ECON207 Session 8

Instrumental Variable and GMM

Anthony Tay

This Version: 22 Oct 2024

Agenda

Session 8

- Recall endogeneity problem
- Introduction to IV and GMM Estimation

Code for today's class will use the following packages:

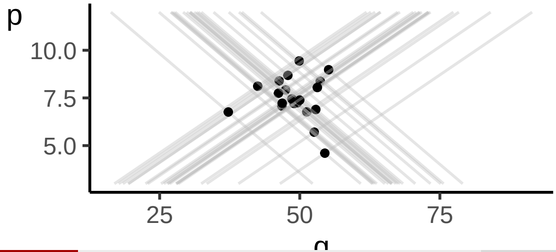
```
library(tidyverse); library(readxl)
library(lmtest); library(sandwich); library(car)
library(ivreg); library(gmm)
```

Recall Endogeneity Problem:

Suppose we want to estimate demand function for a certain good, and suppose

$$Q^d = \delta_0 + \delta_1 P + \epsilon^d \qquad \text{(Demand Eq } \delta_1 < 0\text{)}$$
$$Q^s = \alpha_0 + \alpha_1 P + \epsilon^s \qquad \text{(Supply Eq } \alpha_1 > 0\text{)}$$
$$Q^s = Q^d \qquad \text{(Market Clearing)}$$

Observed quantities and prices occur at intersection of demand and supply eqs



Recall Endogeneity Problem:

- Observed prices and quantities reflect neither demand nor supply eqs
- Regression of Q on P produces an estimate that is some average of α_1 and δ_1
- Issue is that both prices and quantity are simultaneously determined
- Price, in particular, is not exogenous
  - Variation in the data comes from demand and supply shocks
  - Demand/supply shocks shift demand/supply functions
  - Shifting dd/ss function changes both quantity *and* price
  - The regressor (price) is correlated with the regression noise term.

Recall Endogeneity Problem:

Demonstration of inconsistency of OLS estimator (from Session 2)

Solving for quantity and prices gives

$$P = \frac{\alpha_0 - \delta_0}{\delta_1 - \alpha_1} + \frac{\epsilon^s - \epsilon^d}{\delta_1 - \alpha_1}$$

$$Q = \left(\delta_0 + \delta_1 \frac{\alpha_0 - \delta_0}{\delta_1 - \alpha_1}\right) + \frac{\delta_1 \epsilon^s - \alpha_1 \epsilon^d}{\delta_1 - \alpha_1}$$

which implies

$$Var(P) = \frac{\sigma_s^2 + \sigma_d^2}{(\delta_1 - \alpha_1)^2} \quad \text{and} \quad Cov(P,Q) = \frac{\delta_1 \sigma_s^2 + \alpha_1 \sigma_d^2}{(\delta_1 - \alpha_1)^2}$$

Recall Endogeneity Problem:

The OLS estimator of β_1 in the regression $Q_i = \beta_0 + \beta_1 P_i + \epsilon_i$:

$$\hat{\beta}_1^{ols} = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(P_i - \overline{P})}{\sum_{i=1}^{n}(P_i - \overline{P})^2} \xrightarrow{p} \frac{\text{cov}(Q,P)}{\text{var}(P)} = \frac{\delta_1\sigma_s^2 + \alpha_1\sigma_d^2}{\sigma_s^2 + \sigma_d^2}$$

Recall Endogeneity Problem:

Now **suppose** there is some observable variable r_t that shifts the supply function but not the demand function. That is

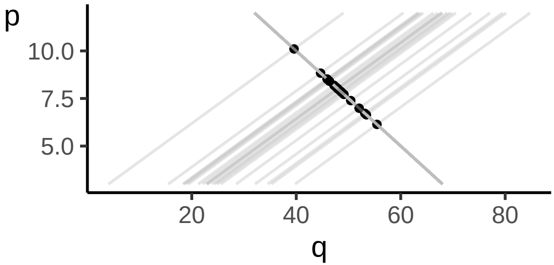
$$Q^d = \delta_0 + \delta_1 P + \epsilon^d \qquad \text{(Demand Eq } \delta_1 < 0\text{)}$$
$$Q^s = \alpha_0 + \alpha_1 P + \alpha_2 r + \epsilon^s \qquad \text{(Supply Eq } \alpha_1 > 0\text{)}$$
$$Q^s = Q^d \qquad \text{(Market Clearing)}$$

where $\alpha_2 \neq 0$ and r is uncorrelated with the demand shocks, i.e.,

Instumental Variables (IV) Intro

Imagine that we can "shut down" the demand and supply shocks, but allow r to change

- Only the supply curve changes as r changes
- Intersections of dd and ss will "map" out the demand function



IV Intro

Lesson:

- We can use variation in r to map out demand function
- In practice we *cannot* "shut down" demand and supply shock (or any other factors driving demand and supply)
- Big question: *how do we isolate variation in P due to r only*?

Answer:

- Regress P_i on r_i, and then regress Q_i on *fitted* P_i, i.e., $\widehat{P}_i$
- $\widehat{P}_i$ contains variation in Y_i due to r_i only
- Use r_i as "instrument" to identify demand function

IV Intro

- Step 1: regress the endogenous regressor P_i onto the exogenous instrument r_i:

$$P_i = \phi_0 + \phi_1 r_i + u_i$$

and collect the fitted values

$$\hat{P}_i = \hat{\phi}_0 + \hat{\phi}_1 r_i$$

where $\hat{\phi}_1 = \dfrac{\sum_{i=1}^{n}(P_i - \overline{P})(r_i - \overline{r})}{\sum_{i=1}^{n}(r_i - \overline{r})^2}$ and $\hat{\phi}_0 = \overline{P} - \hat{\phi}_1 \overline{r}$

IV Intro

- Step 2: regress Q_i on $\hat{P}_i$. We have

$$\hat{\delta}_1^{iv/2sls} = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(\hat{P}_i - \overline{\hat{P}})}{\sum_{i=1}^{n}(\hat{P}_i - \overline{\hat{P}})^2}$$

where "iv" ~ "instrumental variable" and $2sls$" ~ "2-stage least squares"

Since

$$\hat{P}_i = \hat{\phi}_0 + \hat{\phi}_1 r_i \Rightarrow \hat{P}_i - \overline{\hat{P}} = \hat{\phi}_1(r_i - \overline{r})$$

$$\Rightarrow \sum_{i=1}^{n}(\hat{P}_i - \overline{\hat{P}})^2 = \hat{\phi}_1^2 \sum_{i=1}^{n}(r_i - \overline{r})^2 = \hat{\phi}_1 \sum_{i=1}^{n}(P_i - \overline{P})(r_i - \overline{r})$$

IV Intro

Last equality holds because $\hat{\phi}_1 = \dfrac{\sum_{i=1}^{n}(P_i - \overline{P})(r_i - \overline{r})}{\sum_{i=1}^{n}(r_i - \overline{r})^2}$

Therefore

$$\begin{aligned}\hat{\delta}_1^{iv/2sls} &= \frac{\hat{\phi}_1 \sum_{i=1}^{n}(Q_i - \overline{Q})(r_i - \overline{r})}{\sum_{i=1}^{n}(\hat{P}_i - \overline{\hat{P}})^2} \\ &= \frac{\hat{\phi}_1 \sum_{i=1}^{n}(Q_i - \overline{Q})(r_i - \overline{r})}{\hat{\phi}_1 \sum_{i=1}^{n}(P_i - \overline{P})(r_i - \overline{r})} = \frac{\sum_{i=1}^{n}(Q_i - \overline{Q})(r_i - \overline{r})}{\sum_{i=1}^{n}(P_i - \overline{P})(r_i - \overline{r})}\end{aligned}$$

