

10.5 GMM 的假定

在球型扰动项的假定下，2SLS 最有效率。如果存在异方差或自相关，则存在更有效的方法，即“广义矩估计”(Generalized Method of Moments, 简记 GMM)。

GMM 之于 2SLS，正如 GLS 之于 OLS。

假定 10.1 线性假定(linearity)

$$y_i = \underbrace{\mathbf{x}_i' \boldsymbol{\beta}} + \varepsilon_i \quad (i = 1, \dots, n)$$

其中， $\mathbf{x}_i \equiv (x_{i1} \ x_{i2} \ \cdots \ x_{iK})'$ 为第 i 个观测数据。

iid.

ergodic stationary

假定 10.2 渐近独立的平稳过程

记 L 维工具变量为 $\mathbf{z}_i$ (可能与 $\mathbf{x}_i$ 重叠), $\mathbf{w}_i$ 由 $\{y_i, \mathbf{x}_i, \mathbf{z}_i\}$ 中不重复的变量构成且不含常数项。随机过程 $\{\mathbf{w}_i\}$ 为渐近独立的平稳过程。

假定 10.3 工具变量的正交性

所有工具变量 $\mathbf{z}_i$ 均为“前定”，即与同期扰动项正交。定义 L 维列向量 $\mathbf{g}_i \equiv \mathbf{z}_i \varepsilon_i$, 则 $E(\mathbf{g}_i) = E(\mathbf{z}_i \varepsilon_i) = \mathbf{0}$ 。

假定 10.4 秩条件

$L \times K$ 维矩阵 $E(\mathbf{z}_i \mathbf{x}_i')$ 满列秩, 即 $\text{rank}[E(\mathbf{z}_i \mathbf{x}_i')] = K$ 。记 $\Sigma_{ZX} \equiv E(\mathbf{z}_i \mathbf{x}_i')$ 。

Central Limit Theorem.

假定 10.5 $\{g_i\}$ 为鞅差分序列，其协方差矩阵 $S \equiv E(g_i g_i') = E(\varepsilon_i^2 z_i z_i')$ 为非退化矩阵。

假设 10.6 四阶矩 $E[(x_{ik} z_{ij})^2]$ 存在且有限， $\forall i, j, k$ (finite fourth moments)。

按本性，帮助估计
夹心估计量中
同阶量。

10.6 GMM 的推导

Moments

与总体矩条件 $E(g_i) = E(z_i \varepsilon_i) = 0$ 相对应的样本矩条件为

$$g_n(\hat{\beta}) \equiv \frac{1}{n} \sum_{i=1}^n z_i (y_i - x_i' \hat{\beta}) = 0$$

$L=K$?

此联立方程组，未知数 $\hat{\beta}$ 有 K 个，方程个数为 L 个 (z_i 的维度)。

如果 $L < K$ ，为不可识别，则 $\hat{\beta}$ 有无穷多解。

“OLS”

如果 $L = K$ ，为恰好识别，则 $\hat{\beta}$ 有唯一解，即 $\hat{\beta}_{IV}$ 。

如果 $L > K$ ，为过度识别，则 $\hat{\beta}$ 无解。传统的矩估计法行不通。

虽然无法找到 $\hat{\beta}$ 使得 $\mathbf{g}_n(\hat{\beta}) = \mathbf{0}$ ，总可以找到 $\hat{\beta}$ ，使得向量 $\mathbf{g}_n(\hat{\beta})$ 尽可能地接近 $\mathbf{0}$ ，比如，使二次型 $(\mathbf{g}_n(\hat{\beta}))'(\mathbf{g}_n(\hat{\beta}))$ 最小。

更一般地，用“权重矩阵” (weighting matrix) $\mathbf{W}$ 来构成二次型。

OLS

$$\begin{cases} N = K \\ N > K \end{cases}$$

$$y_i = x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ik}\beta_k$$

$$\min (y - \mathbf{x}'\beta)'(y - \mathbf{x}'\beta)$$

假设 $\hat{W}$ 为 $L \times L$ 维对称正定矩阵(可依赖于样本), 且 $\text{plim}_{n \rightarrow \infty} \hat{W} = W$,

其中 W 为非随机的对称正定矩阵。定义最小化的目标函数为

n 用来推了极限分布.

