Economics 102C: Advanced Topics in Econometrics 8 - Panel Data: Fixed Effects and Differences in Differences

Michael Best

Spring 2015

Introduction

- The methods we've discussed so far are designed for the analysis of cross-sectional data: data on n individuals, firms, students at *given point in time*.
- However, many datasets have a *time dimension* too.
- **Repeated cross-sections** or pseudo-panels are cross-sections drawn at different points in time.
  - e.g. random samples of individuals in the US taken in 1990, 1995, 2000.
  - Each cross-section is representative, but unlikely that any given individual appears in more than one cross section
- **Panel Data** contains repeated observations on the *same* units over time.
  - data tracks individuals over time. each individual i observed at time $t =, 1, \ldots, T_i$
  - a *balanced panel* contains the same number of observations for each individual $T_i = T \; \forall i$

Introduction

- In principle, we sould just treat the data as a big cross-section: a **pooled cross-section**
- However, this ignores the time dimension:
  - Panel data can be useful for eliminating omitted variables that don't vary over time
  - This leads to the **fixed effects** panel estimator
- We can also use panel data for policy analysis
  - Comparing changes over time in a group of individuals that are affected (treated) by a policy change, to a control group that isn't affected.
  - This leads to the **difference-in-differences** estimator.

Framework

- In a panel of individuals $i = 1, \ldots, n$ observed at times $t = 1, \ldots, T$ we write each of the nT observations as

$$
\begin{aligned}
y_{it} &= \mathbf{x}_{it}'\beta + \mathbf{z}_i'\alpha + \varepsilon_{it} \\
&= \mathbf{x}_{it}'\beta + c_i + \varepsilon_{it}
\end{aligned}
$$

- $\mathbf{x}_{it}$ contains k regressors that vary over time for each individual.
- $\mathbf{z}_i$ are individual-specific variables that don't change over time (ethnicity, gender etc.)
  - Some of these variables are observed. If they're all observed, then we can just use OLS.
  - There are almost always some unobserved elements of $\mathbf{z}_i$ though

Framework

- Since for each individual we have multiple $\mathbf{x}_i$ we need to extend the exogeneity assumption
- Static panel data models use the **strict exogeneity** assumption

$$\mathsf{E}\left[\varepsilon_{it}|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}\right] = 0$$

- We also need to worry about the unobserved parts of $\mathbf{z}_i$. One (extreme) assumption is **mean independence**:

$$\mathsf{E}\left[c_i|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\right] = \alpha$$

  the assumption that the unobserved heterogeneity is uncorrelated with the regressors leads to the *random effects* model, a version of the generalized regression model. We won't cover it here, as we almost never believe the mean independence assumption.
- An alternative is to assume that

$$\mathsf{E}\left[c_i|\mathbf{x}_{i1}, \ldots, \mathbf{x}_{iT}\right] = h\left(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT}\right) = h\left(\mathbf{X}_i\right)$$

  This leads to the **fixed effects** model.

Framework

- Often the treatment we are interested in studying varies at the group level so for individual i in group s at time t we write

$$y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \mathbf{x}'_{ist}\beta + \varepsilon_{ist}$$

  where γ_s are group-specific, time invariant effects, λ_t are time-specific, group-invariant effects, D_{st} is a treatment indicator that varies across time and groups, and $\mathbf{x}_{ist}$ are individual-specific variables.
- Here we will use the assumption that conditional on γ_s, λ_t and $\mathbf{x}_{ist}$ the treatment is independent of the potential outcomes for y,

$$\mathsf{E}\left[y_{0ist}|s, t, \mathbf{x}_{ist}, D_{st}\right] = \mathsf{E}\left[y_{0ist}|s, t, \mathbf{x}_{ist}\right] = \gamma_s + \lambda_t + \mathbf{x}'_{ist}\beta$$

  the ***parallel trends*** assumption
- With this, we can use the **difference in differences** estimator to identify the causal effects of the treatment.

Outline

Pooled OLS

- 1 option is simply to ignore the time dimension of the data and estimate the **pooled model**

$$y_{it} = \alpha + \mathbf{x}_{it}'\beta + \varepsilon_{it},\ i = 1, \ldots, n,\ t = 1, \ldots, T_i$$

- Then if we add to our usual assumptions, that

$$\begin{aligned} \mathsf{E}\left[\varepsilon_{it} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT_i}\right] &= 0 \\ \mathsf{Var}\left(\varepsilon_{it} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT_i}\right) &= \sigma_\varepsilon^2 \\ \mathsf{Cov}\left(\varepsilon_{it}, \varepsilon_{js} | \mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots, \mathbf{x}_{iT_i}\right) &= 0\ \forall i \neq j \text{ or } t \neq s \end{aligned}$$

- Then we can simply estimate the model by OLS, and OLS is even the efficient estimator

Pooled OLS

- In a panel data context, it is particularly unlikely that these assumptions all hold.
- In the fixed effects model, ε_{it} contains an fixed (across time) individual component that is correlated with $\mathbf{x}_{it}$. We'll come back to this
- First focus on a random effects model:

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

where $\mathrm{E}\left[c_i|\mathbf{X}_i\right] = \alpha$ (which doesn't depend on $\mathbf{X}_i$)

Pooled OLS

- We can rewrite the model as

$$
\begin{aligned}
y_{it} &= \alpha + \mathbf{x}_{it}'\beta + \varepsilon_{it} + (c_i - \mathsf{E}[c_i|\mathbf{X}_i]) \\
&= \alpha + \mathbf{x}_{it}'\beta + \varepsilon_{it} + u_i \\
&= \alpha + \mathbf{x}_{it}'\beta + w_{it}
\end{aligned}
$$

- The unobserved heterogeneity c_i gives the error term **autocorrelation:** $\mathsf{E}[w_{it}w_{is}] = \sigma_u^2 \; \forall t \neq s$
- The error terms are correlated within individuals across time.
- As in the generalized regression model, OLS is still consistent, but the conventional asymptotic variance estimator underestimates the true variance.

