多元线性回归：广义矩法估计及 MATLAB 实现①

董晨昱[1]，刘维奇[2]

（1. 山西大学 数学科学学院，太原 030006；2. 山西大学 管理与决策研究所，太原 030006）

摘 要：线性回归模型中，最小二乘法估计参数所具有的优良性质源于特定的假设，当某个或多个假设不成立时，如何有效给出回归系数的估计？广义最小二乘、工具变量法和广义矩法估计均是针对这一问题应运而生的，且普通最小二乘法、广义最小二乘法和工具变量法均是广义矩法的特例。从矩法原理出发，针对线性回归模型的经典假设，由回归系数的矩法估计到广义矩法估计进行全面分析，并通过数值模拟算例说明广义矩法估计的 MATLAB 实现。

关键词：广义最小二乘；工具变量法；广义矩估计

中图分类号：O212.4；F224.0　**文献标识码**：A　**文章编号**：（2014）01-0082-09

0 引言

矩法和极大似然法是参数估计问题中最常用的两类方法，极大似然估计要求总体的精确分布形式，但矩法估计则可以在更宽松的条件下进行。在线性回归模型中，普通最小二乘法（ordinary least squares，OLS）和工具变量法都属于矩法估计的特例，当经典假设都满足时，普通最小二乘法所得参数估计满足一致性和渐进正态性，且是最佳线性无偏估计[1]。当解释变量中存在内生变量时，普通最小二乘估计不再是最优估计，而使用工具变量法仍可得到参数的一致估计[2]。每一新方法的产生便是为了进步的需求，广义最小二乘法允许残差项存在异方差和序列相关，两阶段最小二乘法和广义矩法则都可以克服内生变量问题，且广义矩法也允许残差的异方差性和序列自相关性，同时它不要求残差的分布形式，因此广义矩法估计（generalized method of moments，GMM）作为一个计量经济方法被广泛应用于现代经济、金融以及社会科学的研究中。但需要注意的是，广义矩法估计的一致性是建立在大样本基础上的[3]。可以证明，极大似然法、普通最小二乘法、广义最小二乘法和工具变量法均是广义矩法的特例，且它不仅适用于线性回归的参数估计问题，同样也适用于非线性参数估计问题[1]。文中主要以线性回归模型为例，阐述广义矩法估计的原理，并给出基于 MATLAB 的广义矩法估计程序。

1 文献索引分析

通过分析中国知网（CNKI）2004-2013 年十年期间中文 *SCI* 来源期刊和 *CSSCI* 来源期刊关于广义矩法估计的应用研究，从期刊发表年度、学科类型和发表机构三个方面给出了广义矩法估计的研究与应用概况，分析结果见表 1。

虽然广义矩法估计源于 20 世纪 80 年代初，由于计算工具的限制，工具变量及权重矩阵选取的困难，广义矩法估计并没有得到广泛的应用。近年来，在计量软件 EViews 及统计软件 STATA 上都有可用的广义矩法估计命令，从而为广义矩法估计的推广和应用提供了便利。表 1 的文献索引分析中，第一列是近十年应用广义矩法估计在中文 *SCI* 来源期刊和 *CSSCI* 来源期刊上发表的文献数量，第二列和第三列分别是广义矩法估计应用最广的学科和相关文献发表最多的机构。从文献发表时间可以看出广义矩法估计的应用呈逐

---

① 作者简介：董晨昱（1980-），女，山西晋城人，博士生，山西大学数学科学学院讲师，研究方向：金融时间序列等，E-mail：bailange@sxu.edu.cn；刘维奇（1963-），男，山西忻州人，博士，山西大学管理与决策研究所教授，博士生导师，研究方向：金融工程与风险管理、时间序列分析，E-mail：liuwq@sxu.edu.cn。

