Econometrics II

Lecture 8: Fixed Effects and Panel Data

David Schönholzer

Stockholm University

April 25, 2024

Plan for Today

1. Fixed Effects
   - Definitions
   - Identification and Estimation
   - Empirical Bayes
2. Panel Data
   - Structure and Definitions
3. Applications
   - Connected-Set Fixed Effects (AKM)
4. Appendix
   - Group Fixed Effects

Table of Contents

1 Fixed Effects
Definitions
Identification and Estimation
Empirical Bayes

2 Panel Data
Structure and Definitions

3 Applications
Connected-Set Fixed Effects (AKM)

4 Appendix
Group Fixed Effects

Dual Indexation

- Consider groups $\mathbf{J} : \{1, ..., N\} \to \{1, ..., J\}$ and let $X_{ij} = 1[\mathbf{J}(i) = j]$, e.g.
    - individuals across villages
    - regions over time
    - race-by-gender-by-grade cells
- Recall model for group means:

$$\begin{aligned} Y_i &= \sum_{j=1}^{J} X_{ij} \alpha_j + \varepsilon_i \\ &= \alpha_{\mathbf{J}(i)} + \varepsilon_i \end{aligned}$$

- Can re-index with dual indices: $i = 1, ..., N_j$ and

$$Y_{ij} = \alpha_j + \varepsilon_{ij}$$

- Two indices is often enough (rarely need e.g. Y_{imst})

Random Effects

- How to interpret $Y_{ij} = \alpha_j + \varepsilon_{ij}$?
- Two ways: *random effects* and *fixed effects*

- Random effects: $(\alpha_j, \varepsilon_{ij}) \overset{\text{iid}}{\sim} F\left(\mu, \sigma_\alpha^2\right) \times G\left(0, \sigma_\varepsilon^2\right)$ with
  - $\mathbb{E}\left[Y_{ij}\right] = \mathbb{E}\left[\alpha_j\right] = \mu$ for all j
  - $Var\left(\alpha_j\right) = \sigma_\alpha^2$ and $Cov\left(\alpha_j, \alpha_{j'}\right) = 0$ for $j \neq j'$
  - $Var\left(\varepsilon_{ij}\right) = \sigma_\varepsilon^2$ and $Cov\left(\varepsilon_{ij}, \varepsilon_{i'j'}\right) = 0$ for all i, i' and j, j' except if $i = i'$ and $j = j'$
  - $Var\left(Y_{ij}\right) = \sigma_\alpha^2 + \sigma_\varepsilon^2$ and $Cov\left(Y_{ij}, Y_{i'j}\right) = \sigma_\alpha^2$ for $i \neq i'$

- Parameters of interest: $\theta = \left(\mu, \sigma_\alpha^2, \sigma_\varepsilon^2\right)$
- Primary interest: *variance decomposition*

Fixed Effects

- Alternative interpretation of $Y_{ij} = \alpha_j + \varepsilon_{ij}$
- Assume $\varepsilon_{ij} \overset{\text{iid}}{\sim} G\left(0, \sigma_\varepsilon^2\right)$ but no distributional assumption on α_j:
  - $\mathbb{E}\left[Y_{ij}\right] = \alpha_j$ for all i, j
  - $Var\left(\varepsilon_{ij}\right) = \sigma_\varepsilon^2$ and $Cov\left(\varepsilon_{ij}, \varepsilon_{i',j'}\right) = 0$ for $(i,j) \neq (i',j')$
  - $Var\left(Y_{ij}\right) = \sigma_\varepsilon^2$
- Parameters of interest: $\theta = \left(\alpha_1, ..., \alpha_J, \sigma_\varepsilon^2\right)$
- Primary interest: estimating *unobserved heterogeneity*

- Why estimate unobserved heterogeneity? Two reasons:
  1. Defend CIA (or other research design) against omitted variable bias (OVB):

$$\begin{aligned} \tau_j &\equiv \mathbb{E}\left[Y_{ij}(1) - Y_{ij}(0) \,|X_{ij}\right] \\ &= \mathbb{E}\left[Y_{ij}|D_{ij} = 1, X_{ij}\right] - \mathbb{E}\left[Y_{ij}|D_{ij} = 0, X_{ij}\right] \end{aligned}$$

  2. Direct interest in group means

Twins Days in Twinsburg, Ohio

Example of FE to Support CIA: Twins

- Ashenfelter and Krueger (1994): return to schooling using twins
- Earnings Y_{ij} as function of schooling D_{ij} for $i \in \{1, 2\}$:

$$Y_{ij} = \alpha_j + D_{ij}\beta + \varepsilon_{ij}$$

  where α_j captures unobserved family background

- Let D_{ij}^k denote schooling reported by $k \in \{1, 2\}$
- With sample of twins, several ways to estimate β:
  1. Include α_j as FE
  2. Difference between twins:

$$Y_{1j} - Y_{2j} = \left(D_{1j}^1 - D_{2j}^2\right)\beta + u_j,$$

     which is algebraically equivalent to FE
  3. Instrument $\left(D_{1j}^1 - D_{2j}^2\right)$ with $\left(D_{1j}^2 - D_{2j}^1\right)$ since D_{ij}^i is noisy

Controlling for Unobserved Family Background

TABLE 3—ORDINARY LEAST-SQUARES (OLS), GENERALIZED LEAST-SQUARES (GLS), INSTRUMENTAL-VARIABLES (IV), AND FIXED-EFFECTS ESTIMATES OF LOG WAGE EQUATIONS FOR IDENTICAL TWINS[a]

