ECON 6200: Section 7

3/7/25

Goal: Show that all these joint estimation GMM cases are related to single-equation GMM

Agenda

- Multiple-Eq GMM
- Eq-by-Eq GMM
- Special Cases
- Common Coefficients
- Problems ~
- Survey

Multiple Equation GMM

$$Y_m = X_m' \beta_m + \varepsilon_m , \quad m = 1, \dots, M$$

$$E[Z_m \varepsilon_m] = 0, \quad m = 1, \dots, M$$

Notation

For each observation i, define

$$\bar{Y}_i = \begin{pmatrix} Y_{i1} \\ \vdots \\ Y_{iM} \end{pmatrix}, \quad \bar{X}_i = \begin{pmatrix} X_{i1} & 0 & \cdots & 0 \\ 0 & X_{i2} & & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & X_{iM} \end{pmatrix}, \quad \bar{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_M \end{pmatrix}$$

$$\bar{Z}_i = \begin{pmatrix} Z_{i1} & 0 & \cdots & 0 \\ 0 & Z_{i2} & & \vdots \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & Z_{iM} \end{pmatrix}$$

Example : LW69 & KWW

$$
\begin{aligned}
LW69 &= \alpha_1 + \beta_1 \; schooling69 + \gamma_1 \; IQ + \delta_1 \; experience69 + \varepsilon_1 \\
KWW &= \alpha_2 + \beta_2 \; schooling69 + \gamma_2 \; IQ + \varepsilon_2
\end{aligned}
$$

For each observation i

$$
\overline{Y}_i = \begin{pmatrix} LW69_i \\ KWW_i \end{pmatrix}, \quad
\overline{X}_i = \begin{pmatrix} 1 & 0 \\ schooling69_i & 0 \\ IQ_i & 0 \\ experience69_i & 0 \\ 0 & 1 \\ 0 & schooling69_i \\ 0 & IQ_i \end{pmatrix}, \quad
\overline{\beta} = \begin{pmatrix} \alpha_1 \\ \beta_1 \\ \gamma_1 \\ \delta_1 \\ \alpha_2 \\ \beta_2 \\ \gamma_2 \end{pmatrix}
$$

2×1 $\qquad$ 7×2 $\qquad$ 7×1

Assumptions

1) Linear Model: $Y_m = X_m'\beta + \varepsilon_m \quad , \quad m = 1, .., M$

2) $(Y_1, .., Y_M, X_1, ..., X_M, Z_1, ..., Z_M)$ iid

↳ stronger than (Y_i, X_i, Z_i) iid

3) Moment Conditions: $E[Z_m(Y_m - X_m'\beta)] = 0 \quad \forall m$

4) Rank Condition

$$E[\bar{Z}\bar{X}] = E\begin{pmatrix} Z_1X_1' & 0 & \cdots & 0 \\ 0 & Z_2X_2' & \cdots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & Z_MX_M' \end{pmatrix} \text{ has full rank}$$

$$\Longleftrightarrow \quad E(Z_mX_m') \text{ has full rank for each } m$$

5) Regularity Condition:

$$\Omega = E[g_m g_m'] \text{ not singular}$$

Under the assumptions above + finite 4^{th} moments,

$$\hat{\beta}(W) = \left(\mathbb{E}_n(\bar{X}_i\bar{Z}_i')\, W\, \mathbb{E}_n(\bar{Z}_i\bar{X}_i')\right)^{-1}\left(\mathbb{E}_n(\bar{X}_i\bar{Z}_i')\, W\, \mathbb{E}_n(\bar{Z}_i\bar{Y}_i)\right)$$

is consistent and asymptotically normal.

Eq-by-Eq GMM

- setup model such that we assume there is no correlation between errors
  ↳ each equation independent from each other

$$\text{ie: } E[\varepsilon_m \varepsilon_n'] = 0 \quad \forall m \neq n$$

- case when $\hat{W}$ are block diagonals

$$\hat{W} = \begin{bmatrix} w_{11} & 0 & \cdots & 0 \\ 0 & w_{22} & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \cdots & \cdots & w_{MM} \end{bmatrix}$$

$M \times M$

(*) When is Eq-by-Eq GMM = Multiple Eq GMM?

- $\hat{W}$ is a block diagonal
- Just-identified ($l = k$)

Under these 2 cases, there is no efficiency gain from joint estimation.

⊛ Issues with Joint Estimation

- In general, joint estimation is asymptotically more efficient

- Finite sample behavior might cause estimation error, bias, ...