Pooled OLS

- More generally, let's allow the error term to be arbitrarily autocorrelated within individuals. Stack the $T_i$observations for individual i:

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT_i} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}'_{i2} \\ \vdots \\ \mathbf{x}'_{iT_i} \end{pmatrix} \beta + \begin{pmatrix} w_{i1} \\ w_{i2} \\ \vdots \\ w_{iT_i} \end{pmatrix}$$

$$\mathbf{y}_i = \mathbf{X}_i \beta + \mathbf{w}_i$$

- Then we can express the variance of the error term as

$$\text{Var}\left(\mathbf{w}_i | \mathbf{X}_i\right) = \sigma_\varepsilon^2 \mathbf{I}_{T_i} + \mathbf{\Sigma}_i = \mathbf{\Omega}_i$$

Pooled OLS

- Then stacking the observations for the n individuals on top of eachother

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix} \beta + \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_n \end{pmatrix}$$

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{w}$$

- The OLS estimator of β is

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1} \mathbf{X}'\mathbf{y}$$

Pooled OLS

- We can break this in to the sums of each individual's $\mathbf{X}_i'\mathbf{X}_i$ and $\mathbf{X}_i'\mathbf{y}_i$.

$$\mathbf{X}'\mathbf{X} = \begin{pmatrix} \mathbf{X}_1' & \mathbf{X}_2' & \cdots & \mathbf{X}_n' \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix} = \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{X}_i$$

$$\mathbf{X}'\mathbf{y} = \begin{pmatrix} \mathbf{X}_1' & \mathbf{X}_2' & \cdots & \mathbf{X}_n' \end{pmatrix} \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} = \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{y}_i$$

- So that

$$\hat{\beta} = \left(\sum_{i=1}^{n} \mathbf{X}_i'\mathbf{X}_i \right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i'\mathbf{y}_i$$

Pooled OLS

- Plugging in the model for $\mathbf{y}_i$

$$\hat{\beta} = \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i' \left(\mathbf{X}_i \beta + \mathbf{w}_i\right)$$
$$= \beta + \left(\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i\right)^{-1} \sum_{i=1}^{n} \mathbf{X}_i' \mathbf{w}_i$$

- Then, under exactly the same conditions as in the generalized regression model, the asymptotic variance of $\hat{\beta}$ is

$$\text{aVar}\left(\hat{\beta}\right) = \frac{1}{n}\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i\right)^{-1} \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{w}_i \mathbf{w}_i' \mathbf{X}_i\right) \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i\right)^{-1}$$
$$= \frac{1}{n}\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i\right)^{-1} \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i' \boldsymbol{\Omega}_i \mathbf{X}_i\right) \text{plim}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i' \mathbf{X}_i\right)^{-1}$$

Pooled OLS

- As before, we need to estimate this middle matrix, and the natural estimator is

$$\widehat{\text{aVar}}\left(\hat{\beta}\right) = \frac{1}{n}\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\hat{\mathbf{w}}_i\hat{\mathbf{w}}_i'\mathbf{X}_i\right)\text{plim}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i'\mathbf{X}_i\right)^{-1}$$

  where

$$\hat{\mathbf{w}}_i = \begin{pmatrix} \hat{w}_{i1} \\ \hat{w}_{i2} \\ \vdots \\ \hat{w}_{iT_i} \end{pmatrix} = \begin{pmatrix} y_{i1} - \mathbf{x}_{i1}'\hat{\beta} \\ y_{i2} - \mathbf{x}_{i2}'\hat{\beta} \\ \vdots \\ y_{iT_i} - \mathbf{x}_{iT_i}'\hat{\beta} \end{pmatrix}$$

  is each individual's vector of T_i residuals.
- Note that this looks a lot like the clustered standard errors, where the clusters are the individuals.

Outline

Introduction

Pooled OLS

Fixed Effects

Difference in Differences

Which Control Group?

Endogeneity: Panel IV

Fixed Effects

- In the fixed effects model we have unobserved heterogeneity c_i too

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

- but now we assume (reasonably!) that it's correlated with $\mathbf{X}_i$:

$$\mathsf{E}\left[c_i|\mathbf{X}_i\right] = h\left(\mathbf{X}_i\right)$$

- Which, note, doesn't vary over time, so

$$\begin{aligned} y_{it} &= \mathbf{x}_{it}'\beta + h\left(\mathbf{X}_i\right) + \varepsilon_{it} + \left[c_i - h\left(\mathbf{X}_i\right)\right] \\ &= \mathbf{x}_{it}'\beta + \alpha_i + \varepsilon_{it} + \left[c_i - h\left(\mathbf{X}_i\right)\right] \end{aligned}$$

Fixed Effects

- but by construction $c_i - h(\mathbf{X}_i)$ is uncorrelated with $\mathbf{X}_i$ so we can put it into the error term

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

- And now we treat the n α_i terms as parameters to be estimated together with β
- All the assumptions we previously made about ε_{it} are maintained.
- Note that any time-invariant variables will be soaked up into the α_i so the fixed effects model is no good if we are interested in time-invariant characteristics of individuals
- Conversely, the fixed effects estimator will be robust to any (un)observed, time-invariant omitted variables

Fixed Effects

- Assume there are T observations on each individual, and let ι be a $T \times 1$ vector of 1s, then for each individual

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{pmatrix} = \begin{pmatrix} \mathbf{x}'_{i1} \\ \mathbf{x}_{i2} \\ \vdots \\ \mathbf{x}'_{iT} \end{pmatrix} \beta + \begin{pmatrix} \alpha_i \\ \alpha_i \\ \vdots \\ \alpha_i \end{pmatrix} + \begin{pmatrix} \varepsilon_{i1} \\ \varepsilon_{i2} \\ \vdots \\ \varepsilon_{iT} \end{pmatrix}$$

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

- And stacking the n individuals we get

$$\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix} = \begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \\ \vdots \\ \mathbf{X}_n \end{pmatrix} \beta + \begin{pmatrix} \iota & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \iota & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \iota \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix}$$

Fixed Effects

- Let $\mathbf{d}_i$ be a dummy variable indicating the ith individual, then

$$\mathbf{y} = \begin{pmatrix} \mathbf{X} & \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_n \end{pmatrix} \begin{pmatrix} \beta \\ \boldsymbol{\alpha} \end{pmatrix} + \boldsymbol{\varepsilon}$$

  where $\boldsymbol{\alpha} = \begin{pmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_n \end{pmatrix}'$
- If we put all the individual dummies together into an $nT \times n$ matrix $\mathbf{D} = \begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_n \end{pmatrix}$ then

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

- This is the **least squares dummy variable (LSDV)** model

Fixed Effects

- There's nothing special about this model, when n is relatively small, we can just estimate this by OLS. In stata: `reg y x1 x2 x3 i.indiv [,vce(cluster(indiv)]`; or `reg y x1 x2 x3 ibn.indiv, noc [vce(cluster(indiv)]`
- However, often n is large (several thousand). This means that the computer has to invert $\begin{pmatrix} \mathbf{X} & \mathbf{D} \end{pmatrix}' \begin{pmatrix} \mathbf{X} & \mathbf{D} \end{pmatrix}$ which is $(n+k) \times (n+k)$ which small computers can't handle, and even fast computers can take a long time to compute.
- Luckily, there's a shortcut, but to analyze it we need to take a quick detour to derive a result on *partitioned regression*

Partitioned Regression: The Frisch-Waugh-Lovell Theorem

- Imagine a regression involving two sets of variables $\mathbf{X}_1$ and $\mathbf{X}_2$

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon = \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \varepsilon = \mathbf{X}_1\beta_1 + \mathbf{X}_2\beta_2 + \varepsilon$$