Variable	OLS (i)	GLS (ii)	GLS (iii)	IV[a] (iv)	First difference (v)	First difference by IV (vi)
Own education	0.084 (0.014)	0.087 (0.015)	0.088 (0.015)	0.116 (0.030)	0.092 (0.024)	0.167 (0.043)
Sibling's education	—	—	−0.007 (0.015)	−0.037 (0.029)	—	—
Age	0.088 (0.019)	0.090 (0.023)	0.090 (0.023)	0.088 (0.019)	—	—
Age squared (÷100)	−0.087 (0.023)	−0.089 (0.028)	−0.090 (0.029)	−0.087 (0.024)	—	—
Male	0.204 (0.063)	0.204 (0.077)	0.206 (0.077)	0.206 (0.064)	—	—
White	−0.410 (0.127)	−0.417 (0.143)	−0.424 (0.144)	−0.428 (0.128)	—	—
Sample size:	298	298	298	298	149	149
R^2:	0.260	0.219	0.219	—	0.092	—

Example of Random Effects: Project STAR

- Chetty et al. (2011): effect of kindergarten classrooms on adult earnings Y_{ij}
- Students & teachers randomly assigned to classrooms
- Can study observable classroom characteristics D_j, e.g. class size:

$$Y_{ij} = \mu + D_j \beta + \varepsilon_{ij}$$

  with $j \perp \varepsilon_{ij}$ due to randomization

- But can also study unobservable classroom characteristis: $\alpha_j = D_j \beta + \xi_j$
- Two approaches:
  1. α_j as FE: F-test of $H_0 : \alpha_j = \alpha_k$ for all j, k
  2. α_j as RE: Estimate σ_α^2 using ANOVA (e.g. Searle et al. 2009)

Testing for Classroom Effects

TABLE VII

KINDERGARTEN CLASS EFFECTS: ANALYSIS OF VARIANCE

Dependent variable	(1) Grade K scores	(2) Grade 8 scores	(3)	(4) Wage earnings	(5)	(6)
p-value of F test on KG class fixed effects	0.000	0.419	0.047	0.026	0.020	0.040
p-value from permutation test	0.000	0.355	0.054	0.029	0.023	0.055
SD of class effects (RE estimate)	8.77%	0.000%	\$1,497	\$1,520	\$1,703	\$1,454
Demographic controls	x	x		x	x	x
Large classes only					x	
Observable class chars.						x
Observations	5,621	4,448	6,025	6,025	4,208	5,983

Notes. Each column reports estimates from an OLS regression of the dependent variable on school and class fixed effects, omitting one class fixed effect per school. The p-value in the first row is for an F test of the joint significance of the class fixed effects. The second row reports the p-value from a permutation test, calculated as follows: we randomly permute students between classes within each school, calculate the F-statistic on the class dummies, repeat the previous two steps 1,000 times, and locate the true F-statistic in this distribution. The third row reports the estimated standard deviation of class effects from a model with random class effects and school fixed effects. Grade 8 scores are available for students who remained in Tennessee public schools and took the eighth-grade standardized test any time between 1990 and 1997. Both KG and eighth-grade scores are coded using within-sample percentile ranks. Wage earnings is the individual's mean wage earnings over years 2005–2007 (including 0s for people with no wage earnings). All specifications are estimated on the subsample of students who entered a STAR school in kindergarten. All specifications except (3) control for the vector of demographic characteristics used in Table IV: a quartic in parent's household income interacted with an indicator for whether the filing parent is ever married between 1996 and 2008, mother's age at child's birth, indicators for parent's 401(k) savings and home ownership, and student's race, gender, free-lunch status, and age at kindergarten. Column (5) limits the sample to large classes only; this column identifies pure KG class effects because students who were in large classes were rerandomized into different classes after KG. Column (6) replicates column (4), adding controls for the following observable classroom characteristics: indicators for small class, above-median teacher experience, black teacher, and teacher with degree higher than a BA, and classmates' mean predicted score. Classmates' mean predicted score is constructed by regressing test scores on school-by-entry-grade fixed effects and the vector of demographic characteristics listed above and then taking the mean of the predicted scores.

Table of Contents

1 Fixed Effects
- Definitions
- Identification and Estimation
- Empirical Bayes

2 Panel Data
- Structure and Definitions

3 Applications
- Connected-Set Fixed Effects (AKM)

4 Appendix
- Group Fixed Effects

Identification of Fixed Effects

- Define $J \times 1$ vector $\boldsymbol{x}_i = [X_{i1}, \ldots, X_{iJ}]'$
- Single-observation model: $Y_i = \alpha_{\mathbf{J}(i)} + \varepsilon_i = \boldsymbol{x}_i'\boldsymbol{\alpha} + \varepsilon_i$
- Interested in $\boldsymbol{\alpha} = [\alpha_1, \ldots, \alpha_J]'$, the $J \times 1$ vector of FE parameters

- Construct $N \times 1$ vector $\mathbf{X}_j = [X_{1j}, \ldots, X_{Nj}]'$ for each j
- Stacked model: $\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$ where
  - $\mathbf{Y} = [Y_1, \ldots, Y_N]'$ is $N \times 1$ outcome vector
  - $\mathbf{X} = [\mathbf{X}_1, \ldots, \mathbf{X}_J]$ is $N \times J$ design matrix of group indicators
  - $\boldsymbol{\varepsilon} = [\varepsilon_1, \ldots, \varepsilon_N]'$ is $N \times 1$ vector of errors

