Simultaneous Equations

EC420 MSU

Justin Kirkpatrick
Last updated March 22, 2020

Last week

Endogeneity in regression

- u correlated with an explanatory variab [redacted] 1

Instrumental Variables

- A [redacted]
  - [redacted] 1 [redacted] u
  - [redacted] *exclusion restriction*
- Allowed us [redacted] *ident* [redacted] the coefficient of intere [redacted] 1

Estimated with 2SLS

- Th [redacted]
  - But we won't cover them here

Today

- More endogeneity!
- *Simultaneous* equations
- Structural equations
  - Reduced form equations
- When can we identify?
- How do we estimate
  - 2SLS
- In panel data
  - First diff
  - Fixed effects

We think of the variables in our data as being either "endogenous" or "exogenous"

This tells us whether or not we should be worried about correlation with u.

Exogenous

Exogenous means "determined outside the system".

- Things like *rainfall* in ag production and *winning the KIPP lottery* are *exogenous*
    - There is usually nothing *inside* the system that helps determine them.
    - Although...we could think of times that even rainfall is endogenous.
    - Any ideas about how *rainfall* could be endogenous?
    - What about a model that includes the selection of land for farming?
    - How could we fix this?

Endogenous

Endogenous means "determined within the system"

- Things like "education" in a wage equation and "moving my child to an (open-enrollment) charter school" are *endogenous*

We have had a few cases like this.

- First, when x_1 and x_2 are correlated
  - Not a problem when x_1 and x_2 are observable - we just include both in the regression or we "partial out".
- Next, when the variable of interest, x_1 is correlated with an unobservable in u
  - For this we, need an instrument

$$wage = B_0 + B_1 \, educ + u$$

Endogeneous continuted...

For instance:

- When (Y_{0i}, Y_{1i}) are different for some people, and *those* people choose the treatment based on that difference
  - Zuckerberg's Y_{1i} (wages) for dropping out of Harvard was extraordinarily high. His choice was *endogenous*.
  - If we included him and people like him in a regression of $wages = \beta_0 + \beta_1 1(drop - out - of - college) + u$, we'd most certainly get the result that everyone should drop out of college!
- Any idea of how to solve the endogeneity problem in the Zuckerberg example?

(1) Selection on observables in regression

(2) instrument for dropping out of college:

Simultaneity

In the "Zuckerberg-dropped-out-of-college" example, we have an omitted variable in the error (the *special, unique circumstance of having just invented facebook*) which is related to an explanatory variable, $\mathbf{1}(drop - out - of - college)$

- The endogeneity is between an omitted variable and the variable of interest.
  - These are both on the right hand side
  - These are both explanatory variables

Simultaneity

Simultaneity occurs when the *dependent variable*, the y, the left-hand-side, is determined jointly with one or more right-hand-side variables.

- Of course, it's always the case that the dependent variable y is *determined* by one or more right hand side explanatory variables.
- $y = \beta_0 + \beta_1 x_1 + u$ shows this.
- But *simultaneity* is unique in that x_1 itself is *jointly determined* with y.

An example of a county-level labor supply function

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1$$

- h_s is the hours supplied each week by workers in the county
- w is the wage
- z_1 is anything that affects hours supplied
- u_1 is the error term for hours supplied

This equation stands on its own

- It has a causal interpretation (if α_1 can be estimated without bias)
- It is derived from economic theory (higher wages cause people to substitute out of leisure and into labor)

$\Rightarrow$ **So we call this a structural equation**

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1$$

It suffers from simultaneity because:

- A county's w will be determined, in part, by h_s, the supply.
- Wage is determined jointly by the interaction of h_s, w, and h_d, the hours demanded.
- Thus, **simultaneity**.

