Session 7

Instrumental Variables and GMM

Anthony Tay

This Version: 13 Oct 2025

Session 7

- “Endogeneity” problem
- Instrumental variables, method of moments, two stage least squares (2SLS)
  - 1 regressor 1 variable Case, without matrix algebra
  - Simultaneity bias example, unavailable control example, others
  - 1 regressor 1 variable Case, with matrix algebra
- Extensions to regressions with endogenous and exogenous regressors, multiple IVs
  - 2SLS, optimal generalized method of moments (GMM)

The Endogeneity Problem (Omitted Variables)

If $E(Y \mid X) = \beta_0 + \beta_1 X$, then OLS estimation of SLR

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

provides unbiased estimator of β_1, $E(\hat{\beta}_1^{ols}) = \beta_1$

- Key reason: $E(\epsilon \mid X) = 0$

But perhaps $E(Y \mid X)$ is not what you want to estimate

- E.g. You want to estimate $E(Y \mid X, W) = \beta_0 + \beta_1 X + \beta_2 W$ with interest in β_1
- "Effect of X on Y controlling for W"

The Endogeneity Problem (Omitted Variables)

If $E(Y \mid X, W) = \beta_0 + \beta_1 X + \beta_2 W$, we can write

$$Y = \beta_0 + \beta_1 X + \beta_2 W + u\,,\;\; E(u \mid X, W) = 0$$

But then

$$Y = \beta_0 + \beta_1 X + \epsilon\,,\;\; \epsilon = \beta_2 W + u$$

If $Cov(W, X) \neq 0$, then

- $E(\epsilon \mid X)$ will be a function of X
- $\hat{\beta}_1^{ols}$ will be biased for β_1

The Endogeneity Problem (Omitted Variables)

If we assume $E(W \mid X) = \delta_0 + \delta_1 X$, then

$$\delta_0 = E(W) - \delta_1 E(X) \quad \text{and} \quad \delta_1 = \frac{Cov(W, X)}{Var(X)}$$

so we have

$$E(Y \mid X) = (\beta_0 + \beta_2 \delta_0) + \left(\beta_1 + \beta_2 \frac{Cov(W, X)}{Var(X)}\right) X$$

$\widehat{\beta}_1^{ols}$ from simple linear regression of Y on X will provide unbiased estimator for $\beta_1 + \beta_2 Cov(W, X)/Var(X)$, not β_1

Solution is multiple linear regression of Y on X and W, but **what if W is unobservable or unavailable?**

The Endogeneity Problem (Simultaneity Bias)

Consider a demand-and-supply example

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

Suppose you want to estimate market demand

Market Clearing $\Longrightarrow \quad \delta_0 + \delta_1 P_t + \epsilon_t^d = \alpha_0 + \alpha_1 P_t + \epsilon_t^s$

Solving:

$$P_t = \frac{\alpha_0 - \delta_0}{\delta_1 - \alpha_1} + \frac{\epsilon_t^s - \epsilon_t^d}{\delta_1 - \alpha_1}$$
$$Q_t = \left(\delta_0 + \delta_1 \frac{\alpha_0 - \delta_0}{\delta_1 - \alpha_1}\right) + \frac{\delta_1 \epsilon_t^s - \alpha_1 \epsilon_t^d}{\delta_1 - \alpha_1}.$$

The Endogeneity Problem (Simultaneity Bias)

This implies

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

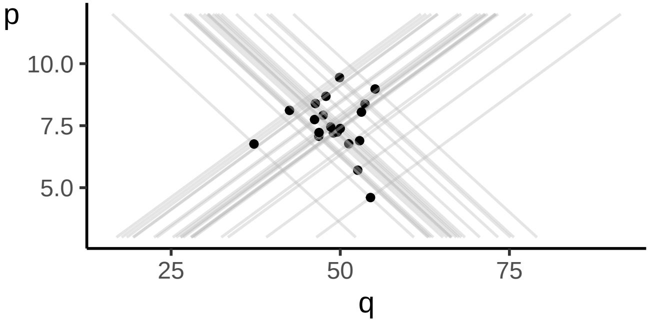
A regression of $Q_t = \beta_0 + \beta_1 P_t + \epsilon_t$, we will get

$$\hat{\beta}_1 \xrightarrow{p} \frac{Cov(Q_t, P_t)}{Var(P_t)} = \frac{\delta_1 \sigma_s^2 + \alpha_1 \sigma_d^2}{\sigma_s^2 + \sigma_d^2} \tag{1}$$

which is neither the price elasticity of demand nor the price elasticity of supply, but a linear combination of the two.