IV Intro

This estimator is consistent for δ_1:

$$\hat{\delta}_1^{iv/2sls} = \frac{\sum_{i=1}^n (Q_i - \overline{Q})(r_i - \overline{r})}{\sum_{i=1}^n (P_i - \overline{P})(r_i - \overline{r})} = \frac{\sum_{i=1}^n (r_i - \overline{r}) Q_i}{\sum_{i=1}^n (r_i - \overline{r}) P_i} = \frac{\sum_{i=1}^n (r_i - \overline{r})(\delta_0 + \delta_1 P_i + \epsilon_i^d)}{\sum_{i=1}^n (r_i - \overline{r}) P_i}$$

$$= \delta_1 + \frac{\sum_{i=1}^n (r_i - \overline{r}) \epsilon_i^d}{\sum_{i=1}^n (r_i - \overline{r}) P_i} \xrightarrow{p} \delta_1 + \frac{Cov(r, \epsilon^d)}{Cov(r, P)}$$

Since

- r is uncorrelated with the demand shock
- r is correlated with price (r shifts the supply function, which changes price), therefore $\hat{\delta}_1^{2sls} \xrightarrow{p} \delta_1$

Method of Moments Perspective

We show another way of looking at this problem

Produces same result in simple dd-ss example, but can be different (depending on implementation) when we extend to "GMM" in more complicated examples

We switch notations (to transition to general case) and write the problem as:

- Estimate
$$Y = \beta_0 + \beta_1 X + \epsilon$$
where $E(\epsilon \mid X) \neq 0$ but is correlated with X
- there exists Z such that $Cov(X, Z) \neq 0$ and $Cov(Z, \epsilon) = 0$ ("Valid Instrument")

In dd-ss example, $Y = Q$, $X = P$, $\epsilon \sim$ demand shock, $Z = r$, the βs are the δs

Method of Moments Perspective

Since

- $E(\epsilon) = 0$
- $Cov(\epsilon, Z) = E(\epsilon Z) = 0$

we have the following "Population Moment Conditions":

$$E(\epsilon) = E(Y - \beta_0 - \beta_1 X) = 0$$

$$E(\epsilon Z) = E((Y_i - \beta_0 - \beta_1 X)Z) = 0$$

Method of Moments Perspective

Replace with sample moments:

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm}X_i) = 0 \qquad [A]$$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm}X_i)Z_i = 0 \qquad [B]$$

Solving gives the method of moments (mm) estimator:

[A] $\Rightarrow \overline{Y} = \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm}\overline{X} \Rightarrow \hat{\beta}_0^{mm} = \overline{Y} - \hat{\beta}_1^{mm}\overline{X}$

Method of Moments Perspective

Substituting into [B] gives

$$\sum_{i=1}^{n}(Y_i - (\overline{Y} - \hat{\beta}_1^{mm}\overline{X}) - \hat{\beta}_1^{mm}X_i)Z_i = 0$$

$$\sum_{i=1}^{n}((Y_i - \overline{Y}) - \hat{\beta}_1^{mm}(X_i - \overline{X}))Z_i = 0$$

$$\sum_{i=1}^{n}(Y_i - \overline{Y})Z_i - \hat{\beta}_1^{mm}\sum_{i=1}^{n}(X_i - \overline{X})Z_i = 0$$

Method of Moments Perspective

Therefore

$$\hat{\beta}_0^{mm} = \overline{Y} - \hat{\beta}_1^{mm}\overline{X}$$

$$\hat{\beta}_1^{mm} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})Z_i}{\sum_{i=1}^{n}(X_i - \overline{X})Z_i} = \frac{\sum_{i=1}^{n}(Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(Z_i - \overline{Z})(X_i - \overline{X})}$$

which is the same as $iv/2sls$ estimator in dd-ss example (with notational substitution)

For the moment, we will refer to the estimators as "iv" estimators

Consistency

Consistency (two arguments):

$$\hat{\beta}_1^{iv} = \frac{\sum_{i=1}^n (Y_i - \overline{Y})Z_i}{\sum_{i=1}^n (X_i - \overline{X})Z_i} = \frac{\sum_{i=1}^n (Z_i - \overline{Z})Y_i}{\sum_{i=1}^n (Z_i - \overline{Z})X_i} = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(\beta_0 + \beta_1 X_i + \epsilon_i)}{\sum_{i=1}^n (Z_i - \overline{Z})X_i}$$

$$= \beta_1 + \frac{\sum_{i=1}^n (Z_i - \overline{Z})\epsilon_i}{\sum_{i=1}^n (Z_i - \overline{Z})X_i} \xrightarrow{p} \beta_1 + \frac{Cov(Z,\epsilon)}{Cov(Z,X)} = \beta_1$$

Alternatively, note that

$$E(Y - \beta_0 - \beta_1 X) = 0 \text{ and } E((Y - \beta_0 - \beta_1 X)Z) = 0$$

implies $\beta_1 = Cov(Z,Y)/Cov(Z,X)$ (Exercise). We have

$$\hat{\beta}_1^{iv} = \frac{\sum_{i=1}^n (Y_i - \overline{Y})(Z_i - \overline{Z})}{\sum_{i=1}^n (X_i - \overline{X})(Z_i - \overline{Z})} \xrightarrow{p} \frac{Cov(Y,Z)}{Cov(X,Z)} = \beta_1$$

Bias

Note that $iv/2sls/mm$ estimators are *biased* estimators

$$\hat{\beta}_1^{iv} = \beta_1 + \frac{\sum_{i=1}^n (Z_i - \overline{Z})\epsilon_i}{\sum_{i=1}^n (Z_i - \overline{Z})X_i}$$

$$E(\hat{\beta}_1^{iv} \mid X_1, \dots, X_n, Z_1, \dots, Z_n) = \beta_1 + \frac{\sum_{i=1}^n (Z_i - \overline{Z})E(\epsilon_i \mid X_1, \dots, X_n, Z_1, \dots, Z_n)}{\sum_{i=1}^n (Z_i - \overline{Z})X_i}$$

but

$$E(\epsilon_i \mid X_1, \dots, X_n, Z_1, \dots, Z_n) \neq 0$$

since ϵ is correlated with X

Standard Errors

If $Var(\epsilon \mid X, Z) = \sigma^2$, then $Var(\epsilon_i \mid X_1, \dots, X_n, Z_1, \dots, Z_n) = \sigma^2$ and we have

$$
\begin{aligned}
Var(\hat{\beta}_1^{iv} \mid \dots) &= \frac{\sum_{i=1}^n (Z_i - \overline{Z})^2 Var(\epsilon_i \mid \dots)}{(\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X}))^2} \\
&= \frac{\sigma^2 \sum_{i=1}^n (Z_i - \overline{Z})^2}{(\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X}))^2} \\
&= \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2 \left(\frac{(\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X}))^2}{\sum_{i=1}^n (Z_i - \overline{Z})^2 \sum_{i=1}^n (X_i - \overline{X})^2} \right)} \\
&= \frac{\sigma^2}{R_{X|Z}^2 \sum_{i=1}^n (X_i - \overline{X})^2}
\end{aligned}
$$

Standard Error

$$Var(\hat{\beta}_1^{iv} \mid \dots) = \frac{\sigma^2}{R^2_{X|Z} \sum_{i=1}^{n} (X_i - \overline{X})^2}$$

demonstrates that the trade-off in obtaining consistent estimators is larger estimator variance

Note: since $0 \leq R^2_{X|Z} < 1$

- If $R_{X|Z} = 1$, then we are back to simple linear regression estimator variance
- But in this case, X and Z are perfectly correlated, and Z must also be endogenous, so not a valid instrument
- If $R^2_{X|Z}$ is near zero, Z is a "weak instrument", variance will be very large

Standard Error

To estimate σ^2, use:

$$\widehat{\sigma^2_{iv}} = \frac{1}{n-2}\sum_{i=1}^{n}\hat{\epsilon}^2_{i,iv}$$

where $\hat{\epsilon}_{i,iv} = Y_i - \hat{\beta}_0^{iv} + \hat{\beta}_1^{iv} X_i$

- Note the definition of $\hat{\epsilon}_{i,iv}$, which uses X_i rather that $\widehat{X}_i$
- Conventional to divide by $n-2$ even though IV/MM/2SLS only has large sample justifications

