

GMM Method (Single-equation)

Pongsa Pornchaiwiseskul
Faculty of Economics
Chulalongkorn University

Stochastic X (1)

Given that, for some k , X_k is random

$$\begin{aligned}\text{COV}(X_k, \varepsilon) &= E((X_k - \mu_k)\varepsilon) \\ &= E(X_k \varepsilon) - \mu_k E(\varepsilon) \\ &= E(X_k \varepsilon)\end{aligned}$$

Covered Topics

- Violation of $C(\mathbf{X}, \varepsilon) = 0$
- Instrument Variable (IV) Method
- Generalized Method of Moments (GMM)

Stochastic X (2)

If some or all X 's are random, additional assumptions about \mathbf{X} are needed. One of them is

$$E(X_{ki} \varepsilon_i) = 0 \quad \forall k, i \implies \hat{\beta} \text{ is consistent}$$

or $C(X_{ki}, \varepsilon_i) = 0$

Violation of $C(X, \varepsilon) = 0$

OLS estimator is inconsistent (invalid).

Why?

$$\begin{aligned}\hat{\beta}_{OLS} &= (X^T X)^{-1} X^T Y \\ &= (X^T X)^{-1} X^T (X\beta + \varepsilon) \\ &= \beta + (X^T X)^{-1} X^T \varepsilon\end{aligned}$$

Violation of $C(X, \varepsilon) = 0$

Take conditional variance on both sides.

$$\begin{aligned}V(\hat{\beta}_{OLS} | X) &= (X^T X)^{-1} X^T V(\varepsilon | X) X (X^T X)^{-1} \\ &= \sigma^2 (X^T X)^{-1}\end{aligned}$$

$$V(\sqrt{n}\hat{\beta}_{OLS} | X) = \sigma^2 \left(\frac{X^T X}{n} \right)^{-1}$$

Unconditional variance=??

Violation of $C(X, \varepsilon) = 0$

Take expectation on both sides.

$$\begin{aligned}E(\hat{\beta}_{OLS}) &= \beta + E\left((X^T X)^{-1} X^T \varepsilon \right) \\ &\neq \beta\end{aligned}$$

$\hat{\beta}_{OLS}$ is biased in general

Proxy of X_k

Definition

X'_k is proxy of X_k if

$$X_k = X'_k + \xi_k \text{ and } E(\xi_k) = 0$$

$$\text{Cov}(X'_k, \varepsilon) = 0 \text{ and } \text{Cov}(X'_k, \xi_k) = 0$$

$$\implies E(X'_k) = E(X_k)$$

$$X'_{ki} + \xi_{ki}$$

Regression Model w/ violation

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X_{Ki} + \varepsilon_i$$

$Cov(X_K, \varepsilon) \neq 0 \Rightarrow OLS$ is inconsistent

Regression Model w/o violation

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K X'_{Ki} + \varepsilon'_i$$

where $\varepsilon'_i = \varepsilon_i + \beta_K \xi_{Ki}$

$Cov(X'_K, \varepsilon') = 0 \Rightarrow OLS$ is consistent

How to get Proxy of X_k (2)

Note that

- 1) it is equivalent to splitting the variable X_k into 2 parts, one part ($Z\gamma$) with no correlation with the original error term (ε) and the other part (ξ) with the correlation.

How to get Proxy of X_k (1)

Assuming that the set of variables

Z are uncorrelated with ε , we

can define the proxy as follows:

$$X_{ki} = Z_i \gamma + \xi_{ki}$$

where $Z\gamma$ is a proxy of X_k

How to get Proxy of X_k (3)

Note that

- 2) $Z\gamma$ will be a good proxy if it can explain most variation of X_k .
Need high R^2 .
- 3) Z is called the set (vector) of instrument variables.

Instrument Variables (1)

Candidates

- all X's which is non-random or random but uncorrelated with \mathcal{E} .
- any variable outside the regression model which is uncorrelated to \mathcal{E} .

