# Econ 312

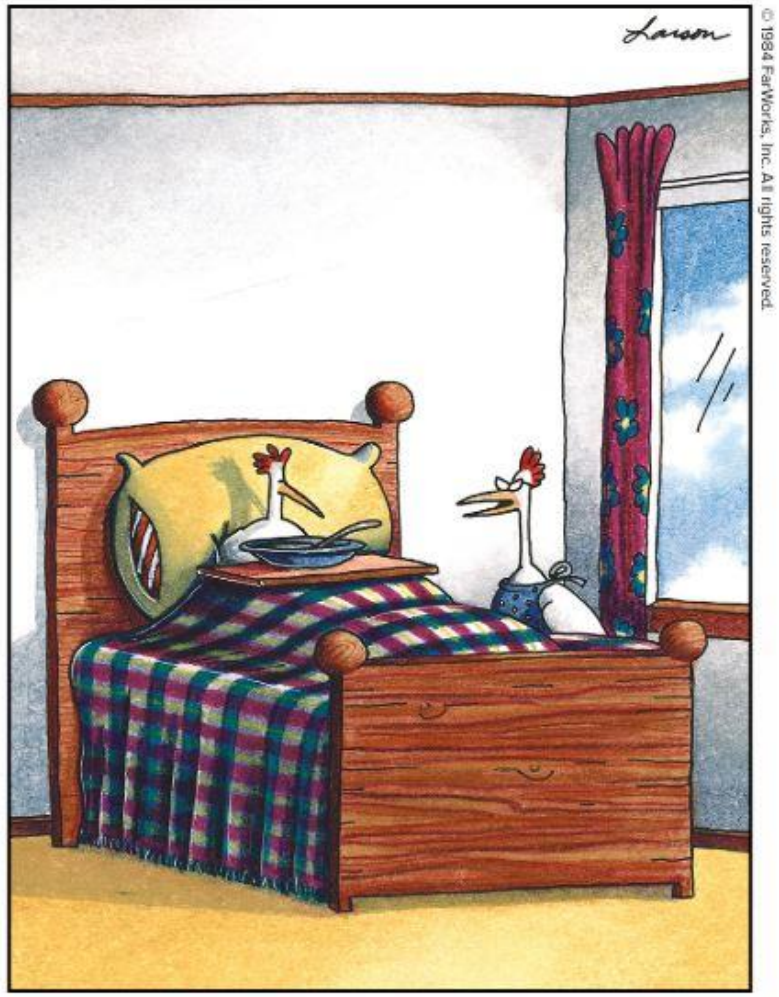
## Monday, April 13

## Method of Moments and Instrumental Variables

Reading: Wooldridge, Chapter 15

Class notes: 133 - 136

# Today's Far Side offering




"Quit complaining and eat it! ... Number one, chicken soup is good for the flu, and number two, it's nobody we know."

A good reason not to eat chickens!

# Context and overview

- OLS is biased and inconsistent if the **error is correlated with a regressor**
- This happens with measurement error in $x$, with omitted regressors, with simultaneous equations, and in other cases
- The **instrumental-variables** (IV) estimator can be derived using the method of moments
- IV is consistent as long as suitable instruments can be found
- The most common application of IV is the **two-stage least-squares estimator**

# Bias in and inconsistency of OLS

- If $u$ is correlated with one of the $x$ variables, then that variable will pick up the effects of the error in addition to its own effects
- Omitted-variable bias leads to exactly this problem
  - A problem when omitted regressor is correlated with included regressor
  - Omitted variable is part of the error term and correlated with included $x$
  - Coefficient of included $x$ picks up the effects of both variables instead of just the included one
- Bias depends on the sign of corr ($x$, $u$)

# Measurement error example

- Random measurement error in $x$: $\tilde{x}_i = x_i + \eta_i$
- True model is $y_i = \beta_0 + \beta_1 x_i + u_i$
- We estimate $y_i = \beta_0 + \beta_1 \tilde{x}_i + (u_i - \beta_1 \eta_i)$
- $\eta_i$ is part of $\tilde{x}_i$ so they are correlated and composite error term is correlated with our regressor
- Assuming $u$ and $\eta$ are independent and normal, $\hat{\beta}_1$ is biased toward zero

$$\operatorname{plim} \hat{\beta}_1 = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_\eta^2} \beta_1$$

- More bias if measurement error is larger part of variance of $\tilde{x}$

# Simultaneous-equations bias

- $x$ and $y$ are jointly determined as part of a system of equations:

$$y = \beta_0 + \beta_1 x + \ldots + u$$

$$x = \gamma_0 + \gamma_1 y + \ldots + v$$

- Similar to problem we analyzed in VAR models
  - Both variables are affected by the other
  - OLS is inconsistent in both equations because of correlation with error
- We will study a very detailed example of this in a supply-demand model when we talk about identification