$$\min_{\hat{\beta}} J(\hat{\beta}, \hat{W}) \equiv n \left(\mathbf{g}_n(\hat{\beta}) \right)' \hat{W} \left(\mathbf{g}_n(\hat{\beta}) \right)$$

其中, 因子 n 不影响最小化。定义“GMM 估计量”为此无约束二次型最小化问题的解(Hansen, 1982):

$$\hat{\beta}_{\text{GMM}}(\hat{W}) \equiv \underset{\hat{\beta}}{\text{argmin}} J(\hat{\beta}, \hat{W})$$

GMM 估计量取决于权重矩阵 $\hat{W}$ 。对 $\hat{W}$ 的自由选择是 GMM 的最大优点之一, 可通过最优地选择 $\hat{W}$ 使 $\hat{\beta}_{\text{GMM}}$ 最有效。

不同矩条件的强弱程度一般不同，强的矩条件意味着其对应的方差较小(矩阵 $\mathbf{S} = E(\mathbf{g}_i \mathbf{g}_i')$ 的主对角线元素)，是比较紧的约束，故会通过 $\hat{\mathbf{W}}$ 得到较大的权重。

$\mathbf{g}_n(\hat{\boldsymbol{\beta}})$ 是 $\hat{\boldsymbol{\beta}}$ 的一次函数，故 $J(\hat{\boldsymbol{\beta}}, \hat{\mathbf{W}})$ 是 $\hat{\boldsymbol{\beta}}$ 的二次(型)函数，通过向量微分可得到其最小化问题的解：

$$\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\mathbf{W}}) = (\mathbf{S}'_{\text{ZX}} \hat{\mathbf{W}} \mathbf{S}_{\text{ZX}})^{-1} \mathbf{S}'_{\text{ZX}} \hat{\mathbf{W}} \mathbf{S}_{\text{ZY}}$$

其中， $\mathbf{S}_{\text{ZX}} \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i \mathbf{x}_i'$ ， $\mathbf{S}_{\text{ZY}} \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{z}_i y_i$ 。

(Handwritten notes: L x K for S_ZX, L x 1 for S_ZY, and E(zx') with an arrow pointing to S_ZX)

秩条件 $\text{rank}[E(\mathbf{z}_i \mathbf{x}_i')] = K$ 及 $\hat{\mathbf{W}}$ 正定保证在大样本下， $(\mathbf{S}'_{\text{ZX}} \hat{\mathbf{W}} \mathbf{S}_{\text{ZX}})^{-1}$ 存在。

$$g_n(\hat{\boldsymbol{\beta}}) = \mathbf{S}_{\text{ZY}} - \mathbf{S}_{\text{ZX}} \hat{\boldsymbol{\beta}}$$

(Handwritten notes: 58 above the beta hat, K x 1 next to the beta hat)

$$\begin{aligned}
 J(\hat{\beta}, \hat{W}) &= n g_n(\hat{\beta})' \hat{W} g_n(\hat{\beta}) \\
 &= n (S_{zy} - S_{zx} \tilde{\beta})' \hat{W} (S_{zy} - S_{zx} \tilde{\beta}) \\
 &= n (S'_{zy} \hat{W} S_{zy} - 2 \tilde{\beta}' S'_{zx} \hat{W} S_{zy} \\
 &\quad + \tilde{\beta}' S'_{zx} \hat{W} S_{zx} \tilde{\beta})
 \end{aligned}$$

$$\frac{\partial J}{\partial \tilde{\beta}} = -2 S'_{zx} \hat{W} S_{zy} + 2 S'_{zx} \hat{W} S_{zx} \tilde{\beta} = 0$$

$$\Rightarrow S'_{zx} \hat{W} S_{zx} \tilde{\beta} = S'_{zx} \hat{W} S_{zy}$$

Pos. def. $\hat{W}$ 正定, $S_{zx} \rightarrow \Sigma_{zx}$ 满秩,
故 $S'_{zx} \hat{W} S_{zx}$ full rank.