- Pop OLS: $\boldsymbol{\alpha} = \mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i']^{-1}\,\mathbb{E}[\boldsymbol{x}_i Y_i]$
- OLS requires two conditions for identification of $\boldsymbol{\alpha}$:
  1. $\mathbb{E}[\boldsymbol{x}_i\varepsilon_i] = 0$ (j conditions): exogenous group assignment
  2. $\mathbb{E}[\boldsymbol{x}_i\boldsymbol{x}_i']$ (and hence $\mathbf{X}'\mathbf{X}$) is full rank, i.e. no multicollinearity

Endogenous Group Assignment

- First consider violation of exogenous group assignment
- Important distinction: if we run

$$Y_{ij} = \alpha_j + D_{ij}\beta + \varepsilon_{ij}$$

  and $\mathbb{E}\left[\alpha_j \varepsilon_{ij}\right] \neq 0$ for some j (e.g. selection into group)
- This implies the estimates of α_j are not identified
- And so OLS estimates $\widehat{\alpha}_j$ would be biased

- However, this is *not* necessarily a problem for identifying β
- As long as $\mathbb{E}\left[D_{ij}\varepsilon_{ij}\right] = 0$, we have that $\widehat{\beta} \xrightarrow{p} \beta$
- Follows from linearity of OLS

- For consistent $\widehat{\alpha}_j$, need CIA (or design) to argue groups form exogenously

Rank Violations in Fixed Effects

Three prominent causes of multicollinearity:

1. Including grand mean and J FE dummies
   - Consider $Y_{ij} = \alpha_j + \varepsilon_{ij}$ with $J = 3$ and $N_j = 2$:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\alpha} + \boldsymbol{\varepsilon}$$

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{12} \\ Y_{22} \\ Y_{13} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{bmatrix}$$

Over-Parametrization By Including Constant

- However, if we also include constant

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{12} \\ Y_{22} \\ Y_{13} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{bmatrix}$$

- Then $\theta = (\mu, \alpha_1, \alpha_2, \alpha_3)$ is not identified – why?
- Intuition: if you tell me $(\alpha_1, \alpha_2, \alpha_3)$, I can compute μ
- Similarly, if I know $(\mu, \alpha_1, \alpha_2)$, I can compute α_3
- Linear algebra: rank $(\mathbf{X}'\mathbf{X}) = 3$ but $\underset{4\times 4}{\mathbf{X}'\mathbf{X}}$, so below full rank
- This determines whether `STATA` drops coefficients!

Group-Invariant Covariates

2. Including group-invariant covariates:
   - Let $Y_{ij} = \alpha_j + W_j\gamma + \varepsilon_{ij}$
   - Then $\theta = (\gamma, \alpha_1, \alpha_2, \alpha_3)$ is not identified in:

$$\begin{bmatrix} Y_{11} \\ Y_{21} \\ Y_{12} \\ Y_{22} \\ Y_{13} \\ Y_{23} \end{bmatrix} = \begin{bmatrix} W_1 & 1 & 0 & 0 \\ W_1 & 1 & 0 & 0 \\ W_2 & 0 & 1 & 0 \\ W_2 & 0 & 1 & 0 \\ W_3 & 0 & 0 & 1 \\ W_3 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma \\ \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} + \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{21} \\ \varepsilon_{12} \\ \varepsilon_{22} \\ \varepsilon_{13} \\ \varepsilon_{23} \end{bmatrix}$$

   - Note that W_j is just a rescaled version of α_j
3. Nested fixed effects:
   - For example, consider $Y_{ijt} = \alpha_j + \xi_{jt} + \varepsilon_{ijt}$
   - α_j is group constant; ξ_{jt} are within-group FE

Estimation: Least-Square Dummy Variables as Frisch-Waugh-Lovell

- Imagine $Y_{ij} = \alpha_j + D_{ij}\beta + \varepsilon_{ij}$ and we want $\widehat{\beta}$
- Let $X_{ij} = 1[i \in j]$ as before
- Consider the following two linear projections:
  - Long regression: $\mathbb{E}^*\left[Y_{ij}|D_{ij}, X_{i1}, ..., X_{iJ}\right] = D_{ij}\beta + \sum_j X_{ij}\alpha_j$
  - Flipped auxiliary: $\mathbb{E}^*\left[D_{ij}|X_{i1}, ..., X_{iJ}\right] = \sum_j X_{ij}\alpha_j$
- Then by FWL: $\beta = \mathbb{E}\left[\tilde{D}_{ij}\tilde{D}'_{ij}\right]^{-1}\mathbb{E}\left[\tilde{D}_{ij}\tilde{Y}_{ij}\right]$ where
  - $\tilde{D}_{ij} = D_{ij} - \mathbb{E}^*\left[D_{ij}|X_{i1}, ..., X_{iJ}\right]$
  - $\tilde{Y}_{ij} = Y_{ij} - \mathbb{E}^*\left[Y_{ij}|X_{i1}, ..., X_{iJ}\right]$
- But note that these are just group-means regressions:
  - $\mathbb{E}^*\left[D_{ij}|X_{i1}, ..., X_{iJ}\right] = \bar{D}_j = \frac{1}{N_j}\sum_i D_{ij}$
  - $\mathbb{E}^*\left[Y_{ij}|X_{i1}, ..., X_{iJ}\right] = \bar{Y}_j = \frac{1}{N_j}\sum_i Y_{ij}$
- So can do $Y_{ij} - \bar{Y}_j = \left(D_{ij} - \bar{D}_j\right)\beta + \xi_{ij}$ – often much faster!
- This is exactly what `reghdfe` does in `STATA`