What is the conditional mean $E(Q_t \mid P_t)$?

The Endogeneity Problem (Simultaneity Bias)



Data looks neither like demand or supply curve

$Q_t^d = \delta_0 + \delta_1 P_t + \epsilon_t^d$ but demand shock correlated with prices

The Endogeneity Problem (Measurement Error)

Suppose $Y = \beta_0 + \beta_1 X + \epsilon$

X is only observed with error, i.e., you observe $X^* = X + u$

Assuming measurement error u is independent of X, we have

$$\begin{aligned} Y &= \beta_0 + \beta_1 X + \epsilon \\ &= \beta_0 + \beta_1 (X^* - u) + \epsilon \\ &= \beta_0 + \beta_1 X^* + (\epsilon - \beta_1 u) \\ &= \beta_0 + \beta_1 X^* + v \end{aligned}$$

In regression of Y on X^*, assumption $E(v \mid X^*) = 0$ does not hold (depends on X)

Instrumental Variables and IV Estimator

Suppose

$$Y = \beta_0 + \beta_1 X + \epsilon\,, \;\; Cov(X, \epsilon) \neq 0$$

Suppose there is another variable Z such that

- $Cov(X, Z) \neq 0$
- $Cov(Z, \epsilon) = 0$ (Note: this means Z is **not** an omitted variable)

Such a variable is called an instrumental variable

Can be used to provide consistent estimator for β_1

Instrumental Variables and IV Estimator

Given iid sample $\{X_i, Y_i, Z_i\}_{i=1}^n$, the IV estimator

$$\beta_1^{iv} = \frac{\sum_{i=1}^n (Y_i - \overline{Y}) Z_i}{\sum_{i=1}^n (X_i - \overline{X}) Z_i} = \frac{\sum_{i=1}^n (Z_i - \overline{Z}) Y_i}{\sum_{i=1}^n (Z_i - \overline{Z}) X_i}$$

is consistent for β_1

$$\beta_1^{iv} = \frac{\sum_{i=1}^n (Z_i - \overline{Z}) Y_i}{\sum_{i=1}^n (Z_i - \overline{Z}) X_i} = \beta_1 + \frac{\frac{1}{n}\sum_{i=1}^n (Z_i - \overline{Z}) \epsilon_i}{\frac{1}{n}\sum_{i=1}^n (X_i - \overline{X}) X_i} \xrightarrow{p} \beta_1 + \frac{\overbrace{Cov(Z, \epsilon)}^{=0}}{\underbrace{Cov(X, Z)}_{\neq 0}} = \beta_1$$

Instrumental Variables and IV Estimator

A Method of Moments Perspective

$Y = \beta_0 + \beta_1 X + \epsilon,\ Cov(\epsilon, Z) = 0$

Assume $E(\epsilon \mid Z) = 0$ and $E(\epsilon) = 0$ in population

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

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

If sample is representative of population

Choose $\widehat{\beta}_0$ and $\widehat{\beta}_1$ so that corresponding *sample moments* hold

Instrumental Variables and IV Estimator

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

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

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

Instrumental Variables and IV Estimator

Substituting into [B] gives

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

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

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

Solving for $\hat{\beta}_0^{mm}$ and $\hat{\beta}_1^{mm}$ gives

$$\hat{\beta}_0^{mm} = \overline{Y} - \hat{\beta}_1^{mm}\overline{X} \quad \text{and} \quad \hat{\beta}_1^{mm} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})Z_i}{\sum_{i=1}^{n}(X_i - \overline{X})Z_i}$$

Instrumental Variables and IV Estimator

A Two-Stage Least Squares (2SLS) Perspective:

Given X_i and Z_i, we can decompose X_i into two uncorrelated parts by regressing X_i on Z_i by OLS:

$$X_i = \hat{\delta}_0 + \hat{\delta}_1 Z_i + r_{i,\,x|z} = \hat{X}_i + r_{i,\,x|z}$$

- Since $\hat{X}_i$ is a linear function of Z_i, and Z_i is uncorrelated with ϵ_i, so $\hat{X}_i$ is uncorrelated with ϵ_i
- All movements in X_i that are correlated with ϵ_i are "concentrated" into $r_{i,x|z}$
- Think of $\hat{X}_i$ as what's left after filtering out movements in X_i that are correlated with ϵ_i