$$\hat{\beta}(\hat{W}) = (S'_{zx} \hat{W} S_{zx})^{-1} S'_{zx} \hat{W} S_{zy}$$

$$S_{zx}^{-1} \hat{W}^{-1} (S'_{zx})^{-1} (S'_{zx}) \hat{W} S_{zy}$$

$$L=K$$

$W \leftarrow$ 选择有效性

在恰好识别情况下， S_{ZX} 为方阵，GMM 还原为普通的 IV 法：效率最大

$$\hat{\beta}_{\text{GMM}}(\hat{W}) = S_{ZX}^{-1} \underbrace{\hat{W}^{-1} S'_{ZX} S'_{ZX} \hat{W}}_{=I} S_{ZY} = \underbrace{S_{ZX}^{-1} S_{ZY}}_{\hat{\beta}_{\text{IV}}}$$

故 GMM 确实是矩估计的推广。

$$(Z'X)^{-1} Z'Y$$

只有在过度识别的情况下，才有必要使用 GMM。

恰好识别时， W 的选择不影响估计

$$\hat{\beta} = S_{ZX}^{-1} S_{ZY}, \quad g_n(\hat{\beta}) = S_{ZY} - S_{ZX} \hat{\beta} = S_{ZY} - S_{ZX} S_{ZX}^{-1} S_{ZY} = 0$$

$$J = 0$$

Hansen J

10.7 GMM 的大样本性质

定理(GMM 估计量的大样本性质)

(1) ($\hat{\beta}_{\text{GMM}}$ 为一致估计) 在假定 10.1-10.4 之下,
 $\text{plim}_{n \rightarrow \infty} \hat{\beta}_{\text{GMM}}(\hat{W}) = \beta$. 

(2) ($\hat{\beta}_{\text{GMM}}$ 为渐近正态) 如果假定 10.3 (即 $E(\mathbf{g}_i) = \mathbf{0}$) 强化为假定 10.5 (即 $\{\mathbf{g}_i\}$ 为鞅差分序列), 则

$$\sqrt{n}(\hat{\beta}_{\text{GMM}} - \beta) \xrightarrow{d} N(\mathbf{0}, \text{Avar}(\hat{\beta}_{\text{GMM}}))$$

其中, $\text{Avar}(\hat{\beta}_{\text{GMM}}) = (\Sigma'_{\text{ZX}} \mathbf{W} \Sigma'_{\text{ZX}})^{-1} \Sigma_{\text{ZX}} \mathbf{W} \mathbf{S} \mathbf{W} \Sigma_{\text{ZX}} (\Sigma'_{\text{ZX}} \mathbf{W} \Sigma_{\text{ZX}})^{-1}$,
 $\mathbf{S} = E(\mathbf{g}_i \mathbf{g}_i') = E(\varepsilon_i^2 \mathbf{z}_i \mathbf{z}_i')$, $\Sigma_{\text{ZX}} \equiv E(\mathbf{z}_i \mathbf{x}_i')$.

收敛到的值,

$E(\mathbf{z}_i) = \mathbf{0}$.

(3) ($\text{Avar}(\hat{\beta}_{\text{GMM}})$ 的一致估计量)如果 $\hat{S}$ 是 S 的一致估计量, 则在假定 10.2 下, $\text{Avar}(\hat{\beta}_{\text{GMM}})$ 的一致估计量为

$$\widehat{\text{Avar}}(\hat{\beta}_{\text{GMM}}) = (S'_{ZX} \hat{W} S_{ZX})^{-1} S'_{ZX} \hat{W} \hat{S} \hat{W} S_{ZX} (S'_{ZX} \hat{W} S_{ZX})^{-1}$$

证明: (1) 抽样误差可写为

样本值.

$$\begin{aligned} \hat{\beta}_{\text{GMM}}(\hat{W}) - \beta &= (S'_{ZX} \hat{W} S_{ZX})^{-1} S'_{ZX} \hat{W} \left(\frac{1}{n} \sum_{i=1}^n z_i y_i \right) - \beta \\ &= (S'_{ZX} \hat{W} S_{ZX})^{-1} S'_{ZX} \hat{W} \left(\frac{1}{n} \sum_{i=1}^n z_i (\underbrace{x'_i \beta + \varepsilon_i}_{\text{样本值}}) \right) - \beta \\ &= (S'_{ZX} \hat{W} S_{ZX})^{-1} S'_{ZX} \hat{W} \left(S_{ZX} \beta + \frac{1}{n} \sum_{i=1}^n \underbrace{z_i \varepsilon_i}_{\text{样本值}} \right) - \beta \\ &= (S'_{ZX} \hat{W} S_{ZX})^{-1} S'_{ZX} \hat{W} \bar{g} \end{aligned}$$