- Asymptotic result presumes model is correctly specified
  - ↳ ie: all model assumptions are satisfied
  - ↳ if misspecified, cannot guarantee consistency

- Most importantly, model misspecification is "contagious": The estimator's probability limit equals

$$\text{plim}\,\hat{\beta}(\boldsymbol{W}) = \beta + \left(\mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{Z}}')\boldsymbol{W}\mathbb{E}(\bar{\boldsymbol{Z}}\bar{\boldsymbol{X}}')\right)^{-1}\mathbb{E}(\bar{\boldsymbol{X}}\bar{\boldsymbol{Z}}')\boldsymbol{W}\mathbb{E}(\bar{\boldsymbol{Z}}\varepsilon).$$

  If any one entry of $\mathbb{E}(\bar{\boldsymbol{Z}}\varepsilon)$ is nonzero, then (in general) every entry of the r.h. matrix product is...

- ...except if $\boldsymbol{W}$ is block diagonal corresponding to equations. (You can verify this claim along the lines of the preceding slides.)

Special Cases

⊛ FIVE estimator

↳ multiple equation GMM under conditional homoscedasticity

Conditional homoscedasticity

$$E[\varepsilon_{im}\varepsilon_{ih} \mid z_{im}, z_{ih}] = \sigma_{mh}$$

$$\forall m, h = 1, 2, \ldots, M$$

⊛ 3SLS

↳ FIVE when the set of instruments are the same across equations

↳ uses 2SLS residuals to calculate var-cov matrix

⊛ Seemingly Unrelated Regression (SUR)

↳ 3SLS if predetermined regressors satisfy "cross" orthogonalities

$$E[x_{im} \cdot \varepsilon_{ih}] = 0 \qquad \forall m, h = 1, 2, \ldots, M$$

Not only is x_{im} not correlated with its own ε_{im},

$$\Rightarrow E[x_{im}\varepsilon_{im}] = 0$$

but they are also not correlated in other equations

$$\Rightarrow E[x_{im}\varepsilon_{ih}] = 0 \quad \text{for } m \neq h$$

(Hayashi pg 283)

Eq-by-Eq Jointly

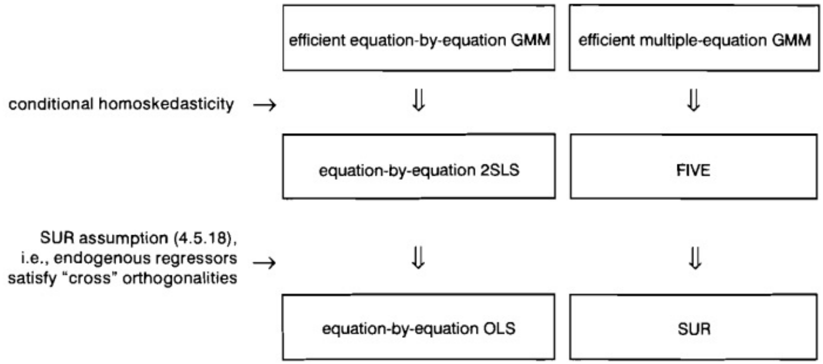

Figure 4.1: OLS and GMM

As we go down this chart, assumptions get stronger.

Common Coefficients

- special case of multiple-equation model where the number of regressors is the same across equations with the same coefficients

If we have M equations

$$Y_m = X_m' \beta + \varepsilon_m \quad \forall m = 1, \ldots, M$$

we have that

$$\beta_1 = \beta_2 = \cdots = \beta_M = \beta$$

$$\bar{\boldsymbol{Y}}_i = \underbrace{\begin{bmatrix} Y_{i1} \\ \vdots \\ Y_{iM} \end{bmatrix}}_{M \times 1}, \bar{\boldsymbol{X}}_i = \underbrace{\begin{bmatrix} \overset{k \times 1}{X_{i1}} & \cdots & \overset{k \times 1}{X_{iM}} \end{bmatrix}}_{K \times M}, \bar{\boldsymbol{Z}}_i = \underbrace{\begin{bmatrix} \overset{l_1 \times 1}{Z_{i1}} & 0 & \cdots & 0 \\ 0 & \overset{l_2 \times 1}{Z_{i2}} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & \cdots & \underset{l_M \times 1}{Z_{iM}} \end{bmatrix}}_{\sum_{m=1}^{M} l_m \times M},$$

- pool information from other equations to help us determine β
  - ↳ It is possible for the system to be identified even if none of the individual equations are identified

$\Rightarrow$ ie:

$$\underbrace{\left[\mathbb{E}(\bar{\boldsymbol{Z}}\bar{\boldsymbol{X}}') \text{ has full rank}\right]}_{\text{system identified}} \Longleftarrow \underbrace{\left[\mathbb{E}(Z_m X_m') \text{ has full rank }, m = 1, \ldots, M\right]}_{\text{equation identified}}.$$

since identification in GMM relies on the rank condition

Panel Data

- Repeated observations on some individuals for multiple time periods
- A form of common coefficients
  - ↳ same covariates
  - ↳ same β (either over i, over t, or over it)

