Section 8: Panel Data and Fixed Effects

Valentine Gilbert

November 4, 2021

Overview

Nonlinear Hypotheses: Motivation

We've learned how to do inference for different kinds of **linear estimates**:

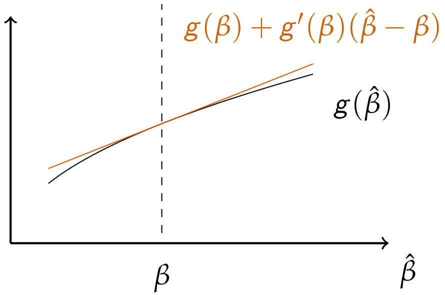
- Single parameters (e.g. $H_0 : \beta_1 = 0$)
- Linear combinations of parameters (e.g. $H_0 : 6\beta_2 + 72\beta_3 = 0$)
- Joint linear hypotheses (e.g. $H_0 : \beta_1 = 0$ AND $6\beta_2 + 72\beta_3 = 0$)

Inference about **nonlinear estimates** is more complicated

- For linear functions of parameter estimates, we use simple rules about the variance
- For example:
  - $var(6\hat{\beta}_1 + 72\hat{\beta}_3) = 6^2 var(\hat{\beta}_1) + 72^2 var(\hat{\beta}_3) + 2 * 6 * 72 cov(\hat{\beta}_1, \hat{\beta}_3)$
  - Then just plug in elements of the variance/covariance matrix to estimate this
- But how do we simplify $var(e^{\hat{\beta}_0 + 12\hat{\beta}_1 + \hat{\sigma}^2/2} - e^{\hat{\beta}_0 + 11\hat{\beta}_1 + \hat{\sigma}^2/2})$!?!?
- To estimate standard errors, we have to use the delta method or the bootstrap

The Delta Method: Intuition

- As we just saw, it's really hard to simplify the variance of a complicated nonlinear function of parameter estimates
- We get around this by using a **linear approximation** of our complicated nonlinear function
- Specifically, use a first-order Taylor-series approximation around the true values of our parameters: $g(\hat{\beta}) \approx g(\beta) + g'(\beta)(\hat{\beta} - \beta)$



The Delta Method: Estimating Variance

- Once we've taken a linear approximation of our nonlinear function, taking the variance of our estimate is just like in the linear case
- With one parameter:

$$g(\hat{\beta}) \approx g(\beta) + g'(\beta)(\hat{\beta} - \beta)$$
$$var(g(\hat{\beta})) \approx g'(\beta)^2 var(\hat{\beta})$$
$$\widehat{var}(g(\hat{\beta})) = g'(\hat{\beta})^2 \widehat{var}(\hat{\beta})$$

- With two parameters:

$$g(\hat{\beta}_0, \hat{\beta}_1) \approx g(\beta_0, \beta_1) + \frac{\partial g(\beta_0, \beta_1)}{\partial \beta_0}(\hat{\beta}_0 - \beta_0) + \frac{\partial g(\beta_0, \beta_1)}{\partial \beta_1}(\hat{\beta}_1 - \beta_1)$$
$$var(g(\hat{\beta}_{,0}\, \hat{\beta}_1)) \approx \left(\frac{\partial g(\beta_0, \beta_1)}{\partial \beta_0}\right)^2 var(\hat{\beta}_0) + \left(\frac{\partial g(\beta_0, \beta_1)}{\partial \beta_1}\right)^2 var(\hat{\beta}_1)$$
$$+ 2\frac{\partial g(\beta_0, \beta_1)}{\partial \beta_0}\frac{\partial g(\beta_0, \beta_1)}{\partial \beta_1} cov(\hat{\beta}_0, \hat{\beta}_1)$$

- Again, estimate this by plugging in sample estimates for β_0, β_1, $var(\cdot)$, and $cov(\cdot)$

The Delta Method: Simple Example

Consider the estimand $\beta_0\beta_1$. The estimator is $\hat{\beta}_0\hat{\beta}_1$. How do we estimate the variance of this with the delta method?

- $\hat{\beta}_0\hat{\beta}_1 \approx \beta_0\beta_1 + \beta_1(\hat{\beta}_0 - \beta_0) + \beta_0(\hat{\beta}_1 - \beta_1)$
- So $var(\hat{\beta}_0\hat{\beta}_1) \approx \beta_1^2 var(\hat{\beta}_0) + \beta_0^2 var(\hat{\beta}_1) + 2\beta_0\beta_1 cov(\hat{\beta}_0, \hat{\beta}_1)$
- $\widehat{var}(\hat{\beta}_0\hat{\beta}_1) = \hat{\beta}_1^2 \widehat{Var}(\hat{\beta}_0) + \hat{\beta}_0^2 \widehat{var}(\hat{\beta}_1) + 2\hat{\beta}_0\hat{\beta}_1 \widehat{cov}(\hat{\beta}_0, \hat{\beta}_1)$

How can we test the hypothesis that $\beta_0\beta_1 = 1$?

