引入 GMM 的目的和两大作用

参数估计方法主要有最小二乘法，极大似然估计法，矩估计法。

GMM 的引入有一定的意义。

首先，处理大样本数据时，如果不知道变量的分布，GMM 估计方法优于 MLE。

极大似然估计法[1]和广义矩估计法[2]被用于大样本情况下模型参数的估计，它们在大样本条件下显示了优良的性质。但是，使用极大似然估计法的前提是已知变量的分布，只是不知道其参数。此时，分布的设定难免有人为的因素，一旦给出了错误的分布假定，MLE 估计量通常是**有偏的**。而广义矩估计方法不需要对变量的分布进行假定，它只需要找到一些矩条件而不是整个密度函数；基于模型实际参数满足一定的矩条件，GMM 估计量总是**一致的**。当然，GMM 有时不能对样本中的全部信息进行有效利用。此外，只有当样本很大时，GMM 估计量才是**渐近有效**的。而在小样本中尽管也是一致的，但却不是有效的。其拥有十分良好的大样本性质，它的小样本性质并不令人满意。

GMM 与 MLE 的差别还在于，在某些情况下，GMM 的计算较 MLE 方便。

其次，GMM 是对工具变量法的补充和完善。当解释变量存在内生性时，研究者是通过寻找工具变量来解决此问题。当工具变量满足的矩条件个数大于待估计参数个数时，方程组的解不唯一。此时，运用广义矩估计法，即使不能够从一阶条件得出解，但它仍能对未知参数进行估计，并且得到渐进有效的参数估计值。

第三，GMM 估计方法允许模型设定中存在异方差和相关性。GMM 方法中权矩阵的选取，考虑了随机误差项的异方差和自相关。即针对异方差或者自相关，GMM 估计方法有不同的权矩阵构造方法。

再次，假设检验。GMM 模型中，设计了多个统计量对模型设定及其参数进行假设检验。

传统的计量经济学估计方法，例如普通最小二乘法、工具变量法和极大似然法等都存在自身的局限性。即其参数估计量必须在满足某些假设时，比如模型的随机误差项服从正态分布或某一已知分布时，才是可靠的估计量。解释变量外生时，估计的参数才是无偏的。而 GMM 不需要知道随机误差项的准确分布信息，允许随机误差项存在异方差和序列相关，因而所得到的参数估计量比其他参数估计方法更有效。

作用：

解决内生变量问题：

解决随机误差项存在异方差和序列相关问题：

参数，一致、渐进正态性

J 统计量，渐进卡方性，可以检验模型及其参数。

变量分布未知问题

---

[1]极大似然估计法(maximumlikelihoodmethod，ML)的应用虽然没有普通最小二乘法广泛，但它是一个具有更强理论性质的点估计方法，它以极大似然原理为基础，通过概率密度函数或者分布律来估计总体参数。对于一些特殊类型的计量经济模型，如我们后面将介绍的 Logit 和 Probit 模型，最小二乘法不再适用，极大似然法成为首选的估计方法。极大似然法的思路 极大似然估计法的出发点是已知被观测现象的分布，但不知道其参数。极大似然法用得到观测值(样本)最高概率的那些参数的值来估计该分布的参数，从而提供了一种用于估计刻画一个分布的一组参数的方法。

[2]广义矩方法确实是一种具有高度概括性的方法。其他的参数估计量可以看做它的特例。比如最小二乘法估计量(OLS)和最大似然估计量(MLE)都是 GMM 估计量的特例。当待估参数较多时。最大似然估计需要较为复杂的数值求解。GMM 估计更加方便。GMM 在时间序列及面板数据分析等许多场合有着广泛的应用

GMM 用于线性模型和非线性模型的区别：

$$\hat{\beta} = \arg\min(g(\beta)'W^{-1}g(\beta))$$

J 统计量的表达式中 g 在线性模型中能够用样本信息表达出来，
而在非线性模型中，只能表达成隐函数。
在 GMM 统计量、参数估计方法、假设检验上，
GMM 在线性模型和非线性模型中运用相同。

$$g(\beta) = \begin{pmatrix} g_1(\beta) \\ g_2(\beta) \\ \vdots \\ g_J(\beta) \end{pmatrix} = \begin{pmatrix} \frac{1}{n}\sum_i z_{1i}e_i \\ \frac{1}{n}\sum_i z_{2i}e_i \\ \vdots \\ \frac{1}{n}\sum_i z_{Ji}e_i \end{pmatrix}$$

2STS vs GMM

$$y_i = X_i'\beta + \mu_i \qquad i = 1, \cdots, n$$

$$X = Z\pi + \mu$$
$$\hat{\pi} = (Z'Z)^{-1}Z'X$$
$$\hat{X} = Z\hat{\pi} = Z(Z'Z)^{-1}Z'X$$

第1阶段估计

$$\boxed{\hat{\beta} = (\hat{X}'X)^{-1}\hat{X}'Y = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z)(Z'Z)^{-1}Z'Y}$$

第2阶段估计

$$\boxed{\hat{\beta} = Arg\ \min\{\frac{1}{n}[Z'(Y - X\beta)]' \cdot \hat{W}^{-1} \cdot \frac{1}{n}[Z'(Y - X\beta)]\}}$$

$$(X'Z)\hat{W}^{-1}(Z'Y - Z'X\hat{\beta}) = 0$$

一阶导

$$(X'Z)\hat{W}^{-1}Z'Y - (X'Z)\hat{W}^{-1}Z'X\hat{\beta} = 0$$
$$(X'Z\hat{W}^{-1}Z'X)^{-1}(X'Z)\hat{W}^{-1}Z'Y = \hat{\beta}$$
$$\hat{W} = \frac{\hat{\sigma}^2}{n^2}Z'Z$$

$$\boxed{\hat{\beta} = (X'Z(Z'Z)^{-1}Z'X)^{-1}(X'Z)(Z'Z)^{-1}Z'Y}$$

GMM 与 OLS 、MLE、IV、 GEE 的联系和区别

1， OLS：选择解释变量作为工具变量构造矩条件，权利矩阵为单位阵，GMM 即为 OLS。参数估计值相同,参数估计量的方差—协方差矩阵一般不相同.

2，MLE：用对数似然函数的导数构造等于 0 的矩条件时，所表示的 GMM 等价于 ML 。

3，IV：GMM 中方程个数等于参数个数时，即等价于工具变量估计法。

4，2STS: 2SLS 是工具变量估计方法的特殊情形，而工具变量估计是 GMM 估计的特殊情形。如果 GMM 中利用了所有先决变量， 2SLS 与 GMM 估计等价。如上图。

5，与 GEE 的对比，GMM 的使用场合是估计方程的个数大于未知参数的个数。而这两者相等时，可以使用 GEE。

GMM 的检验

对于 GMM，关键是两项检验：

一是检验过度识别限制是否有效。 **Ho:E(Z'U)=0** 即 J>k 的那部分是否有效。如果经过检验无效，那么 GMM 在这个意义上就没有优越性。如果拒绝原假设，意味着并非所有的总体矩条件都成立。如果拒绝原假设，而且没有进一步的信息，就不能判断哪个矩条件不成立，或者说哪个工具变量无效。

当 l=K 时，称模型参数“恰好识别”，这时不论总体矩条件是否真的成立，都存在唯一解，意味着当 l=K 时，总体矩条件不可检验。

当 l＞K 时，称模型参数“过度识别”，该检验称为过渡识别约束检验。

过渡识别约束检验也称为 Sargan 检验。

即，软件给出的结果是 J，但是判断时，要使用 nJ. 例如，某个模型计算得到 J=0.029837, 那么，nJ=0.477, 5%的显著性水平下，自由度为 1 的卡方分布的临界值为 3.84。接受过度识别的矩条件为真的假设。

在应用软件中，被称为J统计量，但应该是nJ~$\chi^2(l\text{-}K)$

二是检验构造的矩条件是否成立。如果矩条件不成立，就要从模型设定方面寻找原因。

另外，如果对模型参数施加约束，则需要进行**参数约束检验**。

大数定律及中心极限定理　（依概率收敛）

切比雪夫大数定律，伯努利大数定律，辛钦定理

设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有相同的数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots)$. 作前 n 个随机变量的算术平均

$$\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$$

独立同分布

则对于任意正数 ε，有

$$\lim_{n\to\infty}P\{|\overline{X}-\mu|<\varepsilon\}=\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}=1.$$

这样，上述定理又可叙述为：

定理一　设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，且具有相同的数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots)$，则序列 $\overline{X}=\frac{1}{n}\sum_{k=1}^{n}X_k$ 依概率收敛于 μ，即 $\overline{X}\xrightarrow{P}\mu$.