↓

其中, $\bar{\mathbf{g}} \equiv \frac{1}{n} \sum_{i=1}^n \mathbf{g}_i$, $\mathbf{g}_i \equiv \mathbf{z}_i \varepsilon_i$ 。由于 $(\mathbf{S}'_{ZX} \hat{\mathbf{W}} \mathbf{S}_{ZX})^{-1} \xrightarrow{p} (\mathbf{\Sigma}'_{ZX} \mathbf{W} \mathbf{\Sigma}_{ZX})^{-1}$, $\mathbf{S}'_{ZX} \hat{\mathbf{W}} \xrightarrow{p} \mathbf{\Sigma}'_{ZX} \mathbf{W}$, 而 $\bar{\mathbf{g}} \xrightarrow{p} \mathbf{E}(\mathbf{g}_i) = \mathbf{E}(\mathbf{z}_i \varepsilon_i) = \mathbf{0}$ 。因此, $\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\mathbf{W}}) - \boldsymbol{\beta} \xrightarrow{p} \mathbf{0}$ 。

Slutsky Theorem.

保证 GMM 一致性的最重要条件仍是 $\mathbf{E}(\mathbf{z}_i \varepsilon_i) = \mathbf{0}$, 即工具变量与扰动项正交。

(2) 由于抽样误差 $\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\mathbf{W}}) - \boldsymbol{\beta} = (\mathbf{S}'_{ZX} \hat{\mathbf{W}} \mathbf{S}_{ZX})^{-1} \mathbf{S}'_{ZX} \hat{\mathbf{W}} \bar{\mathbf{g}}$, 故 $\sqrt{n}(\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\mathbf{W}}) - \boldsymbol{\beta}) = (\mathbf{S}'_{ZX} \hat{\mathbf{W}} \mathbf{S}_{ZX})^{-1} \mathbf{S}'_{ZX} \hat{\mathbf{W}} (\sqrt{n} \bar{\mathbf{g}})$ 。

根据假定 10.5 及鞅差分序列的中心极限定理, $\sqrt{n} \bar{\mathbf{g}} \xrightarrow{d} N(\mathbf{0}, \mathbf{S})$, 其中 $\mathbf{S} \equiv \mathbf{E}(\mathbf{g}_i \mathbf{g}_i') = \mathbf{E}(\varepsilon_i^2 \mathbf{z}_i \mathbf{z}_i')$ 。

无自相关
可设独立
NW

由于 $\sqrt{n}(\hat{\beta}_{\text{GMM}}(\hat{W}) - \beta)$ 是 $\sqrt{n}\bar{g}$ 的线性组合，故 $\sqrt{n}(\hat{\beta}_{\text{GMM}}(\hat{W}) - \beta) \xrightarrow{d} N(0, \text{Avar}(\hat{\beta}_{\text{GMM}}))$ 。由于 $(S'_{ZX}\hat{W}S_{ZX})^{-1} \xrightarrow{p} (\Sigma'_{ZX}W\Sigma_{ZX})^{-1}$ ， $S'_{ZX}\hat{W} \xrightarrow{p} \Sigma'_{ZX}W$ ，故

Robust.

$$\text{Avar}(\hat{\beta}_{\text{GMM}}) = (\Sigma'_{ZX}W\Sigma_{ZX})^{-1} \Sigma'_{ZX}W S W \Sigma_{ZX} (\Sigma'_{ZX}W\Sigma_{ZX})^{-1}$$

其中， $(\Sigma'_{ZX}W\Sigma_{ZX})^{-1}$ 为对称矩阵。

(3) 由于 $\hat{S} \xrightarrow{p} S$ ，而且 $S_{ZX} \xrightarrow{p} \Sigma_{ZX}$ ， $\hat{W} \xrightarrow{p} W$ ，故估计量 $\widehat{\text{Avar}}(\hat{\beta}_{\text{GMM}}) = \underbrace{(S'_{ZX}\hat{W}S_{ZX})^{-1}}_{\text{面包}} \underbrace{S'_{ZX}\hat{W}}_{\text{菜}} \underbrace{\hat{S}\hat{W}S_{ZX}}_{\text{面包}} (S'_{ZX}\hat{W}S_{ZX})^{-1}$ 是 $\text{Avar}(\hat{\beta}_{\text{GMM}})$ 的一致估计量。形式上也是一个夹心估计量。

IV, 2SLS 是在K维空间找最优点。2SLS把L维的转换成K维之后找。GMM是在L维空间里找K维子空间的最优点。GMM 的框架更为一般化⁶⁹

Hayashi (2000).