# Instruments

- Suppose that $y = \beta_0 + \beta_1 x + u,$ but that $x$ is correlated with $u$
  - OLS is biased and inconsistent
- We seek a variable $z$ with the following properties:
  - $z$ does not affect $y$ directly
  - $z$ is exogenous (not correlated with $u$)
  - $z$ is strongly correlated with $x$
- These three conditions make $z$ a suitable instrument for $x$
- We can use the assumption that $\text{cov}(z, u) = 0$ in place of the usual $\text{cov}(x, u)$ in deriving an estimator

# Applying method of moments

- Sample moment conditions:

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

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

- Solving yields something similar to OLS normal equation for slope:

$$\hat{\beta}_1^{IV} = \frac{\sum_{i=1}^{n} (z_i - \bar{z})(y_i - \bar{y})}{\sum_{i=1}^{n} (z_i - \bar{z})(x_i - \bar{x})} \qquad \hat{\beta}_1^{OLS} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})(x_i - \bar{x})}.$$

# Properties of IV estimator

- Consistent as long as $z$ is exogenous
- Asymptotically normal (Stata will give $z$ statistic, not $t$)

$$\hat{\beta}_1^{IV} \sim N\left(\beta_1, \frac{\sigma^2}{r_{xz}^2 \sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}\right), \; r_{xz} \equiv \text{corr}\left(x, z\right)$$

- Estimate $\sigma_{IV}^2 = \dfrac{\sum_{i=1}^{n}\left(y_i - \hat{\beta}_0^{IV} - \hat{\beta}_1^{IV} x_i\right)^2}{n-2}$

- **Weak instruments**: $r_{xz} \sim 0$ implies large variance

# Two-stage least squares

- What if we have $L > 1$ valid instruments? Which one to use?
- Any linear combination of $z_1, z_2, \ldots, z_L$ is a valid instrument
- **Two-stage least squares** chooses the linear combination that is most highly correlated with $x$
- Run **first-stage** regression $x = \gamma_0 + \theta_1 z_1 + \ldots + \theta_L z_L + v$
- Use fitted values $\hat{x}$ as the instrument in the **second stage**:

$$y = \beta_0 + \beta_1 \hat{x} + u^*$$

- **ivregress** in Stata: Don't do as two separate regressions

# Multiple endogenous regressors

- $G$ "good" regressors, $B = k - G$ "bad" regressors, $L$ "lucky" instruments
  - $L > B$ → overidentified
  - $L = B$ → just identified
  - $L < B$ → not identified (and cannot be estimated by IV)
- Original model: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_G x_G + \beta_{G+1} x_{G+1} + \ldots + \beta_k x_k + u$
- First stage: $x_{G+j} = \gamma_{0j} + \gamma_{1j} x_1 + \ldots + \gamma_{Gj} x_G + \theta_{1j} z_1 + \ldots + \theta_{Lj} z_L + v_j,\ j = 1, \ldots, B$
- Fitted values: $\hat{x}_{G+j} = \hat{\gamma}_{0j} + \hat{\gamma}_{1j} x_1 + \ldots + \hat{\gamma}_{Gj} x_G + \hat{\theta}_{1j} z_1 + \ldots + \hat{\theta}_{Lj} z_L,\ j = 1, \ldots, B$
- Second stage: $y = \beta_0 + \beta_1 x_1 + \ldots + \beta_G x_G + \beta_{G+1} \hat{x}_{G+1} + \ldots + \beta_K \hat{x}_K + u^*$

# 2SLS in Stata

- To implement in Stata:

  ivregress 2sls *depvar exvars* (*endvars* = *instvars*) , *options*

- *depvar* is dependent variable
- *exvars* are the "good" exogenous regressors
- *endvars* are the "bad" endogenous regressors
- *instvars* are the "lucky" instruments

# Review and summary

- OLS is **biased** and **inconsistent** when $x$ is correlated with $u$
- If we can find enough valid and strong **instruments** we can use **instrumental variables**
- **Two-stage least squares** is the most common implementation of instrumental variables
  - Estimate **first-stage** regression for each endogenous regressor as function of exogenous regressors and instruments
  - Replace each endogenous regressor with its first-stage fitted values in **second-stage** regression
- Stata's **ivregress** command will do both stages and calculate statistics correctly

# Bad economist joke of the day

A woman hears from her doctor that she only has half a year to live. The doctor advises her to marry an economist.

The woman asks, “Will this cure my illness?”

“No,” the doctor answers, “but the six months will seem like a lifetime!”

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

# What's next?

- The next class will examine details of using instrumental variables
  - Tests for instrument strength
  - Tests for instrument validity
  - Tests for endogeneity of regressors

- We will then talk about some famous examples from the economics literature, and do a sample application together