For R^2, use: $R^2 = 1 - \frac{\sum_{i=1}^{n}\hat{\epsilon}^2_{i,iv}}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}$

Using Matrix Algebra

We express this example (both MM and IV/2SLS) using matrix algebra

Suppose the regression equation of interest is:

$$Y = \beta_0 + \beta_1 X_1 + \epsilon$$

where

- X_1 is endogenous (correlated with the noise term)
- Suppose Z_1 is correlated with X_1 and uncorrelated with ϵ
- Note: Previous variables X and Z now labelled as X_1 and Z_1

Moment Conditions

$$\begin{aligned} E(\epsilon) &= E(Y - \beta_0 - \beta_1 X_1) = 0 \\ E(\epsilon Z_1) &= E((Y - \beta_0 - \beta_1 X_1) Z_1) = 0\,. \end{aligned}$$

Using Matrix Algebra

Sample Moment Conditions

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm} X_{i1}) = 0$$

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm} X_{i1})Z_{i1} = 0$$

In matrix algebra: $Z^{\mathrm{T}}(y - X\hat{\beta}^{mm}) = Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta}^{mm} = 0$ where

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix},\ X = \begin{bmatrix} 1 & X_{11} \\ 1 & X_{21} \\ \vdots & \vdots \\ 1 & X_{n1} \end{bmatrix},\ \varepsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{bmatrix},\ Z = \begin{bmatrix} 1 & Z_{11} \\ 1 & Z_{21} \\ \vdots & \vdots \\ 1 & Z_{n1} \end{bmatrix}$$

Using Matrix Algebra

Note that

- $Z^{\mathrm{T}}X$ is 2×2
- Z_{i1} correlated with X_{i1} means that $Z^{\mathrm{T}}X$ is invertible

Solving moment conditions gives

$$\hat{\beta}^{mm} = (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y$$

Using Matrix Algebra

2SLS Approach:

Step 1: Regress X on Z (X is $n \times 2$, so what does this mean?)

- Regress each column of X on Z

$$i_n = Zb_0 + u_{*0} \text{ and } X_{*1} = Zb_1 + u_{*1}$$

We can put into one single matrix:

$$\begin{bmatrix} i_n & X_{*1} \end{bmatrix} = Z \begin{bmatrix} b_0 & b_1 \end{bmatrix} + \begin{bmatrix} u_{*0} & u_{*1} \end{bmatrix}$$

or $X = ZB + U$

We have $\hat{B} = (Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X$

Fitted value is $\hat{X} = Z\hat{B} = Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X$

Using Matrix Algebra

Step 2: Regress y on $\hat{X}$

$$
\begin{aligned}
\hat{\beta}^{iv/2sls} &= (\hat{X}^{\mathrm{T}}\hat{X})^{-1}\hat{X}^{\mathrm{T}}y \\
&= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \\
&= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \\
&= (Z^{\mathrm{T}}X)^{-1}(Z^{\mathrm{T}}Z)(X^{\mathrm{T}}Z)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \\
&= (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y
\end{aligned}
$$

Using Matrix Algebra

Consistency:

$$\begin{aligned}\widehat{\beta}^{iv/2sls} &= (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y = (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}(X\beta+\epsilon)\\ &= \beta + (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}\epsilon\\ &= \beta + (\tfrac{1}{n}Z^{\mathrm{T}}X)^{-1}(\tfrac{1}{n}Z^{\mathrm{T}}\epsilon) \xrightarrow{p} \beta\end{aligned}$$

since $\frac{1}{n}Z^{\mathrm{T}}\epsilon \xrightarrow{p} 0_{2\times 1}$

We omit the arguments for asymptotic normality

Using Matrix Algebra

For asymptotically valid variance-covariance matrices

- Homoskedastic errors

$$\widehat{Var}(\hat{\beta}_{iv}) = \widehat{\sigma^2}(Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}Z(X^{\mathrm{T}}Z)^{-1}$$

  where $\widehat{\sigma^2}$ is as previously defined

- Heteroskedasticity-robust:

$$\widehat{Var}(\hat{\beta}_{iv}) = (Z^{\mathrm{T}}X)^{-1}\left(\sum_{i=1}^{n}\hat{\epsilon}_{i,iv}^2 Z_{i*}^{\mathrm{T}}Z_{i*}\right)(X^{\mathrm{T}}Z)^{-1}$$

  where Z_{i*} are the i-rows of Z

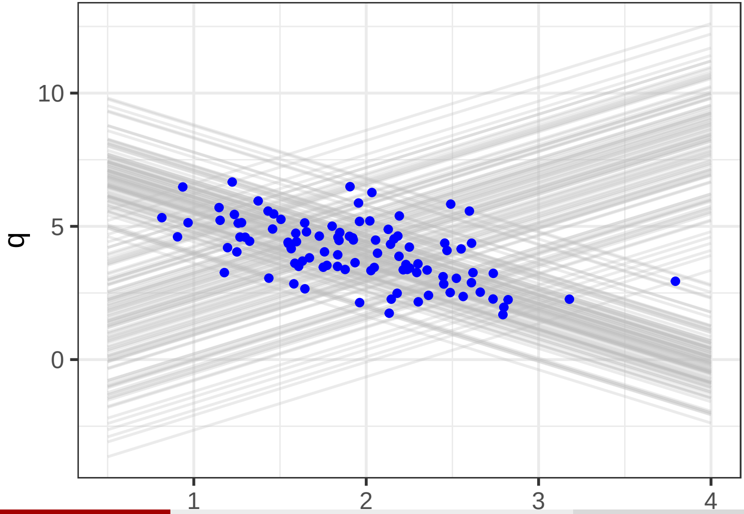
Example 1: Simulated SS/DD

Suppose

$$
\begin{aligned}
Q_i^d &= 8 - 2P_i + \epsilon_i^d && \text{(Demand Eq } \delta_1 < 0\text{)} \\
Q_i^s &= -2 + 2P_i + 0.8r_i + \epsilon_i^s && \text{(Supply Eq } \alpha_1 > 0\text{)} \\
Q_i^s &= Q_i^d && \text{(Market Clearing)}
\end{aligned}
$$

where $r_i \sim N(3, 4)$ observed, $e_i^d \sim N(0, 1)$, $e_i^s \sim N(0, 1)$ unobserved

Example 1: Simulated SS/DD



Example 1: Simulated SS/DD

OLS Estimates

```
dat <- tibble(p=pstar, q=qstar, r=r)
ddss_ols <- lm(q~p, data=dat)
summary(ddss_ols)$coefficients
```

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  6.341252  0.3660495 17.323483 1.366065e-31
p           -1.175239  0.1816905 -6.468357 3.906490e-09
```

Example 1: Simulated SS/DD

IV Estimates (using `ivreg` package, default standard errors)

```
ddss_iv <- ivreg(q ~ p | r, data=dat)
ddss_coef <- summary(ddss_iv)$coef
attr(ddss_coef, "df") <- NULL
attr(ddss_coef, "nobs") <- NULL
ddss_coef
```

```
             Estimate Std. Error   t value     Pr(>|t|)
(Intercept)  8.492688  0.5937838 14.302662 9.954936e-26
p           -2.283065  0.3000241 -7.609603 1.710061e-11
```

Example 1: Simulated SS/DD

IV Estimates (using `ivreg` package, heteroskedasticity-robust standard errors)

```
coeftest(ddss_iv, vcov=vcovHC(ddss_iv, type="HC0"))
```

```
t test of coefficients:

