Fixed-effects regressions

Data Science and Causal Inference Workshop 2025

Dr Christian Engels

START RECORDING NOW

University of St Andrews | FOUNDED 1413

4 September 2025

Learning objectives

- Understand difference between econometrics for prediction and causal analysis

- Understand that endogeneity complicates deduction of causal statements

- Explain key feature of the OLS estimator regarding endogeneity

- Gain knowledge of sources of endogeneity

- Gain ability to speculate on and explain endogeneity concerns in empirical research

1. **Reverse causality**

   Example: Does bank performance drive CEO compensation, or does CEO compensation drive bank performance?

2. **Selection bias**

   Example: Firms choosing to issue bonds vs taking bank loans based on unobservable characteristics

3. **Omitted variable bias**

   Example: Unobservable managerial quality affecting both firm performance and capital structure decisions

- Entity fixed effects control for time-invariant unobservable characteristics by:

    – Removing all between-variation (differences across entities)

    – Focusing only on within-variation (changes within entities over time)

    – Controlling for unobservable characteristics that don't change over time

- Banking example (entity and time fixed effects):

    – Bank fixed effects control for time-invariant bank characteristics like charter value, organizational culture and geographic location

    – Time fixed effects control for common shocks across all banks in each year/quarter (e.g., financial crisis, Covid-19)

Two equivalent approaches:

1. De-meaning method: subtract entity-specific means from all variables

$$Y_{it} - \bar{Y}_i = \beta(X_{it} - \bar{X}_i) + (\varepsilon_{it} - \bar{\varepsilon}_i)$$

2. Least Squares Dummy Variable (LSDV): include dummy variables for each entity

$$Y_{it} = \beta X_{it} + \alpha_i + \varepsilon_{it}$$

Both approaches generalise to settings with high-dimensional fixed effects (e.g., one possibility for corporate loans: *firm* + *bank* + *time* + *location*)

Limitations of fixed effects

- Key considerations:

  - Cannot control for time-varying omitted variables

  - Cannot estimate the effect of time-invariant variables (e.g., bank headquarters location)

  - Requires sufficient within-variation in variables of interest

  - May have lower statistical power due to removed variation

- Important note:

  - Fixed effects are not a panacea - they only address specific sources of endogeneity

- Why cluster at entity level?

  - Observations within each entity (e.g., bank) are likely correlated over time

  - Bank's risk-taking in 2023 is not independent from its risk-taking in 2022

  - Standard errors will be understated without clustering

- Serial correlation in panel data:

  - Firm characteristics persist over time

  - Shocks often affect the same entity for multiple periods

  - Corporate policies show strong autocorrelation

- Panel data best practice: cluster at least at the entity level

Illustration with R and Stata
Fixed-effects regressions with "fixest" (R) and "reghdfe" (Stata)

Positron IDE (by the makers of RStudio)

Seamless switching between R and Python

https://positron.posit.co