# Linear Models Lecture 17: Pooled and Panel Data

Robert Gulotty

University of Chicago

February 26, 2026

# Where We Are: From GMM to Panel Data

- In Lectures 15–16 we developed **GMM** as a unified framework:
  - Specify moment conditions $E[\boldsymbol{g}(\boldsymbol{z}_i, \boldsymbol{\theta}_0)] = \mathbf{0}$
  - Estimate $\hat{\boldsymbol{\theta}}$ by minimizing $\hat{\boldsymbol{g}}' \boldsymbol{W} \hat{\boldsymbol{g}}$
  - Test overidentifying restrictions with the $J$-statistic
- **Panel data** introduces new structure:
  - Repeated observations create natural moment conditions
  - Unobserved heterogeneity $\alpha_i$ must be controlled
  - Different assumptions $\Rightarrow$ different moment conditions $\Rightarrow$ different estimators
- **This lecture:** pooled OLS, fixed effects, between estimators as moment-based estimators. **Next lecture:** random effects, CRE, and dynamic panel GMM (Arellano-Bond).

# Where We Are: From GMM to Panel Data

- In Lectures 15–16 we developed **GMM** as a unified framework:
  - Specify moment conditions $E[\boldsymbol{g}(\boldsymbol{z}_i, \boldsymbol{\theta}_0)] = \boldsymbol{0}$
  - Estimate $\hat{\boldsymbol{\theta}}$ by minimizing $\hat{\boldsymbol{g}}' \boldsymbol{W} \hat{\boldsymbol{g}}$
  - Test overidentifying restrictions with the $J$-statistic
- **Panel data** introduces new structure:
  - Repeated observations create natural moment conditions
  - Unobserved heterogeneity $\alpha_i$ must be controlled
  - Different assumptions $\Rightarrow$ different moment conditions $\Rightarrow$ different estimators
- **This lecture:** pooled OLS, fixed effects, between estimators as moment-based estimators. **Next lecture:** random effects, CRE, and dynamic panel GMM (Arellano-Bond).

# Where We Are: From GMM to Panel Data

- In Lectures 15–16 we developed **GMM** as a unified framework:
  - Specify moment conditions $E[\boldsymbol{g}(\boldsymbol{z}_i, \boldsymbol{\theta}_0)] = \mathbf{0}$
  - Estimate $\hat{\boldsymbol{\theta}}$ by minimizing $\hat{\boldsymbol{g}}'\boldsymbol{W}\hat{\boldsymbol{g}}$
  - Test overidentifying restrictions with the $J$-statistic
- **Panel data** introduces new structure:
  - Repeated observations create natural moment conditions
  - Unobserved heterogeneity $\alpha_i$ must be controlled
  - Different assumptions $\Rightarrow$ different moment conditions $\Rightarrow$ different estimators
- **This lecture:** pooled OLS, fixed effects, between estimators as moment-based estimators.
  **Next lecture:** random effects, CRE, and dynamic panel GMM (Arellano-Bond).

# Panel Estimators as Moment Conditions

| Estimator | Moment Condition | Key Assumption |
|---|---|---|
| Pooled OLS | $E[\boldsymbol{X}_{it}' e_{it}] = \mathbf{0}$ | $\alpha_i$ uncorrelated with $\boldsymbol{X}$ |
| Fixed Effects | $E[\tilde{\boldsymbol{X}}_{it}' \tilde{e}_{it}] = \mathbf{0}$ | Strict exogeneity (within-unit) |
| Random Effects | $E[\boldsymbol{X}_{it}' \nu_{it}] = \mathbf{0}$ | $\alpha_i$ uncorrelated with $\boldsymbol{X}$; GLS weighting |
| Arellano-Bond | $E[Y_{is} \cdot \Delta e_{it}] = 0, \; s \leq t-2$ | Sequential exogeneity; lagged levels as instruments |

- Each row is a **GMM estimator** with different instruments and assumptions.
- More moment conditions $\Rightarrow$ overidentification $\Rightarrow$ testable with the $J$-test.

# Panel Estimators as Moment Conditions

| Estimator | Moment Condition | Key Assumption |
|---|---|---|
| Pooled OLS | $E[\boldsymbol{X}'_{it} e_{it}] = \mathbf{0}$ | $\alpha_i$ uncorrelated with $\boldsymbol{X}$ |
| Fixed Effects | $E[\tilde{\boldsymbol{X}}'_{it} \tilde{e}_{it}] = \mathbf{0}$ | Strict exogeneity (within-unit) |
| Random Effects | $E[\boldsymbol{X}'_{it} \nu_{it}] = \mathbf{0}$ | $\alpha_i$ uncorrelated with $\boldsymbol{X}$; GLS weighting |
| Arellano-Bond | $E[Y_{is} \cdot \Delta e_{it}] = 0, \; s \leq t-2$ | Sequential exogeneity; lagged levels as instruments |

- Each row is a **GMM estimator** with different instruments and assumptions.
- More moment conditions $\Rightarrow$ overidentification $\Rightarrow$ testable with the $J$-test.

# Longitudinal Data

- Longitudinal data describes two dimensions, usually time and individuals.
- Examples:
  - Exit Polls/ CPS/ ANES: people are surveyed following each election.
  - Dyadic data in peace science literature.
  - Election forecasting.
- Data can be stacked to make a data matrix with $N_1 + N_2 + \cdots + N_T$ rows.
- This structure can be incorporated into our statistical model.

# Using time in our analysis

- Time can be treated as an indicator variable (as dummy variables or cumulatively), or as a continuous variable.
- Time period interactions can evaluate changes in slope coefficients across time (the effect of a free fedora on employment may be different today than in 1940).

# Differences in Differences

- We might collect data over time to rule out certain confounding factors.
- Consider a study of the effect (D) of a voter ID law on turnout (Y), observing Texas and Oklahoma before and after the law.
  - Comparison one:

| State | Outcome |
|---|---|
| Texas | $Y = M + D$ |
| Oklahoma | $Y = O$ |

  - Comparison two:

| State | Time | Outcome |
|---|---|---|
| Texas | Before | $Y = M$ |
| | After | $Y = M + (T + D)$ |

- Neither comparison on its own recovers D.

# Differences in Differences

- We might collect data over time to rule out certain confounding factors.
- Consider a study of the effect (D) of a voter ID law on turnout (Y), observing Texas and Oklahoma before and after the law.
  - Comparison one:

| State | Outcome |
| --- | --- |
| Texas | $Y = M + D$ |
| Oklahoma | $Y = O$ |

  - Comparison two:

| State | Time | Outcome |
| --- | --- | --- |
| Texas | Before | $Y = M$ |
| | After | $Y = M + (T + D)$ |

- Neither comparison on its own recovers D.

# Differences in Differences

- We might collect data over time to rule out certain confounding factors.
- Consider a study of the effect (D) of a voter ID law on turnout (Y), observing Texas and Oklahoma before and after the law.
  - Comparison one:

| State | Outcome |
|---|---|
| Texas | $Y = M + D$ |
| Oklahoma | $Y = O$ |

  - Comparison two:

| State | Time | Outcome |
|---|---|---|
| Texas | Before | $Y = M$ |
| | After | $Y = M + (T + D)$ |

- Neither comparison on its own recovers D.

# Differences in Differences

- We might collect data over time to rule out certain confounding factors.
- Consider a study of the effect (D) of a voter ID law on turnout (Y), observing Texas and Oklahoma before and after the law.
  - Comparison one:

| State | Outcome |
|---|---|
| Texas | $Y = M + D$ |
| Oklahoma | $Y = O$ |

  - Comparison two:

| State | Time | Outcome |
|---|---|---|
| Texas | Before | $Y = M$ |
| | After | $Y = M + (T + D)$ |

- Neither comparison on its own recovers D.

# Diff in Diff

| State | Time | Outcome | After-Before | dTexas-dOklahoma |
|---|---|---|---|---|
| Texas | Before | $Y = M$ | | |
| | After | $Y = M + T + D$ | T+D | |
| Oklahoma | Before | $Y = O$ | | |
| | After | $Y = O + T$ | T | |
| | | | | D |

$$D = dY_T - dY_O = (Y_{A,T} - Y_{B,T}) - (Y_{A,O} - Y_{B,O})$$

# Potential Outcomes

- Define $D$ as the binary treatment $d \in \{0, 1\}$ which tells us if we will be in the treatment category once we get to the second period.
- Measurements happen twice, $t \in \{0, 1\}$
- $Y_t^d$ is the outcome in period $t$ for treatment status $d$.
- The outcomes we can observe is $Y_t$ (no d).

$$Y_t = dY_t^1 + (1 - d)Y_t^0$$

- Counterfactual outcomes:
  - $Y_t^1 | D_t = 0$ (Outcomes in a world of treatment among those who weren't)
  - $Y_t^0 | D_t = 1$ (Outcomes in a world of control among those who were treated)

$$\delta_i = Y_t^1 - Y_t^0$$

# Parallel Trends

- Identification assumption/Parallel Trends

$$E[Y^0_{t=1}|D=1] - E[Y^0_{t=0}|D=1] = E[Y^0_{t=1}|D=0] - E[Y^0_{t=0}|D=0]$$

- In previous example, T is the same across states, state level differences ($M - O$) don't vary over time.
- The differences in expected potential non-treatment outcomes are unrelated to belonging to the treated or control group in the post treatment period.
- If the treated had not been subjected to treatment, both subpopulations would have experienced the same time trends.
- The composition cannot be affected by treatment

# Example Obama Support

- Suppose we want to study the effect of racial co-identification (non-identification) on feeling thermometers of Obama in 2008.

$$ObamaRating_i = b_0 + d_{08} Black_i + \boldsymbol{b_x'} \boldsymbol{x}_i + e_i$$

- However, we are worried about the two kinds of omitted factors:
  - Time trends (T): Obama ran after a financial crisis and war that might make him popular generally.
  - Group Differences (M-O): Black and white people might differ in their preferences over any president that is a relatively liberal northern antiwar Democratic senator.
- Suppose Obama's appeal (repulsion) among Black (white) people depends on the economic conditions in 2008, his being a liberal Senator, and not only his race?

