第 4 章　　广义矩估计

对于已知的样本数据，计量经济学和统计学最重要的任务之一就是寻找能使人们估计模型未知参数的技术。一般来说，基于最小化（或最大化）问题的估计过程能够用于许多不同类型模型，它们的区别在于对模型的不同设定性假设。例如，广泛使用的 ML 估计设定模型满足回归模型的绝大多数基本假定，并且要求误差项的分布是已知的。但是，对于实际经济问题这些设定性假设是难以置信的。Hansen(1982)提出的 GMM 估计并不要求模型具有太多的设定性假设，仅仅要求模型满足一组矩条件。因此，在计量经济学中，GMM 估计逐步得到了广泛的应用。本章重点介绍 GMM 估计及其性质和与之相关的假设检验。

§4.1 矩方法 (Method of Moments)

矩方法是一种用样本矩条件类推(analogy)总体矩条件估计模型未知参数的估计技术。下面首先介绍矩条件和统计类推原理。

1 矩条件

设 $\boldsymbol{X}$ 是 $p\times1$ 的随机向量，f 是 $l\times1$ 的向量值函数，$\boldsymbol{\theta}$ 是 $k\times1$ 的参数向量，Θ 是参数空间，即，$\boldsymbol{\theta}\in\Theta$，则称

$$E[f(\boldsymbol{X},\boldsymbol{\theta})] = \mathbf{0} \tag{4.1}$$

是 $\boldsymbol{X}$ 的矩条件，或者，$\boldsymbol{X}$ 的矩方程。

类似地，对于回归模型

$$\boldsymbol{Y} = \boldsymbol{X\beta}+\boldsymbol{u}$$

设$\{y_i, \boldsymbol{X}_i \mid i=1, 2, \ldots, N\}$是模型变量 y 和 X 的 N 个观测值，$f(y_i, \boldsymbol{X}_i, \boldsymbol{\beta}) = y_i - \boldsymbol{X}_i\boldsymbol{\beta}$ 是向量 $\boldsymbol{\beta}$ 的连续函数，并且，$E[f(y_i, \boldsymbol{X}_i, \boldsymbol{\beta})] < \infty$，$i=1, 2, \ldots, N$，则称

$$E[f(y_i, \boldsymbol{X}_i, \boldsymbol{\beta})] = \mathbf{0} \tag{4.2}$$

是回归模型的矩条件。

如果回归模型的矩条件（4.2）中包含 K 个方程，称参数 $\boldsymbol{\beta}$ 是由矩条件恰好识别的；如果（4.2）中包含的方程少于 K 个，称参数 $\boldsymbol{\beta}$ 是由矩条件不可识别的；如果（4.2）中包含的方程多于 K 个，称参数 $\boldsymbol{\beta}$ 是由矩条件过度识别的。

(1) 最小二乘法

例如：对于随机解释变量的线性回归模型

$$y_i = \boldsymbol{X}_i'\boldsymbol{\beta} + u_i \tag{4.3}$$

常常假设 $E(u_i \mid X_i) = 0$，于是，

$$E(y_i \mid \boldsymbol{X}_i) = \boldsymbol{X}_i'\boldsymbol{\beta}$$

并且，使用期望迭代公式，得

$$E(\boldsymbol{X}_i u_i) = E[E(\boldsymbol{X}_i u_i|\boldsymbol{X}_i)] = E[\boldsymbol{X}_i E(u_i|\boldsymbol{X}_i)] = \mathbf{0}$$

因此，模型（4.3）有矩条件

$$E(\boldsymbol{X}_i u_i) = E(\boldsymbol{X}_i (y_i - \boldsymbol{X}_i'\boldsymbol{\beta})) = \mathbf{0} \tag{4.4}$$

其中，$f(y_i, \boldsymbol{X}_i, \boldsymbol{\beta}) = X_i (y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$. 且，矩条件（4.4）中包含的方程等于未知参数的个数 K，即参数 $\boldsymbol{\beta}$ 是由矩条件恰好识别。

(2) 工具变量法

对于随机解释变量的线性回归模型（4.3），如果假设 $E(u_i \mid \boldsymbol{X}_i) = 0$ 不成立，但是，存在工具变量 $\boldsymbol{Z}_i$，这时，模型（4.3）有矩条件

$$E(\boldsymbol{Z}_i u_i) = E(\boldsymbol{Z}_i (y_i - \boldsymbol{X}_i'\boldsymbol{\beta})) = \mathbf{0} \tag{4.5}$$

其中，$f(y_i, \boldsymbol{Z}_i, \boldsymbol{X}_i, \boldsymbol{\beta}) = \boldsymbol{Z}_i(y_i - \boldsymbol{X}_i'\boldsymbol{\beta})$. 并且，矩条件（4.5）中包含的方程取决于工具变量的

个数。

2. 类推原理（Analogy principle）

类推原理是统计学中重要的推断原理之一，它的基本思想是对于总体随机变量 Y 和 X 的典型样本$\{y_i,\boldsymbol{X}_i\}_{i=1}^{N}$，则有适当假设的大数定理和连续函数定理可得，样本矩方程

$$f_N(\boldsymbol{\beta})=\frac{1}{N}\sum_{i=1}^{N}f(y_i \quad \boldsymbol{X}_i \quad \boldsymbol{\beta})=0 \tag{4.6}$$

的解 $\boldsymbol{\beta}_{MM}$ 是对总体矩方程

$$E[f(Y，X，\boldsymbol{\beta})] = \mathbf{0}$$

解的良好估计。

3. 矩估计

矩估计是利用矩条件估计模型的未知参数 $\boldsymbol{\beta}$ 的方法，这里只讨论模型参数由矩条件恰好识别的情况，对于过度识别的情形，在下一节广义矩估计 GMM 中研究；对于不可识别的情况，需要添加识别约束条件，本章不予讨论。由统计类推原理，解样本矩方程

$$f_N(\boldsymbol{\beta})=\frac{1}{N}\sum_{i=1}^{N}f(y_i \quad \boldsymbol{X}_i \quad \boldsymbol{\beta})=0$$

即可得到模型未知参数的良好估计。并称该估计量 $\boldsymbol{\beta}_{MM}$ 为模型参数 $\boldsymbol{\beta}$ 的矩估计量。