IV Estimation (1)

Steps

- run OLS regression for

$$X_{ki} = \mathbf{Z}_i \boldsymbol{\gamma} + \xi_{ki}$$

- get the fitted values for X_k

$$\hat{X}_{ki} = \mathbf{Z}_i \hat{\boldsymbol{\gamma}}$$

$$\text{where } \hat{\boldsymbol{\gamma}} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X}_K$$

Instrument Variables (2)

Choices of IV's

- availability of data
- underlying theories
- pure assumption

IV Estimation (2)

- 3) use them as the proxy for X_{ki}

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \dots + \beta_K \hat{X}_{Ki} + \varepsilon'_i$$

- 4) run OLS on the above regression model to get the estimate for $\boldsymbol{\beta}$

IV Estimation (3)

Matrix Notation

$$\begin{aligned} \mathbf{Y} &= \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \\ &= \mathbf{U}\boldsymbol{\alpha} + \mathbf{V}\boldsymbol{\delta} + \boldsymbol{\varepsilon} \end{aligned}$$

where $\mathbf{X} = [\mathbf{U} \ \mathbf{V}]$ and $\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\alpha} \\ \boldsymbol{\delta} \end{bmatrix}$

$$\text{Cov}(\mathbf{U}, \boldsymbol{\varepsilon}) = \mathbf{0} \text{ but } \text{Cov}(\mathbf{V}, \boldsymbol{\varepsilon}) \neq \mathbf{0}$$

IV Estimation (5)

IV Estimator

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{IV} &= (\mathbf{X}^T \mathbf{A}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T \mathbf{Y} \\ &= (\mathbf{X}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T \mathbf{Y} \\ &= (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T \mathbf{Y} \\ &= (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T (\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{\beta} + (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T \boldsymbol{\varepsilon} \end{aligned}$$

Note that $\mathbf{A}^T \mathbf{A} = \mathbf{A}^2 = \mathbf{A}$ ▲ Asymptotically zero

IV Estimation (4)

Proxy of \mathbf{V}

$$\hat{\mathbf{V}} = \mathbf{Z}\hat{\boldsymbol{\gamma}} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{V}$$

$$\mathbf{Y} = \mathbf{U}\boldsymbol{\alpha} + \mathbf{A}\mathbf{V}\boldsymbol{\delta} + \boldsymbol{\varepsilon}'$$

where $\mathbf{A} = \mathbf{Z}(\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T = \mathbf{A}^T$

Note that $\mathbf{U} = \hat{\mathbf{U}} = \mathbf{A}\mathbf{U}$ because \mathbf{U} is a subset of \mathbf{Z}

$$\mathbf{Y} = \mathbf{A}\mathbf{U}\boldsymbol{\alpha} + \mathbf{A}\mathbf{V}\boldsymbol{\delta} + \boldsymbol{\varepsilon}'$$

$$\mathbf{Y} = \mathbf{A}\mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}'$$

IV Estimation (6)

Take variance on both sides

$$\begin{aligned} \mathbf{V}(\hat{\boldsymbol{\beta}}_{IV}) &= \sigma^2 (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T \mathbf{A} \mathbf{X} (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \\ &= \sigma^2 (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1} \end{aligned}$$

$$\hat{\mathbf{V}}(\hat{\boldsymbol{\beta}}_{IV}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{A} \mathbf{X})^{-1}$$

where

$$\hat{\sigma}^2 = \frac{1}{n - K} (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{IV})^T (\mathbf{Y} - \mathbf{X}\hat{\boldsymbol{\beta}}_{IV})$$

IV Estimation (7)

Equivalence OLS

$$\mathbf{AY} = \mathbf{AX}\boldsymbol{\beta} + \mathbf{A}\boldsymbol{\varepsilon}$$

$$\hat{\boldsymbol{\beta}}_{EQOLS} = (\mathbf{X}^T \mathbf{A}^T \mathbf{A} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{A}^T \mathbf{A} \mathbf{Y} = \hat{\boldsymbol{\beta}}_{IV}$$

In EViews, IV method is the same as Two-stage Least Square (2SLS)

X's not in Instrument list have non-zero correlation with error terms

IV Estimation (9)

The model with proxy or transformed with matrix A may not have one.

- 2) IV does not need normality of the error term.
- 3) a constant term is also an IV

IV Estimation (8)

Note that

- 1) 2SLS estimation may yield negative R-squared even if there is a constant term or equivalent in the original regression equation. Why?

IV Estimation (10)

- 4) If Z=all X plus other IV's from outside, IV method is exactly the same as OLS.

With $\hat{\mathbf{X}} = \mathbf{A}\mathbf{X} = \mathbf{Z}$

$$\begin{aligned} \hat{\boldsymbol{\beta}}_{IV} &= (\hat{\mathbf{X}}^T \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}^T \mathbf{Y} \\ &= (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{Y} = \hat{\boldsymbol{\beta}}_{OLS} \end{aligned}$$

Weighted 2SLS (1)