- Use the F-statistic: $F = (\hat{\beta}_0\hat{\beta}_1 - 1)^2 / \widehat{var}(\hat{\beta}_0\hat{\beta}_1) \xrightarrow{d} \chi_1^2$

The Delta Method: Trickier Example

Now let's return to the example of estimating the average effect of an increase in schooling from 11 to 12 years on the level of earnings

- We said that $\theta = e^{\beta_0+12\beta_1+\sigma^2/2} - e^{\beta_0+11\beta_1+\sigma^2/2}$
- So we can write $\theta \equiv g(\beta_0, \beta_1, \sigma^2)$
  - $\partial g(\cdot)/\partial\beta_0 = e^{\beta_0+12\beta_1+\sigma^2/2} - e^{\beta_0+11\beta_1+\sigma^2/2}$
  - $\partial g(\cdot)/\partial\beta_1 = 12e^{\beta_0+12\beta_1+\sigma^2/2} - 11e^{\beta_0+11\beta_1+\sigma^2/2}$
  - $\partial g(\cdot)/\partial\sigma^2 = .5e^{\beta_0+12\beta_1+\sigma^2/2} - .5e^{\beta_0+11\beta_1+\sigma^2/2}$
- Then plug these into:

$$var(\hat{\theta}) \approx \left(\frac{\partial g(\cdot)}{\partial\beta_0}\right)^2 var(\hat{\beta}_0) + \left(\frac{\partial g(\cdot)}{\partial\beta_1}\right)^2 var(\hat{\beta}_1) + \left(\frac{\partial g(\cdot)}{\partial\sigma_2}\right)^2 var(\hat{\sigma}^2)$$
$$+ 2\frac{\partial g(\cdot)}{\partial\beta_0}\frac{\partial g(\cdot)}{\partial\beta_1} cov(\hat{\beta}_0, \hat{\beta}_1) + 2\frac{\partial g(\cdot)}{\partial\beta_0}\frac{\partial g(\cdot)}{\partial\sigma^2} cov(\hat{\beta}_0, \hat{\sigma}^2) + 2\frac{\partial g(\cdot)}{\partial\beta_1}\frac{\partial g(\cdot)}{\partial\sigma_2} cov(\hat{\beta}_1, \hat{\sigma}^2)$$

- This is equivalent to $\frac{\partial g}{\partial\gamma}(\gamma)'\Omega\frac{\partial g}{\partial\gamma}(\gamma)/N$, where $\frac{\partial g}{\partial\gamma}(\gamma)$ is the vector of partial derivatives and Ω is the asymptotic variance/covariance matrix for my parameters

Panel Data 1/7: Motivation

- Suppose you want to estimate the effect of the minimum wage on employment
- Your strategy is to exploit variation across states in the timing of minimum wage increases using panel data on county-level employment from 1990-2006
- One idea is to estimate the following regression equation:

$$\ln employment_{ict} = \alpha + \eta \ln minwage_{it} + \varepsilon_{ict}$$

- What's wrong with this regression?
  - Does the conditional mean independence assumption hold? No! What are some omitted variables we might worry about?
    - Political party in control of state legislature
    - National economic conditions
    - State demographics
- How can panel data help us? We can control for any characteristics that vary by state but are constant over time or for characteristics that vary over time but are constant across states (including unobservables!)

Panel Data 2/7: Estimation

How does panel data help us control for unobservables?

