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)

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

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

假定 10.2　渐近独立的平稳过程

记 L 维工具变量为 $\boldsymbol{z}_i$(可能与 $\boldsymbol{x}_i$ 重叠)，$\boldsymbol{w}_i$ 由 $\{y_i, \boldsymbol{x}_i, \boldsymbol{z}_i\}$ 中不重复的变量构成且不含常数项。随机过程 $\{\boldsymbol{w}_i\}$ 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 $\boldsymbol{z}_i$ 均为“前定”，即与同期扰动项正交。定义 L 维列向量 $\boldsymbol{g}_i \equiv \boldsymbol{z}_i \varepsilon_i$，则 $\mathrm{E}(\boldsymbol{g}_i) = \mathrm{E}(\boldsymbol{z}_i \varepsilon_i) = \boldsymbol{0}$。

假定 10.4　秩条件

$L \times K$ 维矩阵 $\mathrm{E}(\boldsymbol{z}_i \boldsymbol{x}_i')$ 满列秩，即 $\mathrm{rank}\left[\mathrm{E}(\boldsymbol{z}_i \boldsymbol{x}_i')\right] = K$。记 $\boldsymbol{\Sigma}_{ZX} \equiv \mathrm{E}(\boldsymbol{z}_i \boldsymbol{x}_i')$。

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

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

10.6 GMM 的推导

与总体矩条件 $\mathrm{E}(\boldsymbol{g}_i) = \mathrm{E}(\boldsymbol{z}_i \varepsilon_i) = \boldsymbol{0}$ 相对应的样本矩条件为

$$\boldsymbol{g}_n(\hat{\boldsymbol{\beta}}) \equiv \frac{1}{n} \sum_{i=1}^{n} \boldsymbol{z}_i (y_i - \boldsymbol{x}_i' \hat{\boldsymbol{\beta}}) = \boldsymbol{0}$$

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

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

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

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

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

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

假设$\hat{\boldsymbol{W}}$为$L\times L$维对称正定矩阵(可依赖于样本)，且$\operatorname*{plim}_{n\to\infty}\hat{\boldsymbol{W}}=\boldsymbol{W}$，其中$\boldsymbol{W}$为非随机的对称正定矩阵。定义最小化的目标函数为

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

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)：

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

GMM 估计量取决于权重矩阵$\hat{\boldsymbol{W}}$。对$\hat{\boldsymbol{W}}$的自由选择是 GMM 的最大优点之一，可通过最优地选择$\hat{\boldsymbol{W}}$使$\hat{\boldsymbol{\beta}}_{\text{GMM}}$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$\boldsymbol{S}=\mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}_i')$的主对角线元素)，是比较紧的约束，故会通过$\hat{\boldsymbol{W}}$得到较大的权重。

$\boldsymbol{g}_n(\hat{\boldsymbol{\beta}})$是$\hat{\boldsymbol{\beta}}$的一次函数，故$J(\hat{\boldsymbol{\beta}},\hat{\boldsymbol{W}})$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：

$$\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}(\hat{\boldsymbol{W}})=(\boldsymbol{S}_{ZX}'\hat{\boldsymbol{W}}\boldsymbol{S}_{ZX})^{-1}\boldsymbol{S}_{ZX}'\hat{\boldsymbol{W}}\boldsymbol{S}_{Zy}$$

其中，$\boldsymbol{S}_{ZX}\equiv\dfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\boldsymbol{x}_i'$，$\boldsymbol{S}_{Zy}\equiv\dfrac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i y_i$。

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

在恰好识别情况下，$\boldsymbol{S}_{ZX}$ 为方阵，GMM 还原为普通的 IV 法：

$$\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}(\hat{\boldsymbol{W}}) = \boldsymbol{S}_{ZX}^{-1} \underbrace{\hat{\boldsymbol{W}}^{-1} \boldsymbol{S}_{ZX}'^{-1} \boldsymbol{S}_{ZX}' \hat{\boldsymbol{W}}}_{=\boldsymbol{I}} \boldsymbol{S}_{Zy} = \boldsymbol{S}_{ZX}^{-1} \boldsymbol{S}_{Zy} = \hat{\boldsymbol{\beta}}_{\mathrm{IV}}$$

故 GMM 确实是矩估计的推广。

只有在过度识别的情况下，才有必要使用 GMM。

10.7 GMM 的大样本性质

定理(GMM 估计量的大样本性质)