Unweighted Model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{v} = \mathbf{U}\boldsymbol{\alpha} + \mathbf{V}\boldsymbol{\gamma} + \mathbf{v}$$

where $\mathbf{v}_i = \frac{\varepsilon_i}{w_i}$ or $V(\mathbf{v}_i) = \left(\frac{\sigma}{w_i}\right)^2$

and $Cov(\mathbf{V}, \mathbf{v}) \neq \mathbf{0}$

Weighted 2SLS (3)

$$\hat{\boldsymbol{\beta}}_{WIV} = (\mathbf{X}^T \mathbf{W}^T \mathbf{A} \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}^T \mathbf{A} \mathbf{W} \mathbf{Y}$$

where $\mathbf{A} = \mathbf{W} \mathbf{Z} (\mathbf{Z}^T \mathbf{W}^T \mathbf{W} \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{W}^T$

$$\hat{V}(\hat{\boldsymbol{\beta}}_{WIV}) = \hat{\sigma}^2 (\mathbf{X}^T \mathbf{W}^T \mathbf{A} \mathbf{W} \mathbf{X})^{-1}$$

$$\hat{\sigma}^2 = \frac{1}{n - K} (\mathbf{W} \mathbf{Y} - \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}_{WIV})^T (\mathbf{W} \mathbf{Y} - \mathbf{W} \mathbf{X} \hat{\boldsymbol{\beta}}_{WIV})$$

Weighted 2SLS (2)

Weighted Model

$$\mathbf{W} \mathbf{Y} = \mathbf{W} \mathbf{U} \boldsymbol{\alpha} + \mathbf{W} \mathbf{V} \boldsymbol{\gamma} + \mathbf{W} \mathbf{v} = \boldsymbol{\varepsilon}$$

where $\mathbf{W} = \begin{bmatrix} w_1 & 0 & \cdots & 0 \\ 0 & w_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & w_n \end{bmatrix}$

GMM (1)

As an improvement over IV

- IV estimation method does not fully utilize the assumption or knowledge of zero correlation between \mathbf{Z} and the error term ($\boldsymbol{\varepsilon}$).

GMM (2)

- IV method employs only one proxy from the whole set of Instrument Variables.
- A different subset of Z's can yield a different proxy. We don't have to use the whole set of Z's. We can give different weight to each Z.

GMM (4)

For each IV Z_m , it is expected that $E(Z_m \epsilon) = 0$ for all m. That is,

$$E(Z_1(Y - X\beta)) = 0$$

$$E(Z_2(Y - X\beta)) = 0$$

⋮

$$E(Z_M(Y - X\beta)) = 0$$

where M = #of IV's

$$\mathbf{Z} = [Z_1, Z_2, \dots, Z_M]$$



**Moment
Conditions**

GMM (3)

Concept

- Giving different weights to each Z is equivalent to mixing different subset of Z's
- For each Z, there will be associated matrix A. Need at least K Instrument Variables

GMM (5)

Sample Analogy

$$\hat{\beta}_{GMM} = \arg \min \left(w_1 \mathbf{Z}_1^T (\mathbf{Y} - \mathbf{X}\beta) \right)^2$$

$$+ \left(w_2 \mathbf{Z}_2^T (\mathbf{Y} - \mathbf{X}\beta) \right)^2$$

⋮

$$+ \left(w_M \mathbf{Z}_M^T (\mathbf{Y} - \mathbf{X}\beta) \right)^2$$

where w_m = weight for Z_m

GMM (6)

More general weighting

$$\hat{\beta}_{GMM} = \arg \min (\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{Z}\mathbf{W}\mathbf{Z}^T (\mathbf{Y} - \mathbf{X}\beta)$$

where $\mathbf{W} = \mathbf{M} \times \mathbf{M}$ symmetric PD
weight matrix for \mathbf{Z}

Note that general \mathbf{W} allows not only
own weights but also cross weights

GMM (8)

Estimation

Step 1 $\mathbf{W} = \mathbf{I}$ ($\mathbf{M} \times \mathbf{M}$ identity matrix)

Step 2 Minimize $(\mathbf{Y} - \mathbf{X}\beta)^T \mathbf{Z}\mathbf{W}\mathbf{Z}^T (\mathbf{Y} - \mathbf{X}\beta)$

Step 3 Estimate \mathbf{W} for next iteration

$$\mathbf{W} = \hat{\Omega}^{-1}$$

$$\text{where } \hat{\Omega} = \frac{1}{n-K} \sum_{i=1}^n (Y_i - \mathbf{X}_i \hat{\beta})^2 \mathbf{Z}_i^T \mathbf{Z}_i$$

Back to Step 2 until convergence
occurs

GMM (7)

Ideal Weight matrix

$$\mathbf{W} = \Omega^{-1}$$

where $\Omega = \text{Var}(\mathbf{Z}^T [\mathbf{Y} - \mathbf{X}\beta])$

Why? Similar reason to GLS.