            Estimate Std. Error t value  Pr(>|t|)
(Intercept)  8.49269    0.57835 14.6845 < 2.2e-16 ***
p           -2.28306    0.28389 -8.0421 2.063e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example 1: Simulated SS/DD

```
y <- as.matrix(dat$q)                                        #
X <- as.matrix(cbind(rep(1, length(dat$p)), dat$p))  # Assemble data
Z <- as.matrix(cbind(rep(1, length(dat$r)), dat$r))  #
b_iv <- solve(t(Z) %*% X) %*% (t(Z) %*% y)     # calculate IV estimate
# var-cov assuming homoskedasticity
y_iv_hat <- X %*% b_iv
e_iv <- y - y_iv_hat
s2 <- sum(e_iv^2) / (n-2)
var_b <- s2 * solve(t(Z) %*% X) %*% (t(Z) %*% Z) %*% solve(t(X) %*% Z)
# Robust variance-covariance
S <- matrix(c(0,0,0,0), nrow=2)
for (i in 1:n){
  zi <- matrix(Z[i,], nrow=2)
  S <- S + e_iv[i]^2 * zi %*% t(zi)
}
var_b_rb <- solve(t(Z) %*% X) %*% S %*% solve(t(X) %*% Z)
```

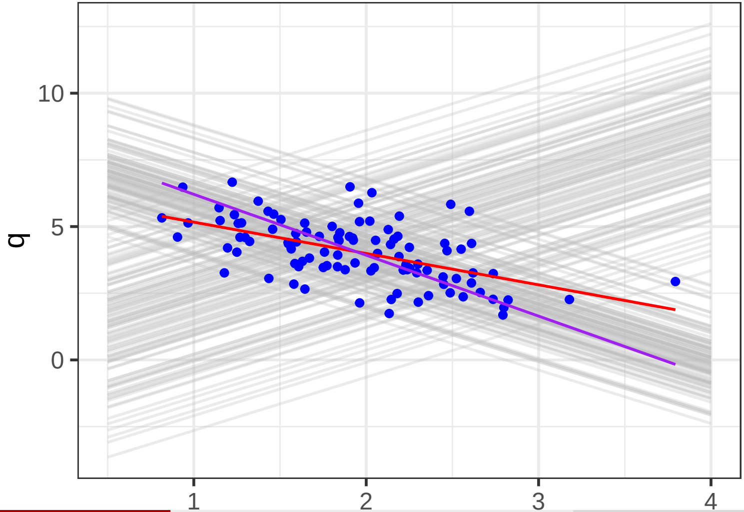
Example 1: Simulated SS/DD

```
iv_coefs <- cbind(b_iv, sqrt(diag(var_b)), sqrt(diag(var_b_rb)))
colnames(iv_coefs) <- c("Estimate", "default s.e.", "robust s.e.")
rownames(iv_coefs) <- c("(Intercept)", "X")
iv_coefs
```

```
             Estimate default s.e. robust s.e.
(Intercept)  8.492688    0.5937838   0.5783452
X           -2.283065    0.3000241   0.2838908
```

- Successfully replicated estimates from `ivreg` package
- Default and heteroskedastic estimates for IV estimates are similar (not surprising, since data is homoskedastic)
- IV s.e. is larger than OLS s.e.
- IV estimate much closer to true value than OLS estimates
- Next slide contains plot of OLS (red) and IV (purple) estimated regression lines

Example 1: Simulated SS/DD



Regressions with Exog. and Endog. Regressors, Multiple IVs

Session 8.3

- Suppose you have both endogenous and exogenous regressors
- With possibly more IVs than endogenous regressors

E.g., 1 exog regressor, 1 endog. regressor, 2 valid instruments for endog. regressor

$$Y = \beta_0 + \beta_1 X_1^k + \beta_2 X_2^g + \epsilon$$

- X_1^k is exogenous (not correlated with the noise term)
- X_2^g is endogenous (correlated with the noise term)
- There exists Z_2 and Z_3 both correlated with X_2^g and uncorrelated with ϵ
- "Over-identification"

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

Method of Moments Approach:

For ease of exposition, we stick to our specific example (1 exog, 1 endog, 2 instruments)

Population: $E(\epsilon) = 0,\ Cov(X_1^k, \epsilon) = 0,\ Cov(Z_2, \epsilon) = 0,\ Cov(Z_3, \epsilon) = 0$

That is,

$$E(\epsilon) = E(Y - \beta_0 - \beta_1 X_1^k - \beta_2 X_2^g) = 0$$

$$E(\epsilon X_1^k) = E((Y - \beta_0 - \beta_1 X_1^k - \beta_2 X_2^g) X_1^k) = 0$$

$$E(\epsilon Z_2) = E((Y - \beta_0 - \beta_1 X_1^k - \beta_2 X_2^g) Z_2) = 0$$

$$E(\epsilon Z_3) = E((Y - \beta_0 - \beta_1 X_1^k - \beta_2 X_2^g) Z_3) = 0$$

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

Data sample: $\{X_{i1}^k, X_{i2}^g, Z_{i2}, Z_{i3}, Y_i\}_{1=i}^n$

Sample Moments:

$$\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^{mm} = \frac{1}{n}\sum_{i=1}^n (Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm} X_{i1}^k - \hat{\beta}_2^{mm} X_{i2}^g) = 0$$

$$\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^{mm} X_{i1}^k = \frac{1}{n}\sum_{i=1}^n ((Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm} X_{i1}^k - \hat{\beta}_2^{mm} X_{i2}^g) X_{i1}^k) = 0$$

$$\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^{mm} Z_{i2} = \frac{1}{n}\sum_{i=1}^n ((Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm} X_{i1}^k - \hat{\beta}_2^{mm} X_{i2}^g) Z_{i2}) = 0$$

$$\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^{mm} Z_{i3} = \frac{1}{n}\sum_{i=1}^n ((Y_i - \hat{\beta}_0^{mm} - \hat{\beta}_1^{mm} X_{i1}^k - \hat{\beta}_2^{mm} X_{i2}^g) Z_{i3}) = 0$$

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

"Solve" sample moments for $\hat{\beta}_0^{mm}$, $\hat{\beta}_1^{mm}$, $\hat{\beta}_2^{mm}$

But 4 equations in 3 unknowns — cannot solve

Choose $\hat{\beta}_0^{mm}$, $\hat{\beta}_1^{mm}$, $\hat{\beta}_2^{mm}$ to make "LHS as close to zero" as possible

- In particular, choose $\hat{\beta}_0^{mm}$, $\hat{\beta}_1^{mm}$, $\hat{\beta}_2^{mm}$ to minimize:

$$\left(\sum_{i=1}^{n} \hat{\epsilon}_i^{mm}\right)^2 + \left(\sum_{i=1}^{n} \hat{\epsilon}_i^{mm} X_{i1}^{k}\right)^2 + \left(\sum_{i=1}^{n} \hat{\epsilon}_i^{mm} Z_{i2}\right)^2 + \left(\sum_{i=1}^{n} \hat{\epsilon}_i^{mm} Z_{i3}\right)^2$$

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

In matrix algebra terms: let

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \; X = \begin{bmatrix} 1 & X_{11}^k & X_{12}^g \\ 1 & X_{21}^k & X_{22}^g \\ \vdots & \vdots & \vdots \\ 1 & X_{n1}^k & X_{n2}^g \end{bmatrix}, \; \widehat{\beta}^{mm} = \begin{bmatrix} \widehat{\beta}_0^{mm} \\ \widehat{\beta}_1^{mm} \\ \widehat{\beta}_2^{mm} \end{bmatrix}, \; Z = \begin{bmatrix} 1 & X_{11}^k & Z_{12} & Z_{13} \\ 1 & X_{21}^k & Z_{22} & Z_{23} \\ \vdots & \vdots & \vdots & \vdots \\ 1 & X_{n1}^k & Z_{n2} & Z_{n3} \end{bmatrix}$$

Sample moments (dropping the $1/n$) are:

$$\underbrace{Z^{\mathrm{T}}}_{4\times n} \underbrace{\left(y - X\widehat{\beta}^{mm}\right)}_{n\times 1} = Z^{\mathrm{T}} y - Z^{\mathrm{T}} X \widehat{\beta}^{mm} = 0$$