- Consider demeaning the long regression that controls for a variable that's constant within state:

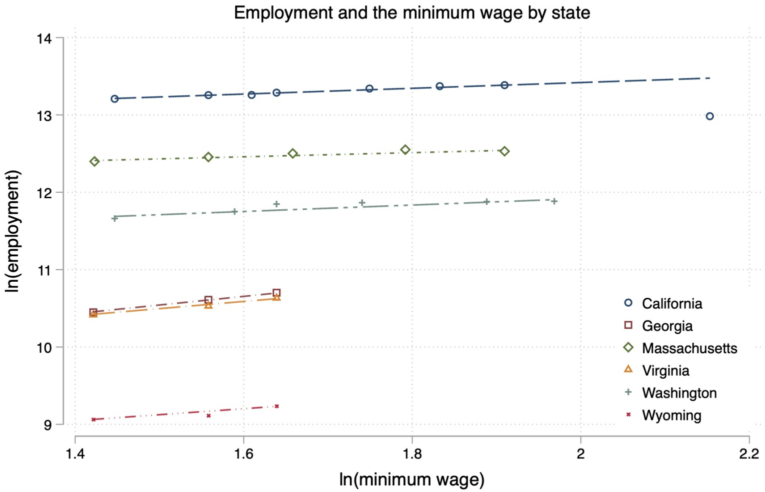
$$\begin{aligned} \ln y_{ict} &= \beta_0 + \beta_1 \ln minwage_{it} + \beta_2 Z_i + u_{ict} \\ \implies \overline{\ln y}_i &= \beta_0 + \beta_1 \overline{\ln minwage}_i + \beta_2 Z_i \\ \implies \ln y_{ict} - \overline{\ln y}_i &= \beta_1 [\ln minwage_{it} - \overline{\ln minwage}_i] + u_{ict} \end{aligned}$$

- The last regression is one we can estimate without controlling for Z_i!
- From the Frisch-Waugh-Lovell theorem, we know we can also estimate this regression by including fixed effects in our model:

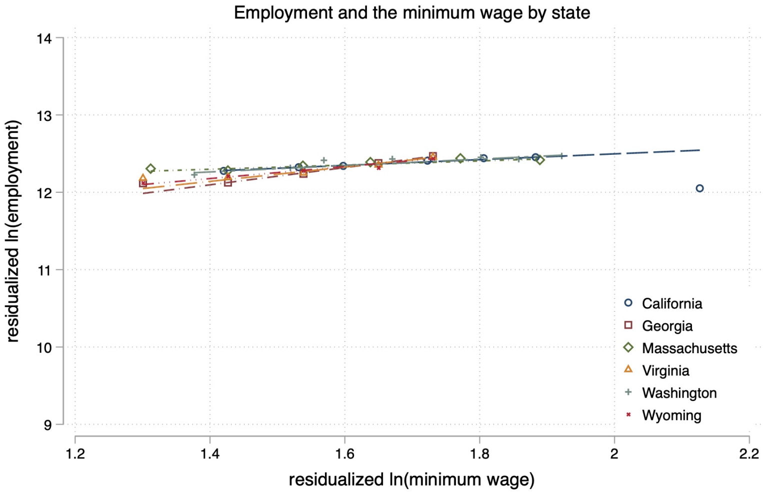
$$\ln y_{ict} = \delta_i + \beta_1 \ln minwage_{it} + u_{ict}$$

- The fixed effects effectively demean the data for us, which takes care of any omitted variables that are constant within states
- The same logic applies to time fixed effects, which difference out time shocks/trends that are shared across states

- This binned scatter plot shows the relationship between ln(employment) and ln(minwage) for the states with the highest and lowest minimum wages in 2018
- Notice that states with high levels of employment also tend to have high minimum wages.

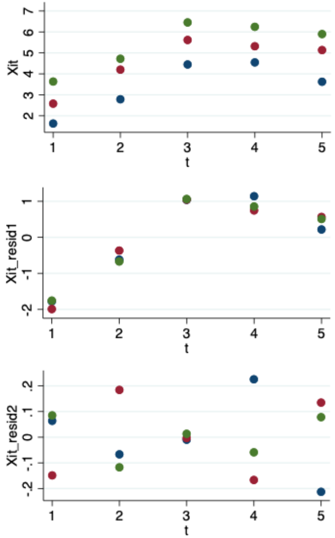


- This binned scatter plot shows the relationship between ln(employment) and ln(minwage) for the same states once we've partialled out state effects.
- State effects essentially discard between-state variation in employment and the minimum wage, keeping only the within-state variation.