命题 在假定 10.1、假定 10.2 与假定 10.6 下(四阶矩存在), 对于 β 的任何一致估计量 $\hat{\beta}$, 定义残差 $e_i \equiv y_i - \mathbf{x}_i' \hat{\beta}$, 则 $s^2 \equiv \frac{1}{n} \sum_{i=1}^n e_i^2$ 是 $\sigma^2 \equiv E(\varepsilon_i^2)$ 的一致估计, 而且 $\hat{\mathbf{S}} \equiv \frac{1}{n} \sum_{i=1}^n e_i^2 \mathbf{z}_i \mathbf{z}_i'$ 是 $\mathbf{S} \equiv E(\varepsilon_i^2 \mathbf{z}_i \mathbf{z}_i')$ 的一致估计。

命题 使 $\text{Avar}(\hat{\beta}_{\text{GMM}})$ 最小化的“最优权重矩阵”(optimal weighting matrix)为 $\hat{\mathbf{W}} = \hat{\mathbf{S}}^{-1}$ 。

定义 使用 $\hat{\mathbf{S}}^{-1}$ 为权重矩阵的 GMM 估计量被称为“效率 GMM”(efficient GMM)或“最优 GMM”(optimal GMM)。

方差协方差矩阵最小。

2825,

$$\hat{W} = \hat{S}^{-1} \text{ 时, } \hat{\text{Avar}}(\hat{\beta}_{\text{GMM}}) = (\Sigma'_{zx} S^{-1} \Sigma_{zx})^{-1}$$

$$\hat{W} \rightarrow S^{-1}$$

$$\text{证明 } (\Sigma'_{zx} W \Sigma_{zx})^{-1} \Sigma'_{zx} W S W \Sigma_{zx} (\Sigma'_{zx} W \Sigma_{zx})^{-1}$$

$$\geq (\Sigma'_{zx} S^{-1} \Sigma_{zx})^{-1}$$

$\forall W$ 正定对称.

证: $A \geq B$, $A - B$ 是 p.s.d. 半正定.

只需证 $B^{-1} - A^{-1}$ p.s.d.

放缩法, C 可逆.

$$\exists C' C = S^{-1} \text{ (因 } S \text{ p.d.)}$$

$$S = (C' C)^{-1} = C^{-1} (C^{-1})' \quad (C \text{ 是 } L \times L)$$

$$B^{-1} - A^{-1} = \Sigma'_{zx} C' C \Sigma_{zx}$$

$$\text{令 } H = C \Sigma_{zx}$$

$$\text{故 } \Sigma_{zx} = C^{-1} H$$

4' c 3'

$$- \Sigma'_{zx} W \Sigma_{zx} (\Sigma'_{zx} W C^{-1} (C^{-1})' W \Sigma_{zx})^{-1} \Sigma'_{zx} W \Sigma_{zx}$$

$$= H' H - H' (C^{-1})' W \Sigma_{zx} (\Sigma'_{zx} W C^{-1} C^{-1}' W \Sigma_{zx})^{-1} \Sigma'_{zx} W C^{-1} H$$

$$\text{定义 } G \equiv C^{-1}' W \Sigma_{zx}$$

Projection Matrix

$$B^{-1} - A^{-1} = H' H - H' G (G' G)^{-1} G' H$$

$$= H' M_G H$$

$$M_G = I_L - G(G' G)^{-1} G'$$

M : Annihilator.

M_G 是 p.s.d. 因为 $x' M_G x$

清永矩阵.

$$= x' M_G' M_G x \geq 0.$$

Idempotent

因此 $B^{-1} - A^{-1}$ p.s.d.

