Econ 312

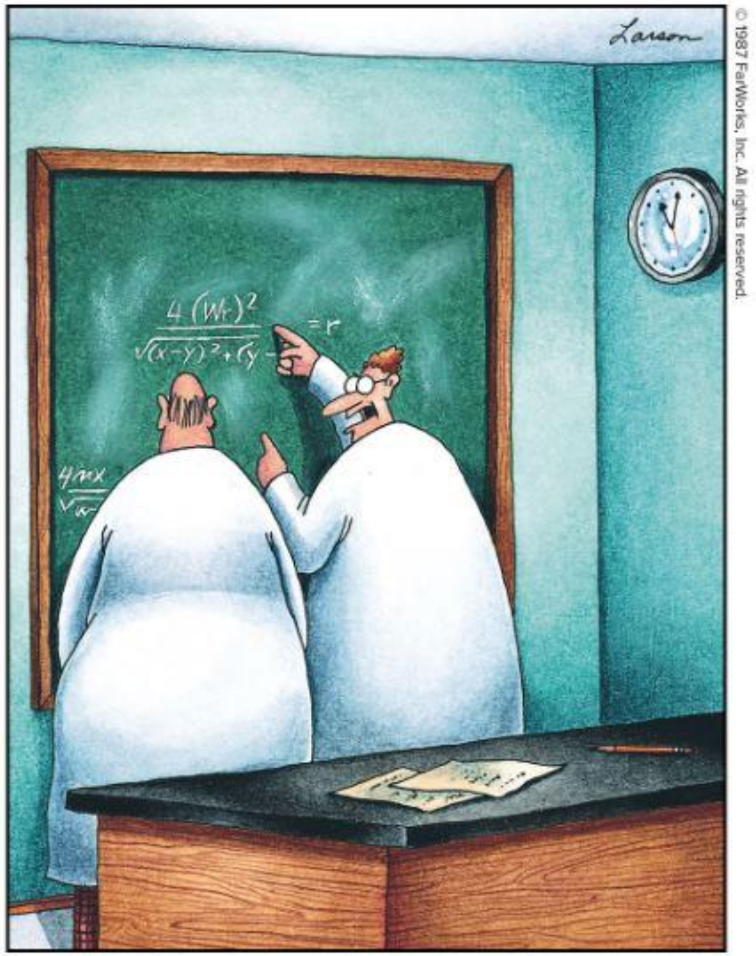
Monday, April 20

Estimating Simultaneous Equations

Readings: Wooldridge, Chapter 16

Class notes: 149 - 153

Today’s Far Side offering



“Yes, yes, I *know* that, Sidney—everybody knows *that!* ... But look: Four wrongs *squared,* minus two wrongs to the fourth power, divided by this formula, *do* make a right.”

Use high-powered math cautiously in social sciences

Context and overview

- We examined the identification problem in simultaneous-equation models in the previous class
- Today we talk about how to estimate identified models as entire systems of equations
- There are several reasons we might want to do this, including efficiency and the ability to test coefficient restrictions across equations
- **Seemingly unrelated regressions** have no endogenous variables on the right, but we still may want to estimate them together
- **Three-stage least squares** extends the two-stage least-squares estimators of individual equations to estimate as a system

System estimation or individual equations?

Two reasons for estimating equations jointly rather than individually:

1. **More efficient** because you can take account of autocorrelation between the error terms of the same observation in different equations
   - This is like panel-data model where the error terms of one state in the different periods might be correlated or the error terms of different stats in same period might be correlated
2. **Can impose or test restrictions** between coefficients of one regression and those of another

Seemingly unrelated regressions (SUR)

- **No endogenous regressors**, but we still want to do joint estimation
- Sometimes called "Zellner-efficient" estimator
- Use `sureg` command in Stata

- Implemented by "stacking" regressions for individual equations on top of each other (stacked regression)

Setup of SUR

- Three individual equations with n observations:

$$\mathbf{y}_1 = \mathbf{X}_1\boldsymbol{\beta}_1 + \mathbf{u}_1,$$
$$\mathbf{y}_2 = \mathbf{X}_2\boldsymbol{\beta}_2 + \mathbf{u}_2,$$
$$\mathbf{y}_3 = \mathbf{X}_3\boldsymbol{\beta}_3 + \mathbf{u}_3.$$

- $\mathbf{X}_1$, $\mathbf{X}_2$, and $\mathbf{X}_3$ can be the same or different exogenous variables for the three equations, but cannot include any $\mathbf{y}$ variables
- We will create stacked matrices out of these vectors and matrices