Instrumental Variables and IV Estimator

Idea: use only the movements in X_i that are uncorrelated with ϵ_i when determining the effect of X_i on Y_i, i.e.,

- (Stage 1): Regress X_i on Z_i, compute $\hat{X}_i$
- (Stage 2): Regress Y_i on $\hat{X}_i$ instead of X_i

$$\hat{\beta}_1^{2sls} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})(\hat{X}_i - \overline{\hat{X}})}{\sum_{i=1}^{n}(\hat{X}_i - \overline{\hat{X}})^2}$$

Instrumental Variables and IV Estimator

This turns out to be equivalent to the IV/MM estimator

- Since $\hat{X}_i = \hat{\delta}_0 + \hat{\delta}_1 Z_i$, we have

$$\overline{\hat{X}} = \hat{\delta}_0 + \hat{\delta}_1 \overline{Z} \quad \text{and} \quad \hat{X}_i - \overline{\hat{X}} = \hat{\delta}_1 (Z_i - \overline{Z})$$

Squaring and summing over i gives

$$\sum_{i=1}^{n} (\hat{X}_i - \overline{\hat{X}})^2 = \hat{\delta}_1^2 \sum_{i=1}^{n} (Z_i - \overline{Z})^2$$

Instrumental Variables and IV Estimator

Substituting into 2sls estimator gives

$$\hat{\beta}_1^{2sls} = \frac{\hat{\delta}_1 \sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\hat{\delta}_1^2 \sum_{i=1}^n (Z_i - \overline{Z})^2} = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\hat{\delta}_1 \sum_{i=1}^n (Z_i - \overline{Z})^2}$$

Since $\hat{\delta}_1$ is the OLS estimator for Z_i coefficient in a regression of X_i and Z_i, we have

$$\hat{\delta}_1 = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})}{\sum_{i=1}^n (Z_i - \overline{Z})^2}$$

Therefore

$$\hat{\beta}_1^{2sls} = \frac{\sum_{i=1}^n (Z_i - \overline{Z})(Y_i - \overline{Y})}{\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X})}$$

Instrumental Variables and IV Estimator

Where do instruments come from?

E.g., Suppose there is observable R that shifts the supply function but not the demand function, i.e.,

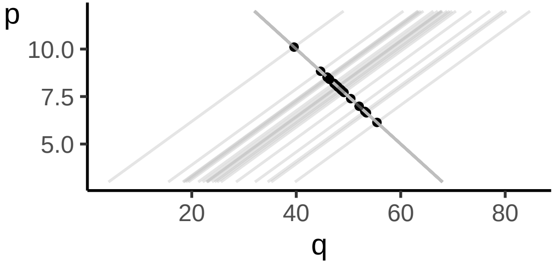
$$
\begin{aligned}
Q^d &= \delta_0 + \delta_1 P + \epsilon^d && \text{(Demand Eq } \delta_1 < 0\text{)} \\
Q^s &= \alpha_0 + \alpha_1 P + \alpha_2 R + \epsilon^s && \text{(Supply Eq } \alpha_1 > 0\text{)} \\
Q^s &= Q^d && \text{(Market Clearing)}
\end{aligned}
$$

- R does not shift demand, $Cov(R, \epsilon_t^d) = 0$
- R shifts supply, supply shift changes prices, $Cov(R, P) \neq 0$

R is a valid instrument for P in the demand equation

Instrumental Variables and IV Estimator

As a mental experiment, imagine "shutting down" demand and supply shocks, allowing only R to change



- Variation in R helps to "identify" the demand function

Instrumental Variables and IV Estimator

- In practice, cannot "shut down" the demand shocks
- How do we isolate variation in P due to R only?