(1) OLS 估计的矩估计表示

显然，在满足回归基本假定 7 时，对于随机解释变量的线性回归模型（4.3），参数 $\boldsymbol{\beta}$ 的 OLS 估计

$$\hat{\boldsymbol{\beta}}=\frac{1}{N}\left(\sum_{i=1}^{N}\boldsymbol{X}_i\boldsymbol{X}_i^{'}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{X}_i y_i \tag{4.7}$$

是矩条件

$$E\left[\boldsymbol{X}\left(y-\boldsymbol{X}^{'}\boldsymbol{\beta}\right)\right]=\mathbf{0} \tag{4.8}$$

的样本矩方程

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{X}_i\left(y_i-\boldsymbol{X}_i^{'}\boldsymbol{\beta}\right)=\mathbf{0} \tag{4.8′}$$

的解。并且，在满足回归基本假定 1-5 和假定 7 时，OLS 估计（4.7）是矩方程（4.4）解的一致估计。

另外，对于线性回归模型（4.3），OLS 估计（4.7）使得 $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ 是 $\boldsymbol{Y}$ 的最佳线性预测（best linear predictor）。并且，如果 $E(\boldsymbol{Y}\,|\,\boldsymbol{X})$ 是非线性的，则在

$$\hat{\boldsymbol{\beta}}=arg\min_{\beta}E\left[\left(E(\mathbf{Y}\,|\,\mathbf{X})-\boldsymbol{X}\boldsymbol{\beta}\right)^2\right]$$

的意义下，$\boldsymbol{Y}$ 的线性预测 $\boldsymbol{X}\hat{\boldsymbol{\beta}}$ 是对 $\boldsymbol{Y}$ 的最佳线性逼近。

(2) 工具变量估计的矩估计表示

类似地，在模型（4.3）仅有与解释变量相同的工具变量时，工具变量估计

$$\boldsymbol{\beta}_{IV}=\frac{1}{N}\left(\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{X}_i^{'}\right)^{-1}\sum_{i=1}^{N}\boldsymbol{Z}_i y_i \tag{4.9}$$

也是矩条件

$$E\left[\boldsymbol{Z}\left(y-\boldsymbol{X}^{'}\boldsymbol{\beta}\right)\right]=\mathbf{0} \tag{4.10}$$

的样本矩方程

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\left(y_i-\boldsymbol{X}_i^{'}\boldsymbol{\beta}\right)=\mathbf{0} \tag{4.10′}$$

的解。同样，在满足回归基本假定 7′和假定 8 时，工具变量估计（4.9）是矩方程（4.10）解的渐近正态一致估计。

(3) ML 估计的矩估计表示

类似于 OLS 估计和工具变量估计，极大似然估计也是矩估计的一种。

假设$\{\boldsymbol{X}_1, \boldsymbol{X}_2,\dots, \boldsymbol{X}_N\}$是服从分布 $P_{\boldsymbol{\theta}}\in \mathcal{P}$ 的随机总体独立抽取的 N 个样本，其中，$\boldsymbol{\theta}$ 是参数空间 $\boldsymbol{\Theta}$ 中未知参数，$\mathcal{P}=\left\{N\left(u,\sigma^2\right)|\boldsymbol{\theta}=\left(u,\sigma^2\right)\in\boldsymbol{\Theta}\right\}$，则 ML 估计 $\hat{\boldsymbol{\theta}}$ 是样本$\{\boldsymbol{X}_1, \boldsymbol{X}_2,\dots, \boldsymbol{X}_N\}$对数似然函数

$$\log L\left(\boldsymbol{\theta}\,|\,\boldsymbol{X}\right)=\sum_{i=1}^{N}\log L_i\left(\boldsymbol{\theta}\,|\,\boldsymbol{X}_i\right)$$

的极大值。即满足一阶条件

$$N^{-1}\sum_{i=1}^{N}\left.\frac{\partial\log L_i\left(\boldsymbol{\theta}\,|\,\boldsymbol{X}_i\right)}{\partial\boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}=\mathbf{0} \tag{4.11}$$

显然，对数似然函数的一阶条件（4.11）是矩条件

$$E\left[\frac{\partial\log L\left(\boldsymbol{\theta}\,|\,\boldsymbol{X}\right)}{\partial\boldsymbol{\theta}}\right]=\mathbf{0}$$

的样本矩条件。所以，如果模型满足回归基本假定 1-7，极大似然估计 $\tilde{\boldsymbol{\theta}}$ 也是一种矩估计。

§4.2 广义矩估计 (Generalized Method of Moments Estimation)

1 广义矩问题

对于工具变量估计的矩方程(4.10)，即，$E\left[\boldsymbol{Z}_i\left(y_i-\boldsymbol{X}_i^{'}\boldsymbol{\beta}\right)\right]=\mathbf{0}$，如果 $\boldsymbol{Z}$ 中工具变量的个数 l 大于 $\boldsymbol{X}$ 中解释变量的个数 k，则矩方程(4.10)存在解的充分条件是 $l\times k$ 矩阵$E(\boldsymbol{Z}'\boldsymbol{X})$是奇异的，否则，方程组

$$E(\boldsymbol{Z}'\boldsymbol{Y})-E(\boldsymbol{Z}'\boldsymbol{X})\boldsymbol{\beta}=\mathbf{0}$$

无解。

并且，由于样本误差的缘故，即使对于总体而言矩阵$E(\boldsymbol{Z}'\boldsymbol{X})$是奇异的，其样本矩矩阵

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{X}_i^{'}$$

一般也未必是奇异的。于是，l 个样本矩方程组（4.10′）

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\left(y_i-\boldsymbol{X}_i^{'}\boldsymbol{\beta}\right)=0$$

可能不存在解。然而，对于样本矩方程组(4.10′)的 k 个样本矩条件组成的子方程组，均可得到 $\boldsymbol{\beta}$ 的不同估计。

更一般地，Sargan（1958）以 l 个矩方程的 k 个线性组合估计 $\boldsymbol{\beta}$，即，广义矩方程组

$$\frac{1}{N}\boldsymbol{\Gamma}_N\sum_{i=1}^{N}\boldsymbol{Z}_i\left(y_i-\boldsymbol{X}_i^{'}\boldsymbol{\beta}\right)=\boldsymbol{0}$$

的解 $\boldsymbol{\beta}_{GMM}$ 被称为 $\boldsymbol{\beta}$ 的广义矩估计，其中，$\boldsymbol{\Gamma}_N$ 是 k 个线性组合系数的 $k\times l$ 矩阵。

