Lecture 2: Basics of Panel Data

Let’s load some packages so that I can make some better looking plots:

Today's Plan

- Recap OLS and various forms of standard errors
- Standard errors are tedious but I guess you are supposed to know this stuff
- Hopefully first and last time we talk about this

Recap: Asymptotics for OLS and the Linear Model

OLS

$$y_i = \beta_0 + \beta x_i + u_i$$

Recall the three basic OLS assumptions

1. $E(u_i|X_i) = 0$
2. (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d.
3. Large outliers are rare $E[Y^4] < \infty$ and $E[X^4] < \infty$.

Gauss Markov Adds two assumptions:

1. $E(u_i|X_i) = 0$
2. (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d.
3. Large outliers are rare $E[Y^4] < \infty$ and $E[X^4] < \infty$.
4. $Var(u_i) = \sigma^2$ (homoskedasticity)
5. $u_i \sim N(0, \sigma^2)$ (normal errors)

Under these assumptions you learned that OLS is BLUE

Outliers and Leverage

One way to find outliers is to calculate the leverage of each observation i. We begin with the hat matrix:

$$P = X(X'X)^{-1}X'$$

and consider the diagonal elements which for some reason are labeled h_{ii}

$$h_{ii} = x_i(X'X)^{-1}x_i'$$

This tells us how influential an observation is in our estimate of $\widehat{\beta}$. Particularly important for $\{0, 1\}$ dummy variables with uneven groups.

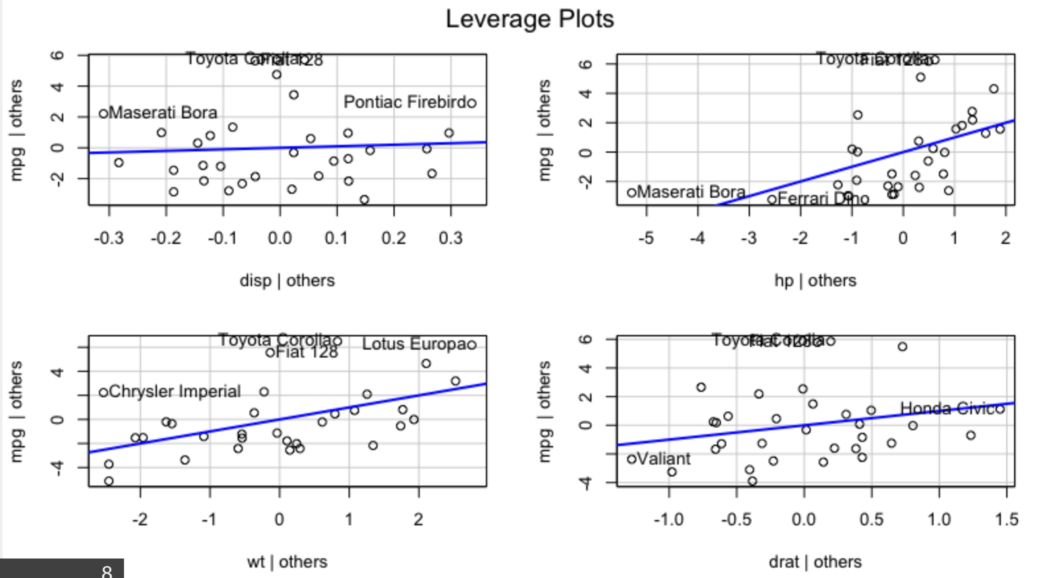
- This is sometimes called the Jackknife
- Sometimes it is helpful to know what would happen if we omitted a single observation i
- Turns out we don't need to run N regressions

$$\begin{aligned}\widehat{\beta}_{-i} &= (X'_{-i}X_{-i})^{-1}X'_{-i}Y_{-i}\\ &= \widehat{\beta} - (X'X)^{-1}x_i\tilde{u}_i \quad \text{where } \tilde{u}_i = (1-h_{ii})^{-1}\hat{u}_i\end{aligned}$$

- $\tilde{u}_i$ has the interpretation of the LOO prediction error.
- high leverage observations move $\widehat{\beta}$ a lot.

You can read more about this in Ch3 of Hansen. [Skip derivation]

Leverage and QQ plots


Leverage Plots
Toyota Corolla
Fiat 128
oMaserati Bora
Pontiac Firebirdo
mpg | others
disp | others
oMaserati Bora
oFerrari Dino
hp | others
Toyota Corollao
oFiat 128
Lotus Europao
oChrysler Imperial
wt | others
oValiant
Honda Civico
drat | others

Start with the variance of the residuals to form a diagonal matrix D:

$$\text{Var}(\mathbf{u}|\mathbf{X}) = \mathbb{E}\left(\mathbf{u}\mathbf{u}'|\mathbf{X}\right) = \mathbf{D}$$

$$\mathbf{D} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right) = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}$$

- $\mathbf{D}$ is diagonal because $\mathbb{E}[u_i u_j|X] = \mathbb{E}[u_i|x_i]\mathbb{E}[u_j|x_j] = 0$ (independence)
- The elements of D_i are given by $\mathbb{E}[u_i^2|X] = \mathbb{E}[u_i^2|x_i] = \sigma_i^2$.
- In the homoskedastic case $\mathbf{D} = \sigma^2\mathbf{I}_n$.