首先要估计 S^{-1} 选择 $\hat{W}$ 只需要一致。

由于 2SLS 一致，用 2SLS 的残差来计算 $\hat{S} \equiv \frac{1}{n} \sum_{i=1}^n e_i^2 z_i z_i'$ 也一致。

两步最优 GMM 估计：
 $\hat{\beta}_{2SLS}$ I_L 也可以 (B1gg')

第一步：使用 2SLS，得到残差，计算 (p_t, q_t) 。
 $\hat{e}_{2SLS}$

第二步：最小化 $J(\hat{\beta}, \hat{S}^{-1})$ ，得到 $\hat{\beta}_{GMM}(\hat{S}^{-1})$ 。
 $W \rightarrow \hat{S}^{-1} \rightarrow \hat{e}_{2SLS}$

实际操作中，常使用“迭代法”(iterative GMM)直至估计值收敛，即用第二步所获残差再来计算 $\hat{S}$ ，然后再求 $\hat{\beta}_{GMM}(\hat{S}^{-1})$ ，以此类推。

$\hat{S}$ 包含 4 阶矩。在小样本时，有人建议用 I_L

$\hat{S}_0 \rightarrow \hat{\beta}_{GMM(1)} \hat{e}_{GMM}$
 $\hat{S}_1 \rightarrow \hat{\beta}_{GMM(2)}$

$$\sigma^2 \quad E[\varepsilon^2 \varepsilon \varepsilon'] \quad \sigma^2 E(\varepsilon \varepsilon')$$

命题 在条件同方差的情况下，最优 GMM 就是 2SLS。

证明：假设 $E(\varepsilon_i^2 | \mathbf{z}_i) = \sigma^2 > 0$ (条件同方差)，则根据迭代期望定律，

$$\mathbf{S} \equiv E(\mathbf{z}_i \mathbf{z}_i' \varepsilon_i^2) = E_{\mathbf{z}_i} E(\mathbf{z}_i \mathbf{z}_i' \varepsilon_i^2 | \mathbf{z}_i) = E_{\mathbf{z}_i} [\mathbf{z}_i \mathbf{z}_i' E(\varepsilon_i^2 | \mathbf{z}_i)] = \sigma^2 E(\mathbf{z}_i \mathbf{z}_i')$$

因此， $\tilde{\mathbf{S}} \equiv s^2 \mathbf{S}_{ZZ}$ 是 $\mathbf{S}$ 的一致估计量，其中 $\mathbf{S}_{ZZ} \equiv \frac{1}{n} \mathbf{Z}'\mathbf{Z}$ 。

使用 $\tilde{\mathbf{S}}^{-1} = (s^2 \mathbf{S}_{ZZ})^{-1}$ 为最优权重矩阵，则最优 GMM 估计量为

$$\begin{aligned} \hat{\beta}_{\text{GMM}}(\tilde{\mathbf{S}}^{-1}) &= \left(\mathbf{S}'_{ZX} (s^2 \mathbf{S}_{ZZ})^{-1} \mathbf{S}_{ZX} \right)^{-1} \mathbf{S}'_{ZX} (s^2 \mathbf{S}_{ZZ})^{-1} \mathbf{S}_{Zy} \\ &= \left(\mathbf{S}'_{ZX} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{ZX} \right)^{-1} \mathbf{S}'_{ZX} \mathbf{S}_{ZZ}^{-1} \mathbf{S}_{Zy} \end{aligned}$$

因此，2SLS的估计，如果存在异方差，此时选择了robust，结果是稳健的，考虑了异方差问题。但是并不是efficient的。存在异方差的时候，用两步GMM，或者多步迭代的GMM，能得到更准确的（efficient）的估计

若2SLS, robust 不是最。
可改成两步GMM, 看能否。
是最好。

由于 $S_{ZX} \equiv \frac{1}{n} Z'X$, $S_{ZZ} \equiv \frac{1}{n} Z'Z$, $S_{Zy} \equiv \frac{1}{n} Z'y$, 故

$$\begin{aligned}\hat{\beta}_{\text{GMM}}(\tilde{S}^{-1}) &= \left(\frac{1}{n} X'Z \cdot n(Z'Z)^{-1} \cdot \frac{1}{n} Z'X \right)^{-1} \frac{1}{n} X'Z \cdot n(Z'Z)^{-1} \frac{1}{n} Z'y \\ &= \left(X'Z(Z'Z)^{-1} Z'X \right)^{-1} X'Z(Z'Z)^{-1} Z'y \equiv \hat{\beta}_{2\text{SLS}}\end{aligned}$$