于是，广义矩估计 $\boldsymbol{\beta}_{GMM}$ 的优劣，取决于 $k\times l$ 矩阵 $\boldsymbol{\Gamma}_N$ 的选择，$\boldsymbol{\Gamma}_N$ 的最优选择问题是广义矩估计的主要问题之一，被称为广义矩问题。

2 广义矩估计的一般形式

设 $\boldsymbol{X}$ 是 $p\times1$ 的随机向量，f 是 $l\times1$ 的向量值函数，$\boldsymbol{\theta}$ 是 $k\times1$ 的参数向量，并且 $k\le l$，Θ 是参数空间，即，$\boldsymbol{\theta}\in\Theta$，

$$E[f(\boldsymbol{X},\ \boldsymbol{\theta})]=\boldsymbol{0} \tag{4.12}$$

是 $\boldsymbol{X}$ 的矩方程。

对于 $\boldsymbol{X}$ 的 N 个观测值 $\left\{\boldsymbol{X}_1,\ \cdots,\ \boldsymbol{X}_N\right\}$，则基于矩方程(4.12)的样本矩条件是

$$f_N(\boldsymbol{\theta})=\frac{1}{N}\sum_{i=1}^{N}f\left(\boldsymbol{X}_i,\ \boldsymbol{\theta}\right) \tag{4.13}$$

并且称最小化问题的解

$$\boldsymbol{\theta}_{GMM}=arg\min_{\theta\in\Theta}f_N(\boldsymbol{\theta})^{'}\boldsymbol{A}_N f_N(\boldsymbol{\theta}) \tag{4.14}$$

为参数向量 $\boldsymbol{\theta}$ 的广义矩估计量，其中，$\boldsymbol{A}_N$ 是秩大于等于 k 的 $l\times l$ 非负定的权重矩阵。也称

$$Q_N(\boldsymbol{\theta})\triangleq\left(\frac{1}{N}\sum_{i=1}^{N}f\left(\boldsymbol{X}_i,\ \boldsymbol{\theta}\right)\right)^{'}\boldsymbol{A}_N\left(\frac{1}{N}\sum_{i=1}^{N}f\left(\boldsymbol{X}_i,\ \boldsymbol{\theta}\right)\right) \tag{4.14}$$

是 GMM 估计的估计准则。

显然，如果线性回归模型的未知参数 $\boldsymbol{\beta}$ 是由矩条件恰好识别的，即 $l=k$，选择

$$\boldsymbol{A}_N=N\left(\boldsymbol{X'X}\right)^{-1}$$

则 $\boldsymbol{\beta}$ 的 GMM 估计量 $\boldsymbol{\beta}_{GMM}$ 正好是 OLS 估计量

$$\hat{\boldsymbol{\beta}}=\frac{1}{N}\left(\sum_{i=1}^{N}X_iX_i^{'}\right)^{-1}\sum_{i=1}^{N}X_iy_i=\left(\boldsymbol{X'X}\right)^{-1}\left(\boldsymbol{X'Y}\right)$$

并且，样本矩方程组

$$f_N(\boldsymbol{\beta})=\frac{1}{N}\sum_{i=1}^{N}X_i\left(y_i-X_i'\boldsymbol{\beta}\right)=\mathbf{0}$$

与 *OLS* 估计的正规方程组同解。

类似地，对于过度识别的线性回归模型，即 $l\geq k$，设 $\mathbf{A}_N=N\left(\boldsymbol{Z'Z}\right)^{-1}$，则 $\boldsymbol{\beta}$ 的 GMM 估计量也正好是标准工具变量估计量

$$\boldsymbol{\beta}_{IV}=\left(\boldsymbol{X'Z}(\boldsymbol{Z'Z})^{-1}\boldsymbol{X'Z}\right)^{-1}\boldsymbol{X'Z}(\boldsymbol{Z'Z})^{-1}(\boldsymbol{X'Y}).$$

3 GMM 的例子：2SLS 和 3SLS

(1) 二阶段最小二乘法 2SLS

二阶段最小二乘法（Two Stage Least Squares）是估计联立方程模型参数常用的方法，这里将说明 2SLS 也是一种特殊的 GMM 估计方法。

设 $\mathbf{Z}_i$ 是单方程

$$y_i=\boldsymbol{X}_i'\boldsymbol{\theta}+u_i \tag{4.15}$$

的 $l\times 1$（$l>k$）工具变量向量，所以，满足矩条件

$$E\left(\boldsymbol{Z}_i u_i\right)=0 \tag{4.16}$$

令 $\boldsymbol{W}_i=\left(y_i,\ \boldsymbol{X}_i',\ \boldsymbol{Z}_i'\right)$，$f\left(\boldsymbol{W}_i,\ \theta\right)=\boldsymbol{Z}_i\left(y_i-\boldsymbol{X}_i'\boldsymbol{\theta}\right)$，并且，$\boldsymbol{Y}=\left(y_1,\ \cdots,\ y_N\right)'$，$\boldsymbol{X}=\left(\boldsymbol{X}_1,\ \cdots,\ \boldsymbol{X}_N\right)'$ 和 $\boldsymbol{Z}=\left(\boldsymbol{Z}_1,\ \cdots,\ \boldsymbol{Z}_N\right)'$ 分别是变量 y，X 和 Z 的一个样本数据向量，则单方程（4.15）的样本矩条件是

$$f_N(\boldsymbol{\theta})=\frac{1}{N}\sum_{i=1}^{N}f\left(\boldsymbol{W}_i,\ \boldsymbol{\theta}\right)=\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\left(y_i-\boldsymbol{Z}_i'\boldsymbol{\theta}\right)=\frac{1}{N}\boldsymbol{Z}'\left(\mathbf{Y}-\boldsymbol{X\theta}\right) \tag{4.17}$$

因为 $l>k$，矩方程 $f_N(\boldsymbol{\theta})=\mathbf{0}$ 无解。

若选择权重矩阵 $\boldsymbol{A}_N=\left(\boldsymbol{Z}'\boldsymbol{Z}/N\right)^{-1}$，则 $\boldsymbol{\theta}$ 的 GMM 估计是最小化问题

$$\boldsymbol{\theta}_{GMM}=arg\min_{\boldsymbol{\theta}\in\Theta}\left\{f_N(\boldsymbol{\theta})'\mathbf{A}_N f_N(\boldsymbol{\theta})\right\}=N^{-2}\left(\mathbf{Y}-\boldsymbol{X\theta}\right)'\boldsymbol{Z}\boldsymbol{A}_N\boldsymbol{Z}'\left(\mathbf{Y}-\boldsymbol{X\theta}\right) \tag{4.18}$$