The "link" between h_s and h_d is the equilibrium

- $h_s = h_d = h$. Since this happens in every county, we use h_i.
- We only observe this equilibrium, but we might want to know about the values of α_1 and α_2

So we can take our two equations:

$$h_s = \alpha_1 w + \beta_1 z_1 + u_1$$

$$h_d = \alpha_2 w + \beta_2 z_2 + u_2$$

And impose the equilibrium condition: for every i, $h_s = h_d = h_i$

$$h_i = \alpha_1 w_i + \beta_1 z_{1i} + u_{1i}$$

$$h_i = \alpha_2 w_i + \beta_2 z_{2i} + u_{2i}$$

In this simultaneous system of equations: two different z's #1 & #2

$$h_i = \alpha_1 w_i + \beta_1 z_{1i} + u_{1i}$$

$$h_i = \alpha_2 w_i + \beta_2 z_{2i} + u_{2i}$$

h_i and w_i are the endogenous variables. Why?

Because, given $z_{1i}, z_{2i}, u_{1i}, u_{2i}$, then h_i and w_i are completely determined endog.

- with a few assumptions about α_1 and α_2

The dependent variable and one or more explanatory variables are jointly determined within the system.

This happens often in economics

We have many parties interacting with each other, and equilibriums are the outcomes of those interactions.

Think of *marginal analysis* - how we think of a seller setting a price in a market. It's a lot of expectations about interactions.

p & q

w h

Back to the simultaneous system of equations:

$$h_i = \alpha_1 w_i + \beta_1 z_{1i} + u_{1i}$$

"Shock to Supply / Demand" "perturbation"

$$h_i = \alpha_2 w_i + \beta_2 z_{2i} + u_{2i}$$

Note that the z_{1i} and z_{2i} are different variables, while w_i is the same in both equations.

- u_{1i} and u_{2i} are different as well. And uncorrelated with each other.
- We refer to the u_{1i} and u_{21} as the *structural errors*.

Example W 16.1

incpc likely affects polpc, too. So it might not be a good instrument!!

$$murdpc = \alpha_1 polpc + \beta_{10} + \beta_{11} incpc + u_1$$

$$polpc = \alpha_2 murdpc + \beta_{20} + \cancel{other}\ u_2$$

→ thus, we cannot identify α_1 without an instrument.

- $murdpc$ is murders per capita
- $incpc$ is income per capita, which shifts murder rates
- β_{10} is the intercept for equation 1
- $polpc$ is police per capita
- β_{20} is the intercept for equation 2

Is this simultaneous?

- Yes. Just as hours supplied, hours demanded, and wage are jointly determined, $murdpc$ and $polpc$ are jointly determined.
- The city chooses $polpc$ based, in part, on $murdpc$, while murderers choose $murdpc$ based, in part, on $polpc$.
- Even though we're interested in α_1, we need to understand the second equation to avoid bias.

MICHIGAN STATE UNIVERSITY

Simultaneity bias

We can formally show the bias in simultaneous equations. Remember, bias occurs when an explanatory variable is correlated with u (and thus $E[u|x] \neq 0$)

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

$E(u_1 | y_2) \neq 0$

y_1, y_2 could be $murdpc$ and $polpc$.

But estimating this first equation by OLS would result in a biased α_1. So we can't do it.

- Specifically, we are in trouble on the first if y_2 is correlated with u_1;
- And if y_1 is correlated with u_2 for the second.
- Let's see why this is true...

To see bias, substitute the first equation into the second

$$y_2 = \alpha_2\underbrace{(\alpha_1 y_2 + \beta_1 z_1 + u_1)}_{y_1} + \beta_2 z_2 + u_2$$

$$(1-\alpha_2\alpha_1)y_2 = \alpha_2\beta_1 z_1 + \beta_2 z_2 + \underbrace{\alpha_2 u_1 + u_2}_{uh-oh}$$

$(1-\alpha_2\alpha_1)$ $\qquad$ $(1-\alpha_2\alpha_1)$

Divide both sides by $(1-\alpha_2\alpha_1)$:

$$y_2 = \frac{\alpha_2\beta_1}{(1-\alpha_2\alpha_1)} z_1 + \frac{\beta_2}{(1-\alpha_2\alpha_1)} z_2 + \underbrace{\frac{\alpha_2}{(1-\alpha_2\alpha_1)} u_1 + \frac{1}{(1-\alpha_2\alpha_1)} u_2}_{v_2}$$

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2$$

This is called the *reduced form* equation.