# Example Obama Support

- Suppose we want to study the effect of racial co-identification (non-identification) on feeling thermometers of Obama in 2008.

$$ObamaRating_i = b_0 + d_{08} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- However, we are worried about the two kinds of omitted factors:
  - Time trends (T): Obama ran after a financial crisis and war that might make him popular generally.
  - Group Differences (M-O): Black and white people might differ in their preferences over any president that is a relatively liberal northern antiwar Democratic senator.
- Suppose Obama's appeal (repulsion) among Black (white) people depends on the economic conditions in 2008, his being a liberal Senator, and not only his race?

# Example Obama Support

- Suppose we want to study the effect of racial co-identification (non-identification) on feeling thermometers of Obama in 2008.

$$ObamaRating_i = b_0 + d_{08} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- However, we are worried about the two kinds of omitted factors:
  - Time trends (T): Obama ran after a financial crisis and war that might make him popular generally.
  - Group Differences (M-O): Black and white people might differ in their preferences over any president that is a relatively liberal northern antiwar Democratic senator.
- Suppose Obama's appeal (repulsion) among Black (white) people depends on the economic conditions in 2008, his being a liberal Senator, and not only his race?

# Example Obama Support

- Suppose we want to study the effect of racial co-identification (non-identification) on feeling thermometers of Obama in 2008.

$$ObamaRating_i = b_0 + d_{08} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- However, we are worried about the two kinds of omitted factors:
  - Time trends (T): Obama ran after a financial crisis and war that might make him popular generally.
  - Group Differences (M-O): Black and white people might differ in their preferences over any president that is a relatively liberal northern antiwar Democratic senator.
- Suppose Obama's appeal (repulsion) among Black (white) people depends on the economic conditions in 2008, his being a liberal Senator, and not only his race?

# Example Obama Support

- Suppose we want to study the effect of racial co-identification (non-identification) on feeling thermometers of Obama in 2008.

$$ObamaRating_i = b_0 + d_{08} Black_i + \boldsymbol{b}'_{\boldsymbol{x}} \boldsymbol{x}_i + e_i$$

- However, we are worried about the two kinds of omitted factors:
  - Time trends (T): Obama ran after a financial crisis and war that might make him popular generally.
  - Group Differences (M-O): Black and white people might differ in their preferences over any president that is a relatively liberal northern antiwar Democratic senator.
- Suppose Obama's appeal (repulsion) among Black (white) people depends on the economic conditions in 2008, his being a liberal Senator, and not only his race?

# Example Obama Support

- We can use a repeated cross section to difference out similar attitudes toward another Democratic candidate.

$$Kerry_i = b_0 + d_{04} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- If we assume that
  - All time varying factors other than the candidate's race affected Black and white people's response to Democratic nominees equally.
  - All differences in evaluations of D. nominees between Black and white people did not change between 04 and 08.
- Then $\delta$ is an estimate of the Average Treatment Effect of the candidate's race on attitudes.

$$\delta = d_{08} - d_{04} = (\bar{r}_{08,B} - \bar{r}_{08,W}) - (\bar{r}_{04,B} - \bar{r}_{04,W})$$

# Example Obama Support

- We can use a repeated cross section to difference out similar attitudes toward another Democratic candidate.

$$Kerry_i = b_0 + d_{04} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- If we assume that
  - All time varying factors other than the candidate's race affected Black and white people's response to Democratic nominees equally.
  - All differences in evaluations of D. nominees between Black and white people did not change between 04 and 08.
- Then $\delta$ is an estimate of the Average Treatment Effect of the candidate's race on attitudes.

$$\delta = d_{08} - d_{04} = (\bar{r}_{08,B} - \bar{r}_{08,W}) - (\bar{r}_{04,B} - \bar{r}_{04,W})$$

# Example Obama Support

- We can use a repeated cross section to difference out similar attitudes toward another Democratic candidate.

$$Kerry_i = b_0 + d_{04} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- If we assume that
  - All time varying factors other than the candidate's race affected Black and white people's response to Democratic nominees equally.
  - All differences in evaluations of D. nominees between Black and white people did not change between 04 and 08.
- Then $\delta$ is an estimate of the Average Treatment Effect of the candidate's race on attitudes.

$$\delta = d_{08} - d_{04} = (\bar{r}_{08,B} - \bar{r}_{08,W}) - (\bar{r}_{04,B} - \bar{r}_{04,W})$$

# Example Obama Support

- We can use a repeated cross section to difference out similar attitudes toward another Democratic candidate.

$$Kerry_i = b_0 + d_{04} Black_i + \boldsymbol{b}_{\boldsymbol{x}}' \boldsymbol{x}_i + e_i$$

- If we assume that
  - All time varying factors other than the candidate's race affected Black and white people's response to Democratic nominees equally.
  - All differences in evaluations of D. nominees between Black and white people did not change between 04 and 08.
- Then $\delta$ is an estimate of the Average Treatment Effect of the candidate's race on attitudes.

$$\delta = d_{08} - d_{04} = (\bar{r}_{08,B} - \bar{r}_{08,W}) - (\bar{r}_{04,B} - \bar{r}_{04,W})$$

# Estimating Differences in Differences

- Stack 04 and 08 data and compare two linear models:
- Define $yr08_i = 1$ for the 2008 survey takers and $B_i = \mathbb{I}(\text{Race}=\text{"Black"})$

$$DRating_i = b_{0,08} yr08_i + b_{0,04}(1 - yr08_i) + d_{08} yr08_i B_i + d_{04}(1 - yr08_i) B_i + \mathbf{b}_x' \mathbf{x}_i + e_i$$
$$DRating_i = b_{0,08} yr08_i + b_{0,04}(1 - yr08_i) + dB_i + \mathbf{b}_x' \mathbf{x}_i + e_i$$

- Then an F test tells us the effect of racial co-identification or non-identification.

# Estimating Differences in Differences

- Stack 04 and 08 data and compare two linear models:
- Define $yr08_i = 1$ for the 2008 survey takers and $B_i = \mathbb{I}(\text{Race=''Black''})$

$$DRating_i = b_{0,08} yr08_i + b_{0,04}(1 - yr08_i) + d_{08} yr08_i B_i + d_{04}(1 - yr08_i) B_i + \boldsymbol{b_x'} \boldsymbol{x}_i + e_i$$
$$DRating_i = b_{0,08} yr08_i + b_{0,04}(1 - yr08_i) + dB_i + \boldsymbol{b_x'} \boldsymbol{x}_i + e_i$$

- Then an F test tells us the effect of racial co-identification or non-identification.

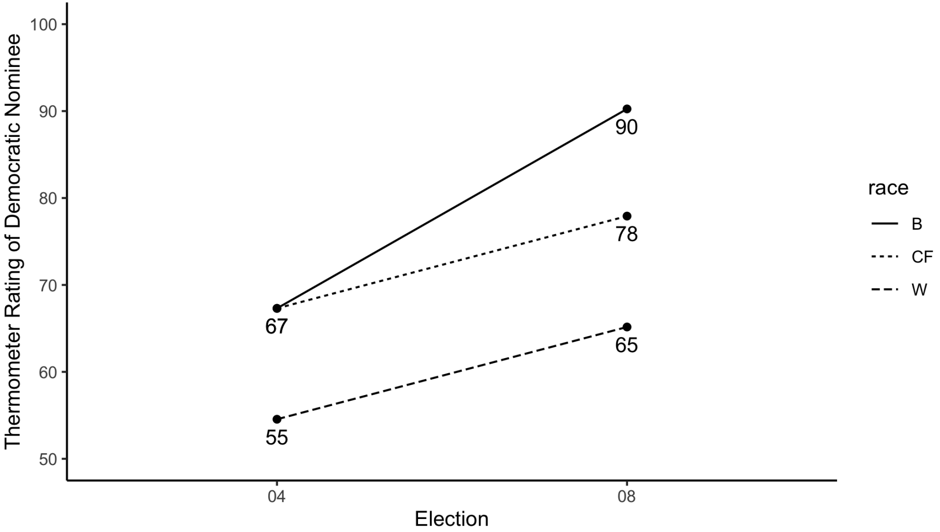
# Estimating Differences in Differences

- Stack 04 and 08 data and compare two linear models:
- Define $yr08_i = 1$ for the 2008 survey takers and $B_i = \mathbb{I}(\text{Race}=\text{"Black"})$

$$DRating_i = b_{0,08} yr08_i + b_{0,04}(1 - yr08_i) + d_{08} yr08_i B_i + d_{04}(1 - yr08_i) B_i + \boldsymbol{b}'_{\boldsymbol{x}} \boldsymbol{x}_i + e_i$$
$$DRating_i = b_{0,08} yr08_i + b_{0,04}(1 - yr08_i) + dB_i + \boldsymbol{b}'_{\boldsymbol{x}} \boldsymbol{x}_i + e_i$$

- Then an F test tells us the effect of racial co-identification or non-identification.