A useful identity for linear algebra:

$$\mathrm{Var}(a\mathbf{Z}) = a^2\,\mathrm{Var}(\mathbf{Z})$$
$$\mathrm{Var}(A\mathbf{Z}) = A\,\mathrm{Var}(\mathbf{Z})A'$$

Recall that $\mathrm{Var}(\mathbf{Y}|\mathbf{X}) = \mathrm{Var}(\mathbf{u}|\mathbf{X})$ and also recall the formula for $\widehat{\beta}$:

$$\widehat{\beta} = \underbrace{(X'X)^{-1}X'}_{A}\,Y = A'Y$$

$$\begin{aligned}\mathbf{V}_{\widehat{\beta}} = \mathrm{Var}(\widehat{\beta}|X) &= (X'X)^{-1}X'\,\mathrm{Var}(Y|X)X(X'X)^{-1} \\ &= (X'X)^{-1}(X'\mathbf{D}X)(X'X)^{-1}\end{aligned}$$

We have that $(X'\mathbf{D}X) = \sum_{i=1}^{N} x_i x_i' \sigma_i^2$. Under homoskedasticity $\mathbf{D} = \sigma^2 \mathbf{I}_n$ and $\mathbf{V}_{\widehat{\beta}} = \sigma^2 (X'X)^{-1}$.

$$\mathbf{D} = \text{diag}\left(\sigma_1^2, \ldots, \sigma_n^2\right) = \mathbb{E}\left(u_i u_i' | \mathbf{X}\right) = \mathbb{E}\left(\widetilde{\mathbf{D}} | \mathbf{X}\right)$$

We can estimate $\widehat{\mathbf{V}}_{\widehat{\beta}}$ by plugging in $\mathbf{D} \rightarrow \widetilde{\mathbf{D}}$:

$$\begin{aligned} \mathbf{V}_{\widehat{\beta}} &= (X'X)^{-1}(X'\widetilde{\mathbf{D}}X)(X'X)^{-1} \\ &= (X'X)^{-1}\left(\sum_{i=1}^{N} x_i x_i' u_i^2\right)(X'X)^{-1} \end{aligned}$$

The expectation shows us this estimator is unbiased:

$$\begin{aligned} E[\mathbf{V}_{\widehat{\beta}} | X] &= (X'X)^{-1}\left(\sum_{i=1}^{N} x_i x_i' E[u_i^2 | X]\right)(X'X)^{-1} \\ &= (X'X)^{-1}\left(\sum_{i=1}^{N} x_i x_i' \sigma_i^2\right)(X'X)^{-1} = (X'X)^{-1}(X'DX)(X'X)^{-1} \end{aligned}$$

Heteroskedasticity Consistent (HC) Variance Estimates

What we need is a consistent estimator for $\hat{u}_i^2$.

$$\mathbf{V}_{\widehat{\beta}}^{HC0} = (X'X)^{-1}\left(\sum_{i=1}^{N} x_i x_i' \hat{u}_i^2\right)(X'X)^{-1}$$

$$\mathbf{V}_{\widehat{\beta}}^{HC1} = (X'X)^{-1}\left(\sum_{i=1}^{N} x_i x_i' \hat{u}_i^2\right)(X'X)^{-1} \cdot \left(\frac{n}{n-k}\right)$$

Could use $\tilde{u}_i$ instead of $\hat{u}_i$ for a better estimate

$$\mathbf{V}_{\widehat{\beta}}^{HC2} = (X'X)^{-1}\left(\sum_{i=1}^{N} (1-h_{ii})^{-1} x_i x_i' \hat{u}_i^2\right)(X'X)^{-1}$$

$$\mathbf{V}_{\widehat{\beta}}^{HC3} = (X'X)^{-1}\left(\sum_{i=1}^{N} (1-h_{ii})^{-2} x_i x_i' \hat{u}_i^2\right)(X'X)^{-1}$$

- We know that $\mathbf{V}_{\widehat{\beta}}^{HC3} > \mathbf{V}_{\widehat{\beta}}^{HC2} > \mathbf{V}_{\widehat{\beta}}^{HC0}$ because $(1 - h_{ii}) < 1$.
- $HC3$ are the most conservative and also place the most weight on potential outliers.
- `Stata` uses $HC1$ as the default and it is what most people refer to when they say robust standard errors.
- These are often called White (1980) SE's or Eicher-Huber-White SE's.
- In small sample some evidence that $HC2$ does better.

Heteroskedasticity Consistent (HC) Variance Estimates

To read about SE's in `estimatr`:
`https://declaredesign.org/r/estimatr/articles/mathematical-notes.html`

What is Clustering?

Suppose we want to relax our i.i.d. assumption:

- Each observation i is a villager and each group g is a village
- Each observation i is a student and each group g is a class.
- Each observation t is a year and each entity i is a state.
- Each observation t is a week and each entity i is a shopper.

We might expect that $\mathrm{Cov}(u_{g1}, u_{g2}, \ldots, u_{gN}) \neq 0 \rightarrow$ independence is a bad assumption.

Clustering: Intuition

The groups (villages, classrooms, states) are independent of one another, but within each group we can allow for arbitrary correlation.