- Say we are only interested in β_2. Do we have to solve for the whole $\beta = \begin{pmatrix} \beta_1' & \beta_2' \end{pmatrix}'$ vector?
- OLS solves the **normal equations**:

$$\begin{aligned} \left(\mathbf{X}'\mathbf{X}\right)\hat{\beta} &= \mathbf{X}'\mathbf{y} \\ \left(\begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix}\right) \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \mathbf{X}_1' \\ \mathbf{X}_2' \end{pmatrix} \mathbf{y} \\ \begin{pmatrix} \mathbf{X}_1'\mathbf{X}_1 & \mathbf{X}_1'\mathbf{X}_2 \\ \mathbf{X}_2'\mathbf{X}_1 & \mathbf{X}_2'\mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} &= \begin{pmatrix} \mathbf{X}_1'\mathbf{y} \\ \mathbf{X}_2'\mathbf{y} \end{pmatrix} \end{aligned}$$

Partitioned Regression: The Frisch-Waugh-Lovell Theorem

- Start with the solution of the first set of normal equations for $\hat{\beta}_1$:

$$
\begin{aligned}
\mathbf{X}_1'\mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_1'\mathbf{X}_2\hat{\beta}_2 &= \mathbf{X}_1'\mathbf{y} \\
\mathbf{X}_1'\mathbf{X}_1\hat{\beta}_1 &= \mathbf{X}_1'\mathbf{y} - \mathbf{X}_1'\mathbf{X}_2\hat{\beta}_2 \\
\hat{\beta}_1 &= \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y} - \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\beta}_2 \\
&= \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\left(\mathbf{y} - \mathbf{X}_2\hat{\beta}_2\right)
\end{aligned}
$$

Theorem (Orthogonal Partitioned Regression)

If $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal, then the coefficients can be obtained by separately regressing $\mathbf{y}$ on $\mathbf{X}_1$ and regressing $\mathbf{y}$ on $\mathbf{X}_2$

Proof.

If $\mathbf{X}_1$ and $\mathbf{X}_2$ are orthogonal, then $\mathbf{X}_1'\mathbf{X}_2 = 0$ (by definition) and so the solution to the normal equations above is simply $\hat{\beta}_1 = \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y}$, the OLS coefficients from a regressino of $\mathbf{y}$ on $\mathbf{X}_1$. The same reasoning applies to $\hat{\beta}_2$. $\square$

Partitioned Regression: The Frisch-Waugh-Lovell Theorem

- If $\mathbf{X}_1$ and $\mathbf{X}_2$ aren't orthogonal, we can still obtain $\hat{\beta}_1$ and $\hat{\beta}_2$ separately:

Theorem (Frisch-Waugh-Lovell Theorem)

In the regression of $\mathbf{y}$ on 2 sets of variables $\mathbf{X}_1$ and $\mathbf{X}_2$, the coefficients $\hat{\beta}_2$ can be obtained by regressing the residuals from a regression of $\mathbf{y}$ on $\mathbf{X}_1$ on the residuals from regressions of each column of $\mathbf{X}_2$ on $\mathbf{X}_1$.

Proof.

The second set of normal equations is

$$\mathbf{X}_2'\mathbf{X}_1\hat{\beta}_1 + \mathbf{X}_2'\mathbf{X}_2\hat{\beta}_2 = \mathbf{X}_2'\mathbf{y}$$

Inserting the solution for $\hat{\beta}_1$

$$\mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{y} - \mathbf{X}_2'\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\mathbf{X}_2\hat{\beta}_2 + \mathbf{X}_2'\mathbf{X}_2\hat{\beta}_2 = \mathbf{X}_2'\mathbf{y}$$

Partitioned Regression: The Frisch-Waugh-Lovell Theorem

Proof.

rearranging

$$\mathbf{X}_2'\left[\mathbf{I}-\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{X}_2\hat{\beta}_2=\mathbf{X}_2'\underbrace{\left[\mathbf{I}-\mathbf{X}_1\left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1}\mathbf{X}_1'\right]\mathbf{y}}_{\text{residuals in regression of } \mathbf{y} \text{ on } \mathbf{X}_1 \text{ only}}$$

$$\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\hat{\beta}_2=\mathbf{X}_2'\mathbf{M}_1\mathbf{y}$$

$$\hat{\beta}_2=\left(\mathbf{X}_2'\mathbf{M}_1\mathbf{X}_2\right)^{-1}\mathbf{X}_2'\mathbf{M}_1\mathbf{y}$$

□

- So we can find $\hat{\beta}_2$ through a series of regressions
  - regress $\mathbf{y}$ on $\mathbf{X}_1$ and store the residuals $\mathbf{M}_1\mathbf{y}$
  - regress each column of $\mathbf{X}_2$ on $\mathbf{X}_1$ and store all the residuals in $\mathbf{M}_1\mathbf{X}_2$
  - regress $\mathbf{M}_1\mathbf{y}$ on $\mathbf{M}_1\mathbf{X}_2$

Partitioned Regression: Practice

1. Using the Frisch-Waugh-Lovell Theorem, show that in a regression with a constant term,

$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon_i$$

the coefficients $\beta_1, \ldots, \beta_k$ can be obtained by regressing y in deviations from its mean $\bar{y} = \sum_{i=1}^{n} y_i$ on the x_k in deviations from their means.

   1.1 Letting $\mathbf{X}_1 = \iota$, the $n \times 1$ vector of 1s, and $\mathbf{X}_2 = \begin{pmatrix} \mathbf{x}_1 & \mathbf{x}_2 & \ldots & \mathbf{x}_k \end{pmatrix}$ be the $n \times k$ matrix of x variables, show that $\mathbf{M}_1\mathbf{y} = \left[\mathbf{I} - \mathbf{X}_1 \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} \mathbf{X}_1'\right]\mathbf{y} = \mathbf{y} - \iota\bar{y}$ and that for each column of $\mathbf{X}_2$
   $\mathbf{M}_1\mathbf{x}_k = \left[\mathbf{I} - \mathbf{X}_1 \left(\mathbf{X}_1'\mathbf{X}_1\right)^{-1} \mathbf{X}_1'\right]\mathbf{x}_k = \mathbf{x}_k - \iota\bar{x}_k$

Partitioned Regression: Practice-Solutions

1. With $\mathbf{X}_1 = \iota$, we have

$$\begin{aligned}
\mathbf{M}_1\mathbf{y} &= \left[\mathbf{I} - \iota\left(\iota'\iota\right)^{-1}\iota'\right]\mathbf{y} \\
&= \mathbf{y} - \iota\frac{1}{n}\iota'\mathbf{y} \\
&= \mathbf{y} - \iota\frac{1}{n}\sum_{i=1}^{n} y_i \\
&= \mathbf{y} - \iota\bar{y}
\end{aligned}$$

which is the $n \times 1$ column vector of the y_i in deviations from their mean $\bar{y}$.
The same argument applies to any $n \times 1$ column vector, such as the columns of $\mathbf{X}_2$

