Econ 312

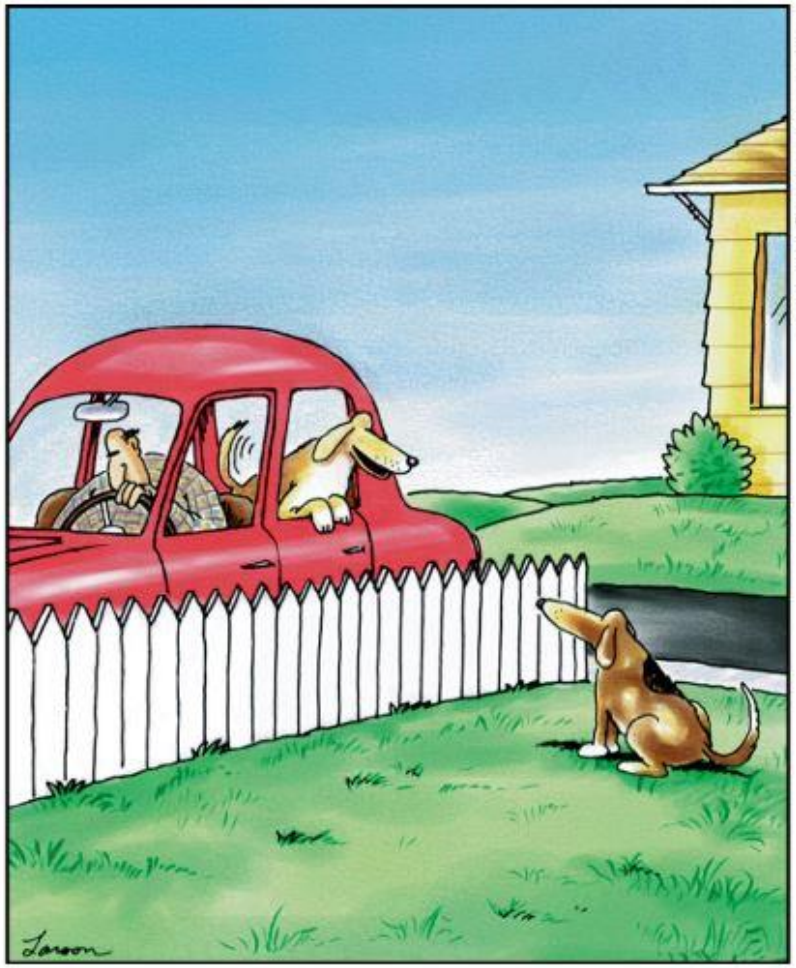
Wednesday, April 15

Implementing Instrumental-Variables Estimation

Readings: Wooldridge, Chapter 15

Class notes: 136 - 141

Today's Far Side offering



"Ha ha ha, Biff. Guess what? After we go to the drugstore and the post office, I'm going to the vet's to get tutored."

Poor pooch!

Context and overview

- This class focuses on how we actually do **instrumental-variables** estimation
- We introduce briefly the **generalized method of moments** (GMM) estimator
- We discuss three tests relevant to IV estimation
  - Test of **instrument strength**
  - Test of **instrument/model validity** (only for overidentified models)
  - Hausman test of **endogeneity of regressors**
- We then review a few famous examples of IV regressions and work through an example from the Hill-Griffiths-Lim text (that is the Daily Problem)

Generalized method of moments (GMM)

- Suppose that model is overidentified (2 instruments z_1 and z_2, and 1 endogenous regressor x)
- There are three valid moment conditions but only two coefficients to estimate:

$$\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \hat{m}_1 = 0$$

$$\sum_{i=1}^{n} z_{i,1}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \hat{m}_2 = 0$$

$$\sum_{i=1}^{n} z_{i,2}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = \hat{m}_3 = 0$$

- All three *should* hold for the true set of coefficients, but they won't hold exactly for any estimates of the coefficients in data

Uses of GMM

1. The degree to which one set of two parameters satisfies *all three* equations is a **measure/test of the validity** of the model and instruments
2. We could estimate the model by **minimizing some function** of these equations such as $c_1\hat{m}_1^2 + c_2\hat{m}_2^2 + c_3\hat{m}_3^2$ where the c coefficients are chosen weights
   - This is the GMM class of estimators; different estimators depending on choice of weights
   - The 2SLS estimator is a GMM estimator with a particular choice of weights

Instrument strength

- Is z strongly enough correlated with x to be a strong instrument?
- Recall that weak instruments lead to large variance in coefficient estimator
- If x_k is the endogenous regressor, do the regression

$$x_k = \gamma_0 + \gamma_1 x_1 + \ldots + \gamma_{k-1} x_{k-1} + \theta_1 z_1 + v_k$$

- Calculate F statistic for null that $\theta_1 = 0$
  - Can be more than one instrument and more than one θ coefficient
- Compare F statistic to 10 rather than usual critical value
  - $F > 10$ means strong instruments
- Appropriate threshold is different for more complicated cases