- If correlation is within an individual overtime we call it serial correlation or autocorrelation
- Just like in time-series$\rightarrow$ we have fewer effective independent observations in our sample.
- Asymptotics now about the number of groups $G \rightarrow \infty$ not observations $N \rightarrow \infty$

Begin by stacking up observations in each group $\mathbf{y}_g = [y_{g1}, \ldots, y_{gn_g}]$, we can write OLS three ways:

$$
\begin{aligned}
y_{ig} &= x_{ig}'\beta + u_{ig} \\
\mathbf{y}_g &= \mathbf{X}_g\beta + \mathbf{u}_g \\
\mathbf{Y} &= \mathbf{X}\beta + \mathbf{u}
\end{aligned}
$$

All of these are equivalent:

$$
\begin{aligned}
\widehat{\beta} &= \left(\sum_{g=1}^{G}\sum_{i=1}^{n_g} x_{ig}'x_{ig}\right)^{-1}\left(\sum_{g=1}^{G}\sum_{i=1}^{n_g} x_{ig}'y_{ig}\right) \\
\widehat{\beta} &= \left(\sum_{g=1}^{G} \mathbf{X}_g'\mathbf{X}_g\right)^{-1}\left(\sum_{g=1}^{G} \mathbf{X}_g'\mathbf{y}_g\right) \\
\widehat{\beta} &= \left(\mathbf{X}'\mathbf{X}\right)^{-1}\left(\mathbf{X}'\mathbf{Y}\right)
\end{aligned}
$$

Clustering (Continued)

The error terms have covariance within each cluster g as:

$$\Sigma_g = \mathbb{E}\left(\mathbf{u}_g\mathbf{u}_g'|\boldsymbol{X}_g\right)$$

In order to calculate $\widehat{V}_{\widehat{\beta}}$ we replace the covariance matrix $\mathbf{D}$ with Ω and consider an estimator $\widehat{\Omega}_n$. We exploit independence across clusters:

$$\text{var}\left(\left(\sum_{g=1}^{G}\boldsymbol{X}_g'\mathbf{u}_g\right)|\boldsymbol{X}\right) = \sum_{g=1}^{G}\text{var}\left(\boldsymbol{X}_g'\boldsymbol{u}_g|\boldsymbol{X}_g\right) = \sum_{g=1}^{G}\boldsymbol{X}_g'\mathbb{E}\left(\boldsymbol{u}_g\boldsymbol{u}_g'|\boldsymbol{X}_g\right)\boldsymbol{X}_g = \sum_{g=1}^{G}\boldsymbol{X}_g'\Sigma_g\boldsymbol{X}_g \equiv \Omega_N$$

And an estimate of the variance:

$$\boldsymbol{V}_{\widehat{\beta}} = \text{var}(\widehat{\beta}|\boldsymbol{X}) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\Omega_n\left(\mathbf{X}'\mathbf{X}\right)^{-1}$$

$$
\begin{aligned}
\widehat{\Omega}_n &= \sum_{g=1}^{G} X_g' \widehat{\mathbf{u}}_g \widehat{\mathbf{u}}_g' X_g \\
&= \sum_{g=1}^{G} \sum_{i=1}^{n_g} \sum_{\ell=1}^{n_g} x_{ig} x_{\ell g}' \widehat{u}_{ig} \widehat{u}_{\ell g} \\
&= \sum_{g=1}^{G} \left(\sum_{i=1}^{n_g} x_{ig} \widehat{u}_{ig} \right) \left(\sum_{\ell=1}^{n_g} x_{\ell g} \widehat{u}_{\ell g} \right)'
\end{aligned}
$$

- First line makes explicit: independence over each of G clusters
- Last line easiest for computer

$$\widehat{\boldsymbol{V}}_{\hat{\beta}}^{\mathrm{CR}1} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{g=1}^{G}\boldsymbol{X}_g'\widehat{u}_g\widehat{\boldsymbol{u}}_g'\boldsymbol{X}_g\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

$$\widehat{\boldsymbol{V}}_{\hat{\beta}}^{\mathrm{CR}3} = \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}\left(\sum_{g=1}^{G}\boldsymbol{X}_g'\widetilde{u}_g\widetilde{\boldsymbol{u}}_g'\boldsymbol{X}_g\right)\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}$$

- Can replace $\hat{\mathbf{u}}_g \to \tilde{\mathbf{u}}_g$ for leave-one out like *HC*3 (these are called *CR*3).

Clustering in R

Most Asked PhD Student Econometric Question

How should I cluster my standard errors?

- Heck if I know.
- This is very problem specific
- It matters a lot → standard errors can get orders of magnitude larger.
- Do you believe across group independence or not? [this is the only thing that matters]
- If you include fixed effects probably you need at least clustering at that level.

Newey West Standard Errors (HAC)

- In serially correlated data we need to account for $\text{Cov}(u_t, u_{t-1}, \ldots) \neq 0$.
- Clustering is one solution, but we may end up throwing away all of our data.
- Instead we could estimate the serial correlation.
- May also want standard errors that are heteroskedasticity AND autocorrelation consistent (HAC).
- Have to select a number of lags L

$$\widehat{\Omega}_{n,L}^{HAC} = \sum_{t=1}^{T} u_t^2 x_t x_t' + \sum_{l=1}^{L} \sum_{t=l+1}^{T} w_l u_t u_{t-l} \left(x_t x_{t-l}' + x_{t-l} x_t' \right)$$

$$w_l = 1 - \frac{l}{L+1}$$

What about β?