4 equations in 3 unknowns — cannot solve

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

Instead, choose $\hat{\beta}_0^{mm}$, $\hat{\beta}_1^{mm}$, $\hat{\beta}_2^{mm}$ to minimize the "sum of squared moments"

$$\underbrace{(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})^{\mathrm{T}}}_{1\times 4}\underbrace{(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})}_{4\times 1}$$

$$= \; y^{\mathrm{T}}ZZ^{\mathrm{T}}y - 2\hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}ZZ^{\mathrm{T}}y + \hat{\beta}^{\mathrm{T}}X^{\mathrm{T}}ZZ^{\mathrm{T}}X\hat{\beta}$$

Minimizing this gives

$$\hat{\beta}_{mm} = (X^{\mathrm{T}}ZZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}y$$

Requires that the 4×3 matrix $Z^{\mathrm{T}}X$ has full column rank

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

Consistency:

$$\begin{aligned}\hat{\beta}_{mm} &= (X^{\mathrm{T}}ZZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}y \\ &= (X^{\mathrm{T}}ZZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}(X\beta+\epsilon) \\ &= (X^{\mathrm{T}}ZZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}X\beta + (X^{\mathrm{T}}ZZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}\epsilon \\ &= \beta + ((\tfrac{1}{n}X^{\mathrm{T}}Z)((\tfrac{1}{n}Z^{\mathrm{T}}X))^{-1}((\tfrac{1}{n}X^{\mathrm{T}}Z)(\tfrac{1}{n}Z^{\mathrm{T}}\epsilon) \quad \xrightarrow{p} \beta\end{aligned}$$

Requires that $\frac{1}{n}Z^{\mathrm{T}}\epsilon \xrightarrow{p} 0_{4\times 1}$ and $\frac{1}{n}Z^{\mathrm{T}}X \xrightarrow{p} \Sigma_{ZX}$ full column rank

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

Variance (Details of proof omitted)

- Homoskedasticity:

$$\widehat{Var}(\hat{\beta}_{mm}) = \sigma^2 (X^{\mathrm{T}} Z Z^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Z (Z^{\mathrm{T}} Z) Z^{\mathrm{T}} X (X^{\mathrm{T}} Z Z^{\mathrm{T}} X)^{-1}$$

  Estimate σ^2 with $\widetilde{\sigma^2} = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_{i,iv}^2$

- Heteroskedasticity-Robust

$$\widehat{Var}(\hat{\beta}_{mm}) = (X^{\mathrm{T}} Z Z^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Z \left(\sum_{i=1}^{n} \hat{\epsilon}_{i,iv}^2 Z_{i*}^{\mathrm{T}} Z_{i*} \right) Z^{\mathrm{T}} X (X^{\mathrm{T}} Z Z^{\mathrm{T}} X)^{-1}$$

Regressions with Exog. and Endog. Regressors, Mult. IVs (MM)

In **just-identified** case, where #endo = #instruments

- e.g., in our case, if we only have Z_2 to instrument for X_2^g,

then $Z^{\mathrm{T}}X$ is square, and

- $\hat{\beta}_{mm} = (X^{\mathrm{T}}ZZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}y = (Z^{\mathrm{T}}X)^{-1}(X^{\mathrm{T}}Z)^{-1}X^{\mathrm{T}}ZZ^{\mathrm{T}}y = (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y$

Variance

- Homoskedasticity: $\widehat{Var}(\hat{\beta}_{mm}) = \widehat{\sigma^2}(Z^{\mathrm{T}}X)^{-1}(Z^{\mathrm{T}}Z)(X^{\mathrm{T}}Z)^{-1}$
- Heteroskedasticity-Robust $\widehat{Var}(\hat{\beta}_{mm}) = (Z^{\mathrm{T}}X)^{-1}\left(\sum_{i=1}^{n}\hat{\epsilon}_{i,iv}^{2}Z_{i*}^{\mathrm{T}}Z_{i*}\right)(X^{\mathrm{T}}Z)^{-1}$

Expressions same as “IV” case

Regressions with Exog. and Endog. Reg., Mult. IVs (2SLS)

Two-Stage Least Squares Approach

Step 1: Regress every column of X on Z

- Estimates are $\hat{B} = (Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X$ (This is 4×3)
- Fitted Values are $\hat{X} = Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X$
  - This is $n \times 3$
  - First column is vector of 1's
  - Second column is X_{i1}^{k}
  - Third column is $\hat{X}_{i2}^{g}$ from regression of X_{i2}^{g} on intercept, X_{i1}^{k}, Z_{i2} and Z_{i3}

Regressions with Exog. and Endog. Reg., Mult. IVs (2SLS)

Step 2: Regress y on $\hat{X}$

$$
\begin{aligned}
\hat{\beta}^{2sls} &= (\hat{X}^{\mathrm{T}}\hat{X})^{-1}\hat{X}^{\mathrm{T}}y \\
&= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \\
&= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y\,.
\end{aligned}
$$

Requires that the 4×3 matrix $Z^{\mathrm{T}}X$ has full column rank and that Z has full column rank

Regressions with Exog. and Endog. Reg., Mult. IVs (2SLS)

Consistency:

$$
\begin{aligned}
\hat{\beta}^{2sls} &= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \\
&= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}(X\beta + \epsilon) \\
&= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X\beta \\
&\qquad + (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}\epsilon \\
&= \beta + (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}\epsilon \\
&= \beta + (\tfrac{1}{n}X^{\mathrm{T}}Z(\tfrac{1}{n}Z^{\mathrm{T}}Z)^{-1}\tfrac{1}{n}Z^{\mathrm{T}}X)^{-1}\tfrac{1}{n}X^{\mathrm{T}}Z(\tfrac{1}{n}Z^{\mathrm{T}}Z)^{-1}\tfrac{1}{n}Z^{\mathrm{T}}\epsilon \xrightarrow{p} \beta
\end{aligned}
$$

Regressions with Exog. and Endog. Reg., Mult. IVs (2SLS)

Variance (details of proofs omitted)

Homoskedasticity

$$\widehat{Var}(\hat{\beta}_{2sls}) = \widehat{\sigma^2}(X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}$$

Heteroskedasticity-Robust

$$\widehat{Var}(\hat{\beta}_{2sls}) =$$

$$(X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}\left[\sum_{i=1}^{n}\hat{\epsilon}_{i,iv}^{2}Z_{i*}^{\mathrm{T}}Z_{i*}\right](Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X(X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}$$

Regressions with Exog. and Endog. Reg., Mult. IVs (2SLS)

In **just-identified** case, where #endo = #instruments, $Z^{\mathrm{T}}X$ is square

Therefore

$$\begin{aligned}\hat{\beta}_{2sls} &= (X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y \\ &= (Z^{\mathrm{T}}X)^{-1}(Z^{\mathrm{T}}Z)(X^{\mathrm{T}}Z)^{-1}X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}y = (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y\end{aligned}$$

Same as "IV" estimator

Variance formulas also converge to IV variance formulas

- Just-identified case: 2SLS and MM same as IV
- Over-identified case: 2SLS $\neq$ MM, IV does not apply

Example 2

Session 8.4

Example using data from `earnings2019.csv`

$$\ln earn = \beta_0 + \beta_1 educ + \beta_2 age + \beta_3 \ln tenure + \epsilon$$

Concern: $Ability$ is omitted. Furthermore, no good proxy for ability available. $Ability$ correlated with $educ$, so $educ$ correlated with ϵ

That is, we assume age and $\ln tenure$ exogenous, $educ$ endogenous

Suppose $feduc$ and $meduc$ are valid instruments

Example 2 OLS Estimates

```
dat <- read_csv("data\\earnings2019.csv", show_col_types=FALSE) %>%
  mutate(ln_earn = log(earn), ln_tenure=log(tenure), const = 1)
cat("Assuming homoskedasticity (default)\n")
mdl2_ols <- lm(ln_earn ~ age + ln_tenure + educ, data=dat) # Estimate OLS
summary(mdl2_ols)$coefficients %>% round(4) # Print default coefficients and standard
```

```
Assuming homoskedasticity (default)
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9699     0.0628 15.4349    0e+00
age           0.0025     0.0008  3.3341    9e-04
ln_tenure     0.1477     0.0090 16.4018    0e+00
educ          0.1268     0.0038 33.1159    0e+00
```