(1) ($\hat{\boldsymbol{\beta}}_{\text{GMM}}$ 为一致估计) 在假定 10.1−10.4 之下，$\operatorname*{plim}_{n\to\infty} \hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\boldsymbol{W}}) = \boldsymbol{\beta}$。

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

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

其中，$\operatorname{Avar}(\hat{\boldsymbol{\beta}}_{\text{GMM}}) = (\boldsymbol{\Sigma}'_{ZX}\boldsymbol{W}\boldsymbol{\Sigma}'_{ZX})^{-1}\boldsymbol{\Sigma}_{ZX}\boldsymbol{WSW}\boldsymbol{\Sigma}_{ZX}(\boldsymbol{\Sigma}'_{ZX}\boldsymbol{W}\boldsymbol{\Sigma}_{ZX})^{-1}$，$\boldsymbol{S} = \mathrm{E}(\boldsymbol{g}_i\boldsymbol{g}'_i) = \mathrm{E}(\varepsilon_i^2\boldsymbol{z}_i\boldsymbol{z}'_i)$，$\boldsymbol{\Sigma}_{ZX} \equiv \mathrm{E}(\boldsymbol{z}_i\boldsymbol{x}'_i)$。

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

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

证明：(1) 抽样误差可写为

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\boldsymbol{W}}) - \boldsymbol{\beta} &= (\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\boldsymbol{S}_{ZX})^{-1}\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i y_i\right) - \boldsymbol{\beta} \\
&= (\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\boldsymbol{S}_{ZX})^{-1}\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\left(\frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i(\boldsymbol{x}'_i\boldsymbol{\beta} + \varepsilon_i)\right) - \boldsymbol{\beta} \\
&= (\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\boldsymbol{S}_{ZX})^{-1}\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\left(\boldsymbol{S}_{ZX}\boldsymbol{\beta} + \frac{1}{n}\sum_{i=1}^{n}\boldsymbol{z}_i\varepsilon_i\right) - \boldsymbol{\beta} \\
&= (\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\boldsymbol{S}_{ZX})^{-1}\boldsymbol{S}'_{ZX}\hat{\boldsymbol{W}}\,\overline{\boldsymbol{g}}
\end{aligned}$$

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

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

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

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

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

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

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

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

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

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

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

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

两步最优 GMM 估计：

第一步：使用 2SLS，得到残差，计算(p_t, q_t)。

第二步：最小化$J(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{S}}^{-1})$，得到$\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\boldsymbol{S}}^{-1})$。

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$\hat{\boldsymbol{S}}$，然后再求$\hat{\boldsymbol{\beta}}_{\text{GMM}}(\hat{\boldsymbol{S}}^{-1})$，以此类推。

命题　在条件同方差的情况下，最优 GMM 就是 2SLS。

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

$$\boldsymbol{S} \equiv \mathrm{E}(\boldsymbol{z}_i \boldsymbol{z}_i' \varepsilon_i^2) = \mathrm{E}_{\boldsymbol{z}_i} \mathrm{E}(\boldsymbol{z}_i \boldsymbol{z}_i' \varepsilon_i^2 \mid \boldsymbol{z}_i) = \mathrm{E}_{\boldsymbol{z}_i} \left[\boldsymbol{z}_i \boldsymbol{z}_i' \mathrm{E}(\varepsilon_i^2 \mid \boldsymbol{z}_i) \right] = \sigma^2 \mathrm{E}(\boldsymbol{z}_i \boldsymbol{z}_i')$$

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

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

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

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

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

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$\hat{\boldsymbol{S}}^{-1}$，而在条件同方差假定下，可直接令$\hat{\boldsymbol{S}}^{-1} = \boldsymbol{S}_{ZZ}^{-1}$。

故 2SLS 也称为“一步 GMM”。

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 $J(\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}, \hat{\boldsymbol{S}}^{-1}) = 0$。

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 $J(\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}, \hat{\boldsymbol{S}}^{-1})$ 应该离 0 不远。

如果 $J(\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}, \hat{\boldsymbol{S}}^{-1})$ 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

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

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

在条件同方差的情况下，J 统计量与 Sargan 统计量相等。

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

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

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

如果$L_1 \ge K$，则可分别用所有L个工具变量$\boldsymbol{z}_i$或前L_1个工具变量$\boldsymbol{z}_{i1}$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$\boldsymbol{z}_{i2}$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：$\mathrm{E}(\boldsymbol{z}_{i2}\varepsilon_i)=0$”。

可以证明：

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

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。

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。