- We *can* estimate π_{21} and π_{22}, but the coefficients lose their structural interpretation.
- Our estimation of π_{21} and π_{22} is unbiased - z's are exogenous, and u_1 itself is uncorrelated wtih y_2.

unbiased: π_{21}, π_{22}

Knowing π_{21}, π_{22} does not get us α_1, α_2!

MICHIGAN STATE UNIVERSITY

Back to checking to see if we have bias in:

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

Remember, we are in trouble if y_2 is correlated with u_1, right?

Well, we know:

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2$$

and v_2, the reduced form error:

$$v_2 = \frac{\alpha_2}{(1 - \alpha_2 \alpha_1)} u_1 + \frac{1}{(1 - \alpha_2 \alpha_1)} u_2$$

Since v_2 has u_1 in it, and since y_2 has v_2 in it, then y_2 **is correlated with** u_1, and OLS of $y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$ is biased.

This is simultanaeity bias

When can we identify α_1 and α_2?

- Our problem here is endogeneity, so we need an instrument.
- Something that shifts y_2 but is not correlated with u_1 (exclusion restriction)
- Do we have something in y_2?

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

Yes. We have z_2, which is exogenous by definition.

It can shift y_2, and is not correlated with u_1. It does not shift y_1 except through y_2 because it is not in the equation for y_1.

$$Y_1 = B_0 + B_1 Y_2 + u$$

$$Z_2$$

Similarly, we can use z_1 to shift y_1.

And both equations can be identified because we have *one exogenous shifter for each endogenous variable in each equation.*

The Rank Condition

In a two-equation system, we can only identify an equation with an endogenous variable if the *other* equation has one or more exogenous variable that does not enter the first equation.

- The instrument must have a non-zero population coefficient

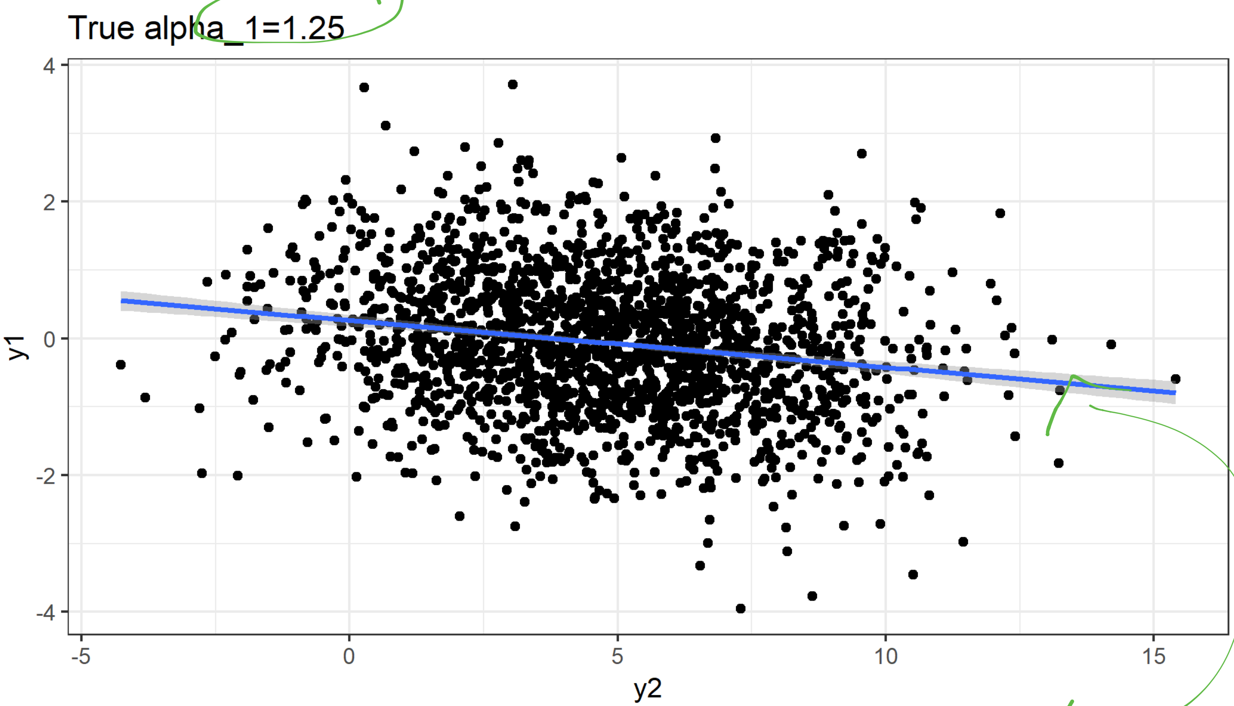
1 instrument per equation
with endogen. problem,
per endog. problem.