The two stage least squares perspective shows that we can regress P on R, and then regress Q on $\hat{P}$, which contains variation in P due to R only

Others: Draft lottery and military service, parent's years of education and child's years of education, years of schooling and distance to nearest college

Instrumental Variables and IV Estimator

Note that IV/MM/2SLS estimators are biased:

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

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

- Since ϵ is correlated with X, $E(\epsilon_i \mid X_1, \ldots, X_n, Z_1, \ldots, Z_n)$ will be some function of X_i, $i = 1, \ldots, n$, and cannot come out of the summation

Instrumental Variables and IV Estimator

Trade-off: Larger standard errors for consistency

Suppose $Var(\epsilon_i \mid X_1, \dots, X_n, Z_1, \dots, Z_n) = \sigma^2$. Then

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

$$= \frac{\sigma^2}{\sum_{i=1}^n (X_i - \overline{X})^2 \left(\frac{(\sum_{i=1}^n (Z_i - \overline{Z})(X_i - \overline{X}))^2}{\sum_{i=1}^n (Z_i - \overline{Z})^2 \sum_{i=1}^n (X_i - \overline{X})^2} \right)}$$

$$= \frac{\sigma^2}{R^2_{X|Z} \sum_{i=1}^n (X_i - \overline{X})^2}$$ where $R^2_{X|Z}$ is R^2 from regression of X_i on Z_i

Instrumental Variables and IV Estimator

After IV estimation

- Report $\widehat{Y} = \widehat{\beta}_0^{iv} + \widehat{\beta}_1^{iv} X$, not $\widehat{Y} = \widehat{\beta}_0^{iv} + \widehat{\beta}_1^{iv} \hat{X}$
- IV residuals are

$$\widehat{\epsilon}_{i,iv} = Y_i - \widehat{\beta}_0^{iv} - \widehat{\beta}_1^{iv} X_i \,, \;\; i = 1, \ldots, n$$

  not $Y_i - \widehat{\beta}_0^{iv} - \widehat{\beta}_1^{iv} \hat{X}$
- R^2 is

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

  which will be less than OLS R^2

IV estimation expressed with matrix algebra

IV estimation of $Y_i = \beta_0 + \beta_1 X_{i1} + \epsilon_i$ with instrument Z_{i1}, $i = 1, \dots, n$

Define

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

Can show

$$\hat{\beta}^{iv} = (Z^{\mathrm{T}} X)^{-1} Z^{\mathrm{T}} y \iff \begin{cases} \hat{\beta}_0^{iv} = \overline{Y} - \hat{\beta}_1^{iv} \overline{X} \\ \hat{\beta}_1^{iv} = \dfrac{\sum_{i=1}^{n} (Z_i - \overline{Z}) Y_i}{\sum_{i=1}^{n} (Z_i - \overline{Z}) X_i} \end{cases}$$

Using Matrix Algebra

MM approach started with the 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, we can write this as

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

Assuming $Z^{\mathrm{T}}X$ is invertible, we have

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

Using Matrix Algebra

The 2SLS approach: In step 1, regress X on Z

- As X is $n \times 2$, this means regressing each column of X on Z:

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

  We can put into this one single matrix:

$$\begin{bmatrix} i_n & X_{*1} \end{bmatrix} = Z \begin{bmatrix} b_0 & b_1 \end{bmatrix} + \begin{bmatrix} u_{*0} & u_{*1} \end{bmatrix} \quad \text{or} \quad X = ZB + U$$

- We have $\hat{B} = (Z^{\mathrm{T}}Z)^{-1}Z^{\mathrm{T}}X$
  - What is the dimension of $\hat{B}$? What are its context?
- The fitted value from this step is

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

Using Matrix Algebra

In step 2, regress y on $\hat{X}$, which gives

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

Using Matrix Algebra

Showing consistency:

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

since $\frac{1}{n}Z^{\mathrm{T}}\epsilon \xrightarrow{p} 0_{2\times 1}$ and we assume $\frac{1}{n}Z^{\mathrm{T}}X$ converges to a non-singular matrix

Using Matrix Algebra

For asymptotically valid variance-covariance matrices

- homoskedastic errors:

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

- The heteroskedasticity-robust version is:

$$\widehat{Var}_{HC0}(\hat{\beta}_{iv}) = (Z^{\mathrm{T}}X)^{-1}\left(\sum_{i=1}^{n}\hat{\epsilon}_{i,iv}^{2}Z_{i*}^{\mathrm{T}}Z_{i*}\right)(X^{\mathrm{T}}Z)^{-1}$$

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

1 exog, 1 endog, 2 instruments

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

where

- X_1^k is exogenous (not correlated with the noise term)
- X_2^g is endogenous (correlated with the noise term), and
- there exists Z_2 and Z_3 both correlated with X_2^g and uncorrelated with ϵ.