FE: Unbiased but Not Generally Consistent

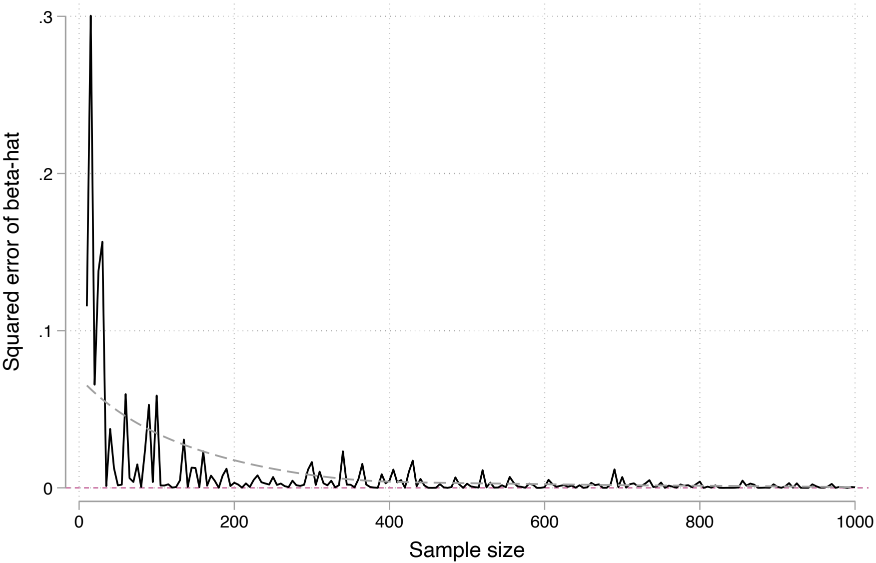
- For fixed $\mathbf{X}_j$ and $\mathbb{E}[\boldsymbol{x}_i \varepsilon_i] = 0$, OLS is unbiased: $\mathbb{E}\left[\widehat{\boldsymbol{\alpha}}_{\text{OLS}}\right] = \boldsymbol{\alpha}$
- But $N_j = J/N$ is often constant as N grows because J grows as well, e.g.
  - Classrooms: $N_j \approx 20$ fixed even if J large
  - Short panels e.g. $N_j \approx 5$
- Then OLS is *not* consistent: $\widehat{\boldsymbol{\alpha}}_{\text{OLS}} \overset{p}{\not\to} \boldsymbol{\alpha}$

- Illustration of this "incidental parameters problem":
  - Consider $Y_{ij} = \alpha_j + D_{ij}\beta + \varepsilon_{ij}$ with $N_j = 2$
  - Assume $\alpha_j \sim N(0,1)$, $\beta = 1$, and $\varepsilon_{ij} \sim N(0,1)$, all iid
  - Do this for $J \in [10, 15, 20, \dots, 1000]$
  - Evaluate for each sample size:

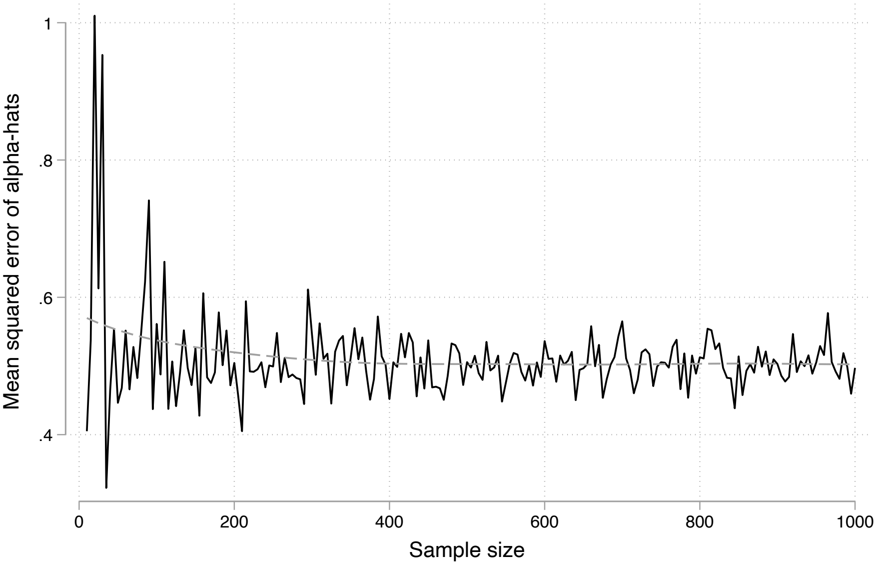
$$\text{Squared error of } \widehat{\beta}_{\text{OLS}} = \left(\widehat{\beta}_{\text{OLS}} - \beta\right)^2$$

$$\text{Mean squared error of } \widehat{\alpha}_{\text{OLS}} = \frac{1}{J}\sum_{j=1}^{J}\left(\widehat{\alpha}_{j,\text{OLS}} - \alpha_j\right)^2$$