Tests of model validity

- Is z correlated with u?
  - Can't just check residuals because no consistent residuals without valid instruments and no test of valid instruments without consistent residuals
- If we have extra instruments, we can test overidentifying restrictions
  1. Get IV residuals
  2. Regress residuals on all z instruments and exogenous regressors
  3. LM test: nR^2 from this regression $\sim \chi^2$ with $L - B$ df
- Intuition: If model is correct, instruments should be correlated with y ONLY through their effect on endogenous x
  - 2SLS residuals are (by construction) the part of y that is orthogonal to the part of z that works through x
  - Residuals should be uncorrelated with z if that assumption is correct and instruments are valid
- If they are correlated (and we reject LM test), then something in model is invalid

Tests of endogeneity

- Do 2SLS and OLS give same values? If so, endogeneity is not a problem
- Application of Hausman test with H_0: $\text{cov}(x, u) = 0$
- Simple procedure:
  1. Add residuals from first-stage regression to OLS of original equation
     - These residuals are the part of x that is not correlated with z and thus includes the endogenous part of x (if there is one)
  2. Use standard t test to see if the residual is significant
     - If so, then endogenous part of x affects y and we need 2SLS to take care of endogeneity
     - If not, then OLS should be fine because only the exogenous part of x affects y

A couple of famous applications

- Angrist
  - Did military service during Vietnam War affect future earnings?
  - Military service might be affected by omitted variables that affect earnings
  - Instrument: Random number for birth date in draft lottery
  - Why it works: Lottery number was random (exogenous) and men with low lottery numbers were much more likely to be drafted into military
- Levitt
  - Do more police reduce crime?
  - Cities might increase police if crime goes up
  - Instrument: Elections
  - Why it works: If cities hire more police before (exogenous) elections, then this is exogenous shock to police staffing

Daily Problem #34

- I'm assuming that you have worked on the daily problem and can refer to the equations and table as we go

Equation: $RENT_i = \beta_0 + \beta_1 MDHOUSE_i + \beta_2 PCTURBAN_i + u_i$

First stage: $MDHOUSE_i = \gamma_0 + \gamma_1 PCTURBAN_i + \gamma_2 FAMINC_i + \gamma_3 REG2_i + \gamma_4 REG3_i + \gamma_5 REG4_i + v_i$

- We are mostly interested in β_1, the effect of house price on rent
  - Endogenous: Higher rent in state rental market should increase house price
  - Proposed instruments: State family income and regional dummies

Instrument strength

- Compare SSR in unrestricted model (2) to restricted model (3) to see how much instruments help predict the endogenous regressor

$$F = \frac{(SSR_R - SSR_U)/4}{SSR_U/44} = 13.3$$

- Threshold for strong instruments is $F > 10$, so our instruments are strong

Hausman test for endogeneity

- Add $\hat{v}$ OLS regression of structural model (1): Column (4)
- *t* statistic for $\hat{v}$ is –1.589 / 0.398 = –3.99
- This is less than the lower critical value for *t* distribution (–2.01) so we reject the null and conclude the MDHOUSE is endogenous
- Remember that these residuals measure the "not surely exogenous" part of MDHOUSE
  - If that part affects RENT, then MDHOUSE is endogenous and we need 2SLS rather than OLS

Test of overidentifying restrictions

- Model is overidentified because $L = 4$ and $B = 1$
  - Three "extra" instruments
  - Testable overidentifying restrictions
- Regress 2SLS residuals on exogenous variables and instruments (Column 6)
- LM test: $nR^2 = 50 \times 0.226 = 11.3 \sim \chi^2(3)$
  - 0.05 critical value of $\chi^2(3)$ is 7.815
  - Test statistic > critical value, so reject null hypothesis
- Something is amiss with the model
  - Invalid instruments?
  - Wrong specification?

Review and summary

- We discussed briefly generalized method of moments, a powerful class of estimators that includes 2SLS as special case
- We have examined issues and tests involving the implementation of two-stage least squares
  - Tests of instrument strength
  - Tests of endogeneity of x
  - Tests of model/instrument validity
- We looked at a couple of ways that IV estimators have been used in famous papers
- We worked a detailed example from the daily problem

Bad economist joke of the day

An economist was crossing a road one day when a frog called out to him and said, "If you kiss me, I'll turn into a beautiful princess." He bent over, picked up the frog, and put it in his pocket.

The frog spoke up again and said, "If you kiss me and turn me back into a beautiful princess, I will stay with you for one week." The economist too the frog out of his pocket, smiled at it, and returned it to his pocket.

The frog then cried out, "If you kiss me and turn me back into a princess, I'll stay with you for a week and do ***anything*** you want." Again the economist took the frog out, smiled at it and put it back into his pocket.

Finally, the frog asked, "What's the matter with you? I've told you I'm a beautiful princess, and that I'll stay with you for a week, and do anything you want. Why won't you kiss me?"

The man said, "Look, I'm an economist. I don't have time for a girlfriend … But a talking frog? Now that's cool!"

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

What's next?

- Our next class looks in detail at identification in simultaneous equation models
- Daily Problem #35 is one of 5 cases of a simple supply-demand model that we will consider
- Depending on the specification of the model it may be possible to identify the structural parameters of the supply equation, of the demand equation, or both (or neither)