定理二（伯努利大数定理）　设 n_A 是 n 次独立重复试验中事件 A 发生的次数. p 是事件 A 在每次试验中发生的概率，则对于任意正数 $\varepsilon>0$，有

$$\lim_{n\to\infty}\left\{\left|\frac{n_A}{n}-p\right|<\varepsilon\right\}=1$$

或

$$\lim_{n\to\infty}\left\{\left|\frac{n_A}{n}-p\right|\geqslant\varepsilon\right\}=0.$$

定理三（辛钦定理）　设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有数学期望 $E(X_k)=\mu (k=1,2,\cdots)$，则对于任意正数 ε，有

$$\lim_{n\to\infty}P\left\{\left|\frac{1}{n}\sum_{k=1}^{n}X_k-\mu\right|<\varepsilon\right\}=1.$$

独立同分布中心极限定理，Liapunov 定理，拉普拉斯极限定理

2. 中心极限定理

（1）独立同分布的中心极限定理

（林德伯格—列维（Lindeberg—Levy）定理）：设随机变量 $X_1, X_2, \cdots, X_n, \cdots$ 相互独立，服从同一分布，且具有相同的数学期望和方差：$E(X_k)=\mu, D(X_k)=\sigma^2 (k=1,2,\cdots)$. 则随机变量

独立同分布

$$Y_n=\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}$$

的分布函数 $F_n(x)$ 收敛到标准正态分布. 即对任意 x 满足

$$\lim_{n\to+\infty}F_n(x)=\lim_{n\to+\infty}P\left\{\frac{\sum_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}dt=\Phi(x).$$

意义：均值为 μ，方差为 $\sigma^2>0$ 的独立同分布的随机变量 $X_1,X_2,\cdots,X_n$ 的和 $\sum\limits_{k=1}^{n}X_k$ 的标准化变量，当 n 充分大时，有

$$Y_n=\frac{\sum\limits_{k=1}^{n}X_k-n\mu}{\sqrt{n}\sigma}\sim N(0,1)$$

(2) 李雅普诺夫(liapunov) 定理

设随机变量 $X_1,X_2,\cdots,X_n,\cdots$ 相互独立，$E(X_k)=\mu_k,D(X_k)=\sigma_k^2(k=1,2,\cdots)$ 存在. 记 $B_n^2=\sum\limits_{k=1}^{n}\sigma_k^2$. 若存在正数 δ，使得当 $n\to+\infty$ 时，

$$\frac{1}{B_n^{2+\delta}}\sum_{k=1}^{n}E\{\mid X_k-\mu_k\mid^{2+\delta}\}\to 0$$

则随机变量之和 $\sum\limits_{k=1}^{n}X_k$ 的标准化变量

$$Z_n=\frac{\sum\limits_{k=1}^{n}X_k-E\left(\sum\limits_{k=1}^{n}X_k\right)}{\sqrt{D\left(\sum\limits_{k=1}^{n}X_k\right)}}=\frac{\sum\limits_{k=1}^{n}X_k-\sum\limits_{k=1}^{n}\mu_k}{\sqrt{B_n}}$$

的分布函数 $F_n(x)$ 对任意 x 满足

$$\lim_{n\to\infty}F_n(x)=\lim_{n\to\infty}P\left\{\frac{\sum\limits_{k=1}^{n}X_k-\sum\limits_{k=1}^{n}\mu_k}{B_n}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt=\Phi(x)$$

(3) 德莫弗—拉普拉斯(De Moivre－Laplace) 定理

设随机变量 $\eta_n(n=1,2,\cdots)$ 服从参数为 $n,p(0<p<1)$ 的二项分布，则对于任意 x 有

$$\lim_{n\to\infty}P\left\{\frac{\eta_n-np}{\sqrt{np(1-p)}}\leqslant x\right\}=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt=\Phi(x)$$

显然 $\eta_n=\sum\limits_{k=1}^{n}X_k$，其中 $X_1,X_2,\cdots,X_n$ 独立同服从 $B(1,p)$ 分布.

二　经典矩方法

1. 经典矩方法的步骤

用样本矩去代替总体矩，得到参数的估计，这种方法称为矩方法（MOM），也称为经典矩方法。

假设有一组随机样本 x_1，x_2，…，x_n，服从 $k\times1$ 维参数向量为 $\boldsymbol{\theta}$ 的一种分布。用经典矩方法来估计 $\boldsymbol{\theta}$ 的步骤如下：

首先，分别计算样本数据的 j 阶总体矩，$j=1$，2，…，k，它们是 $\boldsymbol{\theta}$ 的函数：

$$E(x_i^j)=\mu_j(\boldsymbol{\theta}),\ j=1,2,\cdots,k \tag{2.1}$$

其中 x_i^j 是 x_i 的 j 次方，它们可以不必是前 k 阶总体矩，但它们的形式一定要简单。包含参数 $\boldsymbol{\theta}$ 的总体矩函数就是我们要求解的经济模型。

其次，根据大数定律（LLN），各阶样本矩依概率收敛于各阶总体矩，即：

$$\frac{1}{n}\sum_{i=1}^{n}x_i^j\xrightarrow{p}E(x_i^j),\ j=1,2,\cdots,k \tag{2.2}$$

最后，用样本矩代替总体矩，解 k 个矩方程构成的方程组，解出来的 $\hat{\boldsymbol{\theta}}_{MM}$ 就是 $\boldsymbol{\theta}$ 的矩法估计量：

$$\frac{1}{n}\sum_{i=1}^{n}x_i^j=\mu_j(\hat{\boldsymbol{\theta}}_{MM}),\ j=1,2,\cdots,k\ \Rightarrow\ \hat{\boldsymbol{\theta}}_{MM} \tag{2.3}$$

2. 向量值函数的经典矩方法

假定根据理论，有 k 个基于样本数据的函数，它们是一组未知参数 $\boldsymbol{\theta}$ 的函数。向量值函数的经典矩估计可按照以下步骤进行：

首先，构造向量值函数 $\mathbf{g}(x_i,\ \boldsymbol{\theta})$，它们是随机样本 x_i 和 $k\times1$ 维参数向量 $\boldsymbol{\theta}$ 的函数，$\mathbf{g}(x_i,\ \boldsymbol{\theta})$ 可以是隐函数，使其 j 阶总体矩为零（$j=1$，2，…，k）：

如何构造这些向量值函数．如：$y=x\beta+\varepsilon_i$，x 是外生变量时。

$$\sum_{i=1}^{n}x_i\varepsilon_i=0$$

这个说法有点不对。不是使 j 阶总体矩为零，而是存在 k 个矩条件。如 OLS 方法。

$$E[g_j(x_i,\boldsymbol{\theta})]=0,\ j=1,2,\cdots,k \tag{2.4}$$

包含参数 $\boldsymbol{\theta}$ 的零总体矩函数就是我们要求解的经济模型。

其次，向量值函数 $g_j(x_i,\ \boldsymbol{\theta})$ 的各阶样本矩函数依概率收敛于其各阶总体矩：

$$\frac{1}{n}\sum_{i=1}^{n}g_j(x_i,\boldsymbol{\theta})\xrightarrow{p}E\left[g_j(x_i,\boldsymbol{\theta})\right],\ j=1,2,\cdots,k \tag{2.5}$$

最后，用样本矩函数代替其零总体矩，解 k 个方程构成的方程组，解出来的 $\hat{\boldsymbol{\theta}}_{MM}$ 就是 $\boldsymbol{\theta}$ 的经典矩估计量：

$$\frac{1}{n}\sum_{i=1}^{n}g_j(x_i,\hat{\boldsymbol{\theta}}_{MM})=0,\ j=1,2,\cdots,k\ \Rightarrow\ \hat{\boldsymbol{\theta}}_{MM} \tag{2.6}$$

三　经典矩估计量渐近协方差的计算

这里考虑一般的情况，考虑向量值函数的经典矩估计量的渐近方差和协方差。令样本矩函数向量为：

$$\frac{1}{n}\sum_{i=1}^{n}\underset{(k\times1)}{\mathbf{g}}(x_i,\boldsymbol{\theta}) = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} g_1(x_i,\boldsymbol{\theta}) \\ \frac{1}{n}\sum_{i=1}^{n} g_2(x_i,\boldsymbol{\theta}) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} g_k(x_i,\boldsymbol{\theta}) \end{bmatrix} \tag{2.7}$$