- All of the estimates above should produce identical point estimates
- We have just been talking about adjusting standard errors
- Should the presence of heteroskedasticity change our estimates of $\widehat{\beta}$ as well?

OLS and WLS

A simple extension is Weighted Least Squares (WLS)

- Different motivations
- Suppose we have sampling weights that are not $\frac{1}{n}$ from survey data, etc:
  - If my population is supposed to represent all US residents and my sample is 75% Women...
  - Relax LSA (2) (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d.
- In this case, OLS is still unbiased and consistent, just inefficient

WLS

Can weight each observation as w_i so that $\sum_{i=1}^{N} w_i = 1$ instead of $w_i = \frac{1}{N}$.
Can define a diagonal matrix W with entries w_i.

$$\arg\min_{\beta} \sum_{i=1}^{N} w_i (y_i - X_i\beta)^2 = \arg\min_{\beta} \left\| W^{1/2} | Y - X\beta | \right\|$$

Can also consider a transformation of the data

$$\tilde{y}_i = \sqrt{w}_i y_i, \quad \tilde{x}_i = \sqrt{w}_i x_i$$
$$\tilde{Y} = W^{1/2} Y, \quad \tilde{X} = W^{1/2} X$$

A regression of $\tilde{Y}$ on $\tilde{X}$:

$$\widehat{\beta}_{WLS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = (X'WX)^{-1}X'WY$$

WLS

Also used as a solution to heteroskedasticity

- Relax LSA (2) (X_i, Y_i), $i = 1, \ldots, n$, are i.i.d.
- Relax LSA (4) $Var(u_i) = \sigma^2$ (homoskedasticity)

Why? We are minimizing weighted sum of squared residuals:

$$\sum_{i=1}^{N} w_i (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} w_i \varepsilon_i^2$$

Suppose we have heteroskedasticity so that $Var(\varepsilon_i) = \sigma_i^2$ and $w_i \propto \frac{1}{\sigma_i^2}$.

In this setting WLS is BLUE.

Why does anyone ever run OLS instead of WLS?

- Problem is that σ_i^2 is unknown before we run our regression.
- We can estimate $\widehat{\sigma}_i^2$.

This procedure is known as Iteratively Re-weighted Least Squares IRLS

1. Intialize weights to identity matrix: $W = I$
2. Regress Y on X with weights W
3. Obtain $\widehat{\varepsilon}_i$.
4. Update W with $w_{ii} = \frac{1}{\widehat{\varepsilon}_i^2}$
5. Repeat until parameter estimates don't change

There is no reason to require that W be diagonal. This gives us Generalized Least Squares

$$\widehat{\beta}_{GLS} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\tilde{Y} = (X'\Omega X)^{-1}\Omega' WY$$

The idea is to use the inverse covariance matrix of residuals. But this is high dimensional ($N \times N$) and estimating it is harder than our original problem! Feasible Generalized Least Squares FGLS:

1. Intialize weights to identity matrix: $\widehat{\Omega} = I$
2. Regress Y on X with weighting matrix $\widehat{\Omega}$
3. Obtain $\widehat{\varepsilon}_i$.
4. Construct $E[\varepsilon_i^2|X, Z]$ via (nonlinear) regression: $\exp[\gamma_0 + \gamma_1 x_i + \gamma_2 z_i]$.
5. Update $\widehat{\Omega}$ with $E[\varepsilon_i^2|X, Z]$
6. Repeat until parameter estimates don't change

What is Panel Data?

We can combine cross sectional analysis and time series analysis to form panel data.

- Now y_{it} and x_{it} have two subscripts:
  - i for individual or entity
  - t for time
- It used to be that panel data was rare enough that it was a separate set of topics within econometrics. Now it is the norm.
- The main similarity to time series is that observations within an individual are correlated with one another
- The main similarity to cross sectional econometrics is that individuals are often treated as independent.

Terminology

Longitudinal data another term for panel data (especially in demography/sociology)

Repeated cross section not a panel, but a data structure with multiple individuals observed in each of multiple time periods. In contrast to panel data, we don't observe the same individuals in multiple time periods.

Balanced panel each of n individuals is observed T times, usually over the same time period

Unbalanced panel at least of the individuals are not observed in every period. Sometimes unbalanced panels result from sampling designs, and sometimes they are a result of entry/exit or birth/death

Sparse Panel very little overlap between (i, j). Think about matched firm-worker datasets (nobody works at every firm!)