Fixed Effects

- We had written the fixed effects model as

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

- Often, we're not interested in $\boldsymbol{\alpha}$, just in β, so we can apply the Frisch-Waugh-Lovell theorem to this:

$$\hat{\beta} = \left(\mathbf{X}'\mathbf{M}_D\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{M}_D\mathbf{y}$$

  where

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

Fixed Effects

- Note that the columns of $\mathbf{D}$ are all orthogonal ($\mathbf{d}_i'\mathbf{d}_j = 0\ \forall i \neq j$) so $\mathbf{M}_D$ is a demeaning matrix:

$$\mathbf{D}'\mathbf{D} = \begin{pmatrix} \mathbf{d}_1' \\ \mathbf{d}_2' \\ \vdots \\ \mathbf{d}_n' \end{pmatrix} \begin{pmatrix} \mathbf{d}_1 & \mathbf{d}_2 & \cdots & \mathbf{d}_n \end{pmatrix} = \begin{pmatrix} \mathbf{d}_1'\mathbf{d}_1 & 0 & \cdots & 0 \\ 0 & \mathbf{d}_2'\mathbf{d}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{d}_n'\mathbf{d}_n \end{pmatrix}$$

$$= \begin{pmatrix} T & 0 & \cdots & 0 \\ 0 & T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & T \end{pmatrix}$$

$$\mathbf{M}_D = \begin{pmatrix} \mathbf{M}^0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}^0 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}^0 \end{pmatrix}$$

  where$\mathbf{M}^0 = \mathbf{I}_T - \frac{1}{T}\iota\iota'$ creates deviations from time averages for each individual i
- So, the fixed effects regression is equivalent to regressing $y_{it} - \bar{y}_i$ (where $\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it}$) on $\mathbf{x}_{it} - \bar{\mathbf{x}}_i$
- For this reason, it is sometimes called the *within* estimator

Fixed Effects

- Once we have estimated β we can use the normal equations to recover the estimated $\boldsymbol{\alpha}$

$$\mathbf{D}'\mathbf{D}\hat{\boldsymbol{\alpha}} + \mathbf{D}'\mathbf{X}\hat{\beta} = \mathbf{D}'\mathbf{y}$$
$$\hat{\boldsymbol{\alpha}} = \left(\mathbf{D}'\mathbf{D}\right)^{-1}\mathbf{D}'\left(\mathbf{y} - \mathbf{X}\hat{\beta}\right)$$

i.e. for each individual

$$\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i'\hat{\beta}$$

Fixed Effects

- We can also estimate the asymptotic covariance matrix of $\hat{\beta}$ (i.e. not the full OLS coefficient vector $\begin{pmatrix} \hat{\beta} \\ \hat{\boldsymbol{\alpha}} \end{pmatrix}$) as

$$\widehat{\text{aVar}}\left(\hat{\beta}\right) = s^2 \left(\mathbf{X}'\mathbf{M}_D\mathbf{X}\right)^{-1}$$

where

$$s^2 = \frac{\sum_{i=1}^{n}\sum_{t=1}^{T}\left(y_{it} - \mathbf{x}_{it}'\hat{\beta} - \hat{\alpha}_i\right)^2}{nT - n - k} = \frac{\left(\mathbf{M}_D\mathbf{y} - \mathbf{M}_D\mathbf{X}\hat{\beta}\right)'\left(\mathbf{M}_D\mathbf{y} - \mathbf{M}_D\mathbf{X}\hat{\beta}\right)}{nT - n - k}$$

- Note that the residuals here are

$$\hat{\varepsilon}_{it} = y_{it} - \mathbf{x}_{it}'\hat{\beta} - \hat{\alpha}_i = y_{it} - \mathbf{x}_{it}'\hat{\beta} - \left(\bar{y}_i - \bar{\mathbf{x}}_i'\hat{\beta}\right) = \left(y_{it} - \bar{y}_i\right) - \left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)'\hat{\beta}$$

Fixed Effects: Practice

1. using $\hat{\alpha}_i = \bar{y}_i - \bar{\mathbf{x}}_i' \hat{\beta}$ and the model for y_{it} show that

$$\text{aVar}\left(\hat{\alpha}_i\right) = \frac{\sigma_\varepsilon^2}{T} + \bar{\mathbf{x}}_i' \text{aVar}\left(\hat{\beta}\right) \bar{\mathbf{x}}_i$$

   for which you can use the fact that, by construction $\bar{y}_i$ and $\bar{\mathbf{x}}_i' \hat{\beta}$ are orthogonal

2. If we hold T fixed, and let $n \to \infty$, are the $\hat{\alpha}_i$ consistent estimators of the α_i? To do this, first find the asymptotic variance of $\hat{\beta}$, for which you may use the result that

$$\text{plim}\ \left[\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it} - \bar{\mathbf{x}}_i\right)'\right] = \bar{S}_{xx,i}$$

   where $\bar{S}_{xx,i}$ is a finite matrix

Fixed Effects: Practice

1. For each individual

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_i = \frac{1}{T}\sum_{t=1}^{T} \alpha_i + \mathbf{x}'_{it}\beta + \varepsilon_{it}$$

$$= \alpha_i + \bar{\mathbf{x}}'_i\beta + \frac{1}{T}\sum_{t=1}^{T} \varepsilon_{it}$$

and by construction $\bar{y}_i$ and $\bar{\mathbf{x}}'_i\hat{\beta}$ are orthogonal, so

$$\text{aVar}\left(\hat{\alpha}_i\right) = \text{aVar}\left(\bar{y}_i\right) + \text{aVar}\left(\bar{\mathbf{x}}'_i\hat{\beta}\right)$$

$$= \frac{1}{T^2}\text{aVar}\left(\sum_{t=1}^{T} \varepsilon_{it}\right) + \bar{\mathbf{x}}'_i\text{aVar}\left(\hat{\beta}\right)\bar{\mathbf{x}}_i$$

We still assume the ε_{it} are i.i.d so aVar $\left(\sum_{t=1}^{T} \varepsilon_{it}\right) = T\sigma_{\varepsilon}^2$ and the result follows.

Fixed Effects: Practice

2. Start with aVar $\left(\hat{\beta}\right)$.

$$\begin{aligned}
\text{aVar}\left(\hat{\beta}\right) &= \sigma_\varepsilon^2 \left(\mathbf{X}'\mathbf{M}_D\mathbf{X}\right)^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\left[\frac{1}{n}\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\right]^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\left[T\frac{1}{n}\sum_{i=1}^{n}\frac{1}{T}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\right]^{-1} \\
&= \frac{\sigma_\varepsilon^2}{n}\left(T\bar{S}_{xx,i}\right)^{-1}
\end{aligned}$$

which clearly $\to 0$ as $n \to \infty$. This means that

$$\lim_{n\to\infty} \text{aVar}\left(\hat{\alpha}_i\right) = \frac{\sigma_\varepsilon^2}{T} > 0$$

so that $\hat{\alpha}_i$ is not a consistent estimator of α_i

Fixed Effects

- We argued above that it may be important to allow for the error terms to be correlated over time for each individual, so that $\mathsf{E}\left[\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_i'|\mathbf{X}_i\right] \neq \sigma_\varepsilon^2\mathbf{I}_T$
- We can allow for this, but we have to estimate the middle matrix of the asymptotic covariance matrix:

$$\hat{\beta}_{LSDV} = \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\right)^{-1}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(y_{it}-\bar{y}_i\right)\right)$$