在条件同方差的情况下，两步最优 GMM 可省略为一步。因为两步最优 GMM 中第一步的目的只是得到 $\hat{S}^{-1}$ ，而在条件同方差假定下，可直接令 $\hat{S}^{-1} = S_{ZZ}^{-1}$ 。

故 2SLS 也称为“一步 GMM”。

同方差的情况下，2SLS和两步最优的GMM是一样的。异方差时，2SLS不如两步最优的GMM efficient

但 S^2 的掉了。

GMM 的过度识别检验(Overidentification Test or Hansen's J Test)

在恰好识别的情况下，GMM 最小化的目标函数 $J(\hat{\beta}_{\text{GMM}}, \hat{S}^{-1}) = 0$ 。

在过度识别的情况下，如果所有的过度识别约束都成立，则目标函数 $J(\hat{\beta}_{\text{GMM}}, \hat{S}^{-1})$ 应该离 0 不远。

如果 $J(\hat{\beta}_{\text{GMM}}, \hat{S}^{-1})$ 大于 0 很多，则可倾向于认为某些过度识别约束不成立。

在原假设“ H_0 : 所有矩条件均成立”的情况下，目标函数本身就是检验统计量

$J=0$

$$J(\hat{\beta}_{\text{GMM}}, \hat{S}^{-1}) \xrightarrow{d} \chi^2(L-K)$$

Hansen's J

其中, $(L-K)$ 为过度识别约束的个数, 因为在估计 $\hat{\beta}_{\text{GMM}}$ 的过程中失去了 K 个自由度。

异方差时

在条件同方差的情况下, J 统计量与 Sargan 统计量相等。

就应该用 J

检验部分工具变量的正交性 (Testing Subsets of Orthogonality Condition)

如果过度识别检验拒绝“所有工具变量均为外生”的原假设, 则怀疑部分工具变量不满足正交性(外生性)。

假设在 L 个工具变量 $\mathbf{z}_i$ 中，已知前 L_1 个工具变量 $\mathbf{z}_{i1}$ 满足正交性，而怀疑后 $(L - L_1)$ 个工具变量 $\mathbf{z}_{i2}$ 不满足正交性。

检验原假设“ $H_0 : E(\mathbf{z}_{i2}\varepsilon_i) = 0$ ”。

进行此检验的前提条件是 $L_1 \geq K$ ，保证即使仅用前 L_1 个工具变量 $\mathbf{z}_{i1}$ 进行估计，该模型也为恰好识别。

如果 $L_1 \geq K$ ，则可分别用所有 L 个工具变量 $\mathbf{z}_i$ 或前 L_1 个工具变量 $\mathbf{z}_{i1}$ 进行 GMM 估计，记相应的 J 统计量为 J 与 J_1 。

如果将后 $(L - L_1)$ 个工具变量 $\mathbf{z}_{i2}$ 也用于 GMM 估计，使得 J 统计量大大增加，则倾向于拒绝原假设“ $H_0 : E(\mathbf{z}_{i2}\varepsilon_i) = 0$ ”。

如果所有的工具变量都是外生的，那么选择其中的 L 个，或者 L_1 个，差别不应该很大。前提条件是 L_1 个是外生的。因为假如 L_1 个是内生的，同时剩下的其他也是内生的，也有可能 $J-J_1$ 不会很大。所以前提条件总是在那里的。

可以证明：

$$C \equiv J - J_1 \xrightarrow{d} \chi^2(L - L_1)$$

L_2 有问题？

$$E(CC) = 0 ?$$

不可以这样
统计上错误的。

其中， C 统计量称为“GMM 距离” (GMM distance) 统计量或“Sargan 差” (difference-in-Sargan) 统计量，因为它是两个 GMM 估计的 Sargan-Hansen 统计量之差。

L_1 个 OK：主观判断，

自由度为 $(L - L_1)$ ，即怀疑正交性不成立的工具变量个数。

在存在自相关的情况下使用 GMM

在时间序列数据中，即使存在自相关，也仍可使用 GMM，只要采用异方差自相关稳健的标准误来进行统计推断就行。