年上升趋势，从发表期刊所属学科来看广义矩法估计作为计量方法主要被应用于经济和金融学科的研究。

表 1　文献分析

Table 1　Literature index analysis

发表年度	学科		发表机构	
2013（54）	宏观经济管理与可持续发展（186）	财政与税收（19）	南开大学（17）	中国人民大学（7）
2012（54）	金融（97）	人才学与劳动科学（15）	厦门大学（17）	重庆大学（7）
2011（55）	经济体制改革（66）	农业经济（9）	中南财经政法大学（11）	武汉大学（7）
2010（39）	数学（50）	环境科学与资源利用（8）	西南财经大学（11）	华南师范大学（5）
2009（23）	投资（38）	人口学与计划生育（6）	复旦大学（10）	西安交通大学（5）
2008（13）	企业经济（32）	经济理论及经济思想史（5）	湖南大学（9）	湖北经济学院（5）
2007（5）	贸易经济（28）	保险（4）	东北财经大学（9）	南京大学（5）
2006（8）	市场研究与信息（24）	旅游（4）	上海财经大学（9）	中山大学（5）
2005（3）	证券（20）	行政学与国家行政管理（2）	暨南大学（8）	清华大学（5）
2004（2）	工业经济（19）	会计（2）	华中科技大学（8）	安徽财经大学（4）

2　矩法原理及矩估计

用矩法求参数估计源于 20 世纪初，以大数定律为支撑，Pearson 提出了矩法原理，即用样本矩估计总体矩，用样本矩的函数估计总体矩的函数。

设 X 为一总体（随机变量），X_1，X_2，…，X_n 是来自该总体的一个简单随机样本，记 μ_r 为总体的 r 阶原点矩，记 m_r 为样本的 r 阶原点矩，即 $\mu_r=\mathrm{E}X^r$，$m_r=\frac{1}{n}\sum_{i=1}^{n}X_i^r$，若总体分布依赖于参数向量 $\boldsymbol{\theta}=(\theta_1,\theta_2,\cdots,\theta_k)$，且总体分布的 l（$l>k$）阶原点矩存在，则根据矩法原理，可建立如下方程：

$$\mu_r(\theta_1,\theta_2,\cdots,\theta_k)=m_r\Leftrightarrow \mathrm{E}X^r=\frac{1}{n}\sum_{i=1}^{n}X_i^r,\qquad r=1,2,\cdots,l \tag{1}$$

通常情况下，k 个参数用前 k 阶矩建立方程即可，若方程组（1）有解，则称其解 $\hat{\boldsymbol{\theta}}$（$X_1$，$X_2$，…，$X_n$）为 $\boldsymbol{\theta}$ 的矩估计。一般地，我们可以将 $\boldsymbol{m}_r(\boldsymbol{X};\boldsymbol{\theta})=\mathrm{E}(\boldsymbol{X}^r-\boldsymbol{\mu}_r)=0$（$r=1,2,\cdots,l$）称为总体矩条件，而称 $\hat{\boldsymbol{m}}_r(\boldsymbol{X}_1,\boldsymbol{X}_2,\cdots,\boldsymbol{X}_n;\boldsymbol{\theta})=\frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{X}_i^r-\boldsymbol{\mu}_r)=0(r=1,2,\cdots,l)$ 为样本矩条件。

3　线性回归模型

设被解释变量 y 关于解释变量 x_1，x_2，…，x_k 的线性回归数据结构式如下：

$$y=\beta_0+\beta_1x_1+\beta_2x_2+\cdots+\beta_kx_k+u \tag{2}$$

式中，β_i（$i=0,1,\cdots,k$）为回归系数（未知参数）；u 为随机误差（残差）。在给定 n 个独立的观测后，可以得到线性回归的统计模型：

$$y_i=\beta_0+\beta_1x_{1i}+\beta_2x_{2i}+\cdots+\beta_kx_{ki}+u_i,\qquad i=1,2,\cdots,n \tag{3}$$

线性回归模型的矩阵表示为

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{u} \tag{4}$$

式中，$\boldsymbol{Y}=(y_1,y_2,\cdots,y_n)'$，$\boldsymbol{\beta}=(\beta_0,\beta_1,\cdots,\beta_k)'$，$\boldsymbol{u}=(u_1,u_2,\cdots,u_n)'$，$\boldsymbol{X}=\begin{bmatrix}1 & x_{11} & \cdots & x_{k1}\\ \vdots & \vdots & & \vdots\\ 1 & x_{1n} & \cdots & x_{kn}\end{bmatrix}$。