| | demft | | |
|---|---|---|---|
| I(1 - y08) | 54.5*** | 53.0*** | |
| | (1.4) | (1.4) | |
| y08 | 65.2*** | 65.9*** | 10.6*** |
| | (1.3) | (1.3) | (1.0) |
| black | | 21.7*** | 12.8*** |
| | | (1.0) | (1.9) |
| I(1 - y08):black | 12.8*** | | |
| | (1.9) | | |
| y08:black | 25.1*** | | 12.3*** |
| | (1.2) | | (2.3) |
| age | 0.02 | 0.02 | 0.02 |
| | (0.02) | (0.02) | (0.02) |
| Constant | | | 54.5*** |
| | | | (1.4) |


Thermometer Rating of Democratic Nominee
100
90
80
70
60
50
04
08
Election
90
78
67
65
55
race
B
CF
W

# Policy Evaluation

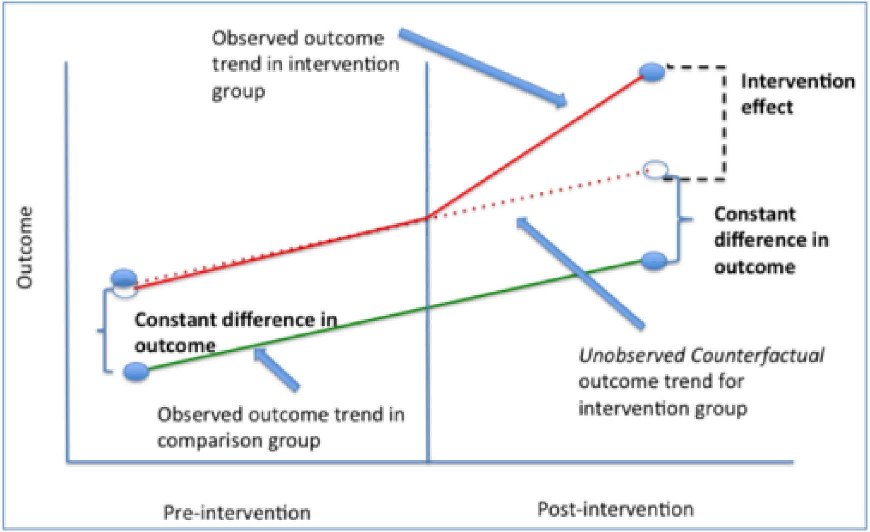
- Diff-in-diff is often used to study binary policies when we theorize that selection bias is driven by fixed group features.
- Key assumption for causal inference: parallel trends:
  - The amount of uncontrolled selection bias is not changing over time.
  - Time trends are the same across groups.
- Example of violation of parallel trends: Suppose that Texas and Oklahoma are trending blue, but Texas is closer to flipping, causing both voter ID laws and more investment by Democrats.

# Policy Evaluation

- Diff-in-diff is often used to study binary policies when we theorize that selection bias is driven by fixed group features.
- Key assumption for causal inference: parallel trends:
  - The amount of uncontrolled selection bias is not changing over time.
  - Time trends are the same across groups.
- Example of violation of parallel trends: Suppose that Texas and Oklahoma are trending blue, but Texas is closer to flipping, causing both voter ID laws and more investment by Democrats.

# Policy Evaluation

- Diff-in-diff is often used to study binary policies when we theorize that selection bias is driven by fixed group features.
- Key assumption for causal inference: parallel trends:
  - The amount of uncontrolled selection bias is not changing over time.
  - Time trends are the same across groups.
- Example of violation of parallel trends: Suppose that Texas and Oklahoma are trending blue, but Texas is closer to flipping, causing both voter ID laws and more investment by Democrats.


Observed outcome trend in intervention group
Intervention effect
Outcome
Constant difference in outcome
Constant difference in outcome
Unobserved Counterfactual outcome trend for intervention group
Observed outcome trend in comparison group
Pre-intervention
Post-intervention

# Alternative: Synthetic Control

- When we are studying aggregates like states, we might not have parallel trends.
- However, in some cases we can construct a plausible counterfactual from combinations of untreated units.
- The synthetic control approach uses covariates to construct these counterfactual weighted combinations.
- Assumes that we have a long time series and the treated unit outcomes lie within the convex hull of untreated outcomes.

# Panel data

- The above analysis assumes that group compositions don't change.
- Suppose we can measure the same individuals.
- We can then account for variation within groups, across groups, and across time.

# Panel Example 1: Voting behavior

- Consider a population of families. Choose one at random.
- For each family member $t$, we observe voting decisions $Y_t$ and exposure to a persuasive messages $Z_t$.
- Each family has a household wealth $W$ and a set of stories passed down from their grandparents $A$.
- We observe $Y_t$, $Z_t$ and $W$, but not $A$.
- We want to estimate the effect of $Z_t$ on $Y_t$, holding constant $W$ and $A$.

# Panel Example 1: Voting behavior

- Consider a population of families. Choose one at random.
- For each family member $t$, we observe voting decisions $Y_t$ and exposure to a persuasive messages $Z_t$.
- Each family has a household wealth $W$ and a set of stories passed down from their grandparents $A$.
- We observe $Y_t$, $Z_t$ and $W$, but not $A$.
- We want to estimate the effect of $Z_t$ on $Y_t$, holding constant $W$ and $A$.

# Panel Example 1: Voting behavior

- Consider a population of families. Choose one at random.
- For each family member $t$, we observe voting decisions $Y_t$ and exposure to a persuasive messages $Z_t$.
- Each family has a household wealth $W$ and a set of stories passed down from their grandparents $A$.
- We observe $Y_t$, $Z_t$ and $W$, but not $A$.
- We want to estimate the effect of $Z_t$ on $Y_t$, holding constant $W$ and $A$.

# Panel Example 1: Voting behavior

- Consider a population of families. Choose one at random.
- For each family member $t$, we observe voting decisions $Y_t$ and exposure to a persuasive messages $Z_t$.
- Each family has a household wealth $W$ and a set of stories passed down from their grandparents $A$.
- We observe $Y_t$, $Z_t$ and $W$, but not $A$.
- We want to estimate the effect of $Z_t$ on $Y_t$, holding constant $W$ and $A$.

# Panel Example 1: Voting behavior

- Consider a population of families. Choose one at random.
- For each family member $t$, we observe voting decisions $Y_t$ and exposure to a persuasive messages $Z_t$.
- Each family has a household wealth $W$ and a set of stories passed down from their grandparents $A$.
- We observe $Y_t$, $Z_t$ and $W$, but not $A$.
- We want to estimate the effect of $Z_t$ on $Y_t$, holding constant $W$ and $A$.

# Panel Example 2: Airline cost structure

- Consider a population of airlines. Choose one at random.
- In each year, $t$, each airline pays $Y_t$ dollars to fly $Z_t$ passengers.
- Each airline uses a proprietary technology $A$ which does not vary over time.
- We observe N randomly selected Airlines but do not measure $A$.
- We want to estimate the way that $Y_t$ responds to $Z_t$, holding constant $A$.

# Panel Example 2: Airline cost structure

- Consider a population of airlines. Choose one at random.
- In each year, $t$, each airline pays $Y_t$ dollars to fly $Z_t$ passengers.
- Each airline uses a proprietary technology $A$ which does not vary over time.
- We observe N randomly selected Airlines but do not measure $A$.
- We want to estimate the way that $Y_t$ responds to $Z_t$, holding constant $A$.

# Panel Example 2: Airline cost structure

- Consider a population of airlines. Choose one at random.
- In each year, $t$, each airline pays $Y_t$ dollars to fly $Z_t$ passengers.
- Each airline uses a proprietary technology $A$ which does not vary over time.
- We observe N randomly selected Airlines but do not measure $A$.
- We want to estimate the way that $Y_t$ responds to $Z_t$, holding constant $A$.

# Panel Example 2: Airline cost structure

- Consider a population of airlines. Choose one at random.
- In each year, $t$, each airline pays $Y_t$ dollars to fly $Z_t$ passengers.
- Each airline uses a proprietary technology $A$ which does not vary over time.
- We observe N randomly selected Airlines but do not measure $A$.
- We want to estimate the way that $Y_t$ responds to $Z_t$, holding constant $A$.

# Panel Example 2: Airline cost structure

- Consider a population of airlines. Choose one at random.
- In each year, $t$, each airline pays $Y_t$ dollars to fly $Z_t$ passengers.
- Each airline uses a proprietary technology $A$ which does not vary over time.
- We observe N randomly selected Airlines but do not measure $A$.
- We want to estimate the way that $Y_t$ responds to $Z_t$, holding constant $A$.

# Panel Example 3: Candidate Electoral Performance

- Consider a population of candidates. Choose one at random.
- For each year $t$, the candidate receives vote share $Y_t$.
- We know each candidate has an underlying charisma $A$ which does not vary over time.
- We observe N randomly selected candidates, but do not measure $A$.
- We want to estimate the effect of $Y_{t-1}$ on $Y_t$, holding constant $A$.

# Panel Data

- The voters in a family, the airlines and candidates over time are all examples of *panel data*.
- Our observations are double indexed, usually divided into $N$ units and $T$ time periods/groups.
- If there are many $T$ but few $N$, we call the data long and narrow, (time series cross section).
- If every unit is observed in every period/ group, the data are called balanced.

# Panel Data

- The voters in a family, the airlines and candidates over time are all examples of *panel data*.
- Our observations are double indexed, usually divided into $N$ units and $T$ time periods/groups.
- If there are many $T$ but few $N$, we call the data long and narrow, (time series cross section).
- If every unit is observed in every period/ group, the data are called balanced.

# Panel Data

- The voters in a family, the airlines and candidates over time are all examples of *panel data*.
- Our observations are double indexed, usually divided into $N$ units and $T$ time periods/groups.
- If there are many $T$ but few $N$, we call the data long and narrow, (time series cross section).
- If every unit is observed in every period/ group, the data are called balanced.