Example of "overidentification"

1 exog, 1 endog, 2 instruments

Take MM approach?

Population satisfies $E(\epsilon) = 0$, $Cov(X_1^k, \epsilon) = 0$, $Cov(Z_2, \epsilon) = 0$, $Cov(Z_3, \epsilon) = 0$, i,e.,

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

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

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

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

1 exog, 1 endog, 2 instruments

MM approach

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

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

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

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

But how do we solve four equations in three unknowns?

1 exog, 1 endog, 2 instruments

Solution:

Choose $\widehat{\beta}_0^{mm}$, $\widehat{\beta}_1^{mm}$, $\widehat{\beta}_2^{mm}$ to minimize the sum of squared moments:

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

1 exog, 1 endog, 2 instruments

Define

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

sample moments (dropping the $1/n$) can be written as

$$\underbrace{Z^{\mathrm{T}}}_{4\times n} \underbrace{(y - X\hat{\beta}^{mm})}_{n\times 1} = \underbrace{Z^{\mathrm{T}}}_{4\times n} \underbrace{y}_{n\times 1} - \underbrace{Z^{\mathrm{T}}}_{4\times n} \underbrace{X}_{n\times 3} \underbrace{\hat{\beta}^{mm}}_{3\times 1} = \underbrace{0}_{4\times 1}$$

1 exog, 1 endog, 2 instruments

We 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

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

which requires that the 3×3 matrix $X^{\mathrm{T}}ZZ^{\mathrm{T}}X$ be invertible

1 exog, 1 endog, 2 instruments

To show consistency of the MM estimator:

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

which 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.

1 exog, 1 endog, 2 instruments

The variance-covariance matrix

- under homoskedasticity is:

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

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

- heteroskedasticity-robust version

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

1 exog, 1 endog, 2 instruments

What happens if we apply these formulas to the **just-identified** case

Suppose we only have Z_2 to instrument for X_2^g

Then $Z^{\mathrm{T}}X$ is square 3×3, and we get

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

1 exog, 1 endog, 2 instruments

The corresponding variance-covariance matrices

- under homoskedasticity reduces to:

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

- heteroskedasticity-robust version reduces to

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

1 exog, 1 endog, 2 instruments

2SLS approach for this example

Stage 1, regress X on Z.

- coefficient 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 a vector of 1's, the second column is X_{i1}^{k}, and the third column is $\hat{X}_{i2}^{g}$ obtained from a regression of X_{i2}^{g} on intercept, X_{i1}^{k}, Z_{i2} and Z_{i3}

1 exog, 1 endog, 2 instruments

Stage 2, we regress y on $\hat{X}$, which gives

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

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

1 exog, 1 endog, 2 instruments

The proof of consistency is as follows:

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

1 exog, 1 endog, 2 instruments

The variance-covariance matrices are (details of proofs omitted)

- under 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 case:

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

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

1 exog, 1 endog, 2 instruments

Note that $\widehat{\beta}^{2SLS} \neq \widehat{\beta}^{mm}$

In the **just-identified** case, where the number of endogenous variables is equal to the number of instruments, $Z^{\mathrm{T}}X$ is square, the 2SLS estimator reduces to:

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

1 exog, 1 endog, 2 instruments

All formulas and results continue to apply to general case

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

- Z is $n \times (K + M + 1)$ containing column of ones, all Z (incl. X^k variables)
- X is $n \times (K + G + 1)$ containing column of ones, all X variables (exog and endog)

Example

Use data from `earnings2019.csv` to estimate the equation

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

- Measure of $ability$ is unavailable and omitted, resulting in endogeneity of $educ$
- Suppose age and $\ln tenure$ are exogenous
- Suppose $feduc$ and $meduc$ are valid instruments

Example

OLS results

```
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

OLS results

```
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

Using the MM formulas derived in this section:

```
## 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]}
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)))
```

Example

Using the 2SLS formulas derived in this section:

```
#--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

```
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
```

```
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

We can get the 2SLS estimates from the 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)
coeftest(mdl2_iv,vcov=vcovHC(mdl2_iv,type="HC0"))  %>% 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