- so we can construct a robust covariance matrix estimator using

$$\widehat{\mathsf{aVar}}_{\mathsf{Robust}}\left(\hat{\beta}_{LSDV}\right) = \begin{array}{c}\left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\right)^{-1}\times \\ \left[\sum_{i=1}^{n}\left(\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\hat{\varepsilon}_{it}\right)\left(\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\hat{\varepsilon}_{it}\right)'\right]\times \\ \left(\sum_{i=1}^{n}\sum_{t=1}^{T}\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)\left(\mathbf{x}_{it}-\bar{\mathbf{x}}_i\right)'\right)^{-1}\end{array}$$

Fixed Effects

- We can also extend the model to allow for time fixed effects as well as individual fixed effects:

$$y_{it} = \mathbf{x}'_{it}\beta + \alpha_i + \delta_t + \varepsilon_{it}$$

but we'd have to drop one time period to avoid the dummy variable trap. Or, we can specify this as

$$y_{it} = \mathbf{x}'_{it}\beta + \mu + \alpha_i + \delta_t + \varepsilon_{it}$$

where

$$\sum_{i=1}^{n} \alpha_i = \sum_{t=1}^{T} \delta_t = 0$$

Fixed Effects

- Then we can estimate the model by double-demeaning the data:

$$y_{*it} = y_{it} - \bar{y}_i - \bar{y}_t + \bar{\bar{y}}$$
$$\mathbf{x}_{*it} = \mathbf{x}_{it} - \bar{\mathbf{x}}_i - \bar{\mathbf{x}}_t + \bar{\bar{\mathbf{x}}}$$

where

$$\bar{y}_i = \frac{1}{T}\sum_{t=1}^{T} y_{it} \quad \bar{y}_t = \frac{1}{n}\sum_{i=1}^{n} y_{it} \quad \bar{\bar{y}} = \frac{1}{nT}\sum_{i=1}^{n}\sum_{t=1}^{T} y_{it}$$

- We then regress y_{*it} on $\mathbf{x}_{*it}$ to get $\hat{\beta}$ and we can calculate

$$\hat{\mu} = \bar{\bar{y}} - \bar{\bar{\mathbf{x}}}'\hat{\beta} \quad \hat{\alpha}_i = (\bar{y}_i - \bar{\bar{y}}) - (\bar{\mathbf{x}}_i - \bar{\bar{\mathbf{x}}})'\,\beta \quad \hat{\delta}_t = (\bar{y}_t - \bar{\bar{y}}) - (\bar{\mathbf{x}}_t - \bar{\bar{\mathbf{x}}})'\,\beta$$

Outline

Difference in Differences

- Let's consider a concrete example of a policy we might evaluate:
- On April 1 1992 New Jersey raised the state minimum wage from \$4.25 to \$5.05
- Card & Krueger (1994) collected data on wages and employment from fast food chains in 2/92 and 11/92.
- They collect the same data for Pennsylvania, a bordering state that did not change the minimum wage.
- The idea is to compare how wages changed in NJ between Feb and Nov, to how wages changed in PA between Feb and Nov.
- i.e. calculate the *difference* between the difference between Feb and Nov wages in NJ and the difference between Feb and Nov wages in PA: The *difference in the differences*.

Difference in Differences

- Denote employment at restaurant i in state s at time t by y_{ist}
- In the potential outcomes notation we have been using, y_{0ist} is the employment at restaurant i when there is a low minimum wage, and y_{1ist} is employment when there is a high minimum wage.
- i.e. in New Jersey we see y_{1ist} in November, but y_{0ist} in February.
- We assume that

$$\mathsf{E}\left[y_{0ist}|s,t\right] = \gamma_2 + \lambda_t$$

  where $s \in \{NJ, PA\}$ and $t \in \{Feb, Nov\}$
- And, we assume that

$$\mathsf{E}\left[y_{1ist} - y_{0ist}|s,t\right] = \delta$$

  (which, note, doesn't depend on s or t)

Difference in Differences

- Then we can write the model as

$$y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \varepsilon_{ist}$$

where D_{st} is a dummy variable for high minimum-wage states and time periods (in this example, $D_{st} = 1$ in NJ in November, and 0 otherwise), and $\mathsf{E}\left[\varepsilon_{ist}|s,t\right] = 0$

- In this case, the Pennsylvania time-difference is

$$\mathsf{E}\left[y_{ist}|s = PA, t = Nov\right] - \mathsf{E}\left[y_{ist}|s = PA, t = Feb\right]$$
$$= (\gamma_{PA} + \lambda_{Nov}) - (\gamma_{PA} - \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb}$$

- And the New Jersey time-difference is

$$\mathsf{E}\left[y_{ist}|s = NJ, t = Nov\right] - \mathsf{E}\left[y_{ist}|s = NJ, t = Feb\right]$$
$$= (\gamma_{NJ} + \lambda_{Nov} + \delta) - (\gamma_{NJ} + \lambda_{Feb}) = \lambda_{Nov} - \lambda_{Feb} + \delta$$

Difference in Differences

- Combining these

$$\{\mathsf{E}[y_{ist}|s=NJ,t=Nov]-\mathsf{E}[y_{ist}|s=NJ,t=Feb]\}$$
$$-\{\mathsf{E}[y_{ist}|s=PA,t=Nov]-\mathsf{E}[y_{ist}|s=PA,t=Feb]\}$$
$$=(\lambda_{Nov}-\lambda_{Feb}+\delta)-(\lambda_{Nov}-\lambda_{Feb})=\delta$$

- We can easily estimate this by using sample averages::

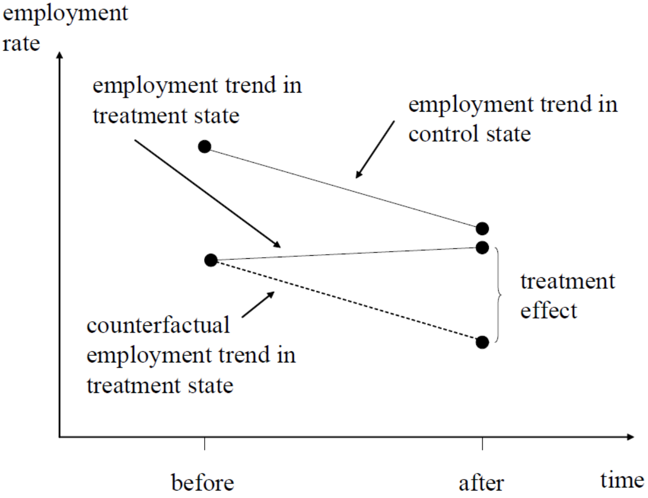
Table 5.2.1: Average employment per store before and after the New Jersey minimum wage increase

Variable	PA (i)	NJ (ii)	Difference, NJ-PA (iii)
1. FTE employment before,	23.33	20.44	-2.89
all available observations	(1.35)	(0.51)	(1.44)
2. FTE employment after,	21.17	21.03	-0.14
all available observations	(0.94)	(0.52)	(1.07)
3. Change in mean FTE	-2.16	0.59	2.76
employment	(1.25)	(0.54)	(1.36)

Notes: Adapted from Card and Krueger (1994), Table 3. The table reports average full-time equivalent (FTE) employment at restaurants in Pennsylvania and New Jersey before and after a minimum wage increase in New Jersey.