Wide Panel has many individuals (large n); a

Long Panel has many time periods (large T). The asymptotic properties of an estimator can be different when $n \to \infty$ as opposed to $T \to \infty$

Basics of Panel Data

Often interested in a regression of the form:

$$y_{it} = \beta_i x_{it} + c_i + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

- With repeated observations on the same individual the assumption that u_{it} is I.I.D. is unrealistic $\rightarrow$ will need to adjust standard errors.
- Why? This year's outcome is likely related to last year's outcome...
- But with repeated observations on an individual we can control for a great deal of unobserved heterogeneity or omitted variables.
- We may want to include lagged $y_{i,t-1}$ as a regressor in dynamic panel models.

$$y_{it} = \beta_i x_{it} + c_i + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

- Full homogeneity (Pooled): $\beta_i = \beta$ $c_i = c$ for all i.
- Individual Effects: $\beta_i = \beta$ for all i.
- Full heterogeneity (β_i, c_i) are all different (potentially).

$$y_{it} = \beta x_{it} + c + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

- Requires that $E[x_{it}' u_{it}] = 0$ or that $E[e_{it} | \mathbf{X}_i] = 0$.
- Is this reasonable? Usually not

$$y_{it} = \beta x_{it} + c_i + u_{it} \quad i = 1, \ldots, N \quad t = 1, \ldots, T$$

Now we assume that $(\mathbf{y_i}, \mathbf{X_i})$ are i.i.d across i but not necessarily t with $\mathbf{X}_i = [X_{i1}, X_{i2}, \ldots, X_{iT}]$ Two well known cases:

Fixed Effects $E[u_{it}|\mathbf{X}_i, c_i] = 0$ conditional on FE, we have conditional mean independence

- Mostly about solving Omitted Variable Bias problem
- Unbiasedness and Consistency

Random Effects $E[c_i|\mathbf{X}_i] = 0$ individual effects are uncorrelated with information about individual i

- These are really about heteroskedasticity and efficiency
- The point estimates $\widehat{\beta}$ still change though (for same reason as WLS or GLS)

Random Effects

$$y_{it} = \beta x_{it} + c_i + u_{it}$$

- Not as popular in econometrics as they used to be
- Efficiency isn't the big concern, unbiasedness is
- We usually have enough data that reducing your SE's by 10% isn't an issue.
- Idea: re-scale the data so that it has spherical variance $\sigma^2 \cdot \mathbf{I}_N$

How are Random Effects Estimated?: FGLS

Step 1: Estimate the pooled regression

$$y_{it} = \beta x_{it} + e_{it}$$

Step 2: calculate means and variances:

$$\hat{\sigma}_e^2 = \frac{1}{NT} \sum_{i=1}^{T} \sum_{t=1}^{T} \hat{e}_{it}^2$$

$$\hat{c}_i = \frac{1}{T} \sum_{t=1}^{T} \hat{e}_{it}, \qquad \hat{u}_{it} = \hat{e}_{it} - \hat{c}_i$$

$$\hat{\sigma}_c^2 = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{c}_i - \frac{1}{N} \sum_{i=1}^{N} \hat{c}_i \right)^2$$

$$\hat{\sigma}_u^2 = \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \left(\hat{u}_{it} - \frac{1}{NT} \sum_{i=1}^{N} \sum_{t=1}^{T} \hat{u}_{it} \right)^2$$

Step 3: Update the Matrix $\widehat{\Omega}$

$$\hat{\Omega}_{RE} = \begin{bmatrix} \hat{\sigma}_c^2 + \hat{\sigma}_u^2 & \hat{\sigma}_c^2 & \cdots & \hat{\sigma}_c^2 \\ \hat{\sigma}_c^2 & \hat{\sigma}_c^2 + \hat{\sigma}_u^2 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_c^2 & \hat{\sigma}_c^2 & \cdots & \hat{\sigma}_c^2 + \hat{\sigma}_u^2 \end{bmatrix}$$

Step 4: Caclulate the (F)GLS estimator:

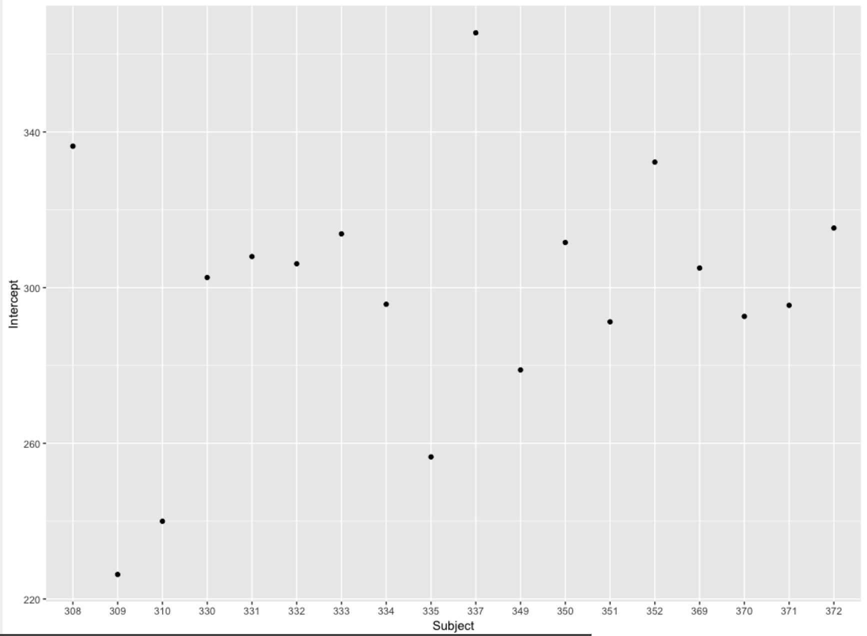
$$\hat{\beta}_{\mathrm{RE}} = \left(\sum_{i=1}^{N} \mathbf{X}_i' \hat{\Omega}_{RE}^{-1} \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{N} \mathbf{X}_i' \hat{\Omega}_{RE}^{-1} \mathbf{Y}_i \right)$$