根据中心极限定理（CLT）有：

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\underset{(k\times1)}{\mathbf{g}}(x_i,\boldsymbol{\theta}) - \mathbf{0}\right) = \sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^{n}\underset{(k\times1)}{\mathbf{g}}(x_i,\boldsymbol{\theta}) \xrightarrow{d} N(0,\underset{(k\times k)}{\boldsymbol{\Omega}}) \tag{2.8}$$

其中协方差矩阵 $\underset{(k\times k)}{\boldsymbol{\Omega}}$ 为：

$$\underset{(k\times k)}{\boldsymbol{\Omega}} = E\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)\cdot\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)'\right] \tag{2.9}$$

即：

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta}) \xrightarrow{d} N\left(0,\frac{1}{n}\boldsymbol{\Omega}\right) \tag{2.10}$$

$$\mathbf{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right) = \frac{1}{n}\boldsymbol{\Omega} \tag{2.11}$$

令 $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{MM}$，在真值 $\boldsymbol{\theta}$ 处，对经典矩估计量 $\hat{\boldsymbol{\theta}}$ 的样本矩函数 $\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})$ 进行一阶泰勒展开，可得：

$$\mathbf{0} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}}) \approx \frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta}) + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) \tag{2.12}$$

令：

$$\underset{(k\times k)}{\mathbf{G}(\boldsymbol{\theta})} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \tag{2.13}$$

则：

$$\begin{aligned}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta}) &= -\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)^{-1}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta}) \\ &= -\mathbf{G}^{-1}(\boldsymbol{\theta})\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta}) \\ &\xrightarrow{d} -\mathbf{G}^{-1}(\boldsymbol{\theta})\cdot N\left(\mathbf{0},\frac{1}{n}\boldsymbol{\Omega}\right) \\ &\xrightarrow{d} N\left\{\mathbf{0},\mathbf{G}^{-1}(\boldsymbol{\theta})\cdot\frac{1}{n}\boldsymbol{\Omega}\cdot[\mathbf{G}^{-1}(\boldsymbol{\theta})]'\right\}\end{aligned} \tag{2.14}$$

因此，经典矩估计量的渐近协方差为：

$$\mathbf{Var}(\hat{\boldsymbol{\theta}}_{MM}) = \frac{1}{n}\mathbf{G}^{-1}\boldsymbol{\Omega}(\mathbf{G}^{-1})' \tag{2.15}$$

广义矩估计

一 广义矩方法

上节我们考虑了 k 个未知参数、k 个矩条件的情形，下面假定我们遇到了 k 个未知参数、m 个矩条件（$m>k$）的情形。

通过构造隐函数，使其各阶总体矩为 0，矩条件为如下形式：

How?如，IV。及此页下面这个例子

$$E[\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})]=E\begin{bmatrix} g_1(x_i,\boldsymbol{\theta}) \\ g_2(x_i,\boldsymbol{\theta}) \\ \vdots \\ g_m(x_i,\boldsymbol{\theta}) \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m\times 1} \qquad (2.16)$$

包含参数 $\boldsymbol{\theta}$ 的零总体矩函数就是我们要求解的经济模型。用样本矩代替总体矩，则样本矩条件为：

$$\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})=\begin{bmatrix} \frac{1}{n}\sum_{i=1}^{n} g_1(x_i,\boldsymbol{\theta}) \\ \frac{1}{n}\sum_{i=1}^{n} g_2(x_i,\boldsymbol{\theta}) \\ \vdots \\ \frac{1}{n}\sum_{i=1}^{n} g_m(x_i,\boldsymbol{\theta}) \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m\times 1} \qquad (2.17)$$

这里有 m 个方程，只有 k 个未知数，当 $m>k$ 时，上述模型是过度识别的。广义矩方法不是像经典矩方法那样，把多余的矩条件舍弃，而是把 m 个样本矩条件的 k 个线性组合设定为零。

找一个 k 行 m 列的矩阵 $\mathbf{A}_{(k\times m)}$，使它满足：

经典矩估计中，只取K阶样本矩

$$\underset{(k\times m)}{\mathbf{A}}\cdot\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right]=\underset{(k\times 1)}{\mathbf{0}} \qquad (2.18)$$

例如，对两个矩条件、一个参数（$m=2$，$k=1$）的情形，式（2.18）可写为：

$$[a_1 \quad a_2]\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} g_1(x_i,\theta) \\ \frac{1}{n}\sum_{i=1}^{n} g_2(x_i,\theta) \end{pmatrix}=0$$

即：

$$a_1\frac{1}{n}\sum_{i=1}^{n} g_1(x_i,\theta)+a_2\frac{1}{n}\sum_{i=1}^{n} g_2(x_i,\theta)=0$$

解之可得广义矩估计量 $\hat{\theta}_{GMM}$。

E（XU）=0，X=X1,X3；
E（U）=0；
E（ZU）=0，Z=Z1,Z2

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\beta_3x_{3i}+\mu_i \qquad i=1,2,\cdots,n$$

- 如果**x2**为随机变量，**z1**、**z2** 为它的工具变量，**GMM**关于参数估计量的矩条件为：

$$\begin{cases} \sum y_i-\sum(\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i})=0 \\ \sum y_ix_{1i}-\sum(\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i})x_{1i}=0 \\ \sum y_iz_{1i}-\sum(\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i})z_{1i}=0 \\ \sum y_iz_{2i}-\sum(\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i})z_{2i}=0 \\ \sum y_ix_{3i}-\sum(\hat{\beta}_0+\hat{\beta}_1x_{1i}+\hat{\beta}_2x_{2i}+\hat{\beta}_3x_{3i})x_{3i}=0 \end{cases}$$

5个等于0的矩条件，求解4个参数

该方程组是如何得到的？

如何求解该方程？

选择不同的矩阵 $\mathbf{A}_{(k\times m)}$，将得到不同的 *GMM* 估计量。矩阵 $\mathbf{A}$ 类似于一个工具变量（*IV*）矩阵，选择不同的矩阵 $\mathbf{A}$，将得到不同的估计量。当 $m=k$ 时，上述模型是恰好识别的，此时选择任何可逆矩阵 $\mathbf{A}$，都将得到相同的 *GMM* 估计量，因为在这种情形下矩阵 $\mathbf{A}$ 是多余的，可以消去。

该问题可以转化为一个最小化目标函数的问题，从而把矩阵 $\mathbf{A}_{(k\times m)}$ 的选择转化为对称权矩阵 $\mathbf{W}_{(m\times m)}$ 的最优选择。令目标函数（或损失函数）为：

损失函数?

对称权矩阵是m*m阶的W 矩阵反映了 g 的重要性

$$Loss = \left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right]'_{(1\times m)}\underset{(m\times m)}{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right] \quad (2.19)$$

最小化该损失函数，解出来的 $\boldsymbol{\theta}$ 就是广义矩估计量 $\hat{\boldsymbol{\theta}}_{GMM}$，即：

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right]'\mathbf{W}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right] \quad (2.20)$$

为了书写方便，令 $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{GMM}$，最小化损失函数的 k 个一阶条件（*FOC*）为：

$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right]'_{(k\times m)}\underset{(m\times m)}{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})\right]_{(m\times 1)} = \underset{(k\times 1)}{\mathbf{0}} \quad (2.21)$$

令：

$$\underset{(m\times k)}{\mathbf{G}(\boldsymbol{\theta})} = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'} \quad (2.22)$$

与权矩阵 $\mathbf{W}$ 相对应的矩阵 $\mathbf{A}$，便可以通过下面的关系式得到：

$$\underset{(k\times m)}{\mathbf{A}} = \left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right]'_{(k\times m)}\underset{(m\times m)}{\mathbf{W}} = \underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\underset{(m\times m)}{\mathbf{W}} \quad (2.23)$$

权矩阵 $\mathbf{W}$ 的选择标准，应使广义矩估计量 $\hat{\boldsymbol{\theta}}_{GMM}$ 最有效。最优权矩阵 $\mathbf{W}$ 的选择将在下一节中专门论述。

二 一般广义矩估计量的性质

假定数据是独立同分布（*i.i.d.*）的，在大样本情况下，一般 *GMM* 估计量具有一致性和渐近正态性。

1. 一致性

$$\text{plim}\ (\hat{\boldsymbol{\theta}}_{GMM}) = \boldsymbol{\theta} \quad (2.24)$$

$$\hat{\boldsymbol{\theta}}_{GMM}\xrightarrow{p}\boldsymbol{\theta} \quad (2.25)$$