Difference in Differences

- We assume the change in the control state (PA) is what the change in the treatment state (NJ) *would have been* if there had been no treatment. The **Parallel Trends** assumption



Difference in Differences

- We can also implement the difference in differences design using a regression:
- Let NJ_s be a dummy for restaurants in New Jersy, and d_t be a dummy for observations from November
- Then we can write the model as

$$y_{ist} = \alpha + \gamma NJ_s + \lambda d_t + \delta \left(NJ_s \times d_t \right) + \varepsilon_{ist}$$

where note that $NJ_s \times d_t = D_{st}$

Difference in Differences

- So in terms of the potential outcomes, and the regression coefficients, the 4 groups have

$$\begin{aligned}
\mathsf{E}[y_{ist}|s=PA,t=Feb] &= \gamma_{PA}+\lambda_{Feb}=\alpha\\
\mathsf{E}[y_{ist}|s=PA,t=Nov] &= \gamma_{PA}+\lambda_{Nov}=\alpha+\lambda\\
\mathsf{E}[y_{ist}|s=NJ,t=Feb] &= \gamma_{NJ}+\lambda_{Feb}=\alpha+\gamma\\
\mathsf{E}[y_{ist}|s=NJ,t=Nov] &= \gamma_{NJ}+\lambda_{Nov}+\delta=\alpha+\gamma+\lambda+\delta
\end{aligned}$$

- And so we can see that

$$\begin{aligned}
\alpha &= \mathsf{E}[y_{ist}|s=PA,t=Feb]=\gamma_{PA}+\lambda_{Feb}\\
\gamma &= \mathsf{E}[y_{ist}|s=NJ,t=Feb]-\mathsf{E}[y_{ist}|s=PA,t=Feb]\\
&= \gamma_{NJ}-\gamma_{PA}\\
\lambda &= \mathsf{E}[y_{ist}|s=PA,t=Nov]-\mathsf{E}[y_{ist}|s=PA,t=Feb]\\
&= \lambda_{Nov}-\lambda_{Feb}\\
\delta &= \{\mathsf{E}[y_{ist}|s=NJ,t=Nov]-\mathsf{E}[y_{ist}|s=NJ,t=Feb]\}\\
&\quad -\{\mathsf{E}[y_{ist}|s=PA,t=Nov]-\mathsf{E}[y_{ist}|s=PA,t=Feb]\}
\end{aligned}$$

Difference in Differences

- We can also extend the model to include individual-level controls and time-varying state level variables:

$$y_{ist} = \gamma_s + \lambda_t + \delta D_{st} + \mathbf{x}'_{it}\beta + \varepsilon_{ist}$$

- Note that time-varying state level variables could be a source of OVB: If some other characteristics of the state changed at the same time as the policy in the treatment state but not in the control state, then $\hat{\delta}$ will conflate the effects of this characteristic with the effects of the policy.
- On the other hand, individual level controls are there only to increase precision.

DiD example 1: Workers' comp and time off work

- Meyer, Viscusi, and Durbin (1995) study a *natural experiment:* Kentucky (1980) & Michigan (1982) increased workers' compensation

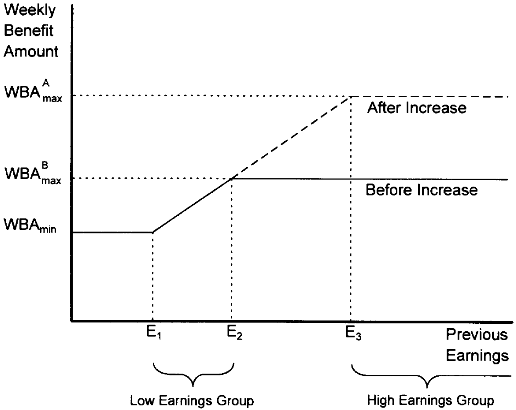


FIGURE 1. TEMPORARY TOTAL BENEFIT SCHEDULE BEFORE AND AFTER AN INCREASE IN THE MAXIMUM WEEKLY BENEFIT

DiD example 1: Workers' comp and time off work

TABLE 1—REPLACEMENT RATES, EARNINGS, AND DEMOGRAPHIC CHARACTERISTICS DURING THE YEARS BEFORE AND AFTER BENEFIT INCREASES

	Kentucky			Michigan		
Variable	Before increase (1)	After increase (2)	Percentage change (3)	Before increase (4)	After increase (5)	Percentage change (6)
Maximum benefit ($)	131.00	217.00	65.65	181.00	307.00	69.61
Replacement rate, high earnings (percent)	32.70 (0.25)	51.02 (0.37)	56.02 (1.65)	30.01 (0.35)	44.15 (0.48)	47.14 (2.33)
Replacement rate, low earnings (percent)	66.42 (0.20)	66.66 (0.22)	0.36 (0.44)	66.64 (0.24)	66.35 (0.30)	−0.45 (0.58)
Average benefit (1983 $), high earnings	151.08 (0.96)	239.09 (1.32)	58.25 (1.33)	220.66 (1.78)	320.48 (2.27)	45.24 (1.56)
Average benefit (1983 $), low earnings	118.58 (0.64)	118.26 (0.74)	−0.27 (0.82)	183.66 (0.78)	182.77 (0.93)	−0.45 (0.58)
Average earnings (1983 $), high earnings	475.31 (2.45)	482.41 (2.73)	1.49 (0.78)	749.72 (7.25)	739.01 (7.49)	−1.43 (1.38)
Average earnings (1983 $), low earnings	179.09 (0.89)	177.54 (0.97)	−0.86 (0.73)	275.83 (0.75)	275.65 (0.83)	−0.07 (0.40)

DiD example 1: Workers' comp and time off work

TABLE 4—KENTUCKY AND MICHIGAN: DURATION AND MEDICAL COSTS OF TEMPORARY TOTAL DISABILITIES DURING THE YEARS BEFORE AND AFTER BENEFIT INCREASES