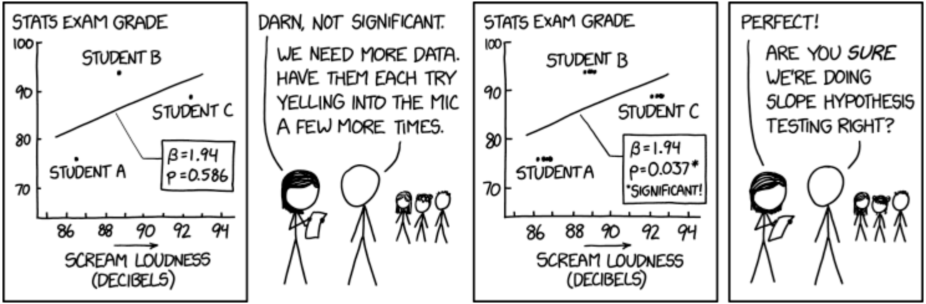


Panel Data 5/7: Identifying Variation

- In models with fixed effects, it's helpful to think about where your "identifying variation" is coming from and how much there is
- This refers to the variation in your independent variable of interest
- The figures show how variation in X_{it} changes when we add group and time fixed effects
- Group fixed effects eliminate variation across groups (colors), while time fixed effects eliminate variation across time t



Panel Data 6/7: Inference



- Uncertainty about your parameter estimates shrinks as you get more information
- New observations that are correlated with existing observations add less information
- Our typical standard error formulas are inconsistent if observations aren't independent

Panel Data 7/7: Inference

- With grouped data, we need to adjust standard errors to account for clustering
- If your observations are correlated within a cluster, then those observations don't tell you as much as if you had the same number of i.i.d. observations.
- Examples:
    - Students' test scores within a classroom are correlated, since they share the same teacher, peers, and classroom environment.
    - Unemployment levels of counties within a state are correlated, since they share the same state-level policy environment.
    - An individual's wages over time are correlated, since they're earned by the same person.
- You should cluster your standard errors at the level of the policy variation.
    - For intuition, imagine randomizing students to small or large classrooms
    - Would you learn much from your experiment if there were only two classrooms?

Fixed Effects 1/4: Omitted variables

- We saw earlier that fixed effects can help us difference out omitted variables that are constant within cross-sectional units or within time
- Going back to the minimum wage example, what are some omitted variables we *do* have to worry about even after controlling for year and state fixed effects?
  - State specific employment trends
  - Demographic shifts within states
  - Changes in a state's political climate
  - Anything that varies at the state-year level and is potentially correlated with the minimum wage and employment
- What if we added state-by-year fixed effects?
  - We would have no variation left in the minimum wage variable! It varies at the state-by-year level.
- Could we add region-by-year fixed effects?

Fixed Effects 2/4: Identification and Collinearity

- An important model in labor economics is the AKM model (after Abowd, Kramarz, and Margolis (ECMA 1999))
- The AKM model uses matched firm and employee data to decompose wages into a part that's attributable to worker ability and a part that's due to firms
- This is done by simply regressing wages on worker and firm fixed effects as well as controls (e.g. education)
- What kind of variation do you need to estimate these fixed effects?

Fixed Effects 3/4: Identification and Collinearity

- We need workers who switch firms
- Consider a case with 2 firms (indexed by $j \in \{1,2\}$) and 2 workers (indexed by $i \in \{1,2\}$). We can estimate the regression:

$$wage_{ij} = \beta_0 + \beta_1 \times \mathbb{1}[i=2] + \beta_2 \times \mathbb{1}[j=2] + u_{ij}$$

- How do we interpret these coefficients?

$$\begin{aligned} \beta_0 &= E[wage_{ij} | i=1, j=1] \\ \beta_0 + \beta_1 &= E[wage_{ij} | i=2, j=1] \\ \beta_0 + \beta_2 &= E[wage_{ij} | i=1, j=2] \\ \beta_0 + \beta_1 + \beta_2 &= E[wage_{ij} | i=2, j=2] \end{aligned}$$

- So what happens if worker 1 only works at firm 1 and worker 2 only works at firm 2?
- What if worker 1 only works at firm 2 and worker 2 only works at firm 1?
- What if worker 1 works at both firms but worker 2 only works at firm 2?

Fixed Effects 4/4: Identification and Collinearity

- This idea generalizes to cases with many workers and many firms.
- In general, fixed effects are identified for a connected set of firms and workers.