2. 渐近正态性

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_{GMM}-\boldsymbol{\theta})\xrightarrow{d}N\{\mathbf{0},[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Omega}\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\} \quad (2.26)$$

$$\text{Var}(\hat{\boldsymbol{\theta}}_{GMM}) = \frac{1}{n}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Omega}\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1} \quad (2.27)$$

证明：

令 $\hat{\boldsymbol{\theta}}=\hat{\boldsymbol{\theta}}_{GMM}$，根据式（2.21），最小化损失函数的一阶条件为：

$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\right]'_{(k\times m)}\underset{(m\times m)}{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\hat{\boldsymbol{\theta}})\right] = \underset{(k\times 1)}{\mathbf{0}}$$

在真值 $\boldsymbol{\theta}$ 处，对一般广义矩估计量 $\hat{\boldsymbol{\theta}}$ 的样本矩函数 $\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})$ 进行一阶泰勒展开，可得：

$$\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\hat{\boldsymbol{\theta}}) \approx \frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta}) + \left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}'}\right)_{(m\times k)}\underset{(k\times 1)}{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})} \quad (2.28)$$

如，以下说明以某一标准选择权矩阵

关于权矩阵的选择，是GMM估计方法的一个核心问题。

$$\hat{\beta} = \arg\min\ (m(\beta)'W^{-1}m(\beta))$$

权矩阵可根据每个样本矩条件估计的精确程度来设置（用方差来度量）。例如，对估计较精确的矩条件给予较大的权重，对估计较不精确的矩条件给予较小的权重。

$$W = \frac{1}{n^2}\sum_i\sum_j Cov[Z_i\varepsilon_i, Z_j\varepsilon_j]$$

如此构造权矩阵体现了上述设置权矩阵的原则。*类似于加权最小二乘法*

权矩阵调整的是 J 个矩条件之间的关系，而不是 n 个样本点之间的关系。

W 应是[(1/n)Var(Z'ε)]$^{-1}$ 的一致估计。

$$\hat{\beta} = \arg\min(\mathbf{m}(\beta)'_{(1\times J)}\mathbf{W}^{-1}_{(J\times J)}\mathbf{m}(\beta)_{(J\times 1)})$$

于是最小化函数的一阶条件可写为：

$$\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]'_{(k\times m)}\underset{(m\times m)}{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\hat{\boldsymbol{\theta}})\right]$$

$$=\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]'_{(k\times m)}\underset{(m\times m)}{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right]$$

$$+\left[\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}'}\right]'_{(k\times m)}\underset{(m\times m)}{\mathbf{W}}\left(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \mathbf{g}(x_i,\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'}\right)_{(m\times k)}\underset{(k\times 1)}{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})}$$

$$=\underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\cdot\underset{(m\times m)}{\mathbf{W}}\cdot\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right]+\underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\cdot\underset{(m\times m)}{\mathbf{W}}\cdot\underset{(m\times k)}{\mathbf{G}(\boldsymbol{\theta})}\cdot\underset{(k\times 1)}{(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})}$$

$$=\underset{(k\times 1)}{\mathbf{0}}$$

故有：

$$\sqrt{n}(\underset{(k\times 1)}{\hat{\boldsymbol{\theta}}}-\boldsymbol{\theta})=-\left[\underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\cdot\underset{(m\times m)}{\mathbf{W}}\cdot\underset{(m\times k)}{\mathbf{G}(\boldsymbol{\theta})}\right]^{-1}\underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\cdot\underset{(m\times m)}{\mathbf{W}}\cdot\left[\sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right]_{(m\times 1)}\tag{2.29}$$

因为：

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})-\mathbf{0}\right)=\sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\xrightarrow{d}N(\mathbf{0},\underset{(m\times m)}{\boldsymbol{\Omega}})\tag{2.30}$$

其中样本矩的协方差矩阵 $\boldsymbol{\Omega}_{(m\times m)}$ 为：

$$\underset{(m\times m)}{\boldsymbol{\Omega}}=E\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)\cdot\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)'\right]\tag{2.31}$$

所以：

$$\sqrt{n}(\underset{(k\times 1)}{\hat{\boldsymbol{\theta}}}-\boldsymbol{\theta})$$

$$=-\left[\underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\cdot\underset{(m\times m)}{\mathbf{W}}\cdot\underset{(m\times k)}{\mathbf{G}(\boldsymbol{\theta})}\right]^{-1}\underset{(k\times m)}{\mathbf{G}'(\hat{\boldsymbol{\theta}})}\cdot\underset{(m\times m)}{\mathbf{W}}\cdot\left[\sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right]$$

$$\xrightarrow{d}-[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\cdot N(\mathbf{0},\boldsymbol{\Omega})\tag{2.32}$$

$$\xrightarrow{d}N\{\mathbf{0},[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Omega}\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\}$$

因此，对一般权矩阵 $\mathbf{W}$，一般广义矩估计量的渐近协方差为：

$$\mathrm{Var}(\hat{\boldsymbol{\theta}}_{GMM})=\frac{1}{n}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\boldsymbol{\Omega}\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\tag{2.33}$$

第三节　最优权矩阵与最优 *GMM*

采用最优权矩阵的 *GMM* 称为“最优 *GMM*”，最优权矩阵 $\mathbf{W}$ 的选择，应能使广义矩估计量 $\hat{\boldsymbol{\theta}}_{GMM}$ 最有效。给定一组矩条件，最优权矩阵规定了如何运用这些矩条件来尽可能精确地估计 $\boldsymbol{\theta}$。“最优 *GMM*”并不能告诉我们在估计过程中使用了哪些矩条件。一般来说，在估计过程中，使用的有效的矩条件越多，*GMM* 估计量越有效。

一　最优权矩阵的选择

1. 最优权矩阵 $\mathbf{W}$ 及最优 *GMM* 估计量的渐近协方差

首先我们给出最优权矩阵 $\mathbf{W}$ 的结果，然后再证明这个结论。

对任意一个非奇异矩阵 $\mathbf{e}$，矩阵 $\mathbf{A}$ 的最优选择是：

$$\underset{(k\times m)}{\mathbf{A}}=\underset{(k\times k)}{\mathbf{e}}\underset{(k\times m)}{\mathbf{G}'}\underset{(m\times m)}{\boldsymbol{\Omega}^{-1}}\tag{2.34}$$

其中 $\boldsymbol{\Omega}_{(m\times m)}$ 是总体矩 $\sqrt{n}E[\mathbf{g}(x_i,\boldsymbol{\theta})]$ 或样本矩 $\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)$ 的渐近

协方差矩阵。

考虑一种简单的情况，当 $\mathbf{e}_{(k\times k)}=\mathbf{I}_{(k\times k)}$ 时，这意味着：

$$\underset{(k\times m)}{\mathbf{A}}=\underset{(k\times m)}{\mathbf{G}'}\underset{(m\times m)}{\mathbf{\Omega}^{-1}} \tag{2.35}$$

是最优的。从上节的定义知 $\mathbf{A}=\mathbf{G}'\mathbf{W}$，因此最优权矩阵 $\mathbf{W}$ 应满足：

$$\mathbf{G}'\mathbf{\Omega}^{-1}=\mathbf{G}'\mathbf{W}\quad\Rightarrow\quad\mathbf{W}=\mathbf{\Omega}^{-1} \tag{2.36}$$

这里，$\mathbf{G}'\mathbf{\Omega}^{-1}=\mathbf{G}'\mathbf{W}$ 是 $\mathbf{W}$ 为最优权矩阵的充分必要条件，而 $\mathbf{W}=\mathbf{\Omega}^{-1}$ 只是其充分条件。即：

$$\mathbf{G}'\mathbf{\Omega}^{-1}=\mathbf{G}'\mathbf{W}\quad\Leftrightarrow\quad\mathbf{W}\text{ 是最优权矩阵} \tag{2.37}$$

$$\mathbf{W}=\mathbf{\Omega}^{-1}\quad\Rightarrow\quad\mathbf{W}\text{ 是最优权矩阵} \tag{2.38}$$

在最优权矩阵 $\mathbf{W}=\mathbf{\Omega}^{-1}$ 条件下，$\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})$ 依分布收敛于 $N\{\mathbf{0},[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}\}$，即：

$$\begin{aligned}\sqrt{n}(\hat{\boldsymbol{\theta}}-\boldsymbol{\theta})&\xrightarrow{d}N\{\mathbf{0},[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\}\\&=N\{\mathbf{0},[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}\}\end{aligned} \tag{2.39}$$

$$\mathbf{Var}(\hat{\boldsymbol{\theta}}_{GMM})=\frac{1}{n}[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1} \tag{2.40}$$