3.1　线性回归模型的基本假设

在普通最小二乘法中，为保证参数估计具有良好的性质，通常对线性回归模型（线性指的是对 $\boldsymbol{\beta}$，而

非 X，因此包含那些可线性化的非线性回归模型）提出如下基本假设：

假设1. 解释变量之间互不相关；

假设2. 残差项均值为0，且同方差。即

$$E(u_j)=0,\ \mathrm{Var}(u_j)=\sigma^2,\ j=1,\ 2,\ \cdots,\ n$$

假设3. 残差项互不相关（序列不相关），即

$$\mathrm{Cov}(u_i,\ u_j)=\sigma_{ij},\ i\neq j,\ i=1,\ 2,\ \cdots,\ n,\ j=1,\ 2,\ \cdots,\ n$$

假设4. 残差项与解释变量之间互不相关。即

$$\mathrm{Cov}(x_i,\ u_j)=0,\ i=1,\ 2,\ \cdots,\ k,\ j=1,\ 2,\ \cdots,\ n$$

当模型满足假设1-4时，称其为标准回归模型，进一步当模型满足 $u_i\sim N(0,\ \sigma^2)$，$i=1,\ 2,\ \cdots,\ n$ 时，称其为标准正态回归模型。下文中均假设解释变量间互不相关，主要针对后面三个假设进行研究。

3.2 线性回归模型的矩法估计

在假设1-3成立的条件下，针对解释变量中是否存在内生变量（与残差项相关的变量）①，即假设4是否成立，分两种情形给出线性回归模型的矩法估计。

3.2.1 不存在内生变量的情形

若自变量中不存在内生变量，则有：$\mathrm{Cov}(x_i,\ u)=E(x_iu)=E(x_i(Y-\beta_0-\beta_1X_1-\cdots-\beta_kX_k))=0$，$i=1,\ 2,\ \cdots,\ k$，且残差项满足 $Eu=E(Y-\beta_0-\beta_1X_1-\cdots-\beta_kX_k)=0$。从而得到总体的 $k+1$ 个矩条件：

$$\begin{aligned}\boldsymbol{m}_0(Y,\ X;\ \boldsymbol{\beta})&=E(Y-\beta_0-\beta_1X_1-\cdots-\beta_kX_k)=0,\ \boldsymbol{m}_i(Y,\ X;\ \boldsymbol{\beta})\\&=E(x_i(Y-\beta_0-\beta_1X_1-\cdots-\beta_kX_k))=0,\ i=1,\ 2,\ \cdots,\ k\end{aligned}\tag{5}$$

其对应的样本矩条件为

$$\hat{\boldsymbol{m}}_0(Y,\ X;\ \boldsymbol{\beta})=\frac{1}{n}\sum_{j=1}^{n}u_j=\frac{1}{n}\sum_{j=1}^{n}(y_j-(\beta_0+\beta_1x_{1j}+\beta_2x_{2j}+\cdots+\beta_kx_{kj}))=0$$

$$\hat{\boldsymbol{m}}_i(Y,\ X;\ \boldsymbol{\beta})=\frac{1}{n}\sum_{j=1}^{n}x_{ij}u_j=\frac{1}{n}\sum_{j=1}^{n}x_{ij}(y_j-(\beta_0+\beta_1x_{1j}+\beta_2x_{2j}+\cdots+\beta_kx_{kj}))=0,\ i=1,\ 2,\ \cdots,\ k\tag{6}$$

综合上述 $k+1$ 个样本矩条件，即

$$\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})=\frac{1}{n}\boldsymbol{X}'(\boldsymbol{Y}-\boldsymbol{X\beta})=\boldsymbol{0}_{(k+1)\times 1}\tag{7}$$

解该方程组，参数向量 $\boldsymbol{\beta}$ 的矩估计为

$$\hat{\boldsymbol{\beta}}_{\mathrm{MM}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}=\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\tag{8}$$