Example 2 OLS Estimates

```
cat("\nUsing heteroskedasticity-robust standard errors")
coeftest(mdl2_ols, vcov=vcovHC, type="HC") %>% round(4)# Robust standard errors
```

```
Using heteroskedasticity-robust standard errors
t test of coefficients:

            Estimate Std. Error t value Pr(>|t|)
(Intercept)   0.9699     0.0655 14.8101   <2e-16 ***
age           0.0025     0.0008  3.1315   0.0017 **
ln_tenure     0.1477     0.0092 16.0494   <2e-16 ***
educ          0.1268     0.0039 32.1078   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Example 2 MM Estimates

```
## Assemble data for MM / 2SLS
y <- dat %>% select(c(ln_earn)) %>% as.matrix()
X <- dat %>% select(c(const, age, ln_tenure, educ)) %>% as.matrix()
Z <- dat %>% select(c(const, age, ln_tenure, feduc, meduc)) %>% as.matrix()
n <- length(y)
Zcol <- dim(Z)[2]
ZTX <- t(Z) %*% X ; XTZ <- t(X) %*% Z ; ZTZ <- t(Z) %*% Z ; ZTy <- t(Z) %*% y

#--MM--
beta_MM <- solve(XTZ %*% ZTX) %*% XTZ %*% ZTy
ehat_IV <- y - X %*% beta_MM
s2hat <- sum(ehat_IV^2)/n
eZZ <- matrix(0, nrow=Zcol, ncol=Zcol)
for (i in 1:n){
  eZZ <- eZZ + ehat_IV[i]^2 * t(Z[i,,drop=F]) %*% Z[i,,drop=F]
}
```

Example 2 MM Estimates

```
vbeta_MM <-  s2hat * solve(XTZ%*%ZTX) %*% XTZ %*% ZTZ %*% ZTX %*% solve(XTZ%*%ZTX)
vbeta_MM_rob <- solve(XTZ%*%ZTX) %*% XTZ %*% eZZ %*% ZTX %*% solve(XTZ%*%ZTX)
MM_results <- cbind(estimates = beta_MM,
                    s.e. = sqrt(diag(vbeta_MM)),
                    s.e.robust = sqrt(diag(vbeta_MM_rob)))
MM_results %>% round(4)
```

```
          ln_earn   s.e. s.e.robust
const     -1.7976 0.5844     0.5911
age        0.0096 0.0026     0.0027
ln_tenure  0.1386 0.0108     0.0111
educ       0.2991 0.0336     0.0341
```

Example 2 2SLS Estimates (from formulas)

```
#--2SLS--
beta_TSLS <- solve(XTZ %*% solve(ZTZ) %*% ZTX) %*% XTZ %*% solve(ZTZ) %*% ZTy
ehat_TSLS <- y - X %*% beta_TSLS
s2hat_TSLS <- sum(ehat_TSLS^2)/n
eZZ_TSLS <- matrix(0, nrow=Zcol, ncol=Zcol)
for (i in 1:n){
  eZZ_TSLS <- eZZ_TSLS + ehat_TSLS[i]^2 * t(Z[i,,drop=F]) %*% Z[i,,drop=F]
}
vbeta_TSLS <-  s2hat_TSLS * solve(XTZ %*% solve(ZTZ) %*% ZTX)
vbeta_TSLS_rob <- solve(XTZ %*% solve(ZTZ) %*% ZTX) %*% XTZ %*% solve(ZTZ) %*%
  eZZ_TSLS %*% solve(ZTZ) %*% ZTX %*% solve(XTZ %*% solve(ZTZ) %*% ZTX)
TSLS_results <- cbind(estimates = beta_TSLS,
                      s.e. = sqrt(diag(vbeta_TSLS)),
                      s.e.robust = sqrt(diag(vbeta_TSLS_rob)))
```

Example 2 2SLS Estimates (from formulas)

```
TSLS_results %>% round(4)
```

```
           ln_earn   s.e. s.e.robust
const      -0.3843 0.1915     0.2088
age         0.0032 0.0008     0.0009
ln_tenure   0.1399 0.0096     0.0098
educ        0.2205 0.0131     0.0143
```

Example 2 2SLS Estimates (from ivreg package)

```
mdl2_iv <- ivreg(ln_earn ~ age + ln_tenure + educ |
                    age + ln_tenure + feduc + meduc, data=dat)
mdl2_iv_coef <- summary(mdl2_iv)$coef
attr(mdl2_iv_coef,"df")<-NULL
attr(mdl2_iv_coef,"nobs")<-NULL
mdl2_iv_coef %>% round(4)
```

```
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.3843     0.1916 -2.0058   0.0449
age           0.0032     0.0008  3.9523   0.0001
ln_tenure     0.1399     0.0096 14.5964   0.0000
educ          0.2205     0.0131 16.8675   0.0000
```

Examples 2 (2SLS, from ivreg package)

```
coeftest(mdl2_iv,vcov=vcovHC(mdl2_iv,type="HC0"))  %>% round(4)
```

```
t test of coefficients:

             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.3843     0.2088 -1.8406   0.0657 .
age           0.0032     0.0009  3.7345   0.0002 ***
ln_tenure     0.1399     0.0098 14.2368   <2e-16 ***
educ          0.2205     0.0143 15.4326   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

General Case

All formulas and results continue to apply to general case with K exogenous regressors, G endogenous regressors and M instruments, $M \geq G$

$$Y = \beta_0 + \beta_1 X_1^k + \cdots + \beta_K X_K^k + \beta_{K+1} X_{K+1}^g + \cdots + \beta_{K+G} X_{K+G}^g + \epsilon$$

with

- instrumental variables $Z_1, \ldots Z_K, Z_{K+1}, \ldots Z_{K+M}$
- $Z_1 = X_1^k, \ldots, Z_K = X_K^k$
- $E(\epsilon) = 0$ and $cov(Z_j, \epsilon) = 0$ for all $j = 1, \ldots, K + M$

Only change is Z is $n \times K + M + 1$ and X is $n \times K + G + 1$

Generalized Method of Moments

Session 8.5 (Optimal) Generalized Method of Moments (GMM)

- Simplified version
- encompasses MM and TSLS as special cases

Generalized Method of Moments

GMM estimator: minimize *weighted* sum of squared moments, i.e.,

$$\hat{\beta}_W^{gmm} = \text{argmin}_{\hat{\beta}} \underbrace{(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})^{\mathrm{T}}W(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})}_{"J(W)"}$$

where

- W is some symmetric positive-definite weight matrix (may change with n and may be data dependent)
- X is the $n \times K + G + 1$ matrix of regressors (exogenous and endogenous)
- Z is the $n \times K + M + 1$ matrix of exogenous variables (exogenous regressors and instruments)

Generalized Method of Moments

Minimizing $J(W)$ gives

$$\hat{\beta}_W^{gmm} = (X^{\mathrm{T}} Z W Z^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Z W Z^{\mathrm{T}} y \qquad \text{(Exercise!)}$$

- Obviously MM is GMM with $W = I_n$ and 2SLS is GMM with $W = (Z^{\mathrm{T}} Z)^{-1}$
- Consistency:

$$\begin{aligned}
\hat{\beta}_W^{gmm} &= (X^{\mathrm{T}} Z W Z^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Z W Z^{\mathrm{T}} y \\
&= \beta + (X^{\mathrm{T}} Z W Z^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Z W Z^{\mathrm{T}} \epsilon \\
&= \beta + [(\tfrac{1}{n} X^{\mathrm{T}} Z) W (\tfrac{1}{n} Z^{\mathrm{T}} X)]^{-1} (\tfrac{1}{n} X^{\mathrm{T}} Z) W (\tfrac{1}{n} Z^{\mathrm{T}} \epsilon) \quad \xrightarrow{p} \quad \beta
\end{aligned}$$

Generalized Method of Moments

Variance:

Under homoskedasticity:

$$\widehat{Var}(\hat{\beta}_W^{gmm}) = \widehat{\sigma^2}(X^{\mathrm{T}}ZWZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZW(Z^{\mathrm{T}}Z)^{-1}WZ^{\mathrm{T}}X(X^{\mathrm{T}}ZWZ^{\mathrm{T}}X)^{-1}$$

Heteroskedasticity-robust

$$\widehat{Var}(\hat{\beta}_W^{gmm}) \\ = (X^{\mathrm{T}}ZWZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZW\left(\sum_{i=1}^{n}\hat{\epsilon}_{i,gmm}^2 Z_{i*}^{\mathrm{T}}Z_{i*}\right)WZ^{\mathrm{T}}X(X^{\mathrm{T}}ZWZ^{\mathrm{T}}X)^{-1}$$

Optimal GMM

Optimal Choice of Weights?