Our two equation system, again (with the problems in red and blue:

$$y_1 = \alpha_1 y_2 + \beta_1 z_1 + u_1$$

$$y_2 = \alpha_2 y_1 + \beta_2 z_2 + u_2$$

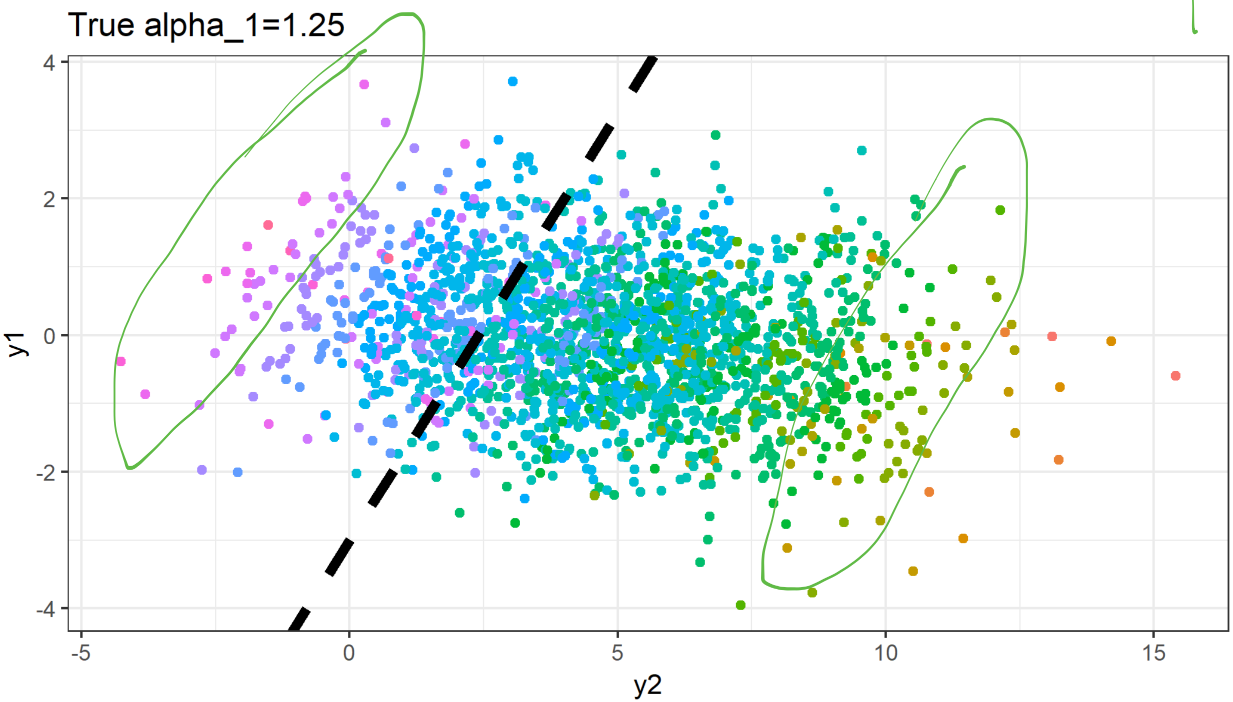
A visual example:



