GMM Method (Single-equation)

Pongsa Pornchaiwiseskul
Faculty of Economics
Chulalongkorn University

Covered Topics

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

Stochastic X (1)

Given that, for some k, X_k is random

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

Stochastic X (2)

If some or all X's are random,

additional assumptions about **X** are needed. One of them is

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

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

Violation of C(X,Ɛ)=0

OLS estimator is inconsistent (invalid).

Why?

$$\hat{\beta}_{OLS} = \left(X^T X\right)^{-1} X^T Y$$

$$= \left(X^T X\right)^{-1} X^T \left(X\beta + \varepsilon\right)$$

$$= \beta + \left(X^T X\right)^{-1} X^T \varepsilon$$

Violation of C(X,Ɛ)=0

Take expectation on both sides.

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

$$\neq \beta$$

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

Violation of C(X,Ɛ)=0

Take conditional variance on both sides.

$$V(\hat{\beta}_{OLS} \mid X) = \left(X^T X\right)^{-1} X^T V(\varepsilon \mid X) X \left(X^T X\right)^{-1}$$
$$= \sigma^2 \left(X^T X\right)^{-1}$$

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

Unconditional variance=??

Proxy of X_k

Definition

X_k' is proxy of X_k if

$X_k = X_k' + \xi_k$ and $E(\xi_k) = 0$

$Cov(X_k', \varepsilon) = 0$ and $Cov(X_k', \xi_k) = 0$

==> $E(X_k') = E(X_k)$

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

Regression Model w/ violation

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

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

Regression Model w/o violation

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

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

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

How to get Proxy of $\mathbf{X_k}$ (1)

Assuming that the set of variables **Z** are uncorrelated with ε, we can define the proxy as follows:

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

where $\mathbf{Z}\boldsymbol{\gamma}$ is a proxy of X_k

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**γ) with no correlation with the original error term(ε) and the other part (ξ) with the correlation.

How to get Proxy of X_k (3)

Note that

2) **Z**γ 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

1) all X's which is non-random or random but uncorrelated with ε.
2) any variable outside the regression model which is uncorrelated to ε.

Instrument Variables (2)

Choices of IV's

1) availability of data
2) underlying theories
3) pure assumption

IV Estimation (1)

Steps

1) run OLS regression for

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

2) get the fitted values for X_k

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

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

IV Estimation (2)

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

$$Y_i = \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_K \hat{X}_{Ki} + \varepsilon_i'$$

4) run OLS on the above regression model to get the estimate for β

IV Estimation (3)

Matrix Notation

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

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

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

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

IV Estimation (4)

Proxy of $\mathbf{V}$

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

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

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

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

$$\mathbf{Y} = \mathbf{AU\alpha} + \mathbf{AV\delta} + \boldsymbol{\varepsilon}'$$

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

IV Estimation (5)

IV Estimator

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

Asymptotically zero

Note that $A^TA=A^2=A$

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

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

$$\hat{\boldsymbol{\beta}}_{EQOLS} = (\mathbf{X}^T\mathbf{A}^T\mathbf{AX})\mathbf{X}^T\mathbf{A}^T\mathbf{AY} = \hat{\boldsymbol{\beta}}_{IV}$$

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

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

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 (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 (10)

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

With $\hat{\mathbf{X}} = \mathbf{AX} = \mathbf{X}$

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

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

Weighted 2SLS (1)

Unweighted Model

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

where $\boldsymbol{\upsilon}_i = \dfrac{\varepsilon_i}{w_i}$ or $\mathrm{V}(\boldsymbol{\upsilon}_i) = \left(\dfrac{\sigma}{w_i}\right)^2$

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

Weighted 2SLS (2)

Weighted Model

$$\mathbf{WY} = \mathbf{WU\alpha} + \mathbf{WV\gamma} + \mathbf{W\boldsymbol{\upsilon}}$$

$=\varepsilon$

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

Weighted 2SLS (3)

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

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

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

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

GMM (1)

As an improvement over IV

- IV estimation method does not fully utilize the assumption or knowledge of zero correlation between **Z** and the error term (ε).

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 (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 (4)

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

$$\left.\begin{array}{l} E(Z_1(Y-X\beta)=0 \\ E(Z_2(Y-X\beta)=0 \\ \quad \vdots \\ E(Z_M(Y-X\beta)=0 \end{array}\right\} \textbf{Moment Conditions}$$

where M = #of IV's

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

GMM (5)

Sample Analogy

$$\hat{\boldsymbol{\beta}}_{GMM} = \arg\min \left(w_1 \mathbf{Z}_1^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)^2 + \left(w_2 \mathbf{Z}_2^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)^2 + \cdots + \left(w_M \mathbf{Z}_M^T (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})\right)^2$$

where w_m = weight for Z_m

GMM (6)

More general weighting

$$\hat{\boldsymbol{\beta}}_{GMM} = \arg\min\ (\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})^T\,\mathbf{ZWZ}^T(\mathbf{Y}-\mathbf{X}\boldsymbol{\beta})$$

where **W** = MxM symmetric PD weight matrix for **Z**

Note that general W allows not only own weights but also cross weights

GMM (7)

Ideal Weight matrix

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

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

Why? Similar reason to GLS.

GMM (8)

Estimation

Step 1 W=I (MxM identity matrix)

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

Step 3 Estimate W for next iteration

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

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

Back to Step 2 until convergence occurs