It turns out (proof omitted) that an optimal choice of weights is

$$W^* = \left(\sum_{i=1}^{n} \hat{\epsilon}_{i,gmm}^2 Z_{i*}^{\mathrm{T}} Z_{i*} \right)^{-1}$$

Optimal GMM

Two-step approach:

- Compute $\hat{\beta}_W^{gmm}$ for some (non-optimal) weighting matrix W.
  - Common choice is to use $W = (Z^{\mathrm{T}} Z)^{-1}$, which gives the (inefficient but consistent) 2SLS estimator $\hat{\beta}^{2sls}$, calculate $\hat{\epsilon}_{i,2sls}$
  - Calculate $W^* = \left(\sum_{i=1}^{n} \hat{\epsilon}_{i,2sls}^2 Z_{i*}^{\mathrm{T}} Z_{i*} \right)^{-1}$
  - Calculate the optimal GMM estimator as

$$\hat{\beta}^{gmm} = (X^{\mathrm{T}} Z W^* Z^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} Z W^* Z^{\mathrm{T}} y.$$

Optimal GMM

Variance:

- Homoskedasticity:

$$\widehat{Var}(\hat{\beta}^{gmm}) = \widehat{\sigma^2}\left(X^{\mathrm{T}}Z(Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X\right)^{-1}$$

where $\widehat{\sigma^2} = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_{i,gmm}^2$

- Heteroskedasticity-Robust:

$$\widehat{Var}(\hat{\beta}^{gmm}) = \left(X^{\mathrm{T}}Z\left(\sum_{i=1}^{N}\hat{\epsilon}_{i,gmm}^2 Z_{i*}^{\mathrm{T}}Z_{i*}\right)^{-1}Z^{\mathrm{T}}X\right)^{-1}$$

Optimal GMM

Form of the variance of the optimal GMM estimator *under conditional homoskedasticity* is the same as that of 2SLS

- 2SLS is as good as optimal GMM under homoskedasticity
- 2SLS and two-step implementation of optimal GMM are not numerically identical, but both are asymptotically efficient

Optimal GMM

Example 3: We calculate GMM estimates and s.e. for `earnings2019` example

Earlier we calculated

- 2SLS residuals as `ehat_2SLS`
- $\sum_{i=1}^{n} \hat{\epsilon}_{i,2sls}^{2} Z_{i*}^{\mathrm{T}} Z_{i*}$ as `eZZ_TSLS`
- calculated $X^{\mathrm{T}}Z$, $Z^{\mathrm{T}}X$, $Z^{\mathrm{T}}Z$ and $Z^{\mathrm{T}}y$ as `XTZ`, `ZTX`, `ZTZ` and `ZTy`

Optimal GMM

```
W <- solve(eZZ_TSLS)
beta_GMM <- solve(XTZ %*% W %*% ZTX) %*% XTZ %*% W %*% ZTy
ehat_GMM <- y - X %*% beta_GMM
s2_GMM <- mean(ehat_GMM^2)
vbeta_GMM <- s2_GMM * solve(XTZ %*% solve(ZTZ) %*% ZTX)
eZZ_GMM <- matrix(0, nrow=Zcol, ncol=Zcol)
for (i in 1:n){
  eZZ_GMM <- eZZ_GMM + ehat_GMM[i]^2 * t(Z[i,,drop=F]) %*% Z[i,,drop=F]
}
vbeta_GMM_rob <- solve(XTZ %*% solve(eZZ_GMM) %*% ZTX)
GMM_results <- cbind(estimates = beta_GMM,
                     s.e. = sqrt(diag(vbeta_GMM)),
                     s.e.robust = sqrt(diag(vbeta_GMM_rob)))
```

Optimal GMM

```
# Computed as in previous slide
GMM_results %>% round(4)
```

```
          ln_earn   s.e. s.e.robust
const     -0.3690 0.1913     0.2086
age        0.0032 0.0008     0.0009
ln_tenure  0.1404 0.0096     0.0098
educ       0.2194 0.0130     0.0143
```

- s.e. is appropriate s.e. for opt. GMM under homoskedasticity
- estimates and s.e. differ from TSLS estimates and s.e., but both are “optimal”
- under heteroskedasticity, opt. GMM with robust s.e. preferred
- in this example, not much difference between TSLS and GMM

```
# Using `gmm` package
GMM_results_pkg <- gmm(
  ln_earn ~ age + ln_tenure + educ,
    ~ age + ln_tenure + feduc + meduc,
  data = dat, wmatrix = "optimal",
  vcov = "MDS", type = "twoStep")
summary(GMM_results_pkg)$coef[,1:2] %>%
  round(4)
```

```
            Estimate Std. Error
(Intercept)  -0.3690     0.2086
age           0.0032     0.0009
ln_tenure     0.1404     0.0098
educ          0.2194     0.0143
```

Testing Linear Restrictions

Session 8.6 Inference after GMM Estimation

We can do usual t- and F-tests after GMM estimation

the "Wald" statistic for jointly testing J number of linear hypotheses $H_0 : R\beta = r$ where R is $J \times (K+1)$ and r is $(K+1) \times 1$,

$$W = (R\hat{\beta}^{gmm} - r)^{\mathrm{T}}(R\,\widehat{Var}(\hat{\beta}^{gmm})R^{\mathrm{T}})^{-1}(R\hat{\beta}^{gmm} - r) \overset{a}{\sim} \chi^2_{(J)}$$

(This is the usual asymptotic version of F-test)

Testing Linear Restrictions

Continuing with our example

$$\ln earn = \beta_0 + \beta_1 age + \beta_2 \ln tenure + \beta_3 educ + \epsilon \,,$$

Test $H_0 : \beta_1 = 0$ and $\beta_2 = \beta_3$ or $(\beta_2 - \beta_3 = 0)$

```
R = matrix(c(0,1,0,0,0,0,1,-1), nrow=2, byrow=TRUE)
r = matrix(c(0,0), ncol=1)
b = beta_GMM
V = vbeta_GMM_rob
F_stat = t(R %*% b-r) %*% solve(R %*% V %*% t(R)) %*% (R %*% b-r)
cat("F:",F_stat,", p-value:", 1-pchisq(F_stat,nrow(R)))
```

```
F: 23.78748 , p-value: 6.833054e-06
```

Testing for Weak Instruments

Weak instruments (those poorly correlated with the endogenous regressors) will result in estimators with poor finite sample properties (high variance, possibly large finite sample biases)

To check for weak instruments, run the “first stage regression” (as though doing 2SLS manually), i.e.,

- Regress each endogenous regressor on all exogenous regressors and instruments
- Test for significance of the instruments in the first stage regressions
- F-statistics should be large (on the order of 20 or so)

Testing for Weak Instruments

The "First Stage Regression" in our example is

$$educ_i = \delta_0 + \delta_1 age_i + \delta_2 \ln tenure_i + \delta_3 feduc_i + \delta_4 meduc_i$$

and the hypothesis of invalid instrument is $H_0 : \delta_3 = \delta_4 = 0$.