# Panel Data

- The voters in a family, the airlines and candidates over time are all examples of *panel data*.
- Our observations are double indexed, usually divided into $N$ units and $T$ time periods/groups.
- If there are many $T$ but few $N$, we call the data long and narrow, (time series cross section).
- If every unit is observed in every period/ group, the data are called balanced.

# Example Long Narrow Data

- Greene analyzes the cost structure in the airline industry from 1970-84.
- Research question: do airlines have positive economies of scale?
- library(AER); data("USAirlines"); analyzes determinants of total costs for 6 airlines.
- The cost function follows a Cobb-Douglas specification. Taking logs yields a linear regression:

$$\ln(\text{cost}) = \beta_0 + \beta_1 \ln(\text{output}) + \beta_2 \ln(\text{fuel price}) + \beta_3 \text{load} + u$$

- *cost*: total costs of labor materials etc in USD 1000.
- *output*: revenue passenger mile index
- input *prices*: fuel prices
- *load*: share of seats sold on flights flown.

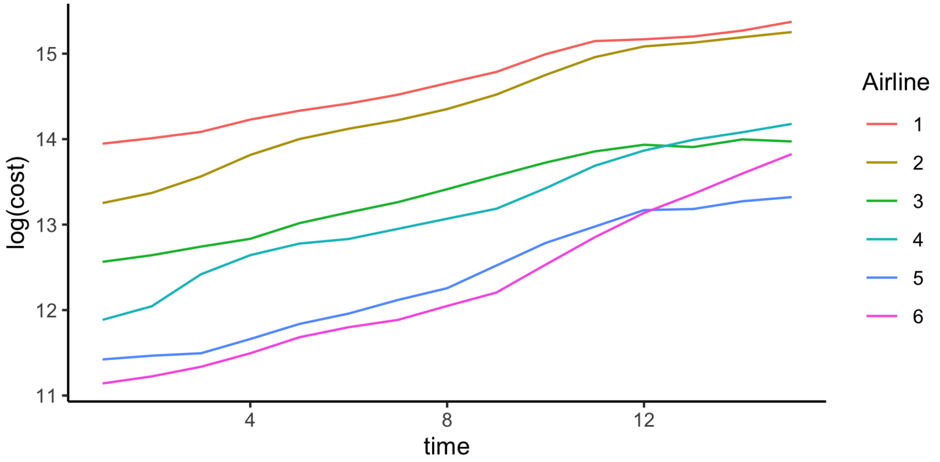
# Example Long Narrow Data

- Greene analyzes the cost structure in the airline industry from 1970-84.
- Research question: do airlines have positive economies of scale?
- library(AER); data("USAirlines"); analyzes determinants of total costs for 6 airlines.
- The cost function follows a Cobb-Douglas specification. Taking logs yields a linear regression:

$$\ln(\text{cost}) = \beta_0 + \beta_1 \ln(\text{output}) + \beta_2 \ln(\text{fuel price}) + \beta_3 \text{load} + u$$

- *cost*: total costs of labor materials etc in USD 1000.
- *output*: revenue passenger mile index
- input *prices*: fuel prices
- *load*: share of seats sold on flights flown.

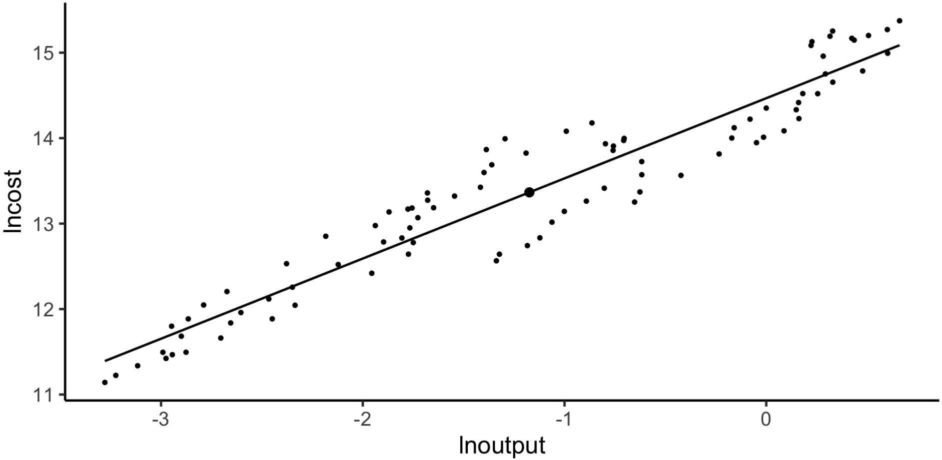
# Example Long Narrow Data: Airline cost



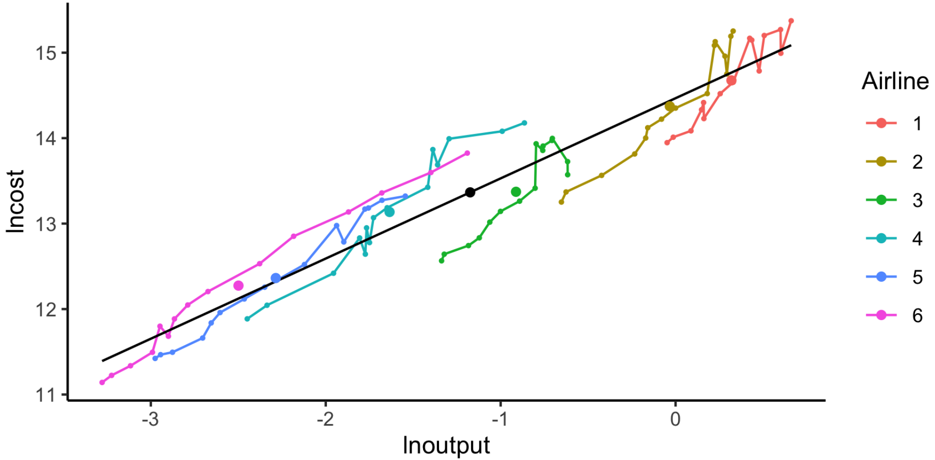
# Pooled Regression of log(costs)

| | Greene 2000 | Greene 2003 |
|---|---|---|
| (Intercept) | $9.52^{***}$ | $9.42^{***}$ |
| | (0.23) | (0.23) |
| lnoutput | $0.88^{***}$ | $0.94^{***}$ |
| | (0.01) | (0.03) |
| lnpfuel | $0.45^{***}$ | $0.46^{***}$ |
| | (0.02) | (0.02) |
| load | $-1.63^{***}$ | $-1.54^{***}$ |
| | (0.35) | (0.34) |
| lnoutput$^2$ | | $0.02^{*}$ |
| | | (0.01) |
| $R^2$ | 0.99 | 0.99 |
| Num. obs. | 90 | 90 |

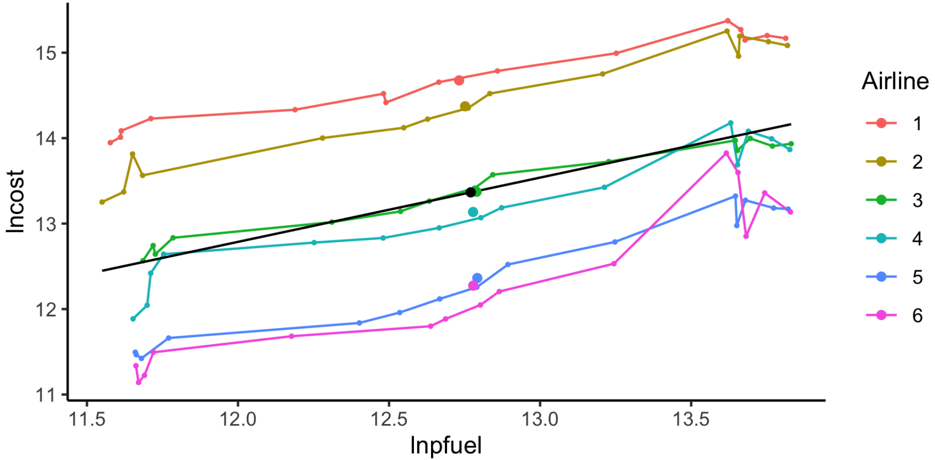
# Cost and Output (Bivariate)



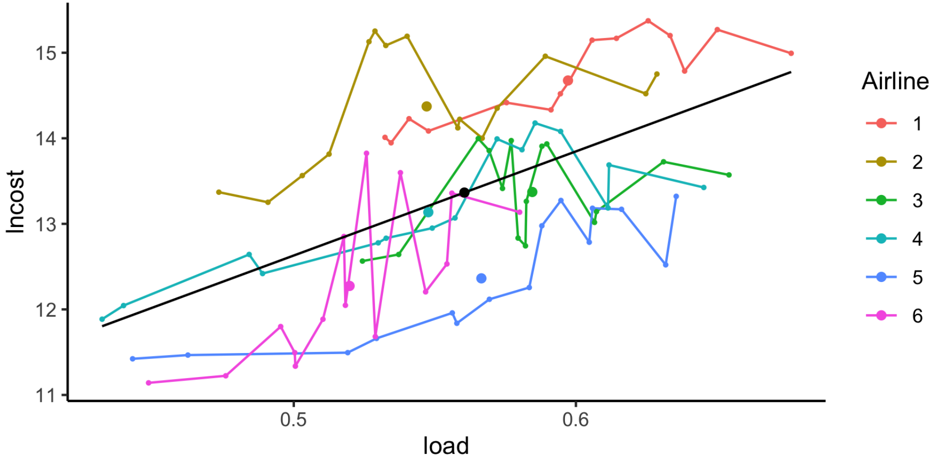
# Cost and Output (Bivariate)