2. 最优 *GMM* 估计量最有效的证明

最优 *GMM* 估计量最有效可以用反证法来证明：选择其他任何的一般权矩阵 $\mathbf{W}\neq\mathbf{\Omega}^{-1}$ 时 $\sqrt{n}\hat{\boldsymbol{\theta}}$ 的渐近协方差与选择最优权矩阵 $\mathbf{W}=\mathbf{\Omega}^{-1}$ 时的 $\sqrt{n}\hat{\boldsymbol{\theta}}$ 渐近协方差之差 $[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}\,[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}-[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}$ 总是半正定的。

证明如下：

$$\begin{aligned}&[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}\,[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}-[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}\\=&[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\{\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}-\mathbf{G}'\mathbf{W}\mathbf{G}\,[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\mathbf{G}\}\,[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\end{aligned} \tag{2.41}$$

设 $\mathbf{Z}_{(m\times1)}$ 为满足：

$$\underset{(m\times m)}{\mathbf{\Omega}}=E(\underset{(m\times1)}{\mathbf{Z}}\underset{(1\times m)}{\mathbf{Z}'}) \tag{2.42}$$

的任意随机向量，令：

$$\underset{(k\times1)}{\mathbf{m}}=\underset{(k\times m)}{\mathbf{G}'}\underset{(m\times m)}{\mathbf{W}}\underset{(m\times1)}{\mathbf{Z}} \tag{2.43}$$

$$\underset{(k\times1)}{\bar{\mathbf{m}}}=\underset{(k\times m)}{\mathbf{G}'}\underset{(m\times m)}{\mathbf{\Omega}^{-1}}\underset{(m\times1)}{\mathbf{Z}} \tag{2.44}$$

$$\mathbf{U}=\mathbf{m}-E[\mathbf{m}\bar{\mathbf{m}}'](E[\bar{\mathbf{m}}\bar{\mathbf{m}}'])^{-1}\bar{\mathbf{m}} \tag{2.45}$$

于是：

$$\begin{aligned}&E[\mathbf{U}\mathbf{U}']\\&=E[\mathbf{m}\mathbf{m}']-E[\mathbf{m}\bar{\mathbf{m}}'](E[\bar{\mathbf{m}}\bar{\mathbf{m}}'])^{-1}E[\bar{\mathbf{m}}\mathbf{m}']\\&\quad-E[\mathbf{m}\bar{\mathbf{m}}'](E[\bar{\mathbf{m}}\bar{\mathbf{m}}'])^{-1}E[\bar{\mathbf{m}}\mathbf{m}']\\&\quad+E[\mathbf{m}\bar{\mathbf{m}}'](E[\bar{\mathbf{m}}\bar{\mathbf{m}}'])^{-1}E[\bar{\mathbf{m}}\bar{\mathbf{m}}'](E[\bar{\mathbf{m}}\bar{\mathbf{m}}'])^{-1}E[\bar{\mathbf{m}}\mathbf{m}']\\&=E[\mathbf{m}\mathbf{m}']-E[\mathbf{m}\bar{\mathbf{m}}'](E[\bar{\mathbf{m}}\bar{\mathbf{m}}'])^{-1}E[\bar{\mathbf{m}}\mathbf{m}']\\&=\mathbf{G}'\mathbf{W}\cdot E(\mathbf{Z}\mathbf{Z}')\cdot\mathbf{W}\mathbf{G}-\mathbf{G}'\mathbf{W}\cdot E(\mathbf{Z}\mathbf{Z}')\cdot\\&\quad\mathbf{\Omega}^{-1}\mathbf{G}[\mathbf{G}'\mathbf{\Omega}^{-1}\cdot E(\mathbf{Z}\mathbf{Z}')\cdot\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{\Omega}^{-1}\cdot E(\mathbf{Z}\mathbf{Z}')\cdot\mathbf{W}\mathbf{G}\\&=\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}-\mathbf{G}'\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]'\mathbf{G}'\mathbf{W}\mathbf{G}\end{aligned} \tag{2.46}$$

从而：

$$\begin{aligned}&[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}\,[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}-[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}\\&\quad=[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\cdot E\,[\mathbf{U}\mathbf{U}']\cdot[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\end{aligned} \tag{2.47}$$

一般地，如果矩阵 $\mathbf{P}$ 是半正定的，矩阵 $\mathbf{Q}$ 是非奇异的，则 $\mathbf{Q}'\mathbf{P}\mathbf{Q}$ 是半正定的。由于 $E[\mathbf{U}\mathbf{U}']$ 是半正定的，因此 $[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\cdot E[\mathbf{U}\mathbf{U}']\cdot[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}$ 也是半正定的，从而 $[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}\mathbf{G}'\mathbf{W}\mathbf{\Omega}\mathbf{W}\mathbf{G}[\mathbf{G}'\mathbf{W}\mathbf{G}]^{-1}-[\mathbf{G}'\mathbf{\Omega}^{-1}\mathbf{G}]^{-1}$ 是半正定的。这就证明了“最优 *GMM*”估计量最有效的结论。

二　最优 GMM 估计示例

设(x_i, y_i)独立同分布，观测值样本为(x_i, y_i)，$i=1, 2, \cdots, n$，要估计的模型为：$Ey_i=\mu$。已知$Ex_i=0$，如何构造一个有效的广义矩估计量来估计参数μ？

总体矩条件为：$E[\mathbf{g}(\mu)]=\mathbf{0}$，其中：

$$\mathbf{g}(\mu)=\begin{pmatrix} y_i-\mu \\ x_i \end{pmatrix} \tag{2.48}$$

样本矩条件为：

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mu)=\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}y_i-\mu \\ \frac{1}{n}\sum_{i=1}^{n}x_i \end{pmatrix}=\mathbf{0} \tag{2.49}$$

这里有两个矩条件，一个参数，因此模型是过度识别的。令总体矩的渐近协方差矩阵为：

$$\begin{aligned}\mathbf{\Omega} &= E\left[\sqrt{n}\,\mathbf{g}(\mu)\cdot\sqrt{n}\,\mathbf{g}'(\mu)\right] \\ &= \mathrm{n}\begin{pmatrix} \mathrm{E}(\mathrm{y_i}-\mu)^2 & \mathrm{E}[\mathrm{x_i}(\mathrm{y_i}-\mu)] \\ \mathrm{E}[\mathrm{x_i}(\mathrm{y_i}-\mu)] & \mathrm{Ex_i^2} \end{pmatrix} \\ &= \mathrm{n}\begin{pmatrix} \sigma_y^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_x^2 \end{pmatrix}\end{aligned} \tag{2.50}$$

由于：

$$\hat{\sigma}_y^2=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\frac{1}{n}\sum_{i=1}^{n}y_i\right)^2$$

$$\hat{\sigma}_x^2=\frac{1}{n}\sum_{i=1}^{n}x_i^2$$

$$\hat{\sigma}_{xy}=\frac{1}{n}\sum_{i=1}^{n}\left(y_i-\frac{1}{n}\sum_{i=1}^{n}y_i\right)x_i$$

因此$\mathbf{\Omega}$的样本估计值为：

$$\begin{aligned}\hat{\mathbf{\Omega}} &= n\begin{pmatrix} \hat{\sigma}_y^2 & \hat{\sigma}_{xy} \\ \hat{\sigma}_{xy} & \hat{\sigma}_x^2 \end{pmatrix} \\ &= n\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\left(y_i-\frac{1}{n}\sum_{i=1}^{n}y_i\right)^2 & \frac{1}{n}\sum_{i=1}^{n}\left(y_i-\frac{1}{n}\sum_{i=1}^{n}y_i\right)x_i \\ \frac{1}{n}\sum_{i=1}^{n}\left(y_i-\frac{1}{n}\sum_{i=1}^{n}y_i\right)x_i & \frac{1}{n}\sum_{i=1}^{n}x_i^2 \end{pmatrix}\end{aligned}$$

取最优权矩阵：

$$\mathbf{W}=\mathbf{\Omega}^{-1}=\frac{1}{n(\hat{\sigma}_y^2\hat{\sigma}_x^2-\hat{\sigma}_{xy}^2)}\begin{pmatrix} \hat{\sigma}_x^2 & -\hat{\sigma}_{xy} \\ -\hat{\sigma}_{xy} & \hat{\sigma}_y^2 \end{pmatrix} \tag{2.51}$$