Testing for Weak Instruments

```
mdl_firststage <- lm(educ ~ age+ln_tenure+feduc+meduc, data=dat)
linearHypothesis(mdl_firststage, c('feduc=0','meduc=0'),
                 vcov=vcovHC(mdl_firststage,type="HC1"))
```

```
Linear hypothesis test

Hypothesis:
feduc = 0
meduc = 0

Model 1: restricted model
Model 2: educ ~ age + ln_tenure + feduc + meduc

Note: Coefficient covariance matrix supplied.

  Res.Df Df      F    Pr(>F)
1   4943
2   4941  2 252.69 < 2.2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Tests of Overidentifying Restrictions

Recall

- GMM objective function: $J(W) = (Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})^{\mathrm{T}}W(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})$
- General GMM estimator: $\hat{\beta}_W^{gmm} = (X^{\mathrm{T}}ZWZ^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}ZWZ^{\mathrm{T}}y$

If $Z^{\mathrm{T}}X$ is square (the just-identified case) and invertible, then

- GMM estimator reduces to $\hat{\beta}_W^{gmm} = (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y$
- Obj. fn. becomes: $J(W) = (Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta}^{gmm})^{\mathrm{T}}W(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta}^{gmm}) = 0$

since $Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta}^{gmm} = Z^{\mathrm{T}}y - Z^{\mathrm{T}}X(Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y = 0$

Tests of Overidentifying Restrictions

- In the over-identified case, we have $J(W) > 0$
- However, if moment conditions **do** hold, then
  - sample moment conditions should hold approximately, and
  - $J(W)$ will be close to zero
  - It can be shown that

$$J \overset{a}{\sim} \chi^2_{(M-G)}$$

    $M - G$ is the number of "overidentifying restrictions" (number of excess instruments)
- "Test of Overidentified Restrictions" or J-test

Tests of Overidentifying Restrictions

- Significant J-stat indicates that one or more of the moment conditions do not hold
  - perhaps one (or more) of the presumed exog. regressors is actually endogenous, or
  - one of the instruments is not exogenous, or
  - some combination of these situations

In our example we have one extra moment restriction so we can carry out the J-test:

```
J <- t(ZTy - ZTX %*% beta_GMM) %*% solve(eZZ_GMM) %*% (ZTy - ZTX %*% beta_GMM)
Jpval <- 1-pchisq(J,ncol(Z)-ncol(X))
cat("J-stat:", J, "   p-value:", Jpval)
```

```
J-stat: 8.168    p-value: 0.004263589
```

The J-statistic indicates some misspecification

Testing Endogeneity

If we have valid instruments, we can test if one or more (or all) of the endogenous regressors can be treated as exogenous

In $Y = X\beta + \epsilon$ suppose

- $X = \begin{bmatrix} 1_n & X^k_{1*} & \dots & X^k_{K*} & X^g_{K+1,*} & \dots & X^g_{K+G,*} \end{bmatrix}$
- $Z = \begin{bmatrix} 1_n & X^k_{1*} & \dots & X^k_{K*} & Z_{K+1,*} & \dots & Z_{K+M,*} \end{bmatrix}$

The population moment conditions are $E(Z^{\mathrm{T}}\epsilon) = 0$

Is X^g_{K+1} actually exogenous? If so, then we can add it to the vector Z

- $\tilde{Z} = \begin{bmatrix} 1_n & X^k_{1*} & \dots & X^k_{K*} & X^g_{K+1,*} & Z_{K+1,*} & \dots & Z_{K+M,*} \end{bmatrix}$

The population moment conditions become $E(\tilde{Z}^{\mathrm{T}}\epsilon) = 0$

Testing Endogeneity

- Estimate the regression equation using instrument set Z, get J_Z
- Estimate the regression equation using instrument set $\tilde{Z}$, get $J_{\tilde{Z}}$
  - If in fact X^g_{K+1} is exogenous, both J-statistics should be close in value (with $J_{\tilde{Z}}$ larger than J_Z since more moment conditions are involved when using $\tilde{Z}$).
  - If X^g_{K+1} is in fact not exogenous, then there should be significant difference between J_Z and $J_{\tilde{Z}}$
- Under the null that $X^g_{K+1,i}$ is exogenous, the difference-in-J statistic is

$$C = J_{\tilde{Z}} - J_Z \overset{a}{\sim} \chi^2_Q$$

where Q is the no. of endog. vars. being tested for exogeneity (here $Q = 1$)

Testing Endogeneity

Continuing with our example, test if $educ$ can be treated as exogenous

One complication is that to ensure $C > 0$, the weight matrix used in computing $J(Z)$ has to be the appropriate submatrix of the weight matrix used in computing $J(\tilde{Z})$

This is implemented in the R code below

```
# C-Statistic, checking if "educ" is endogenous
#-- GMM when "educ" is exogenous
# set up matrices
Zr <- dat %>% select(c(const, age, ln_tenure, educ, feduc, meduc)) %>% as.matrix()
Zrcol <- dim(Zr)[2]
ZrTX <- t(Zr) %*% X ; XTZr <- t(X) %*% Zr ; ZrTZr <- t(Zr) %*% Zr ; ZrTy <- t(Zr) %*% y
```

Testing Endogeneity

```
#-- Get the necessary weight matrices
beta_TSLS_a <- solve(XTZr %*% solve(ZrTZr) %*% ZrTX) %*% XTZr %*% solve(ZrTZr) %*% ZrTy
ehat_TSLS_a <- y - X %*% beta_TSLS_a
eZZ_TSLS_a <- matrix(0, nrow=Zrcol, ncol=Zrcol)
for(i in 1:n){
  eZZ_TSLS_a <- eZZ_TSLS_a + ehat_TSLS_a[i]^2 * t(Zr[i,,drop=F]) %*% Zr[i,,drop=F]
}
W_a <- solve(eZZ_TSLS_a)
W_b <- W_a[-4,-4]  # fourth row and column associated with educ
#--GMM with educ exog
beta_GMM_a <- solve(XTZr %*% W_a %*% ZrTX) %*% XTZr %*% W_a %*% ZrTy
ehat_GMM_a <- y - X %*% beta_GMM_a
J_a <- t(t(Zr) %*% ehat_GMM_a) %*% W_a %*% t(Zr) %*% ehat_GMM_a
#--GMM with educ endo
beta_GMM_b <- solve(XTZ %*% W_b %*% ZTX) %*% XTZ %*% W_b %*% ZTy
ehat_GMM_b <- y - X %*% beta_GMM_b
J_b <- t(t(Z) %*% ehat_GMM_b) %*% W_b %*% t(Z) %*% ehat_GMM_b
```

Testing Endogeneity

```
C_stat <- J_a - J_b
cat("C:",C_stat,", p-value:", 1-pchisq(C_stat,ncol(Zr)-ncol(Z)))
```

```
C: 46.38576 , p-value: 9.711898e-12
```

We soundly reject the null that $educ$ is exogenous

Implementation in Stata

We mention that GMM is easily implemented in econometrics software like Stata

You could use the following Stata code

```
import delimited "data\earnings2019.csv", clear
gen ln_earn = ln(earnings)
gen ln_tenure = ln(tenure)
ivregress gmm ln_earn (educ = feduc meduc) age ln_tenure
estat firststage
estat overid
estat endog
```

Roadmap

- (Previous) Session 1: Statistics Review
- (Previous) Session 2: Simple Linear Regression
- (Previous) Session 3: Estimator Standard Errors; Multiple Linear Regression
- (Previous) Session 4: Matrix Algebra
- (Previous) Session 5: OLS using Matrix Algebra
- (Previous) Session 6: Hypothesis Testing
- (Previous) Session 7: Prediction
- **This Session 8: Instrumental Variable Regression**
- *Next Session 9: Logistic and Other Regressions*
- Session 10: Panel Data Regressions
- Session 11: Introduction to Time Series
- Session 12: Time Series Regressions