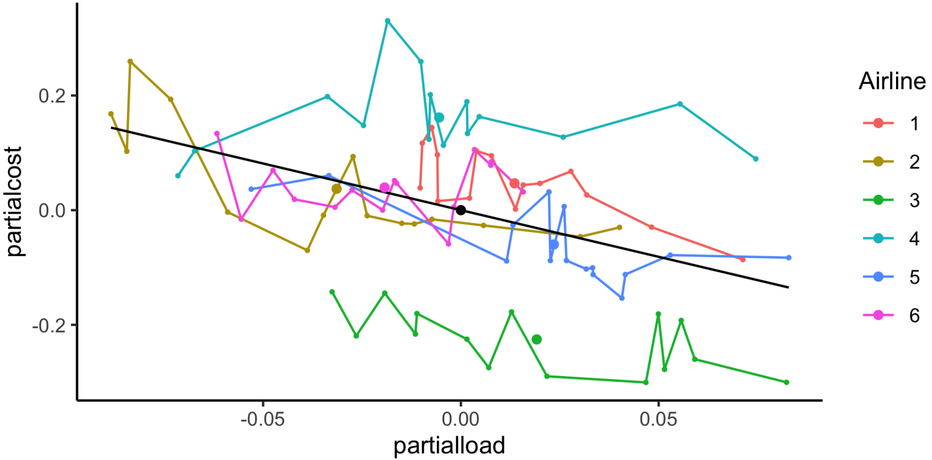
# Cost and price of fuel (Bivariate)



# Cost and Capacity Utilization (load)



# resid(Cost) and resid(load)



# Assumptions of Pooled Regression: Exogeneity

- $\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{e}$
- Strict Exogeneity: $E(e_{it}|\boldsymbol{X}) = \mathbf{0}$ for all time periods of $\boldsymbol{X}$.
- That is, $\boldsymbol{X}$ must not have any long run consequences for $\boldsymbol{Y}$.
- Moreover, $\boldsymbol{X}$ must not itself be a function of past $\boldsymbol{Y}$

# Autocorrelation versus Autoregression

$$Y_t = \alpha Y_{t-1} + \beta X_t + \nu_t$$
$$\nu_t = \rho \nu_{t-1} + e_t$$

- $\rho \neq 0$, $\alpha = 0$ has a "common factor" problem (autocorrelation):
  - Outcomes at $t-1$ are associated with outcomes at $t$ for many unobserved reasons.
  - OLS slope estimates are unbiased but inefficient and produces wrong standard errors.
  - Addressed with GLS/ serial correlation robust standard errors.
- $\alpha \neq 0$ is associated with dynamics:
  - Outcomes at $t-1$ affect outcomes now.
  - OLS is biased, inefficient and produces wrong standard errors.
  - Addressed with lagged variables and other structural time series approaches.

# Autocorrelation versus Autoregression

$$Y_t = \alpha Y_{t-1} + \beta X_t + \nu_t$$
$$\nu_t = \rho \nu_{t-1} + e_t$$

- $\rho \neq 0$, $\alpha = 0$ has a "common factor" problem (autocorrelation):
  - Outcomes at $t-1$ are associated with outcomes at $t$ for many unobserved reasons.
  - OLS slope estimates are unbiased but inefficient and produces wrong standard errors.
  - Addressed with GLS/ serial correlation robust standard errors.
- $\alpha \neq 0$ is associated with dynamics:
  - Outcomes at $t-1$ affect outcomes now.
  - OLS is biased, inefficient and produces wrong standard errors.
  - Addressed with lagged variables and other structural time series approaches.

# Autocorrelation versus Autoregression

$$Y_t = \alpha Y_{t-1} + \beta X_t + \nu_t$$
$$\nu_t = \rho \nu_{t-1} + e_t$$

- $\rho \neq 0$, $\alpha = 0$ has a "common factor" problem (autocorrelation):
  - Outcomes at $t-1$ are associated with outcomes at $t$ for many unobserved reasons.
  - OLS slope estimates are unbiased but inefficient and produces wrong standard errors.
  - Addressed with GLS/ serial correlation robust standard errors.
- $\alpha \neq 0$ is associated with dynamics:
  - Outcomes at $t-1$ affect outcomes now.
  - OLS is biased, inefficient and produces wrong standard errors.
  - Addressed with lagged variables and other structural time series approaches.

# Improvements on Inference

- In panel data we likely violate of iid errors, as observations from the same group/firm are likely associated

$$E[e_{ig}e_{jg}] = \rho\sigma_e^2 > 0$$

- Here $\rho$ is the intra-group/firm correlation.

# Standard error problems.

- If the number of groups is large, we can cluster standard errors using a formula similar to White standard errors.
- If the number of groups is small, we can just inflate our standard errors by $\sqrt{1+(n-1)\rho}$ (the Moulton factor)
- There is a bootstrap solution that has good performance

# Standard error problems.

- If the number of groups is large, we can cluster standard errors using a formula similar to White standard errors.
- If the number of groups is small, we can just inflate our standard errors by $\sqrt{1+(n-1)\rho}$ (the Moulton factor)
- There is a bootstrap solution that has good performance

# Standard error problems.

- If the number of groups is large, we can cluster standard errors using a formula similar to White standard errors.
- If the number of groups is small, we can just inflate our standard errors by $\sqrt{1+(n-1)\rho}$ (the Moulton factor)
- There is a bootstrap solution that has good performance

# White vs Cluster Robust

| | (white) | (cluster by airline) |
|---|---|---|
| (Intercept) | $9.52^{***}$ | $9.52^{***}$ |
| | (0.22) | (0.38) |
| lnoutput | $0.88^{***}$ | $0.88^{***}$ |
| | (0.01) | (0.02) |
| lnpfuel | $0.45^{***}$ | $0.45^{***}$ |
| | (0.02) | (0.03) |
| load | $-1.63^{***}$ | $-1.63^{*}$ |
| | (0.32) | (0.44) |
| $R^2$ | 0.99 | 0.99 |
| N Clusters | | 6 |

## Within-group estimators

- Consider the typical panel model: $y_{it} = \gamma x_{it} + f_i + e_{it}$
- Assume that $E[e_i | f_i, \mathbf{x}_{i1}, \mathbf{x}_{i2}, \dots \mathbf{x}_{iT}] = 0 \quad \forall i$
- But suppose that $E(f_i | x_{it}) \neq 0$. We can dummy for "fixed effects" $f$, producing

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Df} + \boldsymbol{e}$$

- We know from Frisch Waugh Lovell that:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$$
$$\hat{\boldsymbol{f}} = (\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{y}$$

- Where $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}_D = I - \boldsymbol{D}'(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ are the idempotent residual makers.
- $\hat{\beta}$ and $\hat{f}$ are the "within-group" estimators.

# Within-group estimators

- Consider the typical panel model: $y_{it} = \gamma x_{it} + f_i + e_{it}$
- Assume that $E[e_i | f_i, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots \boldsymbol{x}_{iT}] = 0 \quad \forall i$
- But suppose that $E(f_i | x_{it}) \neq 0$. We can dummy for "fixed effects" $f$, producing

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Df} + \boldsymbol{e}$$

- We know from Frisch Waugh Lovell that:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$$
$$\hat{\boldsymbol{f}} = (\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{y}$$

- Where $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}_D = I - \boldsymbol{D}'(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ are the idempotent residual makers.
- $\hat{\beta}$ and $\hat{f}$ are the "within-group" estimators.

## Within-group estimators

- Consider the typical panel model: $y_{it} = \gamma x_{it} + f_i + e_{it}$
- Assume that $E[e_i | f_i, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \dots \boldsymbol{x}_{iT}] = 0 \quad \forall i$
- But suppose that $E(f_i | x_{it}) \neq 0$. We can dummy for "fixed effects" $f$, producing

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{D}\boldsymbol{f} + \boldsymbol{e}$$

- We know from Frisch Waugh Lovell that:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$$
$$\hat{f} = (\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{y}$$

- Where $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}_D = I - \boldsymbol{D}'(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ are the idempotent residual makers.
- $\hat{\beta}$ and $\hat{f}$ are the "within-group" estimators.

## Within-group estimators

- Consider the typical panel model: $y_{it} = \gamma x_{it} + f_i + e_{it}$
- Assume that $E[e_i | f_i, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots \boldsymbol{x}_{iT}] = 0 \quad \forall i$
- But suppose that $E(f_i | x_{it}) \neq 0$. We can dummy for "fixed effects" $f$, producing

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{D}\boldsymbol{f} + \boldsymbol{e}$$

- We know from Frisch Waugh Lovell that:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$$
$$\hat{\boldsymbol{f}} = (\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{y}$$

- Where $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}_D = I - \boldsymbol{D}'(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ are the idempotent residual makers.
- $\hat{\beta}$ and $\hat{f}$ are the "within-group" estimators.

## Within-group estimators

- Consider the typical panel model: $y_{it} = \gamma x_{it} + f_i + e_{it}$
- Assume that $E[e_i | f_i, \boldsymbol{x}_{i1}, \boldsymbol{x}_{i2}, \ldots \boldsymbol{x}_{iT}] = 0 \quad \forall i$
- But suppose that $E(f_i | x_{it}) \neq 0$. We can dummy for "fixed effects" $f$, producing

$$\boldsymbol{y} = \boldsymbol{X}\beta + \boldsymbol{Df} + \boldsymbol{e}$$

- We know from Frisch Waugh Lovell that:

$$\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$$
$$\hat{\boldsymbol{f}} = (\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_X\boldsymbol{y}$$

- Where $\boldsymbol{M}_X = \boldsymbol{I} - \boldsymbol{X}'(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ and $\boldsymbol{M}_D = I - \boldsymbol{D}'(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}'$ are the idempotent residual makers.
- $\hat{\beta}$ and $\hat{f}$ are the "within-group" estimators.