因此，普通最小二乘估计是矩法估计的特例。

3.2.2 存在内生变量的情形

当解释变量与残差项相关时，可引入工具变量，即找到一组变量满足下面两个条件：

条件一：与自变量相关；

条件二：与残差项不相关。

这些工具变量用以解决解释变量与残差项相关的问题。引入工具变量时，需注意以下问题：

（1）工具变量的个数至少与回归系数一样多；

（2）与残差项不相关的解释变量可以看作自身的工具变量；

（3）常数是合适的工具变量。

设线性回归结构式如下：

$$\boldsymbol{y}=\boldsymbol{x\beta}+u=\boldsymbol{v\beta}_1+\boldsymbol{w\beta}_2+u\tag{9}$$

其中，$\boldsymbol{x}=(\boldsymbol{v},\ \boldsymbol{w})$ 是k维的解释变量向量；$\boldsymbol{v}$ 是 k_1 维的外生变量向量（与残差项不相关）；$\boldsymbol{w}$ 是 k_2 维的内生变量向量（与残差项相关），$k_1+k_2=k$，$\boldsymbol{\beta}=(\beta_1',\ \beta_2')'_{k\times 1}$ 是回归系数向量，则

$$E\boldsymbol{v}u=\boldsymbol{0}_{1\times k_1},\ E\boldsymbol{w}u\neq\boldsymbol{0}_{1\times k_2}\tag{10}$$

① 关于解释变量内生性的检验见 Hausman[4]。

设 z_w 是 w 的 k_2 维工具变量向量，由工具变量的两个条件知

$$\mathrm{Cov}(z_w,\ \boldsymbol{w}) \neq \boldsymbol{0}_{k_1\times k_2},\ \mathrm{Cov}(z_w,\ \boldsymbol{u}) = \mathrm{E}z_w\boldsymbol{u} = \boldsymbol{0}_{1\times k_2} \tag{11}$$

$\boldsymbol{v}$ 为外生变量向量，故它可看作是其自身的工具变量，则 $z=(\boldsymbol{v},\ z_w)_{1\times k}$ 是整体解释变量的工具变量向量。由（10）和（11）式可得总体的 k 个矩条件：

$$\boldsymbol{m}\ (\boldsymbol{y},\ z;\ \boldsymbol{\beta}) = \mathrm{E}zu = \mathrm{E}z\ (y-x\boldsymbol{\beta}) = \boldsymbol{0}_{1\times k} \tag{12}$$

在进行 n 次独立的观测后，可得回归模型如下：

$$\boldsymbol{Y} = \boldsymbol{X\beta} + u = \boldsymbol{v\beta}_v + \boldsymbol{w\beta}_W + u \tag{13}$$

其中，$\boldsymbol{Y}=(y_1,\ y_2,\ \cdots,\ y_n)'$，$\boldsymbol{\beta}=(\beta_1,\ \beta_2,\ \cdots,\ \beta_k)'$，$\boldsymbol{u}=(u_1,\ u_2,\ \cdots,\ u_n)'$，

$$\boldsymbol{X}=\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k1} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{kn} \end{bmatrix},\ \boldsymbol{V}=\begin{bmatrix} x_{11} & x_{21} & \cdots & x_{k_1,1} \\ \vdots & \vdots & & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{k_1,n} \end{bmatrix},\ \boldsymbol{W}=\begin{bmatrix} x_{k_1+1,1} & x_{k_1+2,1} & \cdots & x_{k1} \\ \vdots & \vdots & & \vdots \\ x_{k_1+1,n} & x_{k_1+2,n} & \cdots & x_{kn} \end{bmatrix}。$$

同时根据工具变量的观测值可得 $\boldsymbol{Z}=(z_1,\ z_2,\ \cdots,\ z_k)=\begin{bmatrix} z_{11} & z_{21} & \cdots & z_{k1} \\ \vdots & \vdots & & \vdots \\ z_{1n} & z_{2n} & \cdots & z_{kn} \end{bmatrix}$，从而得到相应的 k 个样本矩条件为

$$\hat{\boldsymbol{m}}_i(\boldsymbol{Y},\ \boldsymbol{Z};\ \boldsymbol{\beta}) = \frac{1}{n}\sum_{j=1}^{n} z_{ij}u_j = \frac{1}{n}\sum_{j=1}^{n} z_{ij}(y_j - (\beta_1 x_{1j} + \beta_2 x_{2j} + \cdots + \beta_k x_{kj})) = 0,\ i=1,\ 2,\ \cdots,\ k \tag{14}$$