最小化的目标函数为：

$$Loss=\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mu)\right]'\mathbf{\Omega}^{-1}\left[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(\mu)\right]$$

$$= \frac{1}{n(\hat{\sigma}_y^2\hat{\sigma}_x^2 - \hat{\sigma}_{xy}^2)}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} y_i - \mu & \frac{1}{n}\sum_{i=1}^{n} x_i\end{pmatrix}\begin{pmatrix}\hat{\sigma}_x^2 & -\hat{\sigma}_{xy} \\ -\hat{\sigma}_{xy} & \hat{\sigma}_y^2\end{pmatrix}\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n} y_i - \mu \\ \frac{1}{n}\sum_{i=1}^{n} x_i\end{pmatrix}$$

$$\begin{aligned}&= \frac{1}{n(\hat{\sigma}_y^2\hat{\sigma}_x^2 - \hat{\sigma}_{xy}^2)}\Big[\hat{\sigma}_x^2\Big(\frac{1}{n}\sum_{i=1}^{n} y_i - \mu\Big)^2 - 2\hat{\sigma}_{xy}\Big(\frac{1}{n}\sum_{i=1}^{n} y_i - \mu\Big) \\ &\quad \Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big) + \hat{\sigma}_y^2\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big)^2\Big] \qquad (2.52)\end{aligned}$$

解其一阶条件可得：

$$\begin{aligned}\hat{\mu}_{GMM} &= \arg\min_{\mu} Loss \\ &= \frac{1}{n}\sum_{i=1}^{n} y_i - \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x^2}\Big(\frac{1}{n}\sum_{i=1}^{n} x_i\Big)\end{aligned}$$

三　最优 *GMM* 估计量数值算法的步骤

在经典矩方法中，有 k 个未知参数，k 个矩条件。在广义矩方法中，有 k 个未知参数，m 个矩条件（$m > k$）。假定向量值函数 $\mathbf{g}(x_i, \boldsymbol{\theta})$ 的 j 阶总体矩为零，即：$E[\mathbf{g}_j(x_i, \boldsymbol{\theta})] = 0$，$j = 1, 2, \cdots, m$。用 $\mathbf{g}'(x_i, \boldsymbol{\theta})$ 表示 $\mathbf{g}(x_i, \boldsymbol{\theta})$ 的转置：

$$\mathbf{g}(x_i, \boldsymbol{\theta}) = \begin{bmatrix} g_1(x_i, \boldsymbol{\theta}) \\ g_2(x_i, \boldsymbol{\theta}) \\ \vdots \\ g_m(x_i, \boldsymbol{\theta}) \end{bmatrix} \qquad (2.53)$$

$$\mathbf{g}'(x_i, \boldsymbol{\theta}) = [g_1(x_i, \boldsymbol{\theta}), g_2(x_i, \boldsymbol{\theta}), \cdots, g_m(x_i, \boldsymbol{\theta})] \qquad (2.54)$$

对协方差矩阵 $\boldsymbol{\Omega}_{(m\times m)}$ 进行如下变换：

$$\begin{aligned}\underset{(m\times m)}{\boldsymbol{\Omega}} &= E\Big[\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i, \boldsymbol{\theta})\Big)\cdot\sqrt{n}\Big(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i, \boldsymbol{\theta})\Big)'\Big] \\ &= E\Big[\frac{1}{n}\Big(\sum_{i=1}^{n}\mathbf{g}(x_i, \boldsymbol{\theta})\Big)\cdot\Big(\sum_{i=1}^{n}\mathbf{g}(x_i, \boldsymbol{\theta})\Big)'\Big] \\ &= E\Big\{\frac{1}{n}\sum_{i=1}^{n}[\mathbf{g}(x_i, \boldsymbol{\theta})\cdot\mathbf{g}'(x_i, \boldsymbol{\theta})] + \frac{1}{n}\sum_{i=1}^{n}\sum_{\substack{j=1 \\ j\neq i}}^{n}[\mathbf{g}(x_i, \boldsymbol{\theta})\cdot\mathbf{g}'(x_j, \boldsymbol{\theta})]\Big\} \\ &= \frac{1}{n}\sum_{i=1}^{n}E[\mathbf{g}(x_i, \boldsymbol{\theta})\cdot\mathbf{g}'(x_i, \boldsymbol{\theta})] \qquad (2.55) \\ &\quad + \frac{1}{n}\sum_{l=1}^{n-1}\sum_{i=l+1}^{n}\{E[\mathbf{g}(x_i, \boldsymbol{\theta})\cdot\mathbf{g}'(x_{i-l}, \boldsymbol{\theta})] + E[\mathbf{g}(x_{i-l}, \boldsymbol{\theta})\cdot\mathbf{g}'(x_i, \boldsymbol{\theta})]\}\end{aligned}$$

$$+ \frac{1}{n}\sum_{l=1}^{n-1}\sum_{i=l+1}^{n}\{E[\mathbf{g}(x_i, \boldsymbol{\theta})\cdot\mathbf{g}'(x_{i-l}, \boldsymbol{\theta})] + E[\mathbf{g}(x_{i-l}, \boldsymbol{\theta})\cdot\mathbf{g}'(x_i, \boldsymbol{\theta})]\}$$

一旦给定 $\boldsymbol{\theta}$ 的一个初值 $\hat{\boldsymbol{\theta}}_0$，且 $\hat{\boldsymbol{\theta}}_0$ 是 $\boldsymbol{\theta}$ 的一个一致性估计，基于上面的结果，考虑到 $\mathbf{g}(x_i, \boldsymbol{\theta})$ 之间的序列相关性，纽维和韦斯特（Newey & West, 1987）建议采用下面的算法来估计 $\hat{\boldsymbol{\Omega}}_{(m\times m)}$：

$$\begin{aligned}\hat{\boldsymbol{\Omega}} &= \frac{1}{n}\sum_{i=1}^{n}[\mathbf{g}(x_i, \hat{\boldsymbol{\theta}}_0)\cdot\mathbf{g}'(x_i, \hat{\boldsymbol{\theta}}_0)] \\ &\quad + \frac{1}{n}\sum_{l=1}^{L}w(l)\sum_{i=l+1}^{n}[\mathbf{g}(x_i, \hat{\boldsymbol{\theta}}_0)\cdot\mathbf{g}'(x_{i-l}, \hat{\boldsymbol{\theta}}_0) + \mathbf{g}(x_{i-l}, \hat{\boldsymbol{\theta}}_0)\cdot\mathbf{g}'(x_i, \boldsymbol{\theta}_0)] \\ &\qquad (2.56)\end{aligned}$$

其中：

$$w(l) = 1 - \frac{l}{L+1} \qquad (2.57)$$

最大滞后期 L 必须事先确定，当滞后期大于 L 时，自相关小到可以忽略的程度。L 应满足：

$$\lim_{n\to\infty}\frac{L}{\sqrt[4]{n}}=0 \tag{2.58}$$

于是，最优 *GMM* 估计量可以通过下列数值算法来实现：

第一步：计算 $\boldsymbol{\theta}$ 的初值 $\hat{\boldsymbol{\theta}}_0$ 及 $\boldsymbol{\Omega}$ 的初值 $\hat{\boldsymbol{\Omega}}_0$，初始时令权矩阵为单位矩阵，即 $\mathbf{W}_0=\mathbf{I}$，把此时 *GMM* 估计量作为 $\boldsymbol{\theta}$ 的初值 $\hat{\boldsymbol{\theta}}_0$，即：

$$\begin{aligned}\hat{\boldsymbol{\theta}}_0 &= \arg\min_{\theta}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\Big]'\mathbf{W}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\Big]\\ &= \arg\min_{\theta}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\Big]'\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\Big]\end{aligned} \tag{2.59}$$

在 $\hat{\boldsymbol{\theta}}_0$ 处，运用纽维和韦斯特方法估计 $\hat{\boldsymbol{\Omega}}_0$，并得到最优权矩阵 $\mathbf{W}$ 的一个初始的一致估计量 $\hat{\mathbf{W}}_1=\hat{\boldsymbol{\Omega}}_0^{-1}$。

第二步：通过求解下面的最优化问题，得到与上面的最优权矩阵 $\hat{\boldsymbol{\Omega}}_0^{-1}$ 相对应的 *GMM* 估计量 $\hat{\boldsymbol{\theta}}_1$：

$$\hat{\boldsymbol{\theta}}_1 = \arg\min_{\theta}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\Big]'\hat{\boldsymbol{\Omega}}_0^{-1}\Big[\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\Big] \tag{2.60}$$