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$
\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}
$$

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & 0 \\ 0 & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}$$

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$
\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & 0 \\ 0 & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}
$$

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$
\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & \boldsymbol{0} \\ \boldsymbol{0} & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & \boldsymbol{0} \\ \boldsymbol{0} & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}
$$

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$
\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & 0 \\ 0 & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}
$$

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$
\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & 0 \\ 0 & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}
$$

- Suppose N=2: $\boldsymbol{D} = \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix}$ where $\boldsymbol{i}$ is a $(T \times 1)$ vector of 1s

$$
\begin{aligned}
\boldsymbol{M}_D &= \boldsymbol{I} - \boldsymbol{D}(\boldsymbol{D}'\boldsymbol{D})^{-1}\boldsymbol{D}' \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \left[ \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix}' \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \right]^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} T & 0 \\ 0 & T \end{pmatrix}^{-1} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} & 0 \\ 0 & \boldsymbol{I} \end{pmatrix} - \begin{pmatrix} \boldsymbol{i} & 0 \\ 0 & \boldsymbol{i} \end{pmatrix} \begin{pmatrix} 1/T & 0 \\ 0 & 1/T \end{pmatrix} \begin{pmatrix} \boldsymbol{i}' & 0 \\ 0 & \boldsymbol{i}' \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & 0 \\ 0 & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix}
\end{aligned}
$$

$$\boldsymbol{M}_D\boldsymbol{y} = \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix} \boldsymbol{y}$$

$$= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - \frac{1}{T}\boldsymbol{i}\boldsymbol{i}' \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix}$$

$$= \begin{pmatrix} y_{11} - \frac{\sum_{t=1}^{T} Y_1 t}{T} \\ y_{12} - \frac{\sum_{t=1}^{T} Y_1 t}{T} \\ \cdots \\ y_{1T} - \frac{\sum_{t=1}^{T} Y_1 t}{T} \\ y_{21} - \frac{\sum_{t=1}^{T} Y_2 t}{T} \\ y_{22} - \frac{\sum_{t=1}^{T} Y_2 t}{T} \\ \cdots \\ y_{2T} - \frac{\sum_{t=1}^{T} Y_2 t}{T} \end{pmatrix}$$

$$\boldsymbol{M}_D\boldsymbol{y} = \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{ii}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - \frac{1}{T}\boldsymbol{ii}' \end{pmatrix} \boldsymbol{y}$$

$$= \begin{pmatrix} \boldsymbol{I} - \frac{1}{T}\boldsymbol{ii}' & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{I} - \frac{1}{T}\boldsymbol{ii}' \end{pmatrix} \begin{pmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \end{pmatrix}$$

$$= \begin{pmatrix} y_{11} - \frac{\sum_{t=1}^{T} Y_1 t}{T} \\ y_{12} - \frac{\sum_{t=1}^{T} Y_1 t}{T} \\ \cdots \\ y_{1T} - \frac{\sum_{t=1}^{T} Y_1 t}{T} \\ y_{21} - \frac{\sum_{t=1}^{T} Y_2 t}{T} \\ y_{22} - \frac{\sum_{t=1}^{T} Y_2 t}{T} \\ \cdots \\ y_{2T} - \frac{\sum_{t=1}^{T} Y_2 t}{T} \end{pmatrix}$$

# Within Estimator

Two equivalent ways to estimate:

1. Explicitly controlling for fixed effects $\boldsymbol{f}$ and estimating $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$.
2. Differencing out the group means:

$$y_{it} - \bar{y}_i = \beta'(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) + e_{it} - \bar{e}_i$$

$$\boldsymbol{b}_W = (\boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W$$

$$\boldsymbol{S}_{xx}^W = \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'$$

$$\boldsymbol{S}_{xy}^W = \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(y_{it} - \bar{y}_i)'$$

# Within Estimator

Two equivalent ways to estimate:

1. Explicitly controlling for fixed effects $\boldsymbol{f}$ and estimating $\hat{\beta} = (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{y}$.
2. Differencing out the group means:

$$y_{it} - \bar{y}_i = \beta'(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) + e_{it} - \bar{e}_i$$

$$\boldsymbol{b}_W = (\boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W$$

$$\boldsymbol{S}_{xx}^W = \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)'$$

$$\boldsymbol{S}_{xy}^W = \sum_{i=1}^{N}\sum_{t=1}^{T}(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i)(y_{it} - \bar{y}_i)'$$

# Aside on Stacked Matrix Notation

$\boldsymbol{y}_i$: $T \times 1$, $\boldsymbol{X}_i$: $T \times k$, $\boldsymbol{e}_i$: $T \times 1$, $\boldsymbol{\beta} : k \times 1$

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \beta + \begin{bmatrix} \alpha_1 \mathbf{1} \\ \alpha_2 \mathbf{1} \\ \vdots \\ \alpha_N \mathbf{1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_N \end{bmatrix}$$

$$y = \boldsymbol{X}\beta + \boldsymbol{\alpha} \otimes \mathbf{1}_T + \boldsymbol{e}$$

# Aside on Stacked Matrix Notation

$\boldsymbol{y}_i$: $T \times 1$, $\boldsymbol{X}_i$: $T \times k$, $\boldsymbol{e}_i$: $T \times 1$, $\boldsymbol{\beta} : k \times 1$

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \boldsymbol{\beta} + \begin{bmatrix} \alpha_1 \mathbf{1} \\ \alpha_2 \mathbf{1} \\ \vdots \\ \alpha_N \mathbf{1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_N \end{bmatrix}$$

$$y = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\alpha} \otimes \mathbf{1}_T + \boldsymbol{e}$$

# Aside on Stacked Matrix Notation

$\boldsymbol{y}_i$: $T \times 1$, $\boldsymbol{X}_i$: $T \times k$, $\boldsymbol{e}_i$: $T \times 1$, $\beta : k \times 1$

$$\begin{bmatrix} \boldsymbol{y}_1 \\ \boldsymbol{y}_2 \\ \vdots \\ \boldsymbol{y}_N \end{bmatrix} = \begin{bmatrix} \boldsymbol{X}_1 \\ \boldsymbol{X}_2 \\ \vdots \\ \boldsymbol{X}_N \end{bmatrix} \beta + \begin{bmatrix} \alpha_1 \mathbf{1} \\ \alpha_2 \mathbf{1} \\ \vdots \\ \alpha_N \mathbf{1} \end{bmatrix} + \begin{bmatrix} \boldsymbol{e}_1 \\ \boldsymbol{e}_2 \\ \vdots \\ \boldsymbol{e}_N \end{bmatrix}$$

$$y = \boldsymbol{X}\beta + \boldsymbol{\alpha} \otimes \mathbf{1}_T + \boldsymbol{e}$$

# Software Implementations

- lm(y∼ $x + f_1 + f_2 + \ldots + f_n$)
- plm(y∼ $x$, c("unit", "time"))
- fixest::feols(y∼ x| fe)
- You can also demean by time averages, but you need to correct the standard errors.

# Fixed Effects and unobserved factors

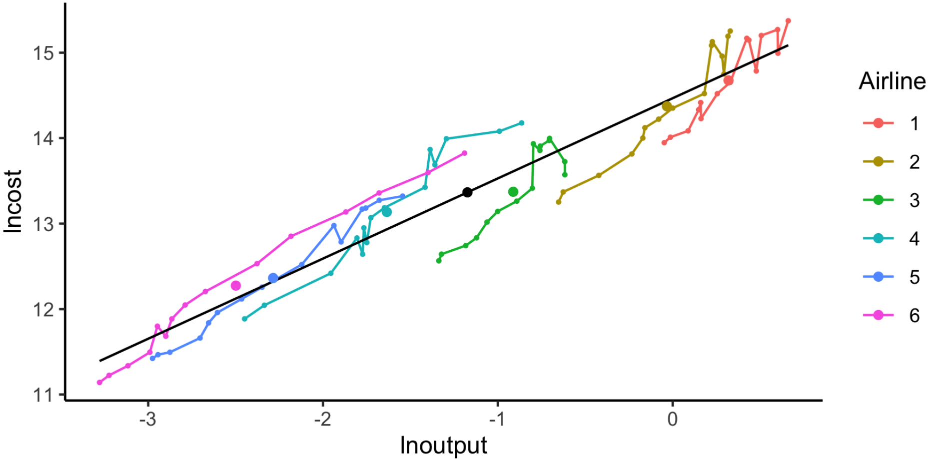
Just as with difference in difference, fixed effects address unobserved effects ($\alpha_i$).

$$y_{it} = \gamma x_{it} + f_i + \alpha_i + e_{it}$$

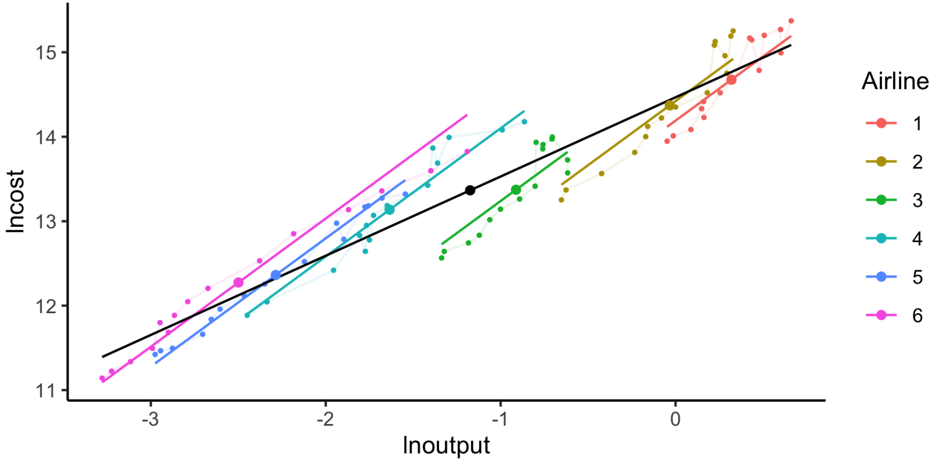
The fixed effects transformation again mean deviates the variables, giving us:

$$y_{it} - \bar{y}_i = \boldsymbol{\beta}'(\boldsymbol{x}_{it} - \bar{\boldsymbol{x}}_i) + \alpha_i - \bar{\alpha}_i + e_{it} - \bar{e}_i$$