综合上述 k 个样本矩条件，即

$$\hat{\boldsymbol{m}}\ (\boldsymbol{Y},\ \boldsymbol{Z};\ \boldsymbol{\beta}) = \frac{1}{n}\boldsymbol{Z}'\ (\boldsymbol{Y}-\boldsymbol{X\beta}) = \boldsymbol{0}_{k\times 1} \tag{15}$$

解方程组（15）可得参数 $\boldsymbol{\beta}$ 的矩估计为

$$\hat{\boldsymbol{\beta}}_{MM} = (\boldsymbol{Z}'\boldsymbol{X})^{-1}\boldsymbol{Z}'\boldsymbol{Y} \tag{16}$$

4 广义矩法估计

1982 年，Hansen 将矩法估计推广为著名的广义矩法估计[3]。矩法估计允许自变量中存在内生变量，除此之外，广义矩法还允许残差项存在异方差和序列相关，且不要求残差项的准确分布信息，因此由广义矩法所得参数估计应比其他估计方法的结果更合乎实际。由上一节的内容知普通最小二乘法和工具变量法都是广义矩法的特例，事实上，广义最小二乘法和极大似然法也都是它的特例。

4.1 广义（加权）最小二乘法

当假设 2 和假设 3（或者其中之一）不满足时，即残差项存在异方差 $\mathrm{E}(u_i^2)=\sigma_{ii}^2$，$i=1,\ 2,\ \cdots,\ n$，($\mathrm{E}(u_i)=0$，$i=1,\ 2,\ \cdots,\ n$ 依然成立)，且残差序列相关，即 $\mathrm{Cov}(u_i,\ u_j)=\sigma_{ij}$，$i\neq j$，$i=1,\ 2,\ \cdots,\ n$，$j=1,\ 2,\ \cdots,\ n$，即

$$\mathrm{Var}(u) = \boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ \sigma_{\Pi} & \sigma_{\Pi 2} & \cdots & \sigma_{nm} \end{pmatrix} \neq \sigma^2\boldsymbol{I}$$

此时普通最小二乘估计仍然满足无偏性和一致性，但不是有效估计。对于正定矩阵 Ω，必存在可逆矩阵 $\boldsymbol{M}$，使得 $\boldsymbol{M\Omega M}'=\boldsymbol{I}\Rightarrow\boldsymbol{M}'\boldsymbol{M}=\boldsymbol{\Omega}^{-1}$。在方程（4）两边同时乘以 $\boldsymbol{M}$，即 $\boldsymbol{Y}=\boldsymbol{X\beta}+u\Rightarrow\boldsymbol{MY}=\boldsymbol{MX\beta}+\boldsymbol{Mu}$，令 $\boldsymbol{Y}^*=\boldsymbol{MY}$，$\boldsymbol{X}^*=\boldsymbol{MX}$，$\boldsymbol{u}^*=\boldsymbol{Mu}$，得到转换后的新模型：

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\boldsymbol{\beta} + \boldsymbol{u}^* \tag{17}$$

由于 $\mathrm{E}(\boldsymbol{u})=\boldsymbol{0}$，则 $\mathrm{E}(\boldsymbol{u}^*)=\mathrm{E}(\boldsymbol{Mu})=\boldsymbol{M}E(\boldsymbol{u})=\boldsymbol{0}$，而新的随机误差项 $\boldsymbol{u}^*$ 的协方差矩阵为 $\mathrm{Var}(\boldsymbol{u}^*)=\mathrm{E}(\boldsymbol{Muu}'\boldsymbol{M}')=\boldsymbol{M}E(\boldsymbol{uu}')\boldsymbol{M}'=\boldsymbol{M\Omega M}'=\boldsymbol{I}$，因此，$\boldsymbol{u}^*$ 均值为 $\boldsymbol{0}$，同方差且无序列相关。即新模型（17）满足线性回归模型假设 1-4。最小二乘法的目标函数，即残差平方和为 $Q=\boldsymbol{u}^{*\prime}\boldsymbol{u}^*=(\boldsymbol{Mu})'(\boldsymbol{Mu})=\boldsymbol{u}'\boldsymbol{M}'\boldsymbol{Mu}=\boldsymbol{u}'$