Stacked regressions

- Define **y**, **X**, **u**, and **β** as

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \mathbf{y}_3 \end{pmatrix}_{3n\times 1} \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_3 \end{pmatrix}_{3n\times(k_1+k_2+k_3+3)} \quad \mathbf{u} = \begin{pmatrix} \mathbf{u}_1 \\ \mathbf{u}_2 \\ \mathbf{u}_3 \end{pmatrix}_{3n\times 1} \quad \boldsymbol{\beta} = \begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \boldsymbol{\beta}_3 \end{pmatrix}_{(k_1+k_2+k_3+3)\times 1}$$

- We can write the system of equations as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$$

- Could estimate by OLS, but not efficient if the error of observation i is **correlated across equations**: $\text{cov}(u_{mi}, u_{li}) = \sigma_{ml} \neq 0$
- **Heteroskedasticity** occurs if $\sigma_m^2 \neq \sigma_l^2$

Efficient estimation of SUR system

- Can use **generalized least squares** to correct for inefficiencies due to autocorrelation and heteroskedasticity
  - Need to estimate the variances and covariances of the error terms across equations: σ_m^2 and σ_{ml} for each m and l
- Two-step procedure (but can be iterated):
  1. **Estimate stacked regression** by OLS and get residuals, then estimate variances and covariances with them

$$\hat{\sigma}_m^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_{mi}^2 \qquad \hat{\sigma}_{ml} = \frac{1}{n}\sum_{i=1}^{n}\hat{u}_{mi}\hat{u}_{li},$$

  2. Use resulting covariance matrix of **u** to **re-estimate system efficiently**
     - Detailed matrix equations are in notes

SUR is more efficient … unless

- If all of the equations have **exactly the same regressors**, so $\mathbf{X}_1 = \mathbf{X}_2 = \mathbf{X}_3$, then SUR is equivalent to OLS of individual equations and there is no efficiency gain from joint estimation
  - Might still want to use SUR to test/impose cross-equation restrictions

- If there is **no correlation among the error terms** across equations, then SUR is exactly the same as OLS of individual equations

SUR in Stata

- sureg (*dvar1 indvars1*) (*dvar2 indvars2*) (*dvar3 indvars3*)
  - Option isure iterates until estimated covariances converge
- To impose cross-equation constraints in estimation use constraints option:
  - constraints ([*dvar1*]*indvar1j* = [*dvar2*]*indvar2j*) imposes the constraint that the *indvar1j* coefficient in the equation for *dvar1* equals the *indvar2j* coefficient in the equation for *dvar2*
- To test such constraints, use the corresponding test command after estimation of sureg, with syntax similar to the constraints above

Three-stage least squares (3SLS)

- If some elements of **x** are endogenous, need to combine 2SLS and SUR to get 3SLS

1. **Estimate reduced forms** by OLS/SUR
   - Calculate fitted values to fill in for endogenous x
2. **Estimate second stage** as in 2SLS for each equation individually
   - Use 2SLS residuals to estimate variances and covariances of error terms within and between equations
3. **Estimate third stage doing SUR** with fitted values of endogenous x replacing actual values

Maximum-likelihood estimators

- 2SLS and 3SLS are NOT maximum-likelihood estimators

- With normal error terms, we can use
  - **Limited-information maximum likelihood** (LIML) ~ 2SLS
    - Change the 2sls specification in Stata to liml
  - **Full-information maximum likelihood** (FIML) ~ 3SLS
    - I cannot find this in Stata, but there is a general-purpose ml command that could be programmed for FIML

Review and summary

- Estimating related equations jointly increases efficiency (usually) and allows cross-equation restrictions on coefficients to be imposed or tested
- **Seemingly unrelated regressions** are jointly estimated systems with no endogenous variables as regressors
- **Three-stage least squares** generalizes two-stage least squares (with endogenous regressors) to allow system estimation
- Limited-information and full-information maximum likelihood are the corresponding ML estimators

Another bad economist joke …

A guy walks into a Washington D.C. curio shop. After browsing, he comes across an exquisite brass rat.

"What a great gag gift," he thinks to himself. After dickering with the shopkeeper over the price, the man purchases the rat and leaves.

As he's walking down the street, he hears scurrying sounds behind him. Stopping and looking around, he see hundreds, then thousands of rats pouring out of alleys and stairwells into the street behind him. In a panic, he runs down the street with the rats not far behind.

The street ends at a pier. He runs to the end of the pier and heaves the brass rat into the Potomac River. All of the rats scurry past him into the river, where they drown.

After breathing a sigh of relief and wiping his brow, the man heads back to the curio shop, finds the shopkeeper, and asks, "Do you have any brass economists?"

--Taken from Jeff Thredgold, *On the One Hand: The Economist's Joke Book*

What's next?

- This class concludes our study of endogenous regressors and systems of regression equations
- Next we turn to regressions in which the dependent variable has a **limited distribution**
  - Dummy dependent variables
  - Restricted dependent variables (*e.g.*, cannot be negative)
  - Ordered or categorical dependent variables
  - Count dependent variables
  - Hazard and duration models