# Cost and Output



# Cost and Output (within estimator)



# Small note on estimation

$$\boldsymbol{b}_{fe} = \beta + (\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{e}$$

Under iid errors, we have: $var(\boldsymbol{b}_{fe}) = \sigma_e^2(\boldsymbol{X}'\boldsymbol{M}_D\boldsymbol{X})^{-1}$. But we need to correct our estimator for the $N$ estimated means:

$$s_e^2 = \frac{\sum_{i=1}^{N}\sum_{t=1}^{T} e_{it}^2}{NT - N - K}$$

## Fixed Effects as GMM Moment Conditions

- Define the within-transformed variables $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$ and $\tilde{\boldsymbol{X}}_{it} = \boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i$.
- The FE (within) estimator solves the moment conditions:

$$E[\tilde{\boldsymbol{X}}_{it}' \tilde{e}_{it}] = \mathbf{0}$$

- This is a **just-identified** GMM problem: $k$ moment conditions for $k$ parameters.
  - No overidentifying restrictions $\Rightarrow$ no $J$-test with FE alone.
- **Preview:**
  - Random effects adds between-group moments $\Rightarrow$ more efficient under stronger assumptions.
  - Arellano-Bond (next lecture) creates *many* moment conditions from lagged levels $\Rightarrow$ overidentification $\Rightarrow$ testable with $J$-statistic.

# Fixed Effects as GMM Moment Conditions

- Define the within-transformed variables $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$ and $\tilde{\boldsymbol{X}}_{it} = \boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i$.
- The FE (within) estimator solves the moment conditions:

$$E[\tilde{\boldsymbol{X}}_{it}' \tilde{e}_{it}] = \mathbf{0}$$

- This is a **just-identified** GMM problem: $k$ moment conditions for $k$ parameters.
  - No overidentifying restrictions $\Rightarrow$ no $J$-test with FE alone.
- **Preview:**
  - Random effects adds between-group moments $\Rightarrow$ more efficient under stronger assumptions.
  - Arellano-Bond (next lecture) creates *many* moment conditions from lagged levels $\Rightarrow$ overidentification $\Rightarrow$ testable with $J$-statistic.

# Fixed Effects as GMM Moment Conditions

- Define the within-transformed variables $\tilde{Y}_{it} = Y_{it} - \bar{Y}_i$ and $\tilde{\boldsymbol{X}}_{it} = \boldsymbol{X}_{it} - \bar{\boldsymbol{X}}_i$.
- The FE (within) estimator solves the moment conditions:

$$E[\tilde{\boldsymbol{X}}_{it}' \tilde{e}_{it}] = \mathbf{0}$$

- This is a **just-identified** GMM problem: $k$ moment conditions for $k$ parameters.
  - No overidentifying restrictions $\Rightarrow$ no $J$-test with FE alone.
- **Preview:**
  - Random effects adds between-group moments $\Rightarrow$ more efficient under stronger assumptions.
  - Arellano-Bond (next lecture) creates *many* moment conditions from lagged levels $\Rightarrow$ overidentification $\Rightarrow$ testable with $J$-statistic.

# Pooled Regression Anatomy

- Call $\boldsymbol{x}_{it}$ the k by 1 vector of right hand side variables.
- Grand means:
- $\bar{\bar{x}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} x_{it}$
- $\bar{\bar{y}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$
- We can write the pooled estimator as:

$$\boldsymbol{b_P} = (\boldsymbol{S}_{xx}^{P})^{-1} \boldsymbol{S}_{xy}^{P}$$

Where

$$\boldsymbol{S}_{xx}^{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})'$$

$$\boldsymbol{S}_{xy}^{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(y_{it} - \bar{\bar{y}})'$$

## Pooled Regression Anatomy

- Call $\boldsymbol{x}_{it}$ the k by 1 vector of right hand side variables.
- Grand means:
- $\bar{\bar{\boldsymbol{x}}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}$
- $\bar{\bar{y}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$
- We can write the pooled estimator as:

$$\boldsymbol{b_P} = (\boldsymbol{S}_{xx}^{P})^{-1} \boldsymbol{S}_{xy}^{P}$$

Where

$$\boldsymbol{S}_{xx}^{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})'$$

$$\boldsymbol{S}_{xy}^{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(y_{it} - \bar{\bar{y}})'$$

# Pooled Regression Anatomy

- Call $\boldsymbol{x}_{it}$ the k by 1 vector of right hand side variables.
- Grand means:
- $\bar{\bar{\boldsymbol{x}}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} \boldsymbol{x}_{it}$
- $\bar{\bar{y}} \equiv \sum_{i=1}^{N} \sum_{t=1}^{T} y_{it}$
- We can write the pooled estimator as:

$$\boldsymbol{b_P} = (\boldsymbol{S}_{xx}^{P})^{-1} \boldsymbol{S}_{xy}^{P}$$

Where

$$\boldsymbol{S}_{xx}^{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})'$$

$$\boldsymbol{S}_{xy}^{P} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\boldsymbol{x}_{it} - \bar{\bar{\boldsymbol{x}}})(y_{it} - \bar{\bar{y}})'$$

```
airlines<-airlines%>% mutate(
  mlnoutput = mean(lnoutput),
  mlnpfuel = mean(lnpfuel),
  mload = mean(load))

Xp<-airlines%>%transmute(
                lnoutputdm = lnoutput-mlnoutput,
                lnpfueldm = lnpfuel-mlnpfuel,
                loaddm = load-mload)%>%
                as.matrix()

Sp_xx = t(Xp)%*%Xp
```

# Between Regression

- Part of the variation in $y_{it}$ and $\boldsymbol{x}_{it}$ is variation across units and part across time.
- The cross sectional part is just:

$$\frac{1}{T}(y_{i1} + y_{i2} + \ldots + y_{iT}) = \beta'(\frac{1}{T}(\boldsymbol{x}_{i1} + \boldsymbol{x}_{i2} + \ldots + \boldsymbol{x}_{iT})) + \frac{1}{T}(e_{i1} + e_{i2} + \ldots + e_{iT})$$
$$\bar{y}_i = \beta' \bar{\boldsymbol{x}}_i + \bar{e}_i$$

Define $\boldsymbol{b_B} = (\boldsymbol{S}^B_{xx})^{-1}\boldsymbol{S}^B_{xy}$.

$$\boldsymbol{S}^B_{xx} = \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})' = \sum_{i=1}^{N} T(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})'$$

$$\boldsymbol{S}^B_{xy} = \sum_{i=1}^{N}\sum_{t=1}^{T}(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{y}_i - \bar{\bar{y}}) = \sum_{i=1}^{N} T(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{y}_i - \bar{\bar{y}})$$

# Between Regression

- Part of the variation in $y_{it}$ and $\boldsymbol{x}_{it}$ is variation across units and part across time.
- The cross sectional part is just:

$$\frac{1}{T}(y_{i1} + y_{i2} + \ldots + y_{iT}) = \beta'(\frac{1}{T}(\boldsymbol{x}_{i1} + \boldsymbol{x}_{i2} + \ldots + \boldsymbol{x}_{iT})) + \frac{1}{T}(e_{i1} + e_{i2} + \ldots + e_{iT})$$
$$\bar{y}_i = \beta' \bar{\boldsymbol{x}}_i + \bar{e}_i$$

Define $\boldsymbol{b_B} = (\boldsymbol{S}^B_{xx})^{-1} \boldsymbol{S}^B_{xy}$.

$$\boldsymbol{S}^B_{xx} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})' = \sum_{i=1}^{N} T(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})'$$

$$\boldsymbol{S}^B_{xy} = \sum_{i=1}^{N} \sum_{t=1}^{T} (\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{y}_i - \bar{\bar{y}}) = \sum_{i=1}^{N} T(\bar{\boldsymbol{x}}_i - \bar{\bar{\boldsymbol{x}}})(\bar{y}_i - \bar{\bar{y}})$$

```
airlines<-airlines%>%
  group_by(airline)%>%
  mutate(
  gmlnoutput=mean(lnoutput),
  gmlnpfuel=mean(lnpfuel),
  gmload=mean(load))

Xb<-airlines%>%transmute(
                blnoutputdm = gmlnoutput-mlnoutput,
                 blnpfueldm = gmlnpfuel-mlnpfuel,
                 bloaddm = gmload-mload) %>%
                 as.matrix()

  Sb_xx = t(Xb)%*%Xb
```

```
airlines<-airlines%>%
  group_by(airline)%>%
  mutate(
  gmlnoutput=mean(lnoutput),
  gmlnpfuel=mean(lnpfuel),
  gmload=mean(load))

  Xw<-airlines%>%transmute(
                  plnoutputdm = lnoutput-gmlnoutput,
                   plnpfueldm = lnpfuel-gmlnpfuel,
                   ploaddm = load-gmload)%>%
                   as.matrix()

  Sw_xx = t(Xw)%*%Xw
```