$\widehat{\beta}_{\text{OLS}}$ Converges to True Value



But $\widehat{\alpha}_{j,\text{OLS}}$ Does Not



And so $\text{Corr}(\widehat{\alpha}_{j,\text{OLS}}, \alpha_j)$ Never Goes to One

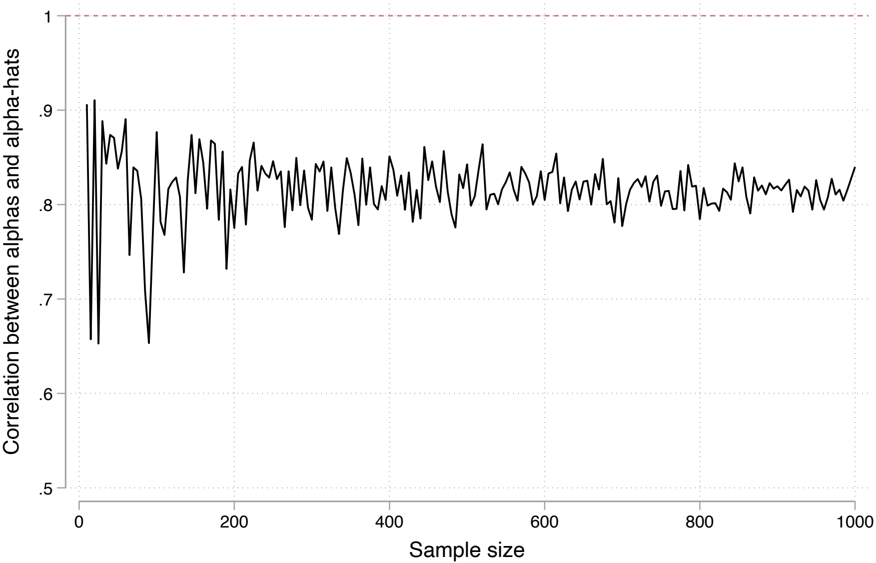


Table of Contents

1 Fixed Effects
- Definitions
- Identification and Estimation
- Empirical Bayes

2 Panel Data
- Structure and Definitions

3 Applications
- Connected-Set Fixed Effects (AKM)

4 Appendix
- Group Fixed Effects

Shrinking Group Means Towards Grand Mean

- How can we make many group means (or other coefficients) less noisy?
- Consider wages Y_{ij} for $N_j = N$ workers in J regions

$$Y_{ij} = \alpha_j + \varepsilon_{ij}$$

- Let $\bar{Y}_j = \frac{1}{N}\sum_{i=1}^{N} Y_{ij}$ be regional average wage
- Want to know α_j that $\bar{Y}_j$ would reach when N large
- Impose parametric restrictions on distribution of $\alpha_j \rightarrow$ random effects
- Assume

$$\bar{Y}_j | \alpha_j \sim N\left(\alpha_j, \frac{\sigma_\varepsilon^2}{N}\right)$$

  where $\sigma_\varepsilon^2 = Var(\varepsilon_{ij} | \alpha_j)$ and

$$\alpha_j \sim N\left(\mu, \sigma_\alpha^2\right)$$

- Sometimes referred to as mixing distribution

Posterior of Random Effects

- It can be shown that in this model:

$$\mathbb{E}\left[\alpha_j | \bar{Y}_j\right] = \mu + \left(\frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{N}}\right)\left(\bar{Y}_j - \mu\right)$$

- Intuition:
  - Linear projection of α_j on $\bar{Y}_j$ (infeasible)
  - Shrink sample average towards prior mean μ
  - More weight on $\bar{Y}_j$ if signal-to-noise ratio $\lambda = \frac{\sigma_\alpha^2}{\sigma_\alpha^2 + \frac{\sigma_\varepsilon^2}{N}}$ is high

- Problem: not useful if we don't know $\left(\sigma_\alpha^2, \sigma_\varepsilon^2, \mu\right)$
- Two approaches:
  1. Bayes: choose $\left(\sigma_\alpha^2, \mu\right)$ based on prior knowledge
  2. Empirical Bayes: estimate these hyperparameters

Empirical Bayes Approach

- Estimate μ with grand mean:

$$\widehat{\mu} = \frac{1}{J}\sum_{j=1}^{J} \bar{Y}_j$$

- Estimate σ_ε^2 from the "within" variance:

$$\widehat{\sigma}_\varepsilon^2 = \frac{1}{J(N-1)}\sum_{j=1}^{J}\sum_{i=1}^{N}\left(Y_{ij} - \bar{Y}_j\right)^2$$

- Estimate σ_α^2 from "overdispersion" in averages:

$$\widehat{\sigma}_\alpha^2 = \frac{1}{J-1}\sum_{j=1}^{J}\left(\bar{Y}_j - \widehat{\mu}\right)^2 - \frac{\widehat{\sigma}_\varepsilon^2}{N}$$

- Thus, the empirical Bayes posterior estimate:

$$\widehat{\mathbb{E}}\left[\alpha_j | \bar{Y}_j\right] = \widehat{\alpha}_j = \widehat{\mu} + \frac{\widehat{\sigma}_\alpha^2}{\widehat{\sigma}_\alpha^2 + \frac{\widehat{\sigma}_\varepsilon^2}{N}}\left(\bar{Y}_j - \widehat{\mu}\right)$$

When is Empirical Bayes Useful?