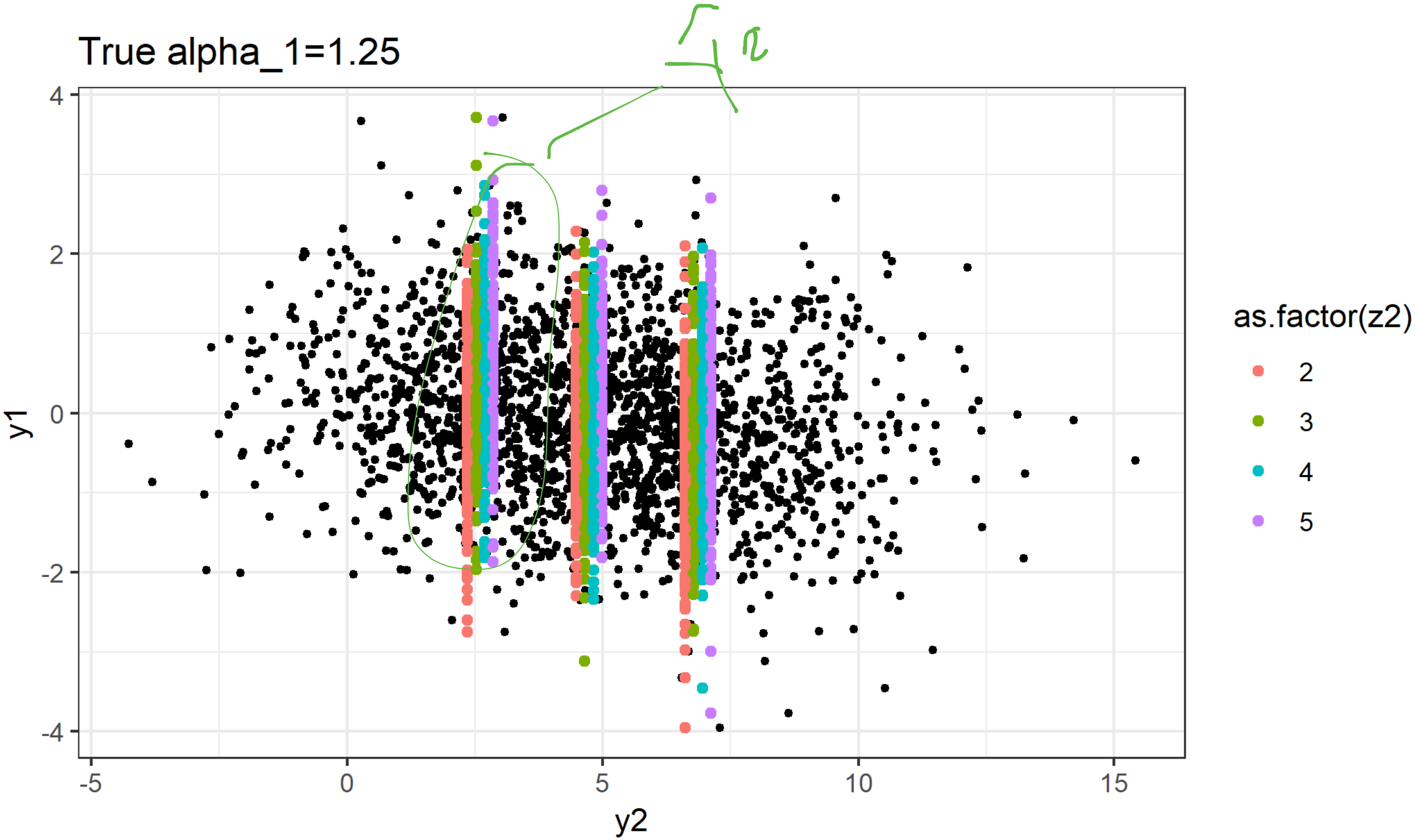
"Naive regression"

Since the problem is that y_2 is correlated with u_1, what if we observed u_1?

$\uparrow y_2$



Of course, we can't control for u_1 since it is unobserved.