$\Omega^{-1}\boldsymbol{u}$，该目标函数是 $\boldsymbol{u}$ 的加权平方和，而权数矩阵是 $\boldsymbol{u}$ 的协方差阵的逆矩阵（因此，广义最小二乘法也称为加权最小二乘法）。而新模型的普通最小二乘估计则是原模型的广义最小二乘法估计（因此广义最小二乘估计是广义矩法估计的特例）。

$$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}=(\boldsymbol{X}^{*\prime}\boldsymbol{X}^{*})^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{Y}^{*}=(\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{Y}=(\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\Omega}^{-1}\boldsymbol{Y} \tag{18}$$

由于变换后的模型满足普通最小二乘的所有假设，因此广义最小二乘法估计量是最佳线性无偏估计。

虽然从理论上讲，广义最小二乘法比普通最小二乘有效，由于通常残差序列的协方差矩阵 $\boldsymbol{\Omega}$ 未知，当使用普通最小二乘估计结果的残差序列的协方差矩阵 $\hat{\boldsymbol{\Omega}}$ 代替广义最小二乘法估计式中的 $\boldsymbol{\Omega}$ 时，估计量虽然仍满足一致性，但却不是最佳线性无偏估计；而且，也很难得到估计量的小样本性质。

当残差存在异方差或序列相关时，普通最小二乘估计仍然满足无偏性，但不满足有效性，因为参数估计的标准误差存在偏差，这时可使用 White[5] 的异方差一致协方差估计方法（残差序列有未知形式的异方差，但序列不相关）和 Newey 和 West[6] 的异方差——自相关一致协方差估计方法（残差序列有未知形式的异方差且存在序列自相关）[6] 来修正 Var（$\hat{\boldsymbol{\beta}}_{\mathrm{GLS}}$），进而修正残差序列异方差和自相关带来的问题。①

White 协方差矩阵：

$$\hat{\sum}_{\mathrm{W}}=\frac{n}{n-k}(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big(\sum_{i=1}^{n}u_i^2x_ix'_i\Big)(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{19}$$

其中，n 是样本点数；k 是回归变量数；u_i 是普通最小二乘残差。

Newey-West 协方差矩阵：

$$\hat{\sum}_{\mathrm{NW}}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\hat{\boldsymbol{\Omega}}(\boldsymbol{X}'\boldsymbol{X})^{-1} \tag{20}$$

其中，$\hat{\boldsymbol{\Omega}}=\dfrac{n}{n-k}\Big[\sum\limits_{i=1}^{n}u_i^2\boldsymbol{x}_i\boldsymbol{x}'_i+\sum\limits_{j=1}^{q}\Big(\Big(1-\dfrac{j}{q+1}\Big)\sum\limits_{i=j+1}^{n}(\boldsymbol{x}_iu_iu_{i-j}\boldsymbol{x}'_{i-j}+\boldsymbol{x}_{i-j}u_{i-j}u_i\boldsymbol{x}'_i)\Big)\Big]$，$q$ 为普通最小二乘残差项的最大自相关阶数，通常取 $q=[n^{1/3}]$，这里 [ ] 表示取整。

由普通最小二乘法推广到广义最小二乘法与由矩法估计推广到广义矩法估计有很多相似之处。下面先从工具变量的阶条件出发，沿着广义最小二乘法的脉络去体会广义矩法估计中“广义”的含义。