- Why might $\widehat{\alpha}_j$ be a "better" estimate for α_j than $\bar{Y}_j$?
- Bias-variance tradeoff: It can be shown for large N (James-Stein 1961):

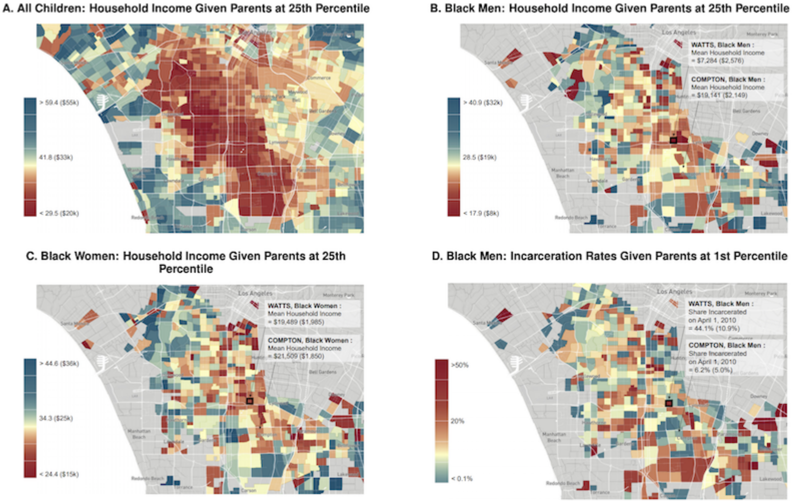
$$\mathbb{E}\left[\left(\widehat{\alpha}_j - \alpha_j\right)^2 |\alpha_j\right] \approx \underbrace{\left(\mu - \alpha_j\right)^2 (1-\lambda)^2}_{\text{bias}} + \underbrace{\lambda^2 \frac{\sigma_\varepsilon^2}{N}}_{\text{variance}}$$

  and if $J \geq 4$, $(\widehat{\alpha}_1, ..., \widehat{\alpha}_J)$ has lower MSE than $(\bar{Y}_1, ..., \bar{Y}_J)$ for $(\alpha_1, ..., \alpha_J)$
- Meaning: worth accepting some bias for lower variance

- Bottom line: shrink when want to estimate many means
- When should I estimate many means? For example:
  - Forecasting / prediction (e.g. map of small area statistics)
  - Input into subsequent regressions
  - Decision making (e.g. which teachers should be fired?)

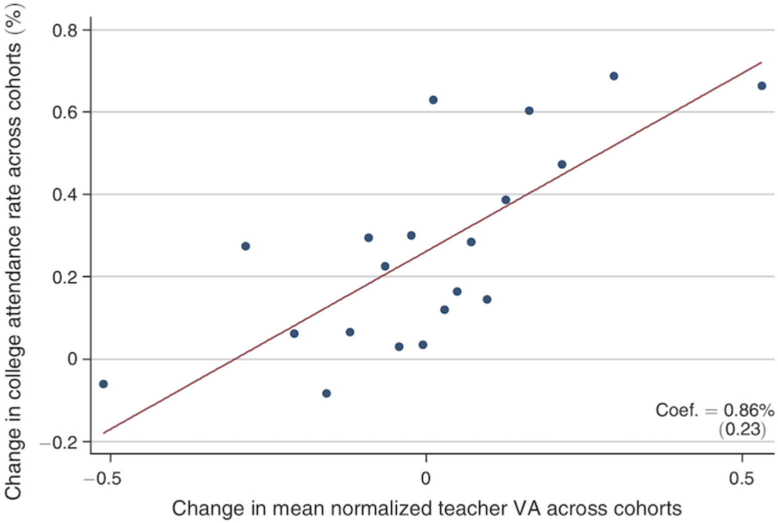
Prediction: Mapping Small Areas

Chetty et al. (2020)



Input into Regressions: Teacher Value Added

Chetty et al. (2014)



Decision Rule: Auditing Discriminating Employers

Kline and Walters (2020)

Figure V: Detection/error tradeoffs, NPRS data

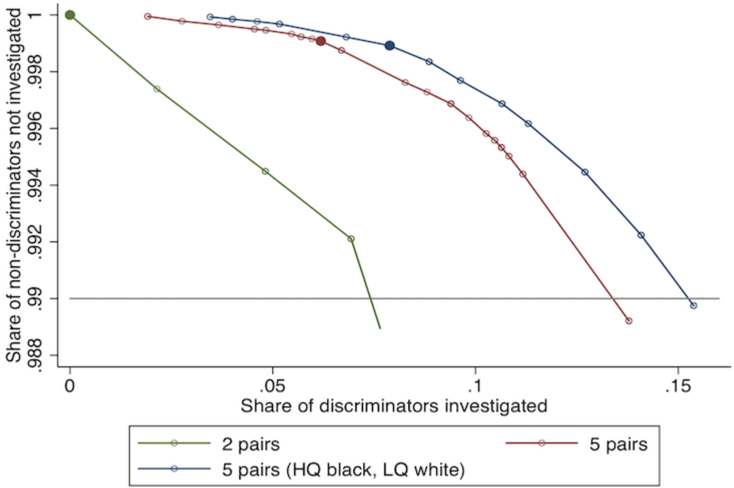


Table of Contents

1 Fixed Effects
   Definitions
   Identification and Estimation
   Empirical Bayes

2 Panel Data
   Structure and Definitions

3 Applications
   Connected-Set Fixed Effects (AKM)

4 Appendix
   Group Fixed Effects

Panel Data Formats

- Careful! We now switch notation:
- Let *units* be indexed by $i = 1, \ldots, N$
- And *periods* indexed by $t = 1, \ldots, T$
- Together, they form *observations* (i, t)