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 ***
```

GMM Estimation

The GMM estimator minimizes *weighted* sum of squared moments, i.e.,

$$\hat{\beta}_W^{gmm} = \text{argmin}_{\hat{\beta}} \underbrace{(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})^{\mathrm{T}}W(Z^{\mathrm{T}}y - Z^{\mathrm{T}}X\hat{\beta})}_{"J(W)"} \tag{3}$$

where W is some symmetric positive-definite weight matrix (may change with n and may be data dependent)

- We will assume W is known and fixed for the moment
- 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).

GMM Estimation

Minimizing $J(W)$ gives

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

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

Consistency:

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

GMM Estimation

The variance-covariance matrix

- under homoskedasticity is:

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

- heteroskedasticity-robust version is

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

GMM Estimation

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

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

Usually implemented with a two-step approach:

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

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

GMM Estimation

The variance-covariance matrix

- under homoskedasticity is

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

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

- heteroskedasticity-robust version is

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

GMM Estimation

The form of the variance of the optimal GMM estimator *under 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).

GMM Estimation

Code below uses GMM package to obtain optimal GMM with heteroskedasticity-robust standard errors

```
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)$coef[,1:2] %>% round(4)
summary(GMM_results_pkg)$stest
```

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

 ##  J-Test: degrees of freedom is 1  ##

                J-test     P-value
Test E(g)=0:    8.1670089  0.0042659
```

Inference after GMM Estimation

Testing Linear Restrictions

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

"Wald" statistic for jointly testing J number of linear hypotheses, $H_0 : \mathcal{R}\beta = r_0$, where $\mathcal{R}$ is $J \times K$ and r_0 is $K \times 1$, is

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

This is the asymptotic the chi-square test (JF)

Inference after GMM Estimation

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)

- 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)

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 $H_0 : \delta_3 = \delta_4 = 0$

Inference after GMM Estimation

```
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:
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
```

It appears that $feduc_i$ and $meduc_i$ are not weak instruments.

Inference after GMM Estimation

Tests of Overidentifying Restrictions

Recall that the

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

If $Z^{\mathrm{T}}X$ is square (the just-identified case) and invertible, then GMM reduces to $\hat{\beta}_W^{gmm} = (Z^{\mathrm{T}}X)^{-1}Z^{\mathrm{T}}y$. Objective function 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

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

Inference after GMM Estimation

In the over-identified case, we will have $J(W) > 0$ in general

However, if moment conditions **do** hold, then sample moment conditions should hold approximately, and $J(W)$ will still be close to zero. It can be shown then 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
- Significant J-stat indicates that one or more of the moment conditions do not hold
  - perhaps one (or more) of the presumed exogenous regressors is actually endogenous
  - perhaps one of the instruments is not exogenous, or
  - some combination of these situations.

Inference after GMM Estimation

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 the regression $Y = X\beta + \epsilon$ 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 $E(Z^{\mathrm{T}}\epsilon) = 0$

If X^g_{K+1} is in fact not endogenous, we can add it to the vector Z, i.e.,

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

and the moment condition $E(\tilde{Z}^{\mathrm{T}}\epsilon) = 0$ will still hold

Inference after GMM Estimation

The idea of the test then is

- Estimate the regression equation using instrument set Z, get J_Z
- Estimate the regression equation using instrument set $\tilde{Z}$, get $J_{\tilde{Z}}$
- X^g_{K+1} is exogenous $\Rightarrow J_Z \approx J_{\tilde{Z}}$ ($J_{\tilde{Z}} > J_Z$ since $\tilde{Z}$ has more moment conditions)
- X^g_{K+1} is in fact not exogenous $\Rightarrow J_{\tilde{Z}} \gg J(Z)$.

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 # of endogenous variables being tested for exogeneity (here $Q = 1$)

Inference after GMM Estimation

```
GMM1 <- gmm(
  ln_earn ~ age + ln_tenure + educ, ~ age + ln_tenure + feduc + meduc,
  data = dat, wmatrix = "optimal", vcov = "MDS", type = "twoStep")
GMM2 <- gmm(
  ln_earn ~ age + ln_tenure + educ, ~ age + ln_tenure + educ + feduc + meduc,
  data = dat, wmatrix = "optimal", vcov = "MDS", type = "twoStep")
JR <- summary(GMM1)$stest[[2]][1]
JU <- summary(GMM2)$stest[[2]][1]
Jdiff <- JU-JR
pval <- 1 - pchisq(Jdiff, 1)
cat("Difference-in-J Test:", Jdiff, " p-val", pval)
```

```
Difference-in-J Test: 47.48337  p-val 5.547229e-12
```