Ex. Pooled OLS

$$\begin{aligned} \boldsymbol{Y}_i &= \boldsymbol{X}_i \beta + \varepsilon_i \\ \mathbb{E}(\boldsymbol{X}_i \varepsilon_i) &= \mathbf{0}. \end{aligned}$$

where the estimator is:

$$\hat{\beta}_{pool} \equiv \left(\sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{X}_i \right)^{-1} \sum_{i=1}^{n} \boldsymbol{X}_i' \boldsymbol{Y}_i = (\boldsymbol{X}' \boldsymbol{X})^{-1} \boldsymbol{X}' \boldsymbol{Y} \xrightarrow{p} \beta$$

- single regression for all observations, ignoring individual and time-specific effects
- same β across all i and all t

2022 Midterm

1 An econometrician wants to use multiple-equation linear GMM with positive definite weighting matrix $\hat{\boldsymbol{W}} \xrightarrow{p} \boldsymbol{W}$ to estimate the three equation

$$\begin{aligned} Y_1 &= \alpha_1 + X'\beta_1 + \varepsilon_1 \\ Y_2 &= \alpha_2 + X'\beta_2 + \varepsilon_2 \\ Y_3 &= \alpha_3 + X'\beta_3 + \varepsilon_3, \end{aligned}$$

where X includes no constant, $\mathbb{E}(XX')$ is invertible, and $\mathbb{E}(X\varepsilon_1) = \mathbb{E}(X\varepsilon_2) = \mathbb{E}(X\varepsilon_3) = 0$. Furthermore, for parts of this question we will assume that $\beta_3 = 0$ and use that information.

1.1 Does the specific choice of $\hat{\boldsymbol{W}}$ matter if $\beta_3 = 0$ is used? What if it is not used?

1.2 Explain how to compute an estimator of $\boldsymbol{\beta} \equiv (\beta_1', \beta_2', \beta_3')'$ that minimizes asymptotic variance among GMM estimators, assuming that the information $\beta_3 = 0$ *is* used.

1.3 Assuming that $\beta_3 = 0$ is imposed, explain how to conduct a specification test of the overall model.

1.4 Provide a test of $H_0 : \beta_3 = 0$ using only (Y_3, X_3) that uses the same asymptotic distribution (hence, critical values) as the test from 1.3. Do you think the tests are asymptotically equivalent? (A reasoned conjecture suffices.)

2012 Midterm

Econometrics II: Prelim Exam

Prof. Jörg Stoye, Spring 2012

This exam consists of 8 questions. Each question carries the same weight. Good luck!

1 Consider the system of equations

$$\begin{aligned} y_{i1} &= \mathbf{x}_{i1}'\boldsymbol{\beta}_1 + \varepsilon_{i1} \\ y_{i2} &= \mathbf{x}_{i2}'\boldsymbol{\beta}_2 + \varepsilon_{i2} \\ y_{i3} &= \mathbf{x}_{i3}'\boldsymbol{\beta}_3 + \varepsilon_{i3} \end{aligned}$$

and the moment conditions

$$\mathbb{E}\mathbf{z}(y_{im} - \mathbf{x}_{im}'\boldsymbol{\beta}_m) = \mathbf{0}, m = 1, 2, 3.$$

Assume that the vectors $(\mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3})$ are all of length 3 and do not have any components in common, whereas $\mathbf{z}$ is of length 5. Assume also homoskedasticity:

$$\mathbb{E}([\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}]'[\varepsilon_{i1}, \varepsilon_{i2}, \varepsilon_{i3}]|\mathbf{x}_{i1}, \mathbf{x}_{i2}, \mathbf{x}_{i3}, \mathbf{z}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix}.$$

Finally, assume that further regularity conditions hold as needed.

We will compare two researchers. Researcher 1 estimates this model by a two-step GMM procedure as discussed in the lecture. Researcher 2 does the same thing but optimizes her choice of weighting matrix $\widehat{\mathbf{W}}$ subject to the constraint that

$$\widehat{\mathbf{W}} = \begin{bmatrix} \underset{[5\times5]}{\widehat{\mathbf{W}}_1} & \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\mathbf{0}} \\ \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\widehat{\mathbf{W}}_2} & \underset{[5\times5]}{\mathbf{0}} \\ \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\mathbf{0}} & \underset{[5\times5]}{\widehat{\mathbf{W}}_3} \end{bmatrix}.$$

1.1 I claim that researcher 2's approach is really a one-step GMM procedure. Give a concise explanation of what I mean, including a brief algebraic demonstration.

1.2 Compare the procedures. Are both estimators consistent? Are they asymptotically normal? Is one of them preferred and why? Explain.

1.3 Assume now that all the above assumptions continue to hold except that

$$\mathbb{E}\mathbf{z}(y_{i1} - \mathbf{x}_{i1}'\boldsymbol{\beta}_1) = [0, 0, 0, a, b]'$$

with $a, b \neq 0$. How does your answer to 1.2 change? How – if at all – does the gist of your answer change if you learn that researchers are only really interested in $\boldsymbol{\beta}_3$?