Violations of the CLNRM

Monday, April 02, 2012
5:09 AM

- Last week, we dealt with three common violations of the CLNRM assumptions:
    - Heteroskedasticity – non-constant error variance
    - Omitted variable bias
    - Multicollinearity – correlation among regressors
- This week, we continue with two more violations
    - Measurement error ✓
    - Endogeneity ✓
- Sadly, these violations are more consequential and harder to fix than the last three that we examined

Measurement error

Monday, April 02, 2012
5:13 AM

What is measurement error?

true DGP: $y^0 = x^0\beta + u^0$

run model: $\underline{y} = \underline{x\beta} + u^0$

$y = \underline{y^0 + v}$ $\qquad$ v, w = measurement error terms

$x = \underline{x^0 + w}$

Substitute & Solve: $y^0 + v = (x^0 + w)\beta + u^0$

$$y^0 = x^0\beta + \underbrace{w\beta + u^0 - v}_{\hat{u}}$$

Measurement error in the OLS context

① inefficiency

$$u = u^0 + w\beta - v$$

$$var(\hat{u}) = \frac{1}{n-k}\hat{u}'\hat{u} > var(\hat{u}^0)$$

SEs of $\hat{\beta}$ are inflated.

Harder to reject the null

② bias in $\hat{\beta}$

CLRM: $E[u|X] = 0$

$$E[u|X] = E[u^0 + w\beta - v \mid x^0 + w]$$

$$E[u|X] = E[u^0 + w\beta - v \mid X^0 + w]$$

$$= \cancel{E[u^0 | X^0 + w]} + E[w\beta | X^0 + w] - \cancel{E[v | X^0 + w]}$$

$$E[u|X] = w\beta \neq 0$$

measurement error in X results in biased $\hat{\beta}$

measurement error in Y results in inefficient $\hat{\beta}$

Remedial measures: multiple measurements

Monday, April 02, 2012
5:21 AM

Could just say "make fewer errors..."

...but if this isn't a possibility, try collecting multiple measurements.

Consider a simple mean of M many measurements of the same concept (on same scale)

- According to the CLT, $\underset{M\to\infty}{\text{plim}}\ \text{var}\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)=0$

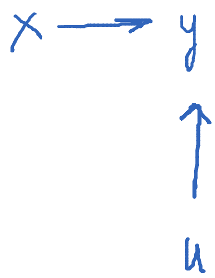
$$\underset{M\to\infty}{\text{plim}}\left(\frac{1}{M}\sum_{i=1}^{M} x_i\right)=\mu_i=x_i^{o}$$

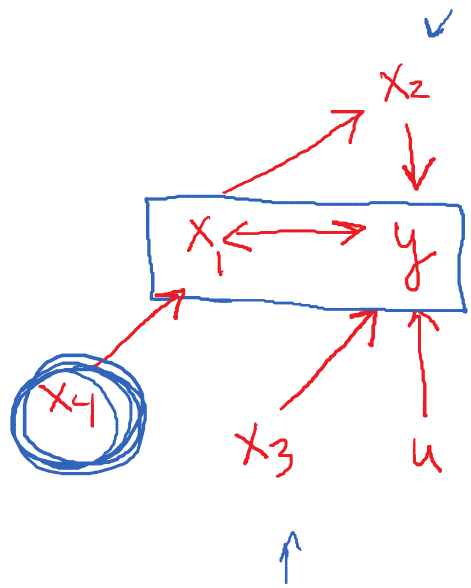
factor analysis

Endogeneity

Monday, April 02, 2012
5:14 AM

What is it?





do we care?

Endogeneity consequences

Monday, April 02, 2012
5:15 AM

Bias : estimates of the relationship between X_1 and y will be biased in the presence of reverse causality between y and X_1.

$$y = x\beta + u_0$$

$$y \leftrightarrow x$$
$$\uparrow$$
$$z$$

$$y = (\alpha_0 + \alpha_1 y + \alpha_2 z + v)\beta + u_0$$

$$y = \alpha_0\beta + \alpha_1\beta y + \alpha_2 z\beta + \underbrace{v\beta + u_0}_{u}$$

CLRM: $E[u|x] = 0$

$$E[u|x] = E[u_0 + v\beta \mid \alpha_0 + \alpha_1 y + \alpha_2 z + v]$$
$$= v\beta \neq 0$$

$\rightarrow$ coefficients β will be biased.

Remedial measures:

Monday, April 02, 2012
5:15 AM

Lag/lead models: A quick and (very) dirty fix

$$y_t = x_t + u$$

$$y_t = x_{t-1} + u$$

x_{t-1}

$x_t \to y_t$

$y_t \to x_t$

won't work unless

$y_t \not\Rightarrow x_{t-1}$.

fails in the presence of strategic anticipation

Two stage least squares (aka instrumental variables regression): Why does it work?

2SLS

instrumental variable

problem: X and u are correlated

we must rid ourselves of the portion of X that is correlated with u.

Procedure for executing 2SLS

Step1: predict X using a variable(s) that are correlated with X but NOT with u (aka not with y).

→ instrumental variable

$$\hat{X} = \hat{\gamma}_0 + \hat{\gamma}_1 Z + \hat{v}$$ || omitted variable: y

Z : instrumental variable

Step two: use $\hat{X}$ in place of X in the model of y.

$$y = X\beta + u_0$$

$$y = (\hat{X} + \hat{v})\beta + u_0 = \hat{X}\beta + \hat{v}\beta + u_0$$

$$y = \hat{X}\beta + u \qquad u \rightarrow \hat{v}\beta + u_0$$

$\hat{X}$ and u are uncorrelated.

in any regression, predicted DV $P_X y$ and predicted residuals $M_X y$ are by definition uncorrelated.

More generally...

$$y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 D + u$$

$$X = \alpha_0 + \alpha_1 y + \alpha_2 Z + \alpha_3 F + v$$

D and F are instruments for y and X, respectively.

Z is a set of exogenous variables that that influence X and y.

When instrumenting X, use <u>all</u> exogenous variables, Z and D and F:

$$\hat{X} = \hat{\gamma}_0 + \hat{\gamma}_1 Z + \hat{\gamma}_2 D + \hat{\gamma}_3 F + \hat{v}$$

Practical Difficulties with 2SLS

Monday, April 02, 2012
5:16 AM

Finding an instrument

correlated w/ X not w/ Y.

No way to test the above.

Z ~ lm(X, Y)

Y ~ lm(Z, X)

Figuring out whether an instrument is good

The consequences of a "weak" instrument

"weak": correlation between Z and X
is small.

Identification concerns

- Unidentified model – fewer instruments than endogenous variable
- Just identified model – exactly one instrument per e.v.
- Overidentified model – >1 instrument per e.v.

Fitting and overfitting

R^2 of first stage is high, then 2SLS estimates will be close to OLS estimates because $\hat{X}$ will be close X.

good if v small.
bad if $\hat{x}$ is overfitted.

2SLS is *consistent*, not unbiased