The colored groupings are $\hat{y}_2$. Each grouping is a different z_1. As z_2 increases, y_2 increases within each grouping. As y_2 increases, in each grouping, y_1 *increases*.

First stage

$$y_2 = \pi_{21} z_1 + \pi_{22} z_2 + v_2$$

```
Call:
lm(formula = y2 ~ z1 + z2, data = df1)

Residuals:
    Min      1Q  Median      3Q     Max 
-6.8495 -1.4853 -0.0776  1.5238  8.6340 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.10099    0.20726  -0.487 0.626124    
z1 π21       2.12757    0.06091  34.932  < 2e-16 ***
z2 π22       0.16575    0.04484   3.696 0.000225 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.234 on 1997 degrees of freedom
Multiple R-squared:  0.3813,    Adjusted R-squared:  0.3807 
F-statistic: 615.4 on 2 and 1997 DF,  p-value: < 2.2e-16
```

Second stage

$$y_1 = \alpha_2 \hat{y}_2 + \beta_1 z_1 + u$$

No z_2 b/c z_2 is instrument for y_2.

```
Call:
lm(formula = y1 ~ y2hat + z1, data = df1)

Residuals:
    Min      1Q  Median      3Q     Max 
-3.6510 -0.6258 -0.0321  0.6163  3.5293 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  0.06252    0.07870   0.794    0.427    
y2hat        1.16183    0.11385  10.205   <2e-16 ***
z1          -2.81419    0.24323 -11.570   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9402 on 1997 degrees of freedom
Multiple R-squared:  0.1254,    Adjusted R-squared:  0.1245 
F-statistic: 143.1 on 2 and 1997 DF,  p-value: < 2.2e-16
```

The second-stage coefficient

We get a pretty accurate estimate for $\alpha_1 = 1.16$ from the second-stage having used z_2 to instrument for y_2.

Do same: z_1 instrument for y_1 $\longrightarrow$ $y_2 = \alpha_2 \hat{y}_1 + z_2 + u_2$

In a panel data setting we'd have a *fixed effect* for each i:

$$y_{it1} = \alpha_1 y_{it2} + \mathbf{z_{it1}}\boldsymbol{\beta_1} + a_{i1} + u_{it1}$$

$$y_{it2} = \alpha_2 y_{it1} + \mathbf{z_{it2}}\boldsymbol{\beta_2} + a_{i2} + u_{it2}$$

$\boldsymbol{a_{i1}}$ is unobserved and potentially correlated with z_{it1}. This presents interesting problems unique to panels.

MICHIGAN STATE UNIVERSITY

First Differencing

One way of handling an unobserved fixed effect in panel data (different from what we've learned on fixed effects) is *first differencing*.

$$y_{it1} - y_{i(t-1)1} = \alpha_1(y_{it2} - y_{i(t-1)2}) + \beta_1(z_{it1} - z_{i(t-1)1}) + a_{i1} - a_{i1} + u_{it1} - u_{i(t-$$

Which can be written using the Δ notation:

$$\Delta y_{it1} = \alpha_1 \Delta y_{it2} + \beta_1 \Delta z_{it1} + \Delta u_{it1}$$

This removes the a_{i1}, and makes it clear that we need an instrument whose *change* is

- Exogenous
- Affects only Δy_{it2} without affecting Δy_{it1} (uncorrelated with Δu_{it1}).
- And *it has to vary within each* i

Using fixed effects

A similar result happens if we include the fixed effect. The fixed effect instruments for itself, and is included as an exogenous variable.

First stage

$$y_{it2} = \pi_{21} z_1 + \pi_{22} z_2 + \gamma_i^1 + v_2$$

Second stage

$$y_{it1} = \alpha_1 \hat{y}_{it2} + \beta_1 z_1 + \gamma_i^2 + u_1$$

γ_i^1 is the fixed effect for each i in the first stage.

γ_i^2 is the fixed effect for each i in the second stage (not a squared term).

Next week

MICHIGAN STATE UNIVERSITY

Difference in Differences

Read MM Ch. 5