第三步：在 $\hat{\boldsymbol{\theta}}_1$ 处，运用纽维和韦斯特公式估计 $\hat{\boldsymbol{\Omega}}_1$，并得到最优权矩阵 $\mathbf{W}$ 的一个更新估计量 $\hat{\mathbf{W}}_2=\hat{\boldsymbol{\Omega}}_1^{-1}$。

在 $\hat{\boldsymbol{\theta}}_1$ 处，计算 $\mathbf{G}$ 的一致估计量：

$$\hat{\mathbf{G}} = E\Big[\frac{\partial\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})}{\partial\boldsymbol{\theta}'}\Big]=\frac{1}{n}\sum_{i=1}^{n}\frac{\partial\mathbf{g}(x_i,\hat{\boldsymbol{\theta}}_1)}{\partial\boldsymbol{\theta}'} \tag{2.61}$$

从而得到：

$$\mathbf{Var}(\hat{\boldsymbol{\theta}}_{GMM})=\frac{1}{n}[\mathbf{G}'\boldsymbol{\Omega}^{-1}\mathbf{G}]^{-1} \tag{2.62}$$

第四步：进行迭代。把新的 $\hat{\boldsymbol{\theta}}_1$ 和新的 $\hat{\boldsymbol{\Omega}}_1$ 作为初值，重复第二步和第三步。这一过程可迭代下去，直至收敛（比如 $\hat{\boldsymbol{\theta}}_{j+1}\approx\hat{\boldsymbol{\theta}}_j$）为止，便得到了最优 *GMM* 估计量及其渐近协方差。

过度识别约束检验 Y=Xβ

假设参数个数为 k 个，当解释变量中包含随机变量即 $E(\mathbf{X}u)\neq 0$ 时，应采用工具变量（*IV*）法。假设工具变量为 $\mathbf{Z}$，它满足：$E(\mathbf{Z}u)=0$

当工具变量个数 m 等于 k 时，可采用普通的工具变量法进行参数估计。当工具变量个数 m 大于 k 时，存在 m 个矩条件，k 个参数。假设这 m 个矩条件都是有效的，可以用 *GMM* 进行估计，目标函数为：

$$\min\Big(\frac{1}{n}\sum_i(y_i-\mathbf{X}_i\boldsymbol{\beta})\mathbf{Z}_i\Big)'\Omega^{-1}\Big(\frac{1}{n}\sum_i(y_i-\mathbf{X}_i\boldsymbol{\beta})\mathbf{Z}_i\Big) \tag{2.63}$$

其中最优权矩阵为：

$$\begin{aligned}\underset{(m\times m)}{\boldsymbol{\Omega}} &= E\Big[\sqrt{n}\Big(\frac{1}{n}\sum_i(y_i-\mathbf{X}_i\boldsymbol{\beta})\mathbf{Z}_i\Big)\cdot\sqrt{n}\Big(\frac{1}{n}\sum_i(y_i-\mathbf{X}_i\boldsymbol{\beta})\mathbf{Z}_i\Big)'\Big]\\ &= E\Big[\frac{1}{n}\sum_i(u_i\mathbf{Z}_i)\cdot\sum_i(u_i\mathbf{Z}_i')\Big]\end{aligned} \tag{2.64}$$

当随机误差项 u 满足同方差假定时，令 $u_i^2=\sigma^2$，则最优权矩阵可简化为：

$$\boldsymbol{\Omega} = E\Big[\frac{1}{n}\sum_i(u_i\mathbf{Z}_i)\cdot\sum_i(u_i\mathbf{Z}_i')\Big]=\frac{\sigma^2}{n}\mathbf{Z}'\mathbf{Z} \tag{2.65}$$

于是目标函数变为：

$$\min \frac{\sigma^2}{n}\left(\frac{1}{n}\mathbf{Z}'\ (Y-\mathbf{X\beta})\right)'\ (\mathbf{Z}'\mathbf{Z})^{-1}\left(\frac{1}{n}\mathbf{Z}'\ (Y-\mathbf{X\beta})\right) \quad (2.66)$$

一阶条件为：

$$(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'(Y-\mathbf{X}\hat{\boldsymbol{\beta}}))=0 \quad (2.67)$$

于是可得：

$$\hat{\boldsymbol{\beta}}_{GMM}=[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{X})]^{-1}[(\mathbf{X}'\mathbf{Z})(\mathbf{Z}'\mathbf{Z})^{-1}(\mathbf{Z}'\mathbf{Y})] \quad (2.68)$$

令 $\mathbf{P_Z}=\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'$，$\hat{\mathbf{X}}'=\mathbf{X}'\mathbf{P_Z}$，则：

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{GMM}&=(\mathbf{X}'\mathbf{P_Z}\mathbf{X})^{-1}\mathbf{X}'\mathbf{P_Z}\mathbf{Y}\\&=(\hat{\mathbf{X}}'\mathbf{X})^{-1}(\hat{\mathbf{X}}'\mathbf{Y})\end{aligned} \quad (2.69)$$

可见，在随机误差项 u 同方差条件下，*GMM* 估计量等价于二阶段最小二乘（2*SLS*）估计量，在经济计量软件中可通过工具变量（*IV*）法进行估计。

二　过度识别约束检验（Hansen 检验或者 *J* 检验）

如果矩条件的个数多于参数的个数，即 $m>k$，则模型是过度识别的，采用 *GMM* 是渐近有效的。问题是，这些拟采用的矩条件或过度识别约束

是否真的成立？我们可以通过检验总体矩条件 $E[\mathbf{g}(x_i,\boldsymbol{\theta})]=0$ 是否为真来判别。过度识别约束检验也称为检验或者 *J* 检验。

一般地，对向量值函数 $\mathbf{g}(x_i,\boldsymbol{\theta})$，若过度识别约束成立，则其各阶总体矩均应为零，即：

$$E[g_j(x_i,\boldsymbol{\theta})]=0,\ j=1,2,\cdots,m \quad (2.70)$$

原假设为：

$$H_0: E[\mathbf{g}(x_i,\boldsymbol{\theta})]=\mathbf{0} \quad (2.71)$$

表示矩条件成立或过度识别约束有效。备择假设为：

$$H_1: E[\mathbf{g}(x_i,\boldsymbol{\theta})]\neq\mathbf{0} \quad (2.72)$$

表示矩条件不成立或过度识别约束无效。如果约束有效，由于样本矩依概率收敛于总体矩，则：

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\xrightarrow{p}E[\mathbf{g}(x_i,\boldsymbol{\theta})]=\mathbf{0} \quad (2.73)$$

因为：

$$\sqrt{n}\cdot\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\xrightarrow{d}N(\mathbf{0},\underset{(m\times m)}{\boldsymbol{\Omega}}) \quad (2.74)$$

其中协方差矩阵 $\boldsymbol{\Omega}_{(m\times m)}$ 为：

$$\underset{(m\times m)}{\boldsymbol{\Omega}}=E\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)\cdot\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\boldsymbol{\theta})\right)'\right] \quad (2.75)$$

令 $\hat{\boldsymbol{\theta}}$ 表示 *GMM* 估计量。如果过度识别约束是有效的，汉森（Hansen，1982）构造了如下的 J 统计量：

$$J = \left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})\right)\right]'\left\{\mathbf{Var}\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})\right)\right]\right\}^{-1}\left[\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})\right)\right] = n\cdot\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})\right)\right]'\hat{\boldsymbol{\Omega}}^{-1}\left[\left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{g}(x_i,\hat{\boldsymbol{\theta}})\right)\right] \quad (2.76)$$

汉森证明，如果在 *GMM* 估计量中权矩阵 $\hat{\boldsymbol{\Omega}}^{-1}$ 是渐近有效的，则：

$$J \xrightarrow{d} \chi^2_{p=m-k} \quad (2.77)$$

自由度为 m-k

如果 $J>\chi^2_p$，则拒绝 H_0，说明矩条件不成立或过度识别约束无效；

如果 $J\leqslant\chi^2_p$，则不能拒绝 H_0：$E[\mathbf{g}(x_i,\boldsymbol{\theta})]=\mathbf{0}$，说明矩条件成立或过度识别约束有效。