	High earnings		Low earnings		Differences		Difference in differences
Variable	Before increase (1)	After increase (2)	Before increase (3)	After increase (4)	[(2)−(1)] (5)	[(4)−(3)] (6)	[(5)−(6)] (7)
Mean duration (weeks)							
Kentucky	11.16	12.89	6.25	7.01	1.72	0.76	0.96
	(0.83)	(0.83)	(0.30)	(0.41)	(1.17)	(0.51)	(1.28)
Michigan	14.76	19.42	10.94	13.64	4.66	2.70	1.96
	(2.25)	(2.67)	(1.09)	(1.56)	(3.49)	(1.90)	(3.97)
Median duration (weeks)							
Kentucky	4.00	5.00	3.00	3.00	1.00	0.00	1.00
	(0.14)	(0.20)	(0.11)	(0.12)	(0.25)	(0.16)	(0.29)
Michigan	5.00	7.00	4.00	4.00	2.00	0.00	2.00
	(0.45)	(0.67)	(0.22)	(0.28)	(0.81)	(0.35)	(0.89)
75th percentile, duration (weeks)							
Kentucky	8.00	10.00	7.00	7.00	2.00	0.00	2.00
	(0.28)	(0.45)	(0.21)	(0.24)	(0.53)	(0.32)	(0.62)
Michigan	10.00	14.00	8.50	9.00	4.00	0.50	3.50
	(0.74)	(1.88)	(0.54)	(0.57)	(2.03)	(0.79)	(2.17)
Mean of log duration							
Kentucky	1.38	1.58	1.13	1.13	0.20	0.01	0.19
	(0.04)	(0.04)	(0.03)	(0.03)	(0.05)	(0.04)	(0.07)
Michigan	1.58	1.87	1.41	1.51	0.29	0.10	0.19
	(0.09)	(0.10)	(0.06)	(0.06)	(0.13)	(0.08)	(0.16)

DiD example 1: Workers' comp and time off work

TABLE 6—REGRESSION EQUATIONS FOR NATURAL LOGARITHM OF DURATION, HIGH- AND LOW-EARNINGS GROUPS POOLED, AND HIGH-EARNINGS GROUP SEPARATELY

	Specification							
	High- and low-earnings groups pooled				High-earnings group only			
	Kentucky		Michigan		Kentucky		Michigan	
Explanatory variable	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)	(viii)
After-increase indicator variable	0.016 (0.045)	−0.004 (0.038)	0.082 (0.084)	0.003 (0.073)	0.228 (0.054)	0.149 (0.044)	0.244 (0.136)	0.260 (0.113)
High-earnings-group indicator variable	−1.522 (1.099)	−0.594 (0.930)	5.577 (4.811)	3.607 (4.162)				
After-increase × high-earnings-group indicator variable	0.215 (0.069)	0.162 (0.059)	0.157 (0.153)	0.203 (0.132)				
ln(Previous earnings)	0.258 (0.104)	0.207 (0.088)	0.901 (0.648)	0.139 (0.562)	0.492 (0.163)	0.229 (0.133)	0.067 (0.496)	−0.335 (0.414)
ln(Previous earnings) × high-earnings group	0.232 (0.187)	0.065 (0.158)	−0.973 (0.803)	−0.587 (0.695)				
Male indicator variable	−0.072 (0.046)	−0.070 (0.039)	−0.303 (0.099)	−0.332 (0.086)	−0.088 (0.133)	0.004 (0.108)	−1.053 (0.631)	−0.489 (0.527)

DiD Example 2: Eissa-Liebman 1996 or Eissa 1995

Outline

Synthetic Control

- What happens when you have many possible control groups?
- E.g. Abadie, Diamond & Hainmueller (2010) want to study the effect of California's 1988 tobacco control progam, Proposition 99.
- As controls, they want to use other states, but which one(s) should they use?
- A model: Suppose we observe $J + 1$ regions, and that only region 1 receives the treatment.
- We observe each region in time periods $t = 1, \ldots, T$
- Region 1 receives the treatment from period $T_0 + 1$ until T

Synthetic Control

- Let y_{it}^N be the potential outcome we *would* observe for region i at time t if region i never receives the treatment
- Let y_{it}^I be the optential outcome we *would* observe for region i at time t if region i receives the treatment from periods $T_0 + 1$ until T
- Assume the treatment has no impact before period $T_0 + 1$: $y_{it}^N = y_{it}^I$ for all $t = 1, \ldots, T_0$ and all $i = 1, \ldots, J + 1$
- Then $\alpha_{it} = y_{it}^I - y_{it}^N$ is the effect of the intervention on region i at time t
- We only observe

$$y_{it} = y_{it}^N + \alpha_{it} D_{it}$$

where D_{it} is a dummy for treatment.

Synthetic Control

- So, for each $t > T_0$ we'd like to estimate

$$\alpha_{1t} = y_{1t}^{I} - y_{1t}^{N}$$

  but we only observe y_{1t}^{I} so we need to estimate y_{1t}^{N}.
- Let's assume the difference in differences model for y_{it}^{N}

$$y_{it}^{N} = \delta_t + \boldsymbol{\theta}_t \mathbf{Z}_i + \mu_i + \varepsilon_{it}$$

- Then, we can use any of the other J regions as a control and construct a series of Diff in Diff estimators. e.g. if we use region $i = 2$,

$$\hat{\alpha}_{1t} = y_{1t} - y_{2t} \, t = T_0 + 1, \ldots, T$$

Synthetic Control

- But, we can do better, we can use a weighted average of all the other regions.
- Let $\mathbf{W} = (w_2, \ldots, w_{J+1})'$ be a $J \times 1$ vector of weights satisfying $w_j \leq 0 \; \forall j = 2, \ldots, J+1$ and $\sum_{j=1}^{J+1} w_j = 1$
- Then any such $\mathbf{W}$ represents a potential *synthetic control*. For a given $\mathbf{W}$, the calue of the outcome at time t is

$$\sum_{j=2}^{J+1} w_j y_{jt} = \delta_t + \boldsymbol{\theta}_t \sum_{j=1}^{J+1} w_j \mathbf{Z}_i + \sum_{j=2}^{J+1} w_j \mu_j + \sum_{j=2}^{J+1} w_j \varepsilon_{jt}$$

- The optimal weights $\mathbf{W}^*$ satisfy

$$\sum_{j=2}^{J+1} w_j^* y_{j1} = y_{11} \quad \sum_{j=2}^{J+1} w_j^* y_{j2} = y_{12} \quad \ldots \quad \sum_{j=2}^{J+1} w_j^* y_{jT_0} = y_{1T_0} \quad \& \quad \sum_{j=2}^{J+1} w_j^* \mathbf{Z}_j = \mathbf{Z}_1$$

Synthetic Control

- It may not be possible to satisfy all of these conditions exactly, but we can come as close as possible.
- For e.g. define the matrices

$$\mathbf{x}_1 = \begin{pmatrix} \mathbf{Z}_1 \\ y_{11} \\ y_{12} \\ \vdots \\ y_{1T_0} \end{pmatrix} \quad \mathbf{X}_0 = \begin{pmatrix} \mathbf{Z}_2 & \mathbf{Z}_3 & & \mathbf{Z}_{J+1} \\ y_{21} & y_{31} & & y_{J+1,1} \\ y_{22} & y_{32} & & y_{J+1,2} \\ \vdots & \vdots & & \vdots \\ y_{2T_0} & y_{3T_0} & & y_{J+1,T_0} \end{pmatrix}$$

- Then we can choose $\mathbf{W}$ to minimize

$$\|\mathbf{x}_1 - \mathbf{X}_0\mathbf{W}\| = \sqrt{(\mathbf{x}_1 - \mathbf{X}_0\mathbf{W})'(\mathbf{x}_1 - \mathbf{X}_0\mathbf{W})}$$