- Two panel data arrangements ("formats"): *Long* and *wide*

Long Panel

i	t	Y_{it}	X_{it}
1	1	5	1
1	2	4	0
1	3	3	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
1	T	4	1
2	1	3	0
2	2	4	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$

Wide Panel

i	Y_{i1}	Y_{i2}	Y_{i3}	$\cdots$	Y_{iT}	X_{i1}	X_{i2}	X_{i3}	$\cdots$
1	5	4	3	$\cdots$	4	1	0	0	$\cdots$
2	3	4	$\cdots$	$\cdots$	$\cdots$	0	1	$\cdots$	$\cdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

- Panel data can be *reshaped* between long and wide
- Each format has its uses

Panel Completeness

- Panel data exist to varying degrees of completeness
- $N \times T$ wide panel $\mathbf{Y}$ with $\circ$ observed and $\bullet$ missing
- In a *strongly balanced* panel, we can observe every entry:

$$\mathbf{Y} = \begin{bmatrix} \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ \\ \circ & \circ & \circ & \circ & \circ & \circ & \circ \end{bmatrix}$$

- *Weakly balanced* panels may or may not have gaps:

$$\begin{bmatrix} \bullet & \bullet & \circ & \circ & \circ & \circ & \bullet \\ \circ & \circ & \circ & \circ & \bullet & \bullet & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \bullet & \bullet \\ \bullet & \bullet & \circ & \circ & \circ & \circ & \bullet \end{bmatrix} \text{ and } \begin{bmatrix} \bullet & \bullet & \circ & \circ & \circ & \circ & \bullet \\ \bullet & \bullet & \circ & \circ & \bullet & \circ & \circ \\ \circ & \circ & \circ & \bullet & \circ & \bullet & \bullet \\ \bullet & \circ & \circ & \circ & \circ & \bullet & \bullet \end{bmatrix}$$

Unbalanced Panels and Repeated Cross-Sections

- *Unbalanced* panels:

$$\begin{bmatrix} \bullet & \bullet & \circ & \bullet & \bullet & \circ & \bullet \\ \bullet & \bullet & \bullet & \circ & \bullet & \circ & \circ \\ \circ & \circ & \circ & \bullet & \circ & \bullet & \bullet \\ \circ & \circ & \circ & \circ & \circ & \bullet & \bullet \end{bmatrix}$$

- *Repeated cross-sections*:

$$\begin{bmatrix} \bullet & \bullet & \bullet & \bullet & \bullet & \circ & \bullet \\ \bullet & \bullet & \bullet & \bullet & \bullet & \circ & \bullet \\ \bullet & \bullet & \circ & \bullet & \bullet & \bullet & \bullet \\ \circ & \bullet & \bullet & \bullet & \bullet & \bullet & \bullet \end{bmatrix}$$

  which we could treat as a *pooled cross-section*

Cohort Panels

- Can often form balanced panel from repeated cross-section
- How? Define *cohorts* $c \in \{1, ..., C\}$ e.g. birth year

i	t	Y_{it}	X_{it}
1	1	5	2
2	1	4	1
3	1	3	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$
i	2	4	2
$\vdots$	$\vdots$	3	3
N	T	4	1

c	t	$\bar{Y}_{ct}$
1	1	4.2
1	2	3.8
$\vdots$	$\vdots$	$\vdots$
1	T	4.1
2	2	3.4
$\vdots$	$\vdots$	$\vdots$
C	T	3.2

Table of Contents

1 Fixed Effects
- Definitions
- Identification and Estimation
- Empirical Bayes

2 Panel Data
- Structure and Definitions

3 Applications
- Connected-Set Fixed Effects (AKM)

4 Appendix
- Group Fixed Effects

Units Moving Across Groups over Time

- Extremely useful to observe i in different groups over t
- Separately identifies unit from group effect

- Consider the following model (Abowd et al. 1999, “AKM”)

$$Y_{it} = \alpha_i + \psi_{\mathbf{J}(i,t)} + \mathbf{X}_{it}'\beta + \varepsilon_{it}$$

where $\mathbf{J} : \mathcal{N} \times \mathcal{T} \to \mathcal{J}$ assigns i in t to $j \in \mathcal{J} = \{1, ..., J\}$

- Example: firm of workers in a panel
- Model is isomorphic to standard model but J treatments
- Interested in ψ_j or $Var(\psi_j)$
- E.g. how much variation in wages explained by firms?

Identification of AKM

- Stacked model:

$$\mathbf{Y} = \mathbf{D}\alpha + \mathbf{F}\psi + \mathbf{X}\xi + \varepsilon$$

- Identification requires

$$\mathbb{E}\left[\mathbf{D}'\varepsilon\right] = 0, \quad \mathbb{E}\left[\mathbf{F}'\varepsilon\right] = 0, \quad \mathbb{E}\left[\mathbf{X}'\varepsilon\right] = 0$$

- Rank condition: set one $\psi_j = 0$ in each “connected set”
- “Exogenous mobility”: $\mathbb{E}\left[\mathbf{F}'\varepsilon\right] = 0$
    - $\Pr\left(\mathbf{J}(i,t) = j|\varepsilon\right) = \Pr\left(\mathbf{J}(i,t) = j\right) = G_{jt}\left(\alpha_i, \psi_1, ..., \psi_J\right)$
    - Does *not* mean that workers must move randomly across firms
    - It does also *not* require turnover to be random across *i*
    - But instead, requires no sorting due to wage trends or match quality