的解。

事实上，(4.18)的一阶条件是

$$\boldsymbol{X}'\boldsymbol{Z}\boldsymbol{A}_N\boldsymbol{Z}'\left(\mathbf{Y}-\boldsymbol{X\theta}\right)=\mathbf{0} \tag{4.19}$$

于是，$\boldsymbol{\theta}$ 的 GMM 估计

$$\boldsymbol{\theta}_{GMM}=\left[\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \tag{4.20}$$

并且，若令 $\hat{\boldsymbol{\delta}}=\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}$，即 $\hat{\boldsymbol{\delta}}$ 是内生变量 $\boldsymbol{X}$ 关于工具变量 $\boldsymbol{Z}$ 回归系数的 OLS 估计

$$\hat{\boldsymbol{X}}=\boldsymbol{Z}\hat{\boldsymbol{\delta}} \tag{4.21}$$

则

$$\boldsymbol{\theta}_{GMM}=\left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\hat{\boldsymbol{X}}'\boldsymbol{Y}=\boldsymbol{\theta}_{2SLS} \tag{4.22}$$

所以，$\boldsymbol{\theta}$ 的 GMM 估计 $\boldsymbol{\theta}_{GMM}$ 也是二阶段 OLS 回归的估计结果 $\boldsymbol{\theta}_{2SLS}$，即，第一步的 *OLS* 回归是解释变量 $\boldsymbol{X}$ 关于工具变量 $\boldsymbol{Z}$ 的多元线性回归；第二步的 *OLS* 回归是被解释变量 $\boldsymbol{Y}$ 关于 $\hat{\boldsymbol{X}}$ 的回归。

与二阶段最小二乘法 2SLS 类似，容易说明联立方程组模型的三阶段最小二乘法（Three Stage Least Squares）也是一种 GMM 估计。

§4.3　GMM 估计的一致性

1959 年，Sargan 首先讨论了非线性工具变量估计的渐近性；更一般地，对于极值估计（或最小化连续函数估计），Amemiya（1985）研究了极值估计的渐近性质。

1 **极值估计**(Extremum estimators)

对 $p\times1$ 随机向量 W 的任意 N 个样本 $\left\{\boldsymbol{W}_1,\ \cdots,\ \boldsymbol{W}_N\right\}$ 和 k 维参数 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$，如果

$Q_N\left(\boldsymbol{\theta}\right)=Q_N\left(\boldsymbol{W}_1,\ \cdots,\ \boldsymbol{W}_N,\ \boldsymbol{\theta}\right)$ 是 $\boldsymbol{\theta}$ 的连续函数，则

$$\hat{\boldsymbol{\theta}}_N=arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}Q_N\left(\boldsymbol{\theta}\right) \tag{4.23}$$

被称为参数 $\boldsymbol{\theta}$ 的极值估计(Extremum estimators)，其中，$\boldsymbol{\Theta}$ 是 k 维参数空间。

显然，*OLS*、*GLS*、*2SLS*、*3SLS* 和 *GMM* 估计皆是极值估计的特例。

2 **极值估计的一致性**

定理：假设下述条件成立：

（1）参数空间 $\boldsymbol{\Theta}$ 是 k 维空间 R^k 的紧致子集[①]，且 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$；

（2）对于 N 个样本 $\left(\boldsymbol{W}_1,\ \cdots,\ \boldsymbol{W}_N\right)$，在参数空间 $\boldsymbol{\Theta}$ 上，估计准则

$$Q_N\left(\boldsymbol{\theta}\right)=Q_N\left(\boldsymbol{W}_1,\ \cdots,\ \boldsymbol{W}_N,\ \boldsymbol{\theta}\right)$$

是 $\boldsymbol{\theta}$ 的连续函数；

（3）在 $\boldsymbol{\Theta}$ 上 $Q_N\left(\boldsymbol{\theta}\right)$ 依概率一致收敛于非随机函数 $Q_\infty\left(\boldsymbol{\theta}\right)$，即，当 $N\to\infty$ 时

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}\left|Q_N\left(\boldsymbol{\theta}\right)-Q_\infty\left(\boldsymbol{\theta}\right)\right|\xrightarrow{p}0 \tag{4.24}$$

并且，假设 $\boldsymbol{\theta}_0$ 是 $Q_\infty\left(\boldsymbol{\theta}\right)$ 的唯一最小值点，　$\boldsymbol{\theta}_0$ 的极值估计是

$$\hat{\boldsymbol{\theta}}_N=arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}Q_N\left(\boldsymbol{\theta}\right)$$

则

$$\operatorname{plim}_{N\to\infty}\hat{\boldsymbol{\theta}}_N=\boldsymbol{\theta}_0 \tag{4.25}$$

证明从略（详见 Amemiya，1985，p.$_{107}$.）。

[①] Engelking,R.(1977), p.164.

对于 GMM 估计，设

$$Q_N(\boldsymbol{\theta}) = f_N(\boldsymbol{\theta})' \boldsymbol{A}_N f_N(\boldsymbol{\theta}) \tag{4.26}$$

其中，$f_N(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} f(\boldsymbol{W}_i, \boldsymbol{\theta})$，则 Amemiya 的一致性定理如下。

3 GMM 估计的一致性

定理：当下列条件成立时，

（1）参数空间 $\boldsymbol{\Theta}$ 是 k 维空间 R^k 的紧致子集，且 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$;

（2）对任意的 $\boldsymbol{W}$ 和 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$，$f(\boldsymbol{W}, \boldsymbol{\theta})$ 是 $\boldsymbol{\theta}$ 的连续函数;

（3）权重矩阵 $\boldsymbol{A}_N \xrightarrow{p} \boldsymbol{A}_0$，$\boldsymbol{A}_0$ 是半正定矩阵;

（4）对所有的 $\boldsymbol{\theta}\in\boldsymbol{\Theta}$，$E\left[f(\boldsymbol{W}, \boldsymbol{\theta})\right]$ 存在，并且，当 $\boldsymbol{\theta}=\boldsymbol{\theta}_0$ 时，

$$\boldsymbol{A}_0 E\left[f(\boldsymbol{W}, \boldsymbol{\theta}_0)\right] = \boldsymbol{0};$$