由于 J 统计量恰好是最优 $\hat{\boldsymbol{\theta}}_{GMM}$ 所对应的最小化损失函数 $\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(1\times m)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right]'\underset{(m\times m)}{\mathbf{W}}\left[\frac{1}{n}\sum_{i=1}^{n}\underset{(m\times 1)}{\mathbf{g}}(x_i,\boldsymbol{\theta})\right]$ 的 n 倍，因此可以把过度识别约束检验统计量 J 看做是 *GMM* 估计过程的一个附产品。不管在什么情况下，只要进行最优 *GMM* 估计，就不难同时给出 J 的结果。这样做，不仅估计出了 $\hat{\boldsymbol{\theta}}_{GMM}$，而且还确定了用于估计 $\hat{\boldsymbol{\theta}}_{GMM}$ 的那些矩条件或过度识别约束是否真的成立，一举两得。

对于线性回归模型 $y=\mathbf{X}\boldsymbol{\beta}+u$，参数个数为 k 个，当解释变量中包含随机变量 $E(\mathbf{X}u)\neq 0$ 时，找到工具变量 $\mathbf{Z}$ 满足 $E(\mathbf{Z}u)=0$。当残差 u 满足同方差假定，且工具变量个数 m 大于 k 时，存在 m 个矩条件，k 个参数。为了检验 m 个矩条件是否都是有效的，过度识别约束检验的 J 统计量可以简化为：

$$J = n\cdot R^2 \xrightarrow{d} \chi^2_{p=m-k} \quad (2.78)$$

其中 R^2 是指残差 $\hat{e}$ 对所有工具变量 $\mathbf{Z}$（包含外生变量）回归的拟合优度，其中 $\hat{e}$ 定义为：

$$\hat{e}_i = y_i - x_i\hat{\beta}_{2SLS} \quad (2.79)$$

在经济计量软件中，$\hat{\beta}_{2SLS}$ 可通过工具变量（*IV*）法估计出来。[3]

广义矩估计方法（generalized method of moments，简写为 GMM）的模型参数直接来自于模型利用的矩条件（moments conditions）（例子见 Greene，2003；[83] Hall，2001；[92] Verbeek，2000；[84] Conley，1999；[87] Kelejian & Prucha，1999；[93] 李子奈、叶阿忠，2000[85]；张晓峒，2003[91]）。这些条件使用的参数可能是线性的，但通常是非线性的。为确保识别，矩条件（moments conditions）的数量应该至少与未知参数的数量一样多。

通常地，我们考虑一个模型，模型的特点在于如下的 R 个矩条件所构成的集合：

$$E[f(X,\theta)]=0 \quad (3.18)$$

其中 f 是包含 R 个元素的向量函数，假设 $R\geqslant K$，则有（$R-K$）个剩余的识别约束。X 是观测变量矩阵，这些变量可能是内生的或者外生的，θ 是包含所有未知参数的 K 维向量。

为了估计 θ，我们考虑方程（3.18）的对应样本：

$$g_N(\theta)=\frac{1}{N}\sum f(X,\theta)=0 \quad (3.19)$$

[3] 在 http://edu.duxiu.com/ 中输入 广义矩估计 广义矩方法 经典矩方法

如果 R 个矩条件的数量等于未知参数 K 的数量，设方程（3.19）中的 R 个元素为零，并且解 θ，能够获得唯一一致的估计量。如果矩条件的数量大于参数的数量，通过设方程（3.19）为零，我们求解出的未知参数不唯一。

相反地，我们选择 θ 为估计量以使样本的矩向量趋于零，在某种意义上把 g_N（θ）的二次方形式最小化，即：

$$\min\Phi(\theta)=\min g_N(\theta)'A_N g_N(\theta) \tag{3.20}$$

其中，A_N 是正定（positive definite）权重矩阵。由此可以解得 GMM 估计量$\tilde{\tilde{\theta}}$（见 Hansen，1982；[94] Conley，1999；[87] Kelejian & Prucha，1999[93]）。虽然我们不能够得到一般情形的 GMM 估计量的分析解，但是能够证明它是一致（consistent）的和渐近正态（asymptotically normal）的。最优化权重矩阵是矩样本的方差一协方差矩阵（variance-covariance matrix）的逆矩阵，由此可以得到最有效的 GMM 估计量。一般地，这个矩阵依赖于未知参数向量 θ，这存在一个线性模型没有遇到的问题，解将采用一个迭代的估计过程。

Hansen 的表述：

汉森的估计问题的系统表述如下。令 $\mathbf{w}_t$ 为一个 t 期观察到的 $(h\times 1)$ 变量向量，令 $\boldsymbol{\theta}$ 为未知 $(a\times 1)$ 系数向量，令 $\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)$ 为 $(r\times 1)$ 向量值函数，$\mathbf{h}:(\mathbb{R}^a\times\mathbb{R}^h)\rightarrow\mathbb{R}^r$。因为 $\mathbf{w}_t$ 是一个随机变量，所以 $\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)$ 也是随机变量。令 $\boldsymbol{\theta}_0$ 表示 $\boldsymbol{\theta}$ 的真实值，假定这一真实值由性质

$$E\{\mathbf{h}(\boldsymbol{\theta}_0,\mathbf{w}_t)\}=\mathbf{0} \tag{14.1.9}$$

表示，则向量方程(14.1.9)的 r 行有时被称作正交性条件。令 $\mathscr{Y}_T\equiv(\mathbf{w}'_T,\mathbf{w}'_{T-1},\cdots,\mathbf{w}'_1)'$ 为包含容量为 T 的样本中全部观察值的 $(Th\times 1)$ 向量，令 $(r\times 1)$ 向量值函数 $\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)$ 表示 $\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)$ 的样本均值：

$$\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)\equiv(1/T)\sum_{t=1}^{T}\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)\ . \tag{14.1.10}$$

注意到 $\mathbf{g}:\mathbb{R}^a\rightarrow\mathbb{R}^r$。在 *GMM* 背后的思想是选取 $\boldsymbol{\theta}$ 使样本矩 $\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)$ 尽可能接近于零总体矩；即 *GMM* 估计量 $\hat{\boldsymbol{\theta}}_T$ 是使数量

$$Q(\boldsymbol{\theta};\mathscr{Y}_T)=[\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)]'\mathbf{W}_T[\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)] \tag{14.1.11}$$

最小的 $\boldsymbol{\theta}$ 值，其中 $\{\mathbf{W}_T\}_{T=1}^{\infty}$ 是一个 $(r\times r)$ 正定权重矩阵序列，W_T 可以是 $\mathscr{Y}_T$ 的函数。通常，以第 5.7 节描述的数值方法求此最小化。

如果要估计的参数个数 (a) 与正交性条件个数 (r) 相同，则一般地，令

$$\mathbf{g}(\hat{\boldsymbol{\theta}}_T;\mathscr{Y}_T)=\mathbf{0}\ , \tag{14.1.12}$$

可使目标函数(14.1.11)最小化。如果 $a=r$，则 *GMM* 估计量是满足这 r 个方程的 $\hat{\boldsymbol{\theta}}_T$ 值。若不然，存在比要估计的参数还多的正交性条件 $(r>a)$，则(14.1.12)并不恰好成立。$\mathbf{g}(\hat{\boldsymbol{\theta}}_T;\mathscr{Y}_T)$ 的第 i 个元素接近于零的程度取决于权重矩阵 $\mathbf{W}_T$ 给予第 i 个正交性条件多大的权重。

对于任意的 $\boldsymbol{\theta}$ 值，$(r\times 1)$ 向量 $\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)$ 的值都是 $(r\times 1)$ 随机向量 $\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)$ 的 T 个实现的样本平均。如果 $\mathbf{w}_t$ 是严格平稳的，且 $\mathbf{h}(\cdot)$ 是连续的，则预

期大数定理成立是合适的：

$$\mathbf{g}(\boldsymbol{\theta};\mathscr{Y}_T)\xrightarrow{p}E\{\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)\}\ .$$

表达式 $E\{\mathbf{h}(\boldsymbol{\theta},\mathbf{w}_t)\}$ 表示的是取决于 $\boldsymbol{\theta}$ 值和 $\mathbf{w}_t$ 的概率规律的总体值。假定这一函数是关于 $\boldsymbol{\theta}$ 连续的，且 $\boldsymbol{\theta}_0$ 是惟一一个使 $\boldsymbol{\theta}$ 满足(14.1.9) 的值，则，在相当一般的平稳性、连续性以及矩条件下，使(14.1.11) 最小化的 $\hat{\boldsymbol{\theta}}_T$ 值提供了 $\boldsymbol{\theta}_0$ 的一个一致估计；细节参见汉森(1982)，加朗和怀特(1988)，以及安德鲁斯和费尔(Fair)(1988)。