Understanding Fixed Effects and Panel Data

Professor Soliman

2025-04-03

Introduction to Fixed Effects and Panel Data

This document provides an overview of fixed effects (FE) models in panel data analysis, using county-level opioid data. We will first review pooled OLS regression and then introduce various fixed effects specifications to illustrate their impact. Note that the data comes from here if you want to download and run/ edit this on your own. Remember to install the fixest package first, but only once, in the console with `install.packages("fixest")`.

Pooled OLS Regression

Before introducing fixed effects, we start with a simple pooled OLS regression, which does not account for unobserved heterogeneity across counties or over time:

```r
lpm <- lm(overdosedeaths ~ percapitapills, data = countyopioids)

# Using fixest for cleaner output
lpm <- feols(overdosedeaths ~ percapitapills, data = countyopioids)
lpm
```

```
## OLS estimation, Dep. Var.: overdosedeaths
## Observations: 26,970
## Standard-errors: IID
##                 Estimate Std. Error  t value  Pr(>|t|)
## (Intercept)    24.394833   0.833155 29.28006 < 2.2e-16 ***
## percapitapills  0.050462   0.017975  2.80725 0.0050002 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 71.6   Adj. R2: 2.551e-4
```

This model estimates the effect of per capita opioid pills on overdose deaths without accounting for time-invariant differences across counties or changes over time.

Introducing Fixed Effects

Fixed effects models help control for unobserved heterogeneity by accounting for time-invariant characteristics at different levels.

Adding Year Fixed Effects

```r
fe_year <- feols(overdosedeaths ~ percapitapills | year,
                 data = countyopioids, cluster = ~county)
fe_year
```

```
## OLS estimation, Dep. Var.: overdosedeaths
## Observations: 26,970
## Fixed-effects: year: 9
## Standard-errors: Clustered (county)
##                Estimate Std. Error  t value Pr(>|t|)
## percapitapills  0.01817   0.035668 0.509404  0.61051
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 71.4     Adj. R2: 0.004433
##              Within R2: 3.606e-5
```

Including year fixed effects accounts for national trends, such as federal policies or overall societal changes in opioid consumption and overdose rates.

Adding State and Year Fixed Effects

```r
fe_yearstate <- feols(overdosedeaths ~ percapitapills | state + year,
                      data = countyopioids, cluster = ~county)
fe_yearstate
```

```
## OLS estimation, Dep. Var.: overdosedeaths
## Observations: 26,970
## Fixed-effects: state: 51,  year: 9
## Standard-errors: Clustered (county)
##                 Estimate Std. Error   t value Pr(>|t|)
## percapitapills -0.011735   0.038689 -0.303326  0.76166
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 67.0     Adj. R2: 0.12336
##              Within R2: 1.298e-5
```

Adding state fixed effects controls for state-level policies and characteristics that do not change over time, further isolating the effect of opioid pills.

Adding County and Year Fixed Effects

```r
fe_yearcounty <- feols(overdosedeaths ~ percapitapills | county + year,
                       data = countyopioids)
fe_yearcounty
```

```
## OLS estimation, Dep. Var.: overdosedeaths
## Observations: 26,970
## Fixed-effects: county: 3,035,  year: 9
## Standard-errors: Clustered (county)
##                 Estimate Std. Error  t value   Pr(>|t|)
## percapitapills -0.118394   0.034026 -3.47956 0.00050931 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## RMSE: 19.0     Adj. R2: 0.92071
##              Within R2: 0.001391
```

Including county fixed effects controls for all time-invariant local characteristics, such as baseline health conditions, demographics, and historical policy influences.