How are Random Effects Estimated?: MLE

- For a number of reasons most software for random effects doesn't do FGLS
- `plm` does this `https://cran.r-project.org/web/packages/plm/vignettes/plmPackage.html`
- It usually assumes that $c_i \sim N(0, \sigma_c^2)$ and $u_{it} \sim N(0, \sigma_u^2)$
- In this world it is easy to do MLE.
- The package I will show you `lme4` does this. `https://cran.r-project.org/web/packages/lme4/vignettes/lmer.pdf`

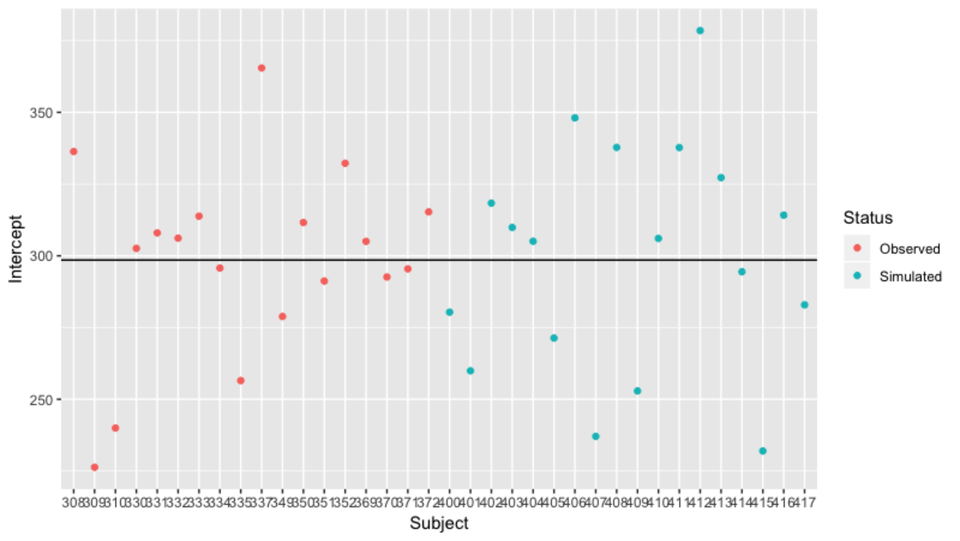
Random Effects in R

Random Effects in R



Intercept
340
300
260
220
308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Subject

Random Effects in R

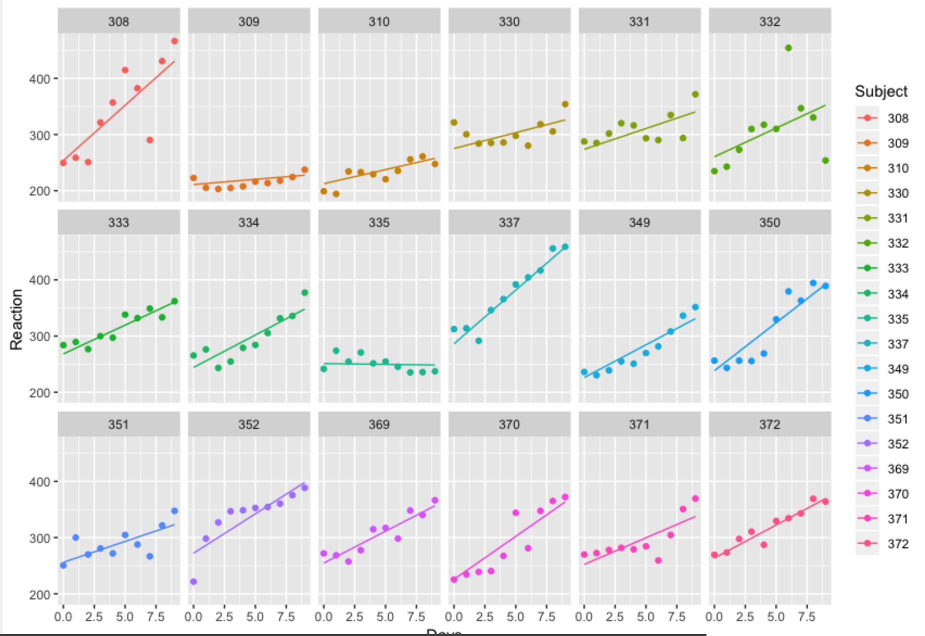


Intercept
350
300
250
Subject
308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
Status
Observed
Simulated

$$y_{it} = \beta_i x_{it} + c_i + u_{it}$$

- Can add a random slope term β_i as well
- This starts to get more useful
- Parametric restrictions $\beta_i \sim N(0, \sigma_b^2)$ prevent β_i realizations from getting too crazy.
- Later we will think about parametrizing this further $\beta_i(z_i)$

Random Slope and Intercept R


308
309
310
330
331
332
333
334
335
337
349
350
351
352
369
370
371
372
Reaction
400
300
200
0.0
2.5
5.0
7.5
Subject

Control Variables vs. Variables of Interest

We call a variable W_i a control variable if:

$$E[u_i|X_i, W_i] = E[u_i|W_i]$$

Consider the regression model

$$Y_i = \beta_0 + \beta_1 X_i + \beta_2 W_i + u_i$$

- β_1 has an interpreation
- $\widehat{\beta}_1$ is unbiased
- $\widehat{\beta}_2$ is potentially biased: something omitted might be correlated with W_i and a determinant of Y_i.