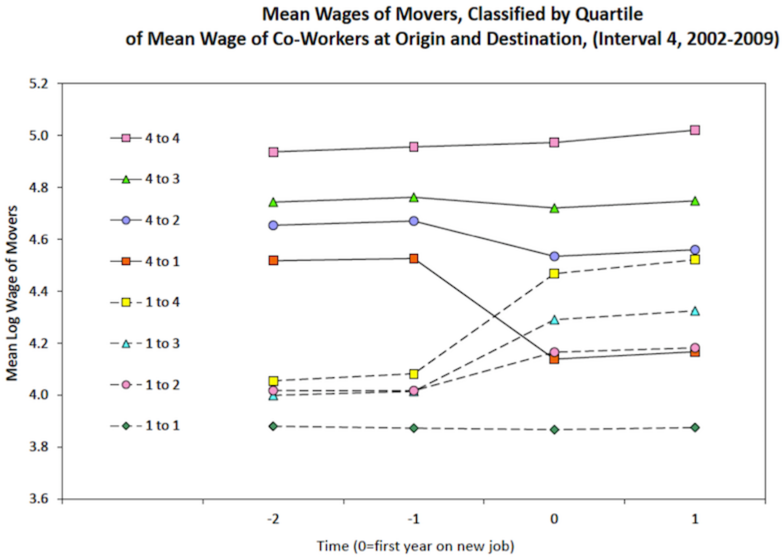
Estimation of AKM

- Inspect stacked model: can estimate via OLS!
  - Worker dummies to estimate α_i
  - Firm dummies to estimate ψ_j
- Recall: $(\widehat{\alpha}, \widehat{\psi})$ are inconsistent unless T gets large
- Can estimate $Var\left(\widehat{\psi}_j\right) = \widehat{\sigma}_\psi^2$ via Method of Moments:

$$\widehat{\sigma}_\psi^2 = \frac{1}{NT - 1}\widehat{\psi}'\mathbf{F}'\mathbf{Q}\mathbf{F}\widehat{\psi}$$

  where **Q** is demeaning matrix
- But this estimator is upward biased
- Unbiased estimate via leave-one-out (Kline et al. 2020)

Ruling Out Moving on Trends or Match Effects: Card et al. (2013)



Rising Share of Wage Variation due to Firms

TABLE III

Estimation Results for AKM Model, Fit by Interval

	(1) Interval 1 1985–1991	(2) Interval 2 1990–1996	(3) Interval 3 1996–2002	(4) Interval 4 2002–2009
Person and establishment parameters				
Number person effects	16,295,106	17,223,290	16,384,815	15,834,602
Number establishment effects	1,221,098	1,357,824	1,476,705	1,504,095
Summary of parameter estimates				
Std. dev. of person effects (across person-year obs.)	0.289	0.304	0.327	0.357
Std. dev. of establ. Effects (across person-year obs.)	0.159	0.172	0.194	0.230
Std. dev. of Xb (across person-year obs.)	0.121	0.088	0.093	0.084
Correlation of person/establ. Effects (across person-year obs.)	0.034	0.097	0.169	0.249
Correlation of person effects/Xb (across person-year obs.)	−0.051	−0.102	−0.063	0.029
Correlation of establ. effects/Xb (across person-year obs.)	0.057	0.039	0.050	0.112
RMSE of AKM residual	0.119	0.121	0.130	0.135
Adjusted R-squared	0.896	0.901	0.909	0.927
Comparison match model				
RMSE of match model	0.103	0.105	0.108	0.112
Adjusted R^2	0.922	0.925	0.937	0.949
Std. dev. of match effect*	0.060	0.060	0.072	0.075
Addendum				
Std. dev. log wages	0.370	0.384	0.432	0.499
Sample size	84,185,730	88,662,398	83,699,582	90,615,841

Table of Contents

1 Fixed Effects
  - Definitions
  - Identification and Estimation
  - Empirical Bayes

2 Panel Data
  - Structure and Definitions

3 Applications
  - Connected-Set Fixed Effects (AKM)

4 Appendix
  - Group Fixed Effects

Grouping Unobserved Heterogeneity

- Consider model with unobserved groups:

$$Y_{it} = \phi_{g(i),t} + D_{it}\beta + \varepsilon_{it}$$

where $g : \{1, ..., N\} \rightarrow \{1, ..., G\}$ is not known

- If we knew β and $\boldsymbol{\phi} = [\phi_{1,1}, ..., \phi_{G,T}]$, we could estimate:

$$\widehat{g}(i; \beta, \boldsymbol{\phi}) = \arg \min_{g=\{1,...,G\}} \sum_{t=1}^{T} (Y_{it} - D_{it}\beta - \phi_{g,t})^2$$

- Since we don't know them, we estimate:

$$\left(\widehat{\beta}, \widehat{\boldsymbol{\phi}}, \widehat{\gamma}\right) = \arg \min_{(\beta, \boldsymbol{\phi}, \gamma)} \sum_{i=1}^{N} \sum_{t=1}^{T} (Y_{it} - D_{it}\beta - \phi_{g,t})^2$$

for $\gamma = [g(1), ..., g(N)]$

Bonhomme and Manresa (2015): Group FE