbf{U}_i\}$, we can add these covariates to our linear model.

Adjusting for Observed Time-Varying Confounders (cont'd)



But of course there are additional assumptions

1. $X_{it'}$ and $Z_{it'}$ cannot directly affect Y_{it}
2. we could solve the latter problem by conditioning on past values of the time-varying confounders but we then would be unable to condition on $\mathbf{U}_i$
3. τ is unidentifiable if $Y_{it'}$ affects X_{it} either directly or through Z_{it}. Implies $corr(\epsilon_{it}, \mathbf{Z}_{it}) \neq 0$

Within-Unit Matching Estimator

Even if no spillover effect and sequential ignorability holds **LM-FE** is inconsistent for τ unless the within-unit ATE or within unit proportion of treated observations is consistent across units (linearity)

We can nonparametrically adjust by comparing treated and control observations across t within i:

$$\hat{\tau}^{match} = \frac{1}{\sum_{i=1}^{N} C_i} \sum_{i=1}^{N} C_i \left(\frac{\sum_{t=1}^{T} X_{it} Y_{it}}{\sum_{t=1}^{T} X_{it}} - \frac{\sum_{t=1}^{T} (1 - X_{it}) Y_{it}}{\sum_{t=1}^{T} (1 - X_{it})} \right) \quad (10)$$

I.e. Take the difference in means between treated and control observations within units and then average across units and the matched set, $\mathcal{M}_{it}$, for each observation (i, t) is:

$$\mathcal{M}_{it}^{match} = \left\{ (i', t') : i' = i, X_{i't'} = 1 - X_{it} \right\} \quad (11)$$

Generalizing the Within-Unit Matching Estimator

For any matched set, $\mathcal{M}_{it}$:

$$\hat{\tau} = \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}(\widehat{Y_{it}(1)} - \widehat{(Y_{it}(0)}) \tag{12}$$

then

$$\widehat{Y_{it}(x)} = \begin{cases} Y_{it} & \text{if} X_{it} = x \\ \frac{1}{\#\mathcal{M}_{it}} \sum_{(i',t')\in\mathcal{M}_{it}} Y_{i't'} & \text{if} X_{it} = 1 - x \end{cases} \tag{13}$$

where $\#\mathcal{M}_{it}$ is the number of observations in the matched set and $D_{it} = \mathbf{1}\{\#\mathcal{M}_{it} > 0\}$

We can add covariates and use any distance metric to match

Matching and First Differences

Assuming no time trend in potential outcomes and no spillover effects, we can create a matching before-and-after or FD estimator:

$$\mathcal{M}_{it}^{BA} = \left\{ (i', t') : i' = i, t' \in \{t-1, t+1\}, X_{i',t'} = 1 - X_{it} \right\} \qquad (14)$$

is equivalent to the FD estimator:

$$\hat{\beta}_{FD} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=2}^{T} \left\{ (Y_{it} - Y_{i,t-1}) - \beta(X_{it} - X_{i,t-1}) \right\}^2 \qquad (15)$$

Unlike fixed effects which is a naive matching estimator (all other observations on i across t), FD matches on observations from subsequent time periods only. Note that past outcomes can still not affect current treatment status

Matching as Weighted Fixed Effects Regression

We can write our within unit matching estimators as a weighted linear regression with unit fixed effects:

$$\hat{\beta}_{WFE} = \underset{\beta}{\operatorname{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} \left\{ (Y_{it} - \bar{Y}_i^*) - \beta(X_{it} - \bar{X}_i^*) \right\}^2 \quad (16)$$

where $\bar{X}_i^*$ and $\bar{Y}_i^*$ are the within-unit averages (over time) weighted by W_{it}:

$$W_{it} = D_{it} \sum_{i'=1}^{N} \sum_{t'=1}^{T} w_{it}^{i't'}; w_{it}^{i't'} = \begin{cases} 1 & \text{if}(i', t') = (i, t) \\ \frac{1}{\#\mathcal{M}_{i't'}} & \text{if}(i, t) \in \#\mathcal{M}_{i't'} \\ 0 & \text{otherwise} \end{cases} \quad (17)$$

Adding in Covariates

$$(\hat{\beta}_{WFEadj}, \hat{\delta}_{WFEadj}) = \text{argmin}_{\beta,\delta} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} \left\{ (Y_{it} - \bar{Y}_i^*) - \beta(X_{it} - \bar{X}_i^*) - \delta^T(\mathbf{Z}_{it} - \hat{\mathbf{Z}}_i^*) \right\}^2 \quad (18)$$

Then we can revisit the potential outcomes under the matching framework and adjust by covatiates:

$$\widehat{Y_{it}(x)} = \begin{cases} Y_{it} - \hat{\delta}^T_{WFEadj}\mathbf{Z}_{it} & \text{if} X_{it} = x \\ \frac{1}{\#\mathcal{M}_{it}} \sum_{(i',t') \in \mathcal{M}_{it}} Y_{i't'} - \hat{\delta}^T_{WFEadj}\mathbf{Z}_{it} & \text{if} X_{it} = 1 - x \end{cases} \quad (19)$$