We could think about the same model but now instead of being part of the residual c_i is a dummy variable that we want to estimate a (fixed) coefficient on

$$y_{it} = \beta x_{it} + c_i + u_{it}$$

Now we require that $E[u_{it}|\mathbf{X}_i, c_i] = 0$

- Conditional on observing c_i we have conditional mean independence property satisfied again.
- c_i is just a conventional omitted variable: without it our estimate is biased, include it in the regression and everything is fine.

Fixed Effects as Controls

$$y_{it} = \beta x_{it} + c_i + u_{it}$$

A weaker condition is $E[u_{it}|\mathbf{X}_i, c_i] = E[u_{it}|c_i]$ but allowing $E[u_{it}|\mathbf{X}_i, c_i] \neq 0$

- Now the fixed effect functions only as a control
- We can't interpret c_i directly, it just proxies for all of the things we don't see
- Our estimates of $\hat{c}_i$ may be biased, but our estimtes of β remain unbiased.
- There is NO causal interpretation of fixed effects unless all variables correlated with Y_i and c_i are in the regression equation.

The fixed effects estimator

$$y_{it} = \beta x_{it} + c_i + u_{it}$$

Is equivalent to the within estimator

$$(y_{it} - \overline{y}_i) = \beta(x_{it} - \overline{x}_i) + (u_{it} - \overline{u}_i)$$

You should have learned about this last semester.
Also known as Absorb / Difference Out / Within Transform

The fixed effects estimator

$$y_{it} = \beta x_{it} + c_i + u_{it}$$

Is also equivalent to the least squres dummy variables (LSDV) regression:

$$y_{it} = \beta x_{it} + \sum_{i=1}^{N} \gamma_i \cdot \mathbf{I}_i + u_{it}$$

You should have learned about this last semester.

Fixed Effects: Multicolinearity

If we include a dummy (or fixed effect for every state we cannot estimate a constant term)

$$y_{it} = \beta_0 + \beta x_{it} + c_i + u_{it}$$

- Most software will drop one fixed effect
- Which fixed effect is dropped matters for c_i but not for $\widehat{\beta}$.

Fixed Effects: Multi-way

Often we want to include multiple dimensions of fixed effects

$$y_{it} = \beta x_{it} + c_i + c_t + u_{it}$$

Two ways to do this

- Within transform the larger dimension $\rightarrow$ Include dummies for the smaller dimension
- Transform the data in both dimensions using Frisch-Lovell-Waugh.
- Former when second dimension is small, latter when both are large

High Dimensional Fixed Effects: Conlon and Rao

- Suppose I want to incorporate store-upc and store-week FE using Nielsen Data.
  - Around 500 weeks since 2006.
  - Around 3000+ UPCs in a category like distilled spirits or breakfast cereal.
  - Can easily find ourselves estimating 50,000+ fixed effects in a single dimension and several thousand in the other.

There are several differencing algorithms for removing the fixed effects. For simplicitly let's assume there are two dimensions of fixed effects N and T where $N >> T$:

$$\tilde{y}_{it} = y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t}$$
$$\tilde{x}_{it} = x_{it} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t}$$

- Could do *iterative demeaning*: easy if $Cov(\bar{x}_{\cdot t}, \bar{x}_{i\cdot}) = 0$. Otherwise hard.
- Depends on graph structure of FE. Sparse FE are very difficult. Balanced Panels are easy.
- LSDV requires inverting the $(N + T) \times (N + T)$ matrix which can be difficult to impossible.

FE in Dimension One

$$\mathbf{Y} = \mathbf{Z}\beta + \mathbf{D}\alpha + \mathbf{u}$$

Partition $\mathbf{X} = [\mathbf{ZD}]$ where

$$\begin{bmatrix} \mathbf{Z'Z} & \mathbf{Z'D} \\ \mathbf{D'Z} & \mathbf{D'D} \end{bmatrix} \begin{bmatrix} \beta \\ \alpha \end{bmatrix} = \begin{bmatrix} \mathbf{Z'Y} \\ \mathbf{D'Y} \end{bmatrix}$$

Can be re-written

$$\begin{bmatrix} Z'\mathbf{Z}\beta + \mathbf{Z'D}\alpha = \mathbf{Z'Y} \\ \mathbf{D'Z}\beta + \mathbf{D'D}\alpha = \mathbf{D'Y} \end{bmatrix}$$

And construct normal equations

$$\begin{bmatrix} \beta = (\mathbf{Z'Z})^{-1}\mathbf{Z'}(\mathbf{Y} - \mathbf{D}\alpha) \\ \alpha = (\mathbf{D'D})^{-1}\mathbf{D'}(\mathbf{Y} - \mathbf{Z}\beta) \end{bmatrix}$$