（5）$f_N(\boldsymbol{\theta})$ 关于 $\boldsymbol{\theta}$ 依概率一致收敛于 $E\left[f(\boldsymbol{W}, \boldsymbol{\theta})\right]$;

则

$$Q_\infty(\boldsymbol{\theta}) = E\left[f(\boldsymbol{W}, \boldsymbol{\theta})\right]' \boldsymbol{A}_0 E\left[f(\boldsymbol{W}, \boldsymbol{\theta})\right] \ge 0 \tag{4.27}$$

且 $\boldsymbol{\theta}_0$ 是 $Q_\infty(\boldsymbol{\theta}) = 0$ 的唯一解；对于 GMM 估计 $\hat{\boldsymbol{\theta}}_N = arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q_N(\boldsymbol{\theta})$，有 $\operatorname{plim}_{N\to\infty}\hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0$.

§4.4 GMM 估计的渐近正态性

1 极值估计的渐近正态性质

类似于极值估计的一致性分析，Amemiya（1985）推导出了关于极值估计的渐近正态性质。

渐近正态定理：在下列条件成立时，

(1) 若 $\hat{\boldsymbol{\theta}}_N = arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} Q_N(\boldsymbol{\theta})$，使得 $\operatorname{plim}_{N\to\infty}\hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}$;

(2) 参数 $\boldsymbol{\theta}$ 是 k 维参数空间 $\boldsymbol{\Theta}$ 的内点[②]；在 $\boldsymbol{\theta}$ 的某邻域内，$Q_N(\boldsymbol{\xi})$ 是二阶可微的;

(3) $Q_N(\boldsymbol{\theta})$ 的梯度是渐近正态的，即

$$\sqrt{N}\frac{\partial Q_N(\boldsymbol{\theta})}{\partial \boldsymbol{\xi}} \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma})^{③} \tag{4.28}$$

(4) 对于任意的 $\tilde{\boldsymbol{\theta}}_N$，如果 $\tilde{\boldsymbol{\theta}}_N \xrightarrow{p} \boldsymbol{\theta}$，Hessian 矩阵依概率收敛于非奇异的非随机矩阵 $\boldsymbol{H}$;

[②] Engelking,R.(1977), p.29.

[③] 这里，令 $\frac{\partial Q_N(\boldsymbol{\theta})}{\partial \boldsymbol{\xi}} = \left.\frac{\partial Q_N(\boldsymbol{\xi})}{\partial \boldsymbol{\xi}}\right|_{\boldsymbol{\xi}=\boldsymbol{\theta}}$

$$\frac{\partial^2 Q_N\left(\tilde{\boldsymbol{\theta}}_N\right)}{\partial \boldsymbol{\xi} \partial \boldsymbol{\xi}^{'}} \xrightarrow{p} \boldsymbol{H} \tag{4.29}$$

则

$$\sqrt{N}\left(\tilde{\boldsymbol{\theta}}_N - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0,\quad \boldsymbol{H}^{-1}\boldsymbol{\Sigma}\boldsymbol{H}^{-1}\right) \tag{4.30}$$

证明：因为参数 $\boldsymbol{\theta}$ 是 $\boldsymbol{\Theta}$ 的内点，$\hat{\boldsymbol{\theta}}_N$ 是 $\boldsymbol{\theta}$ 的一致估计，所以，当 $N \to \infty$ 时，$\hat{\boldsymbol{\theta}}_N$ 不是 $\boldsymbol{\Theta}$ 的内点的概率趋于零。为了简化表述，下面记 $\hat{\boldsymbol{\theta}}_N$ 为 $\hat{\boldsymbol{\theta}}$，$\hat{\theta}_i$ 是 $\hat{\boldsymbol{\theta}}$ 的第 i 个分量元素。

于是，二阶展开式

$$\frac{\partial Q_N\left(\hat{\boldsymbol{\theta}}\right)}{\partial \xi_j} = \frac{\partial Q_N\left(\boldsymbol{\theta}\right)}{\partial \xi_j} + \sum_{l=1}^{k} \frac{\partial^2 Q_N\left(\tilde{\boldsymbol{\theta}}_{[j]}\right)}{\partial \xi_j \partial \xi_l}\left(\hat{\theta}_l - \theta_l\right) = \mathbf{0} \qquad j = 1,\quad \cdots,\quad k \tag{4.31}$$

其中，$\tilde{\boldsymbol{\theta}}_{[j]}$ 是使得 $\left\|\tilde{\boldsymbol{\theta}}_{[j]} - \boldsymbol{\theta}\right\| \le \left\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right\|$ 的 $k\times1$ 向量。

如果 $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$，则 $\tilde{\boldsymbol{\theta}}_{[j]} \xrightarrow{p} \boldsymbol{\theta}$，由条件（4）

$$\frac{\partial^2 Q_N\left(\tilde{\boldsymbol{\theta}}_N\right)}{\partial \xi_j \partial \xi_i} \xrightarrow{p} H_{ji}$$

其中，H_{ji} 是 $\boldsymbol{H}$ 的（j，i）元素。
因此，

$$\sqrt{N}\,\frac{\partial Q_N\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\xi}} + \left[\boldsymbol{H} + o_p\left(1\right)\right]\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) = \mathbf{0}$$

$$\boldsymbol{H}^{-1}\sqrt{N}\,\frac{\partial Q_N\left(\boldsymbol{\theta}\right)}{\partial \boldsymbol{\xi}} = \boldsymbol{H}^{-1}\left[\boldsymbol{H} + o_p\left(1\right)\right]\sqrt{N}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \tag{4.32}$$

最后，由条件(3)和 Cramer-Rao 下界定理[④]即可证得(4.30).