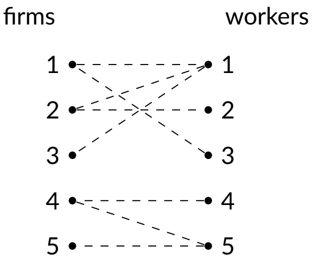


Figure: Two Connected Sets of Firms and Workers

- Suppose we try to estimate this regression equation with worker fixed effects δ_i and firm fixed effects γ_j:

$$wage_{ij} = \alpha + \gamma_2 + \gamma_3 + \gamma_4 + \gamma_5 + \delta_2 + \delta_3 + \delta_4 + \delta_5 + u_{ij}$$

- What will happen? You can't because of multicollinearity: $\mathbb{1}[j = 4] + \mathbb{1}[j = 4] = \mathbb{1}[i = 4] + \mathbb{1}[i = 5]$

Pset Review: Interpreting Racial Disparities

- On problem set 6, you tested a hypothesis about racial disparities in judicial sentencing decisions
- Specifically, you estimated

$$Y_i = \alpha + \beta \text{nonblack}_i + \sum_{j=2}^{9} \left\{ \gamma_j \times \mathbb{1}[\text{calendar}_i = j] + \delta_j \times \left(\mathbb{1}[\text{calendar}_i = j] \times \text{nonblack}_i \right) \right\} + \eta_w +$$

- Then you tested the joint hypothesis that $\delta_j = 0$ for all j
- In plain English, what exactly is this hypothesis testing?
- Why didn't we test the hypothesis that the race gap is zero?
  - Race may be correlated with arrest history or other factors that influence judicial decisions

Pset Review: Interpreting Racial Disparities

- You were able to reject the hypothesis that racial sentencing disparities are constant across judges.
  - Is this evidence of racial discrimination?
  - Could instead reflect judges weighing different offenses differently, for example
  - What if we had been unable to reject our hypothesis? Would that suggest judges aren't racially discriminatory?
- Suppose we hadn't had conditional random assignment – how would this have changed our interpretation of the analysis?
- Arnold, Dobbie, and Yang (QJE 2018) and Arnold, Dobbie, and Hull (NBER 2020) are great papers on racial bias in bail decisions

Pset Review: OVB and Interaction Effects

- On problem set 6, you had to discuss how the predicted effect of increasing incarceration could be biased by OVB

$$\begin{aligned}
\text{laterarr}_i &= \alpha_0 + \alpha_1 \text{probat}_i + \alpha_2 \text{toserve}_i + \alpha_3 \text{probat}_i \times \text{toserve}_i + u_i \\
\text{laterarr}_i &= \beta_0 + \beta_1 \text{probat}_i + \beta_2 \text{toserve}_i + \beta_3 \text{probat}_i \times \text{toserve}_i + \beta_4 Z_i + v_i \\
Z_i &= \gamma_0 + \gamma_1 \text{probat}_i + \gamma_2 \text{toserve}_i + \gamma_3 \text{probat}_i \times \text{toserve}_i + e_i
\end{aligned}$$

- Predicted effect: $6\alpha_2 - 72\alpha_3$
- To recover the effect of omitted variable bias, you can always plug the auxiliary regression into the long regression:

$$\begin{aligned}
\text{laterarr}_i = {} & \beta_0 + \beta_1 \text{probat}_i + \beta_2 \text{toserve}_i + \beta_3 \text{probat}_i \times \text{toserve}_i \\
& + \beta_4 [\gamma_0 + \gamma_1 \text{probat}_i + \gamma_2 \text{toserve}_i + \gamma_3 \text{probat}_i \times \text{toserve}_i + e_i] + v_i \\
= {} & [\beta_0 + \beta_4 \gamma_0] + [\beta_1 + \beta_4 \gamma_1] \text{probat}_i + [\beta_2 + \beta_4 \gamma_2] \text{toserve}_i \\
& + [\beta_3 + \beta_4 \gamma_3] \text{probat}_i \times \text{toserve}_i + [v_i + \beta_4 e_i]
\end{aligned}$$

Pset Review: OVB and Interaction Effects

- We just showed that $\alpha_2 = \beta_2 + \beta_4\gamma_2$ and $\alpha_3 = \beta_3 + \beta_4\gamma_3$
- So our predicted effect is

$$\begin{aligned} 6\alpha_2 + 72\alpha_3 &= 6[\beta_2 + \beta_4\gamma_2] + 72[\beta_3 + \beta_4\gamma_3] \\ &= 6\beta_2 + 72\beta_3 + \underbrace{6\beta_4\gamma_2 + 72\beta_4\gamma_3}_{\text{omitted variable bias}} \end{aligned}$$

- You already know how to think through the sign of β_4 and γ_2
- But what about the sign of γ_3? Signing the coefficient on the interaction term is hard
- To sign γ_3, it helps to consider the marginal effect of toserve$_i$ on Z_i:

$$\frac{\partial Z_i}{\partial \text{toserve}_i} = \gamma_2 + \gamma_3 \text{probat}_i$$

- So $\gamma_3 > 0$ if the relationship between Z_i and toserve$_i$ is more positive when probat$_i$ is higher