The Problem with Two-Way Fixed Effects

The two-way fixed effects estimator attempts to control for (in this case) *both* unit- and time-invariant unobserved confounders:

$$Y_{it} = \alpha_i + \gamma_t + \beta X_{it} + \epsilon_{it} \tag{20}$$

(We now must satisfy strict exogeneity with respect to our unobserved unit-invariant confounders as well) and we can estimate our β_{FE2} as follows:

$$\hat{\beta}_{FE2}, = \underset{\beta,\delta}{\text{argmin}} \sum_{i=1}^{N} \sum_{t=1}^{T} W_{it} \left\{ (Y_{it} - \bar{Y}_i - \bar{Y}_t + \bar{Y}) - \beta(X_{it} - \bar{X}_i - \bar{X}_t + \bar{X}) \right\}^2 \tag{21}$$

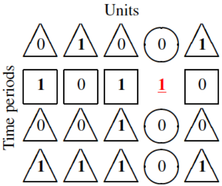
Two-Way Fixed Effects and Matching

The two-way fixed effects estimator can be thought of as a weighted average of each unit fixed effects estimate minus the weighted average of the pooled sample:

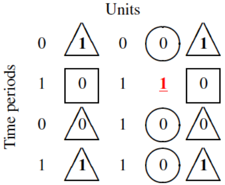
$$\hat{\beta}_{FE2} = \frac{\omega_{FE}\hat{\beta}_{FE} + \omega_{FEtime}\hat{\beta}_{FEtime} - \omega_{pool}\hat{\beta}_{pool}}{\omega_{FE} + \omega_{FEtime} - \omega_{pool}} \tag{22}$$

where the weights (in large samples with many periods) are roughly equal to the average variance of treatment assignment within-unit, within-time, and in the pooled dataset

Two-Way Fixed Effects and Matching



(a) Two-way fixed effects estimator



(b) Two-way matching estimator

$$\mathcal{M}_{it}^{FE2} = \{(i', t') : i' = i, t' \neq t\} \quad \mathcal{M}_{it}^{match} = \{(i', t') : i' = i, X_{i't'} = 1 - X_{it}\}$$

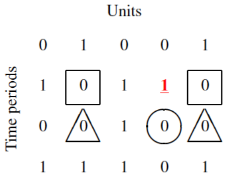
$$\mathcal{N}_{it}^{FE2} = \{(i', t') : i' \neq i, t' = t\} \quad \mathcal{N}_{it}^{match} = \{(i', t') : t' = t, X_{i't'} = 1 - X_{it}\}$$

.

$$\mathcal{A}_{it} = \{(i', t') : i' \neq i, t' \neq t, (i, t') \in \mathcal{M}_{it}, (i', t) \in \mathcal{N}_{it}\}$$

Two-Way Matching and the Differences-in-Differences Estimator

Special case of two-way fixed effects: differences-in-differences estimator



Where:

- the unit-specific matched set is the observation of the previous period if it belongs to the control group or $\{\}$ otherwise
- the time-specific matched set is a group of control observations in the same time period that were also control observations at $t-1$
- the adjustment set are the one-period prior observations of the members of the time-specific matched set

DiD (cont'd)

Then the multiperiod DiD matching estimator is:

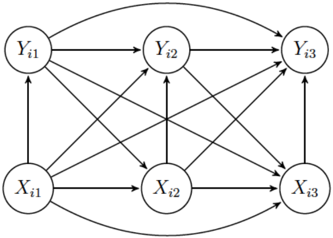
$$\hat{\tau}_{DiD} = \frac{1}{\sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}} \sum_{i=1}^{N}\sum_{t=1}^{T} D_{it}(\widehat{Y_{it}(1)} - \widehat{Y_{it}(0)}) \tag{23}$$

where $D_{it} = X_{it} \cdot \mathbf{1}\left\{\#\mathcal{M}_{it}^{DiD} \cdot \#\mathcal{N}_{it}^{DiD} > 0\right\}$ with $D_{i1} = 0$ and for $D_{it} = 1$:

$$\widehat{Y_{it}(x)} = \begin{cases} Y_{it} & \text{if} X_{it} = \\ Y_{i,t-1} + \frac{1}{\#\mathcal{N}_{it}^{DiD}} \sum_{(i',t)\in(N)_{it}^{DiD}} Y_{i't} - \frac{1}{\#\mathcal{A}_{it}^{DiD}} \sum_{(i',t')\in(A)_{it}^{DiD}} Y_{i't'} & \text{if} X_{it} = \end{cases} \tag{24}$$

What Does Matching Get You?

Marginal Structural Models



Sometimes