2　GMM 估计的渐近正态性质

GMM 估计的渐近正态定理：在下列条件成立时，

(1) 参数 $\boldsymbol{\theta}$ 是 k 维参数空间 $\boldsymbol{\Theta}$ 的内点，$f\left(\boldsymbol{W},\quad \boldsymbol{\theta}\right)$ 是 $\boldsymbol{\Theta}$ 上的连续可微函数；

(2) $D_N\left(\boldsymbol{\theta}\right) = \partial f_N\left(\boldsymbol{\theta}\right) / \partial \boldsymbol{\theta}^{'}$ 依概率收敛于非随机矩阵 $D\left(\boldsymbol{\theta}\right)$，且 $D\left(\boldsymbol{\theta}\right)$ 在 $\boldsymbol{\theta}$ 处连续；

(3) $\sqrt{N} f_N\left(\boldsymbol{\theta}\right)$ 满足中心极限定理

$$\sqrt{N} f_N\left(\boldsymbol{\theta}\right) = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} f\left(\boldsymbol{W}_i,\quad \boldsymbol{\theta}\right) \xrightarrow{d} N\left(0,\quad \boldsymbol{V}_0\right) \tag{4.33}$$

[④] Cramer-Rao 下界定理：如果 $\widehat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ 的无偏估计，则 $Var\left(\widehat{\boldsymbol{\theta}}\right) \ge \left(n\boldsymbol{I}_0\right)^{-1}$.

(4) 对于 $\boldsymbol{D}_0 = D(\boldsymbol{\theta})$，$\boldsymbol{D}_0^{'}\boldsymbol{A}_0\boldsymbol{D}_0$ 是非奇异矩阵[⑤]；

则

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\xrightarrow{d}N\left(0,\quad \boldsymbol{\Omega}_0\right) \tag{4.34}$$

其中，

$$\boldsymbol{\Omega}_0=\left(\boldsymbol{D}_0^{'}\boldsymbol{A}_0\boldsymbol{D}_0\right)^{-1}\boldsymbol{D}_0^{'}\boldsymbol{A}_0V_0\boldsymbol{A}_0\boldsymbol{D}_0\left(\boldsymbol{D}_0^{'}\boldsymbol{A}_0\boldsymbol{D}_0\right)^{-1} \tag{4.35}$$

证明：由 GMM 估计满足一阶条件，

$$\boldsymbol{D}_N\left(\hat{\boldsymbol{\theta}}\right)^{'}\boldsymbol{A}_N f_N\left(\hat{\boldsymbol{\theta}}\right)=0$$

由条件(2)和 GMM 估计 $\hat{\boldsymbol{\theta}}$ 的一致性，

$$\boldsymbol{D}_0^{'}\boldsymbol{A}_0\sqrt{N}f_N\left(\hat{\boldsymbol{\theta}}\right)=o_p\left(1\right)$$

由 $f_N\left(\hat{\boldsymbol{\theta}}\right)$ 在 $\boldsymbol{\theta}$ 处的一阶展开式可得

$$\boldsymbol{D}_0^{'}\boldsymbol{A}_0\left[\sqrt{N}f_N\left(\boldsymbol{\theta}\right)+D_N\left(\boldsymbol{\theta}\right)\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)\right]=o_p\left(1\right)$$

从而

$$\left(\boldsymbol{D}_0^{'}\boldsymbol{A}_0\boldsymbol{D}_0\right)\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=-\boldsymbol{D}_0^{'}\boldsymbol{A}_0\sqrt{N}f_N\left(\boldsymbol{\theta}\right)+o_p\left(1\right)$$

$$\sqrt{N}\left(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}\right)=-\left(\boldsymbol{D}_0^{'}\boldsymbol{A}_0\boldsymbol{D}_0\right)^{-1}\boldsymbol{D}_0^{'}\boldsymbol{A}_0\sqrt{N}f_N\left(\boldsymbol{\theta}\right)+o_p\left(1\right) \tag{4.36}$$

因此，由关于 $\sqrt{N}f_N\left(\boldsymbol{\theta}\right)$ 的中心极限定理即可得到(4.34).

3 GMM 估计的渐近方差及其估计

如果 $\sqrt{N}f_N\left(\boldsymbol{\theta}\right)$ 满足中心极限定理

$$\sqrt{N}f_N\left(\boldsymbol{\theta}\right)=\frac{1}{\sqrt{N}}\sum_{i=1}^{N}f\left(\boldsymbol{W}_i,\quad \boldsymbol{\theta}\right)\xrightarrow{d}\boldsymbol{N}\left(0,\quad \boldsymbol{V}_0\right),$$

且相关的矩存在，则

$$\begin{aligned}V_0&=\lim_{N\to\infty}Var\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}f\left(\boldsymbol{W}_i,\quad \boldsymbol{\theta}\right)\right)\\&=\lim_{N\to\infty}\frac{1}{N}\left(\sum_{i=1}^{N}\sum_{j=1}^{N}E\left[f\left(\boldsymbol{W}_i,\quad \boldsymbol{\theta}\right)f\left(\boldsymbol{W}_j,\quad \boldsymbol{\theta}\right)^{'}\right]\right)\end{aligned} \tag{4.37}$$

于是，当 $\left\{\boldsymbol{W}_1,\quad \cdots,\quad \boldsymbol{W}_N\right\}$ 是 $\boldsymbol{W}$ 的独立样本时，

$$\boldsymbol{V}_0=\lim_{N\to\infty}\frac{1}{N}\left(\sum_{i=1}^{N}E\left[f\left(\boldsymbol{W}_i,\quad \boldsymbol{\theta}\right)f\left(\boldsymbol{W}_i,\quad \boldsymbol{\theta}\right)^{'}\right]\right) \tag{4.38}$$

[⑤] 权重矩阵 $\boldsymbol{A}_N\xrightarrow{p}\boldsymbol{A}_0$，$\boldsymbol{A}_0$ 是半正定矩阵；

特别，当$\left\{\boldsymbol{W}_1, \quad \cdots, \quad \boldsymbol{W}_N\right\}$是 $\boldsymbol{W}$ 的独立同分布的样本时，

$$\boldsymbol{V}_0 = E\left[f\left(\boldsymbol{W}_i, \quad \boldsymbol{\theta}\right) f\left(\boldsymbol{W}_i, \quad \boldsymbol{\theta}\right)'\right] \tag{4.39}$$

$$\boldsymbol{D}_0 = E\left(\frac{\partial f\left(\boldsymbol{W}_i, \quad \boldsymbol{\theta}\right)}{\partial \boldsymbol{\theta}'}\right) \tag{4.40}$$

为了得到$\boldsymbol{\Omega}_0$的一致估计，用$\boldsymbol{A}_N$、

$$\hat{\boldsymbol{D}} = \frac{1}{N}\sum_{i=1}^{N}\frac{\partial f\left(\boldsymbol{W}_i, \quad \hat{\boldsymbol{\theta}}\right)}{\partial \boldsymbol{\theta}'} \tag{4.41}$$