FE in Dimension One

Idea is to Iterate on Normal Equations

$$\begin{bmatrix} \beta = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{Y} - \mathbf{D}\alpha) \\ \alpha = (\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'(\mathbf{Y} - \mathbf{Z}\beta) \end{bmatrix}$$

In one dimension this is silly because we just do the within trnasform. But this idea extends to higher dimensions.

$$\left[\begin{array}{c} \beta = (\mathbf{Z}'\mathbf{Z})^{-1}\,\mathbf{Z}'\,(\mathbf{Y} - \mathbf{D}_1\alpha - \mathbf{D}_2\gamma) \\ \alpha = (\mathbf{D}_1'\mathbf{D}_1)^{-1}\,\mathbf{D}_1'\,(\mathbf{Y} - \mathbf{Z}\beta - \mathbf{D}_2\gamma) \\ \gamma = (\mathbf{D}_2'\mathbf{D}_2)^{-1}\,\mathbf{D}_2'\,(\mathbf{Y} - \mathbf{Z}\beta - \mathbf{D}_1\gamma) \end{array}\right]$$

Note

- We residualize **Y**
- We don't mess with **X** at all
- But we run many regressions

Pass Through Regressions

$$\Delta p_{jt} = \beta_0 + \rho(X_{jt}, \Delta\tau_{jt})\Delta\tau_{jt} + \beta_2\Delta c_{jt} + B\Delta X_{jt} + \alpha_j + \gamma_t + \epsilon_{jt}$$

- γ_t is typically month + year FE.
- α_j is a product-specific trend in price.
- We estimate these in first differences, and vary the time horizon (1 month, 3 months, 6 months).
  - Sometimes we examine only firms which change their prices.

R Code

Pass-Through: Taxes to Retail Prices, CT

	All Retailers			Δ Retail Price$\neq 0$		
Δ Retail Price	1m	3m	6m	1m	3m	6m
	(1)	(2)	(3)	(4)	(5)	(6)
Δ Tax	1.533***	1.257***	1.013***	3.096***	2.301***	2.016***
	(0.271)	(0.202)	(0.264)	(0.706)	(0.479)	(0.553)
Δ Tax*I[size=750mL]	1.168***	1.900***	2.084***	3.191**	3.822***	4.072***
	(0.432)	(0.387)	(0.503)	(1.577)	(0.899)	(1.144)
Δ Tax*I[size=1L]	2.146***	1.833***	1.586***	5.550***	3.376***	3.553***
	(0.650)	(0.383)	(0.470)	(1.663)	(0.920)	(1.132)
Δ Tax *I[size=1.75L]	1.520***	1.154***	1.009***	2.985***	2.191***	2.027***
	(0.309)	(0.227)	(0.263)	(0.718)	(0.502)	(0.570)
Observations	460,116	437,057	410,288	75,227	113,098	142,220
Product FE	Yes	Yes	Yes	Yes	Yes	Yes

Note: All regressions are weighted by 2011 Nielsen units and include month and year fixed effects. Standard errors are clustered at the UPC level.

High Dimensional FE Examples: Backus, Conlon Sinkinson (2019)

$$\kappa_{fg,t} = \beta_1 \log \text{Market Cap}_{f,t} + \beta_2 \frac{1}{\text{Retail Share}_{f,t}} + \beta_3 \text{Indexing}_{f,t} +$$
$$\beta_4 \text{Indexing}^2_{f,t} + \beta_5 \text{Indexing}^3_{f,t} \beta_5 \mathit{BlackRock}_{f,t} + \beta_6 \mathit{Vanguard}_{f,t} + \beta_7 \mathit{StateStreet}_{f,t} + c_{f,g} + c_t + u_{f,g,t}$$

	(1)	(2)	(3)	(4)	(5)
$\frac{1}{1-r_{f,t}}$	0.314^{*}	0.301^{*}	0.304^{*}	0.305^{*}	-0.172^{*}
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\frac{1}{IHHI_{f,t}}$					45.505^{*}
					(0.083)
$\log(\text{market cap})_{f,t}$	0.081^{*}	0.078^{*}	0.077^{*}	0.077^{*}	0.029^{*}
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
$\text{Indexing}_{f,t}$	0.964^{*}	1.079^{*}	237.411^{*}	237.069^{*}	1.253^{*}
	(0.004)	(0.004)	(1.065)	(1.069)	(0.004)
$\text{Indexing}^2_{f,t}$			-68.704^{*}	-70.900^{*}	
			(0.656)	(0.636)	
$\text{Indexing}^3_{f,t}$				-19.064^{*}	
				(0.520)	
$\beta_{f,s,t}$ BlackRock		-0.406^{*}	-0.333^{*}	-0.344^{*}	-0.288^{*}
		(0.007)	(0.007)	(0.007)	(0.006)
$\beta_{f,s,t}$ Vanguard		-0.311^{*}	-0.234^{*}	-0.229^{*}	-1.226^{*}
		(0.016)	(0.016)	(0.016)	(0.014)
$\beta_{f,s,t}$ StateStreet		-0.509^{*}	-0.420^{*}	-0.414^{*}	-0.275^{*}
		(0.014)	(0.014)	(0.014)	(0.012)
N	17,397,247	17,397,247	17,397,247	17,397,247	17,397,247
R^2	0.735	0.735	0.737	0.737	0.797

Employment