4.2 工具变量的阶条件

在回归模型（9），即 $y=x\boldsymbol{\beta}=v\boldsymbol{\beta}_1+w\boldsymbol{\beta}_2+u$ 中，$\boldsymbol{v}$ 是 k_1 维的外生变量向量，$\boldsymbol{w}$ 是 k_2 维的内生变量向量，$k_1+k_2=k$，即有 k 个回归参数。若 z 是 L 维的工具变量向量，则可得总体的 L 个矩条件：

$$\boldsymbol{m}(y,\ z;\ \boldsymbol{\beta})=\mathrm{E}z\boldsymbol{u}=\mathrm{E}z(y-\boldsymbol{x\beta})=\mathrm{E}z(y,\ \boldsymbol{x})\begin{pmatrix}1\\-\beta\end{pmatrix}=\boldsymbol{0} \tag{21}$$

当 $\mathrm{Rank}(\mathrm{E}z\boldsymbol{x})=\mathrm{Rank}(\mathrm{E}z(y,\ \boldsymbol{x}))$ 时，上式有解，否则无解，且当 $\mathrm{Rank}(Ez\boldsymbol{x})=\mathrm{Rank}(\mathrm{E}z(y,\ \boldsymbol{x}))=k$ 时，上式有唯一解（此时，有隐含条件：$L\geqslant k$）。

当 $L>k$ 时，相应的 L 个样本矩条件为

$$\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})=\frac{1}{n}Z'(Y-X\boldsymbol{\beta})=\boldsymbol{0}_{L\times1} \tag{22}$$

即使总体矩条件满足 $\mathrm{E}z(y,\ \boldsymbol{x})_{L\times(k+1)}$ 的秩可以降到 k，由于样本的随机性，$\mathrm{Rank}(\mathrm{E}Z'(Y,\ X))$ 通常不会降，因此没有唯一的 $\boldsymbol{\beta}$ 满足（22）式的 L 个方程，且不同子集的 k 个方程组可得到 $\boldsymbol{\beta}$ 的不同估计值。下面给出广义矩估计的阶条件：

L 个样本矩条件对应 L 个方程，而参数的个数为 k，其解有三种情形：

（1）$L<k$，参数不能识别，$\boldsymbol{\beta}$ 无解；

（2）$L=k$，参数恰好识别，$\hat{\boldsymbol{\beta}}=(Z'X)^{-1}Z'Y$；

（3）$L>k$，参数被过度识别，使用 k 个不同的矩方程可得不同的解，因此，方程组具有多个解。

4.3 过度识别检验

当样本矩条件的个数 $L>k$ 时，参数过度识别。一般情况下，部分矩条件不成立。理论上，若所有矩

① 使用 White 或 Newey-West 异方差一致协方差估计不会改变参数的点估计，只改变参数估计的标准差。

条件都成立，则所有样本矩条件都应在 0 附近，不会偏离很远，Sargan[2] 取检验统计量为

$$J = n\hat{m}(Y,\ Z;\ \hat{\boldsymbol{\beta}})'\hat{S}^{-1}\hat{m}(Y,\ Z;\ \hat{\boldsymbol{\beta}}) \to \chi^2(L-k) \tag{23}$$

给定的显著性水平 α，若检验统计量 J 大于相应的临界值 $\chi^2_{1-\alpha}$ $(L-k)$，说明样本矩条件的约束可能存在问题。

4.4 广义矩方法

当 $L>k$ 时（假设所有的样本矩条件都通过了过度识别检验），矩法估计所得参数的解不唯一，此时可以考虑另外两种方法：两阶段最小二乘法和广义矩法，且正如前面所提到的，两阶段最小二乘法是广义矩法在残差项同方差且序列不相关时的特例。因此，广义矩法估计就是处理过度识别情形的一般方法。

若工具变量的个数恰等于参数的个数 k，线性回归模型的 k 个样本矩条件为 $\hat{\boldsymbol{m}}$ $(\boldsymbol{Y},\ \boldsymbol{Z};\ \boldsymbol{\beta})=\boldsymbol{0}$，等价于：

$$Q=\sum_{j=1}^{k}\hat{m}_j(Y,\ Z;\ \boldsymbol{\beta})^2=\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})'\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})=\boldsymbol{0} \tag{24}$$

当工具变量的个数 $L>k$ 时，可得到 C_L^k 个组合，即可以得到 C_L^k 个矩法估计，但每个矩估计都不能满足（24）式，广义矩法的思想就是使得 $\hat{\boldsymbol{m}}$ $(Y,\ Z;\ \boldsymbol{\beta})$ 的二次距离达到最小，即

$$\hat{\boldsymbol{\beta}}=\arg\min_{\boldsymbol{\beta}}\{\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})'W\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})\} \tag{25}$$