和

$$\hat{\boldsymbol{V}} = \frac{1}{N}\sum_{i=1}^{N} f\left(\boldsymbol{W}_i, \quad \hat{\boldsymbol{\theta}}\right) f\left(\boldsymbol{W}_i, \quad \hat{\boldsymbol{\theta}}\right)' \tag{4.42}$$

分别代替(4.35)中的$\boldsymbol{A}_0$、$\boldsymbol{D}_0$和$\boldsymbol{V}_0$，则$\left\{\boldsymbol{W}_1, \quad \cdots, \quad \boldsymbol{W}_N\right\}$是 $\boldsymbol{W}$ 的独立同分布的样本时，

$$\hat{\boldsymbol{\Omega}}_N = \left(\hat{\boldsymbol{D}}'\boldsymbol{A}_N\hat{\boldsymbol{D}}\right)^{-1}\hat{\boldsymbol{D}}'\boldsymbol{A}_N\hat{\boldsymbol{V}}\boldsymbol{A}_N\hat{\boldsymbol{D}}\left(\hat{\boldsymbol{D}}'\boldsymbol{A}_N\hat{\boldsymbol{D}}\right)^{-1} \tag{4.43}$$

是$\boldsymbol{\Omega}$的一致估计。

于是，由(4.35)可得，

$$\left(\hat{\boldsymbol{\Omega}}_N / N\right)^{-1/2}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}\right) \xrightarrow{d} N\left(\mathbf{0}, \quad \boldsymbol{I}\right) \tag{4.44}$$

并且，矩阵$\hat{\boldsymbol{\Omega}}_N / N$的对角线元素的平方根是随机向量$\hat{\boldsymbol{\theta}}_N$分量的渐近标准差，矩阵$\hat{\boldsymbol{\Omega}}_N / N$也被称为 GMM 估计$\hat{\boldsymbol{\theta}}_N$的渐近方差矩阵估计。

4. 2SLS 估计的渐近方差

(1) 有独立异方差观测值的 2SLS

对于独立异方差观测值的 2SLS，

$$f\left(\boldsymbol{W}_i, \quad \boldsymbol{\theta}\right) = \boldsymbol{Z}_i\left(y_i - \boldsymbol{X}_i'\boldsymbol{\theta}\right) = \boldsymbol{Z}_i u_i \tag{4.45}$$

$$\boldsymbol{A}_0 = \left[E\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)\right]^{-1} \tag{4.46}$$

$$\boldsymbol{D}_0 = E\left(\boldsymbol{Z}_i\boldsymbol{X}_i'\right) \tag{4.47}$$

$$\boldsymbol{V}_0 = E\left(u_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right) \tag{4.48}$$

样本形式为

$$\boldsymbol{A}_N = \left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1} = N\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}$$

$$\hat{\boldsymbol{D}} = \frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{X}_i' = \frac{1}{N}\boldsymbol{Z}'\boldsymbol{X}$$

$$\hat{\boldsymbol{V}} = \frac{1}{N}\sum_{i=1}^{N}u_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i' = \frac{1}{N}\sum_{i=1}^{N}\left(y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\theta}}_{2SLS}\right)^2\boldsymbol{Z}_i\boldsymbol{Z}_i'$$

因此，2SLS 估计 $\hat{\boldsymbol{\theta}}_{2SLS}$ 的被估渐近方差矩阵

$$\begin{aligned}\hat{\boldsymbol{\Omega}}_N/N &= \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\left(\boldsymbol{X}'\boldsymbol{Z}\right)\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\sum_{i=1}^{N}u_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)\left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1} \\ &= \left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\left(\sum_{i=1}^{N}\hat{u}_i^2\hat{\boldsymbol{Z}}_i\hat{\boldsymbol{Z}}_i'\right)\left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1}\end{aligned} \quad (4.49)$$

其中，$\hat{\boldsymbol{X}} = \boldsymbol{Z}\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{X}\right) = \boldsymbol{Z}\hat{\boldsymbol{\delta}}$，$\hat{\boldsymbol{X}} = \boldsymbol{Z}\hat{\boldsymbol{\delta}}$。

(2) 独立同分布的 2SLS

对于独立同分布的 2SLS，条件方差

$$E\left(u_i^2 \mid z_i\right) = \sigma^2 \quad (4.50)$$

不依赖于 $\boldsymbol{Z}_i$，于是，2SLS 矩条件的方差矩阵

$$\boldsymbol{V}_0 = E\left[E\left(u_i^2 \mid \boldsymbol{Z}_i\right)\boldsymbol{Z}_i\boldsymbol{Z}_i'\right] = \sigma^2 E\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right) \quad (4.51)$$

从而，

$$\boldsymbol{\Omega}_0 = \sigma^2\left\{E\left(\boldsymbol{X}_i\boldsymbol{Z}_i'\right)\left[E\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)\right]^{-1}E\left(\boldsymbol{Z}_i\boldsymbol{X}_i'\right)\right\}^{-1} = \sigma^2\left[E\left(\boldsymbol{\delta}\boldsymbol{Z}_i\boldsymbol{Z}_i'\boldsymbol{\delta}'\right)\right]^{-1} \quad (4.52)$$

其中，$\boldsymbol{\delta} = E\left(\boldsymbol{X}_i\boldsymbol{Z}_i'\right)\left[E\left(\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)\right]^{-1}$。

所以，$\boldsymbol{\Omega}_0$ 的一致估计

$$\hat{\boldsymbol{\Omega}}_N/N = \hat{\sigma}^2\left[\left(\boldsymbol{X}'\boldsymbol{Z}\right)\left(\boldsymbol{Z}'\boldsymbol{Z}\right)^{-1}\left(\boldsymbol{Z}'\boldsymbol{X}\right)\right]^{-1} = \hat{\sigma}^2\left(\hat{\boldsymbol{X}}'\hat{\boldsymbol{X}}\right)^{-1} \quad (4.53)$$

其中，$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1}^{N}\hat{u}_i^2$，$\hat{u}_i$ 是 2SLS 估计的残差。

§4.5 最佳加权矩阵（Optimal Weight Matrix）

从前面的讨论可见，GMM 估计的加权矩阵选择对参数估计及其方差估计具有重要的影响。因此，依据 Sargan（1958）的思路，使用 GMM 估计时，在给定的矩条件（或正交条件）中确定使 GMM 估计 $\hat{\boldsymbol{\theta}}_N$ 最优的矩条件线性组合是至关重要的。