Comparing Results

To better understand the differences between these models, we summarize them in a table, which may be helpful for your final projects:

```
etable(lpm, fe_year, fe_yearstate, fe_yearcounty)
```

```
##                                 lpm         fe_year     fe_yearstate
## Dependent Var.:    overdosedeaths  overdosedeaths   overdosedeaths
##
## Constant        24.39*** (0.8332)
## percapitapills  0.0505** (0.0180) 0.0182 (0.0357) -0.0117 (0.0387)
## Fixed-Effects:  ----------------- --------------- ----------------
## year                           No             Yes              Yes
## state                          No              No              Yes
## county                         No              No               No
## _______________ _________________ _______________ ________________
## S.E. type                     IID      by: county       by: county
## Observations               26,970          26,970           26,970
## R2                        0.00029         0.00477          0.12528
## Within R2                      --         3.61e-5           1.3e-5
##
##                     fe_yearcounty
## Dependent Var.:      overdosedeaths
##
## Constant
## percapitapills  -0.1184*** (0.0340)
## Fixed-Effects:  -------------------
## year                            Yes
## state                            No
## county                          Yes
## _______________ ___________________
## S.E. type                by: county
## Observations                 26,970
## R2                          0.92966
## Within R2                   0.00139
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Interpretation of Results

1. Pooled OLS (No Fixed Effects)

   The leftmost column represents a simple regression of overdose deaths on per capita opioid pills, without accounting for any unobserved heterogeneity.

   The coefficient on percapitapills is 0.0505 ($p < 0.01$), suggesting a positive correlation between opioid pill distribution and overdose deaths.

   However, this estimate is likely biased due to omitted variables that vary across counties and over time.

2. Adding Year Fixed Effects

   This model includes year fixed effects, which control for national trends over time (e.g., policy changes, awareness campaigns, and economic conditions).

   The coefficient on percapitapills drops to 0.0182 and is no longer statistically significant.

Interpretation: Part of the relationship observed in the OLS model was driven by time trends, not necessarily a direct causal link between opioid distribution and overdose deaths.

3. Adding State Fixed Effects

   This model introduces state and year fixed effects, accounting for time-invariant state-level factors such as:

   ```
   State-specific policies and regulations

   Differences in healthcare infrastructure

   Demographic and economic conditions that do not change over time
   ```

   The coefficient drops further to -0.0117 and remains insignificant.

   Interpretation: The earlier results were partly influenced by state-level differences, suggesting that high-overdose areas might have also had stronger policy interventions affecting opioid distribution.

4. Adding County Fixed Effects

   The final model includes county and year fixed effects, meaning it controls for local, time-invariant characteristics such as:

   ```
   Socioeconomic conditions

   Healthcare access

   Long-term prescribing practices

   Historical trends in opioid use
   ```

   The coefficient on percapitapills now flips negative (-0.1184, $p < 0.01$).

   Interpretation:

   ```
   Earlier models suffered from omitted variable bias-they failed to account for county-specific
   factors that influenced both opioid distribution and overdose deaths.

   The negative coefficient suggests that counties with initially high overdose rates received fewer
   pills over time, likely due to policy restrictions, enforcement, or local prescribing behaviors.

   This aligns with the idea that opioid pills were reduced in high-overdose counties as a policy
   response, rather than a direct causal link where more pills lead to more deaths.
   ```

Key Takeaways

- The raw correlation (OLS) was positive, but it was likely biased by national, state, and county-level confounding factors.
- After controlling for national trends (year FE) and state-level policies (state FE), the estimated effect weakened and became statistically insignificant.
- Once we account for county-specific factors (county FE), the coefficient becomes negative, highlighting the importance of controlling for unobserved heterogeneity in panel data.
- This suggests that earlier models overstated the positive link between opioid pills and overdose deaths, because they did not account for policy responses that reduced opioid availability in high-risk counties.

Understanding Fixed Effects Intuitively

The Role of Fixed Effects

1. Fixed effects introduce dummy variables for each county (or state, or year), effectively allowing each entity to have its own intercept.
2. The model estimates separate intercepts for each unit (e.g., county and year) but does not report them explicitly.
3. Because every observation belongs to exactly one county and one year, these fixed effects fully absorb the intercept.

Why No Intercept is Reported?

1. In a standard OLS regression, the intercept represents the expected value of the dependent variable when all explanatory variables are zero.
2. With fixed effects, each county has its own intercept, meaning a single, global intercept is unnecessary.
3. Instead of a single intercept, the mean differences across counties and years are captured by the fixed effects themselves.

Summary

- **Pooled OLS** can provide misleading estimates due to omitted variable bias.
- **Fixed effects** models help control for unobserved heterogeneity at different levels (year, state, county).
- **Comparing models** reveals how adding different levels of fixed effects impacts our estimated coefficients.
- **Interpretation** of results should consider how omitted variable bias can affect causal inference.

Fixed effects are a powerful tool in panel data analysis, ensuring that we control for confounding factors that do not change over time, leading to more reliable estimates.