Econ 2120, Section 7 - Random vs. Fixed Effects

October 18th 2021

Table of Contents

A variance decomposition

- Imagine we're interested in understanding the size of wage shocks over time
- We have a panel of households with their earnings. What should we do?
  - Can we look at the variance across all the wages we observe?
- Intuitively, we want a variance decomposition
  - How much of the observed variance is not due to unobserved differences across workers?

Example from Gottschalk and Mofitt (2002)

- Authors look at the rise in income inequality – the rise in variance of income in US from 70s to mid 90s
- Data is a panel of male worker wages
- Permanent earnings variance (price of skill) vs Transitory variance (labor market instability, increase in cometitiveness etc.)

Example from Gottschalk and Mofitt (2002)

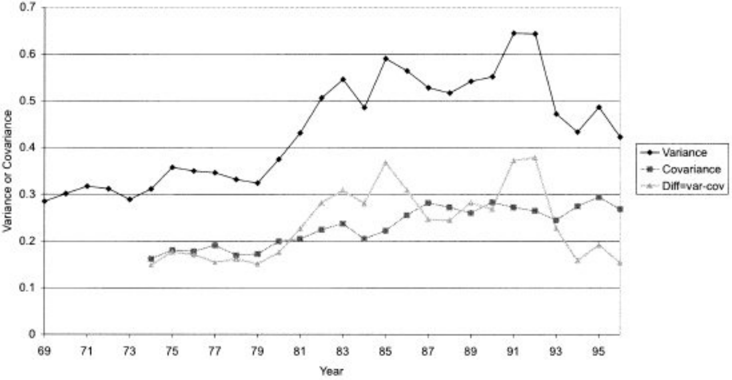

Fig. 1. *Permanent and Transitory Variances of Male Log Earnings,* 1969–1996

Working with Residuals

- If we have a population vector Y_i and predictors Z_i, we can form residuals from the unrestricted predictor:

$$U_i = \begin{pmatrix} U_{i1} \\ \vdots \\ U_{iT} \end{pmatrix} = Y_i - (I \otimes X_i')\pi$$

  - Recall that $Z_{ist} \in span(X_{i1}, X_{i2}, \dots) \;\forall\; s, t$

- To start, let's model U_{it} as a function of some omitted variable and random shocks

$$U_{it} = A_i + V_{it}$$

with $A_i, V_{i1}, \dots, V_{iT}$ mutually uncorrelated

Working with Residuals

- Define
$$\sigma_A^2 = Var(A_i),\ \sigma_{V_t}^2 = Var(V_{it})$$
- Then it follows that
$$\Sigma_U = \begin{pmatrix} (\sigma_A^2 + \sigma_{V_1}^2) & \sigma_A^2 & \dots & \sigma_A^2 \\ \sigma_A^2 & (\sigma_A^2 + \sigma_{V_2}^2) & \dots & \sigma_A^2 \\ \vdots & \vdots & & \vdots \\ \sigma_A^2 & \sigma_A^2 & \dots & (\sigma_A^2 + \sigma_{V_T}^2) \end{pmatrix}$$
- What assumptions have we made to get here? When might they be implausible?

Writing the Variance Decomposition as GLS

- We're interested in learning $\gamma \equiv (\sigma^2_A \;\; \sigma^2_{V_1} \;\; \ldots \;\; \sigma^2_{V_T})'$
- We can simply recast the matrix Σ_U as a vector: $vec(\Sigma_U)$
- Then, there is a known matrix G such that

$$vec(\Sigma_U) = G\gamma$$

- Which suggests running the regression

$$E^*_\Phi[vec(\Sigma_U)|G]$$

  The coefficients of this regression will correspond to γ under the model

Why use GLS?

- The best linear predictor guarantees our prediction is optimal
- We also know what happens when our model is mis-specified.
  - G is 'wrong' in the sense that it doesn't correspond to the unrestricted predictor
- Then we interpret the estimated γ as best approximation to the unrestricted predictor in a minimum-distance sense
  - Where the distance metric is defined for a given Φ

Example from Gottschalk and Mofitt (2002)

Possible extension:

$$U_{it} = \alpha_t A_{it} + V_{it}$$
$$A_{it} = A_{it-1} + \omega_{it}$$
$$V_{it} = \rho_t V_{it-1} + \varepsilon_{it}$$

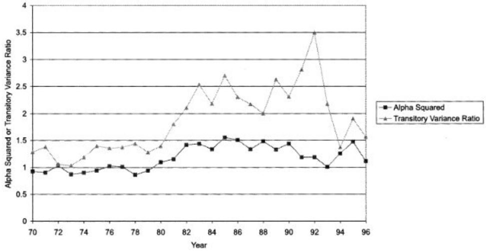

Fig. 2. *Trends in Alpha Squared and the Variance of Transitory Shocks*

Table of Contents

Example of Unit Effects: Teacher Effects

- Question: How much does one year of exposure to a given teacher improve test scores?
- We observe a teacher x year panel of mean test scores as well as classroom demographics

$$Y_i = (Y_{i1} \ \ldots \ Y_{iT})'$$

- What do we do?

Teacher Effects Setup

- As before calculate the residuals U_{it}
- Model errors as $U_{it} = V_i + \epsilon_{it}$
- Define $\sigma_V^2 \equiv E[V_i^2]$ and $\sigma_{st} \equiv E[\epsilon_{is}\epsilon_{it}]$
- Make three substantive restrictions: one interpretation and one statistical
  - Interpretation of the teacher effect: $E[V_i\epsilon_{it}] = 0$
  - No serial correlation: $\sigma_{st} = 0 \ \ \forall \ s \neq t$
  - Homoskedasticity: $\sigma_{tt} = \sigma_\epsilon^2 \ \ \forall \ t$
- What do these assumptions mean in this context?

Approach 1: Unbiased Estimator

- A reasonable estimator for V_i would be

$$\hat{U}_i^{unbiased} = \bar{U}_i = \sum_{t=1}^{T} U_{it}/T$$

- It's nice in that it's both unbiased and consistent in T

$$E_i[\bar{U}_i] = V_i, \quad \lim_{T\to\infty} \bar{U}_i = V_i$$

- But it turns out this estimator is not optimal under squared loss

Approach 2: Best Linear Predictor

- Another estimator of interest might be

$$\hat{U}^{BLP}_{i,T+1} = E^*[U_{i,T+1} | U_{i1}, \ldots, U_{iT}]$$

  Why?

- Under the model (using the variance decomp):

$$\hat{U}_{i,T+1} = \frac{\sigma_V^2}{\sigma_V^2 + \sigma_\epsilon^2 / T} \bar{U}_i$$

- Note that $\frac{\sigma_V^2}{\sigma_V^2 + \sigma_\epsilon^2 / T} < 1$ and goes to one as $T \to \infty$

- We call this a 'shrinkage' estimator because the unbiased estimator is 'shrunk' towards its mean

Comparing the Approaches

- So which estimator should we prefer?
- Which estimator should you use if:
  1. You want to identify the best teacher
  2. You want to know how well teacher effects are predicted by observables

Table of Contents

Summary

Many important points...

- Panel data can be very useful
- So far: dealing with OVB, analysis of variance, predicting the (unobserved) unit effects
- Can look at the variances, not only at the means + minimum distance
- Shrinkage