# Decomposition of Variance

$$y_{it} = \bar{y}_i + (y_{it} - \bar{y}_i)$$

$$\mathbf{S}_{xx}^P = \mathbf{S}_{xx}^B + \mathbf{S}_{xx}^W$$

$$\mathbf{S}_{xy}^P = \mathbf{S}_{xy}^B + \mathbf{S}_{xy}^W$$

# Decomposition of Variance

$$y_{it} = \bar{y}_i + (y_{it} - \bar{y}_i)$$
$$\boldsymbol{S}^P_{xx} = \boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx}$$
$$\boldsymbol{S}^P_{xy} = \boldsymbol{S}^B_{xy} + \boldsymbol{S}^W_{xy}$$

# Relationship between Pooled, Between and Within

$$
\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}_{xx}^P)^{-1}\boldsymbol{S}_{xy}^P \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xy}^B + \boldsymbol{S}_{xy}^W) \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^B\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W - \boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= \lambda\boldsymbol{b}_W + (\boldsymbol{I} - \lambda)\boldsymbol{b}_B
\end{aligned}
$$

# Relationship between Pooled, Between and Within

$$
\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}_{xx}^P)^{-1}\boldsymbol{S}_{xy}^P \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xy}^B + \boldsymbol{S}_{xy}^W) \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^B\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W - \boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= \boldsymbol{\lambda}\boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda})\boldsymbol{b}_B
\end{aligned}
$$

# Relationship between Pooled, Between and Within

$$
\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}_{xx}^P)^{-1}\boldsymbol{S}_{xy}^P \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xy}^B + \boldsymbol{S}_{xy}^W) \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^B\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W - \boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= \boldsymbol{\lambda}\boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda})\boldsymbol{b}_B
\end{aligned}
$$

# Relationship between Pooled, Between and Within

$$\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}_{xx}^P)^{-1}\boldsymbol{S}_{xy}^P \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xy}^B + \boldsymbol{S}_{xy}^W) \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^B\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W - \boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= \boldsymbol{\lambda}\boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda})\boldsymbol{b}_B
\end{aligned}$$

# Relationship between Pooled, Between and Within

$$
\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}^P_{xx})^{-1}\boldsymbol{S}^P_{xy} \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}(\boldsymbol{S}^B_{xy} + \boldsymbol{S}^W_{xy}) \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xy} + (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^B_{xy} \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}\boldsymbol{b}_W + (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^B_{xx}\boldsymbol{b}_B \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}\boldsymbol{b}_W + (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}(\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx} - \boldsymbol{S}^W_{xx})\boldsymbol{b}_B \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx})\boldsymbol{b}_B \\
&= \boldsymbol{\lambda}\boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda})\boldsymbol{b}_B
\end{aligned}
$$

# Relationship between Pooled, Between and Within

$$
\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}^P_{xx})^{-1}\boldsymbol{S}^P_{xy} \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}(\boldsymbol{S}^B_{xy} + \boldsymbol{S}^W_{xy}) \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xy} + (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^B_{xy} \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}\boldsymbol{b}_W + (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^B_{xx}\boldsymbol{b}_B \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}\boldsymbol{b}_W + (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}(\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx} - \boldsymbol{S}^W_{xx})\boldsymbol{b}_B \\
&= (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx})\boldsymbol{b}_B \\
&= \boldsymbol{\lambda}\boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda})\boldsymbol{b}_B
\end{aligned}
$$

# Relationship between Pooled, Between and Within

$$
\begin{aligned}
\boldsymbol{b}_P &= (\boldsymbol{S}_{xx}^P)^{-1}\boldsymbol{S}_{xy}^P \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xy}^B + \boldsymbol{S}_{xy}^W) \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xy}^B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^B\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}(\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W - \boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W\boldsymbol{b}_W + (\boldsymbol{I} - (\boldsymbol{S}_{xx}^B + \boldsymbol{S}_{xx}^W)^{-1}\boldsymbol{S}_{xx}^W)\boldsymbol{b}_B \\
&= \boldsymbol{\lambda}\boldsymbol{b}_W + (\boldsymbol{I} - \boldsymbol{\lambda})\boldsymbol{b}_B
\end{aligned}
$$

# Relationship between Pooled, Between and Within

| | lncost | | |
|---|---|---|---|
| | Pooled | Between | Within |
| lnoutput | 0.883*** | 0.782** | 0.919*** |
| | (0.013) | (0.109) | (0.030) |
| lnpfuel | 0.454*** | −5.524 | 0.417*** |
| | (0.020) | (4.479) | (0.015) |
| load | −1.628*** | −1.751 | −1.070*** |
| | (0.345) | (2.743) | (0.202) |
| Constant | 9.517*** | 85.809 | |
| | (0.229) | (56.483) | |
| Observations | 90 | 6 | 90 |
| $R^2$ | 0.988 | 0.994 | 0.993 |

# What the estimates tell us

- **Output elasticity**

$$\hat{\beta}_1 = 0.883 \text{ (pooled)} \implies \hat{r} = \frac{1}{\hat{\beta}_1} \approx 1.13.$$

  The industry exhibits *mild increasing returns to scale* (costs rise less than proportionally with output).

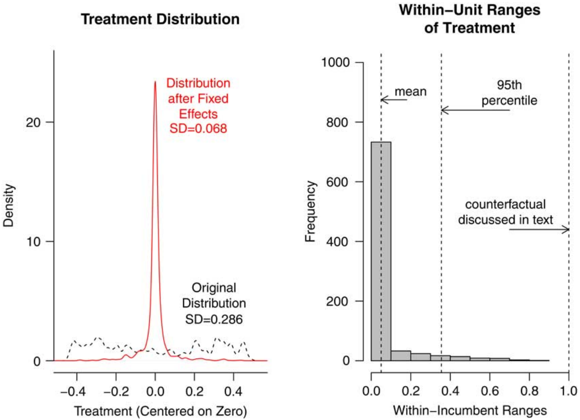
- **Fuel price elasticity**

$$\hat{\beta}_2 = 0.454 \implies \hat{\alpha}_F = \hat{r}\,\hat{\beta}_2 \approx 1.13 \times 0.454 \approx 0.51.$$

  Roughly half of total variable cost is attributable to fuel.

- **Load factor effect** The coefficient on *load* ($-1.63$) means that a 1-percentage-point increase in capacity utilisation lowers total cost by about 1.6 %, consistent with spreading fixed operating expenses over more output.

# Danger




The left panel displays the distribution of media congruence from Snyder and Strömberg (2010) before and after the incumbent and year fixed effects.

# Political Science Application: Tax Progressivity and War

- **Scheve & Stasavage (2010, 2012):** Why did top tax rates rise dramatically in the 20th century?
- **Hypothesis:** Mass military mobilization during WWI/WWII created political pressure for progressive taxation as a "conscription of wealth."
- **Data:** Country-year panel, 20 countries, 1850–1970.
  - $Y_{it}$: Top marginal income tax rate
  - $X_{it}$: War mobilization $\times$ high-mobility indicator
  - $\alpha_i$: Country fixed effects (e.g. institutional differences, colonial legacies)
- **Why fixed effects?** Countries differ systematically in fiscal institutions, political systems, and economic structure. Pooled OLS would confound these with the mobilization effect.
- Strict exogeneity plausible: war mobilization driven by geopolitics, not anticipated tax policy.

# Political Science Application: Tax Progressivity and War

- **Scheve & Stasavage (2010, 2012):** Why did top tax rates rise dramatically in the 20th century?
- **Hypothesis:** Mass military mobilization during WWI/WWII created political pressure for progressive taxation as a "conscription of wealth."
- **Data:** Country-year panel, 20 countries, 1850–1970.
  - $Y_{it}$: Top marginal income tax rate
  - $X_{it}$: War mobilization $\times$ high-mobility indicator
  - $\alpha_i$: Country fixed effects (e.g. institutional differences, colonial legacies)
- **Why fixed effects?** Countries differ systematically in fiscal institutions, political systems, and economic structure. Pooled OLS would confound these with the mobilization effect.
- Strict exogeneity plausible: war mobilization driven by geopolitics, not anticipated tax policy.

# Scheve & Stasavage: R Implementation

```r
library(plm); library(haven); library(fixest)
dta <- read_dta("progressTax.dta")
pdta <- subset(dta, year >= 1900 & year <= 1930)

# Fixed effects with plm
fe_mod <- plm(topratep ~ wwihighmobaft,
              data = pdta,
              index = c("ccode", "year"),
              model = "within")
summary(fe_mod)

# Equivalent with fixest (+ Newey-West SEs)
fe_mod2 <- feols(topratep ~ wwihighmobaft +
    gdppcp + leftseatshp + munsuff + ratiop
    | country,
    data = pdta, panel.id = ~country + year,
    vcov = NW(1) ~ country + year)
```

## Summary

- Difference in difference cancels out group-invariant trends and time invariant group differences (whether of interest or not).
- Fixed effects give the same estimates as demeaning $\boldsymbol{X}$ and $\boldsymbol{y}$ by their over time means, the "within" estimator (so long as you correct the residual variance estimator).
- The within estimator corresponds to the GMM moment condition $E[\tilde{\boldsymbol{X}}'_{it}\tilde{e}_{it}] = \mathbf{0}$, and is just-identified.
- A "between" estimator only uses variation across units, averaged over time.
- Pooled Regression is a weighted average of the between and within estimates, with weights $(\boldsymbol{S}^B_{xx} + \boldsymbol{S}^W_{xx})^{-1}\boldsymbol{S}^W_{xx}$.
- **Next time:** Random effects adds between-group moments (GLS); Arellano-Bond uses lagged levels as instruments for dynamic panels — both fit into our GMM framework from Lectures 15–16.