1 最佳加权矩阵

事实上，使得 $\boldsymbol{A}_0 = m\boldsymbol{V}_0^{-1}$ 的加权矩阵 $\boldsymbol{A}_N$ 是 GMM 估计的最佳加权矩阵，其中，m 是任意正数。这时，渐近方差矩阵

$$\hat{\boldsymbol{\Omega}}_N = \left(\boldsymbol{D}_0'\boldsymbol{V}_0^{-1}\boldsymbol{D}_0\right)^{-1} \qquad (4.54)$$

否则，

$$\left(\boldsymbol{D}_0'\boldsymbol{A}_0\boldsymbol{D}_0\right)^{-1}\boldsymbol{D}_0'\boldsymbol{A}_0\boldsymbol{V}_0\boldsymbol{A}_0\boldsymbol{D}_0\left(\boldsymbol{D}_0'\boldsymbol{A}_0\boldsymbol{D}_0\right)^{-1} - \left(\boldsymbol{D}_0'\boldsymbol{V}_0^{-1}\boldsymbol{D}_0\right)^{-1}$$
$$= \bar{\boldsymbol{D}}\left[\boldsymbol{I} - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right]\bar{\boldsymbol{D}}'$$

其中，$\bar{\boldsymbol{D}} = \left(\boldsymbol{D}_0'\boldsymbol{A}_0\boldsymbol{D}_0\right)^{-1}\boldsymbol{D}_0'\boldsymbol{A}_0\boldsymbol{V}_0^{1/2}$，$\boldsymbol{H} = \boldsymbol{V}_0^{-1/2}\boldsymbol{D}_0$

由于$\left[\boldsymbol{I} - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right]$是幂等矩阵，所以

$$\bar{\boldsymbol{D}}\left[\boldsymbol{I} - \boldsymbol{H}\left(\boldsymbol{H}'\boldsymbol{H}\right)^{-1}\boldsymbol{H}'\right]\bar{\boldsymbol{D}}' \geq \boldsymbol{0}$$ [⑥]

由此可见，为了得到 GMM 的最佳估计$\hat{\boldsymbol{\theta}}_N$，必须构造$\boldsymbol{V}_0$的一致估计，如(4.43)式的估计。

2 2SLS 估计的最佳加权矩阵

对于有独立同分布观测值的 2SLS 估计，由(4.46)和(4.51)，$\boldsymbol{A}_0 = \sigma^{-2}\boldsymbol{V}_0^{-1}$.所以，对于 2SLS 估计，$\boldsymbol{A}_N = \left(\boldsymbol{Z}'\boldsymbol{Z}/N\right)^{-1}$是最佳加权矩阵。

对于有独立异方差观测值的 2SLS 估计，(4.48)式估计的$\hat{\boldsymbol{\Omega}}_N$亦然是 2SLS 渐近方差的一致估计，则用稳健加权矩阵$\left(\sum_{i=1}^{N}u_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}$代替 2SLS 的加权矩阵$\left(\boldsymbol{Z}'\boldsymbol{Z}/N\right)^{-1}$的二阶段最优 GMM 估计（Two-step Optimal GMM Estimators）[⑦]

$$\tilde{\boldsymbol{\theta}} = \left[\boldsymbol{X}'\boldsymbol{Z}\left(\sum_{i=1}^{N}u_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\boldsymbol{Z}'\boldsymbol{X}\right]^{-1}\boldsymbol{X}'\boldsymbol{Z}\left(\sum_{i=1}^{N}u_i^2\boldsymbol{Z}_i\boldsymbol{Z}_i'\right)^{-1}\boldsymbol{Z}'\boldsymbol{Y} \qquad (4.55)$$

当 u_i 关于 $\boldsymbol{Z}_i$ 的条件方差$E\left(u_i^2 \mid \boldsymbol{Z}_i\right)$依赖于 $\boldsymbol{Z}_i$ 时，由于(4.54)式估计的$\tilde{\boldsymbol{\Omega}}_N$并不是 2SLS 渐近方差的一致估计，所以，2*SLS* 估计仅是 GMM 估计中的次优估计。

§4.6 过度识别约束的检验

在 GMM 中，由于工具变量的个数 $l > k$，(4.12)式中存在 l 个识别参数 $\boldsymbol{\theta}$ 的约束，称(4.12)式为 GMM 估计的过度识别约束。将检验过度识别约束是否有效的检验称为过度识别约束检验（Testing the Overidentifying Restrictions）。

按照 Sargan 的意义，GMM 估计方法就是令 l 个样本矩方程 $f_N(\boldsymbol{\theta})$的 k 个线性组合为零来估计参数 $\boldsymbol{\theta}$ 的方法。所以，当模型设定正确时，$f_N(\hat{\boldsymbol{\theta}})$中存在 l-k 个线性独立的组合应该接近于零，但不一定恰好为零。于是，可以构造渐近服从 $\chi^2(l$-$k)$分布的统计量

---

⑥ B≥0 表示 B 是半正定矩阵。

⑦ White,H.(1982)

$$s_{\hat{\theta}} = Nf_N\left(\hat{\boldsymbol{\theta}}\right)^{'}\hat{\boldsymbol{V}}^{-1}f_N\left(\hat{\boldsymbol{\theta}}\right)\xrightarrow{d}\chi^2\left(l-k\right) \tag{4.56}$$

来推断总体矩条件的相应 l -k 个线性组合是否真的等于零。即，检验假设

H_0：l 个总体矩条件成立，即模型设定正确；

H_1：l 个总体矩条件不全成立，即模型设定错误。

其中，$\hat{\boldsymbol{\theta}}$ 是 $\boldsymbol{\theta}$ 的最佳 GMM 估计，$\hat{\boldsymbol{V}}$ 是 $\boldsymbol{V}_0$ 的一致估计。利用统计量 $s_{\hat{\theta}}$ 对过度识别约束有效性的检验被称为 Sargan 检验。显然，拒绝了 Sargan 的过度识别约束检验的零假设意味着模型设定错误，尤其，样本数据不支持 l 个总体矩条件全部成立。

参考文献

Laszlo Matyas, Generalized Method of Moments Estimation, Cambridge University Press, 1999.