- Or, we can minimize the weighted norm

$$\|\mathbf{x}_1 - \mathbf{X}_0\mathbf{W}\|_{\mathbf{V}} = \sqrt{(\mathbf{x}_1 - \mathbf{X}_0\mathbf{W})'\mathbf{V}(\mathbf{x}_1 - \mathbf{X}_0\mathbf{W})}$$

  for a matrix of weights $\mathbf{V}$
- The authors provide the `synth` package in stata to implement this

DiD Example 3: Effect of CA Prop 99

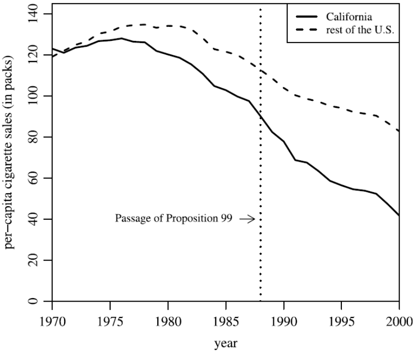


Figure 1. Trends in per-capita cigarette sales: California vs. the rest of the United States.

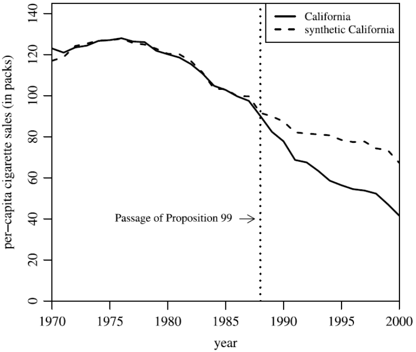


Figure 2. Trends in per-capita cigarette sales: California vs. synthetic California.

Outline

DiD Example 4: Food aid and civil conflict

- All of the mechanics of IV carry over to the panel data setting.
- For e.g. Nunn & Qian (2014) study the effects of US food aid on civil conflict:

$$C_{irt} = \beta F_{irt} + \mathbf{X}_{irt}\Gamma + \delta_r Y_t + \psi_{ir} + \nu_{irt}$$

  $C_{irt} = \{$conflict in country i in region r in year $t\}$ F_{irt} is quantity of food aid, $\mathbf{X}_{irt}$ are covariates, $\delta_r Y_t$ are region-specific time trends, ψ_{ir} are country fixed effects

- Basic idea: use level of wheat production in US one year before, P_{t-1} as an instrument for F_{irt}

$$F_{irt} = \alpha P_{t-1} + \mathbf{X}_{irt}\Gamma + \delta_r Y_t + \psi_{ir} + \varepsilon_{irt}$$

DiD Example 4: Food aid and civil conflict

- Exclusion restriction requires only channel through which P_{t-1} affects C_{irt} is F_{irt}.
  - other time trends in C_{irt} that happen to be correlated with P_{t-1} violate this.
- Develop Diff-in-Diff strategy:

$$
\begin{aligned}
C_{irt} &= \beta F_{irt} + \mathbf{X}_{irt}\Gamma + \varphi_{rt} + \psi_{ir} + \nu_{irt} \\
F_{irt} &= \alpha \left(P_{t-1} \times \bar{D}_{ir}\right) + \mathbf{X}_{irt}\Gamma + \varphi_{rt} + \psi_{ir} + \varepsilon_{irt}
\end{aligned}
$$

  where $\bar{D}_{ir}$ is country i's propensity to receive food aid, and φ_{rt} are region-year fixed effects.
- Countries that tend to receive more food aid are more affected by swings in US wheat production

DiD Example 4: Food aid and civil conflict

Table 2—The Effect of Food Aid on Conflict: Baseline Specification with $P_{t-1} \times D_{ir}$ as the Instrument

	Parsimonious specifications				Baseline specification		
Dependent variable (panels A, B, and C):	Any conflict (1)	Any conflict (2)	Any conflict (3)	Any conflict (4)	Any conflict (5)	Intrastate (6)	Interstate (7)
Panel A. OLS estimates							
US wheat aid (1,000 MT)	−0.00006	−0.00007	−0.00005	−0.00007	−0.00011	−0.00005	−0.00011
	(0.00018)	(0.00018)	(0.00017)	(0.00017)	(0.00017)	(0.00017)	(0.00004)
R^2	0.508	0.508	0.518	0.534	0.549	0.523	0.385
Panel B. Reduced form estimates (× 1,000)**							
Lag US wheat production (1,000 MT) × avg. prob. of any US food aid	0.00829	0.01039	0.01070	0.01133	0.01071	0.00909	−0.00158
	(0.00257)	(0.00263)	(0.00262)	(0.00318)	(0.00320)	(0.00322)	(0.00121)
R^2	0.511	0.512	0.521	0.536	0.551	0.525	0.382
Panel C. 2SLS estimates							
US wheat aid (1,000 MT)	0.00364	0.00303	0.00312	0.00343	0.00299	0.00254	−0.00044
	(0.00174)	(0.00125)	(0.00117)	(0.00106)	(0.00096)	(0.00088)	(0.00033)
Dependent variable (panel D):	US wheat aid (1,000 MT)						
Panel D. First-stage estimates							
Lag US wheat production (1,000 MT) × avg. prob. of any US food aid	0.00227	0.00343	0.00343	0.00330	0.00358	0.00358	0.00358
	(0.00094)	(0.00126)	(0.00120)	(0.00092)	(0.00103)	(0.00103)	(0.00103)
Kleibergen-Paap F-statistic	5.84	7.37	8.24	12.76	12.10	12.10	12.10

DiD Example 5: Schools, Education and Earnings

- Duflo (2001) studies the impact of a large school-building program (INPRES) in Indonesia in the 1970s on education and earnings
- 2-dimensions of variation
  - Intensity of INPRES: Number of schools built inversely proportional to baseline enrolment
  - Birth cohort: Children who were younger when schools built affected more.
- Suggests a Diff-in-Diff specification

$$S_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + (P_j T_i)\,\gamma_1 + (\mathbf{C}_j T_i)\,\delta_1 + \varepsilon_{ijk}$$

where S_{ijk} is schooling of individual i in region j in year k, P_j is program intensity, T_i is a dummy for being young when schools built, and $\mathbf{C}_j$ are region-level variables

DiD Example 5: Schools, Education and Earnings

- Then, we can use the school building program as an instrument for education in a returns to education equation with reduced form

$$y_{ijk} = c_1 + \alpha_{1j} + \beta_{1k} + (P_j T_i)\,\gamma_1 + (\mathbf{C}_j T_i)\,\delta_1 + \varepsilon_{ijk}$$

  where y_{ijk} is the log wage of individual i in region j in year k
- Let's rebuild the main results from this paper.
- `SUPAS.dta` is a slightly modified version of the data used in Duflo (2001), please load it up

DiD Example 5: Schools, Education and Earnings

Here's our to do list:

1. Poke around in the data to get to know it
2. Basic DiD
3. Regression DiD
4. IV results