其中，W 是任意的 $L\times L$ 阶非负定加权矩阵（可能是随机的，但不依赖于参数），且 $\mathrm{Rank}(W)\geqslant k$。在线性回归模型 $Y=X\boldsymbol{\beta}+\boldsymbol{u}$ 中，L 个样本矩条件为 $\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})=\dfrac{1}{n}Z'(Y-X\boldsymbol{\beta})=\boldsymbol{0}_{L\times 1}$，则广义矩估计的目标函数为

$$Q=\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})'W\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})=(Z'(Y-X\boldsymbol{\beta}))'W(Z'(Y-X\boldsymbol{\beta})) \tag{26}$$

由一阶条件可得广义矩估计为

$$\hat{\boldsymbol{\beta}}_{GMM}=[(Z'X)'W(Z'X)]^{-1}[(Z'X)'W(Z'Y)] \tag{27}$$

Hansen[3] 证明对任意的 $L\times L$ 阶非负定加权矩阵 W（可能是随机的，但不依赖于参数），若满足 $\mathrm{Rank}(W)\geqslant k$，则 $\hat{\boldsymbol{\beta}}_{\mathrm{GMM}}$ 是一致估计。

4.5 权重矩阵的选择

在广义最小二乘估计中，目标函数为残差序列的加权平方和，权重矩阵是残差序列协方差矩阵的逆矩阵。以此类似，广义矩法估计的目标函数是各样本矩的加权平方和，因此权重矩阵可考虑样本矩的协方差矩阵的逆矩阵，即

$$W=S^{-1},\ S=\mathrm{Var}(\sqrt{n}\,\hat{\boldsymbol{m}}(Y,\ Z;\ \hat{\boldsymbol{\beta}}))=\frac{1}{n}\sum_{i=1}^{n}\hat{g}_i(Y,\ Z;\ \hat{\boldsymbol{\beta}})\hat{g}_i(Y,\ Z;\ \hat{\boldsymbol{\beta}})' \tag{28}$$

$\hat{g}$ 满足 $\hat{\boldsymbol{m}}(Y,\ Z;\ \hat{\boldsymbol{\beta}})=\dfrac{1}{n}\sum_{i=1}^{n}\hat{g}_i(Y,\ Z;\ \hat{\boldsymbol{\beta}})$。由（27）可知广义矩法估计依赖于加权矩阵 W，（28）式表明权重矩阵依赖于参数值，这是一个循环的问题。Hansen[3] 取权重矩阵 $W=S^{-1}$，S 称为谱密度矩阵，考虑到残差项可能存在异方差性和序列相关性，其估计量 $\hat{S}$ 通过下面关系式得到：

$$\hat{S}=\hat{S}_0+\sum_{j=1}^{q}\left(1-\frac{j}{q+1}\right)(\hat{S}_j+\hat{S}'_j),\ \hat{S}_j=\frac{1}{n}\sum_{i=j+1}^{n}\hat{g}_i(Y,\ Z;\ \hat{\boldsymbol{\beta}})\hat{g}_{i-j}(Y,\ Z;\ \hat{\boldsymbol{\beta}})' \tag{29}$$

q 为残差自相关的滞后阶数，通常取 $q=[n^{1/3}]$，[ ] 表示取整。权重矩阵依赖于参数取值，同样是一个循环问题，在实际中是按如下步骤来实施广义矩法估计的：

第一步，设权重矩阵为单位矩阵，即 $W_0=I$；

第二步，求解优化问题：$\hat{\boldsymbol{\beta}}_1=\arg\min_{\boldsymbol{\beta}}\{\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})'W_0\hat{\boldsymbol{m}}(Y,\ Z;\ \boldsymbol{\beta})\}$；

第三步，由 $\hat{\boldsymbol{\beta}}$ 得到调整后的权重矩阵 W_1；

第四步，使用调整后的权重矩阵重复第二步和第三步直到权重矩阵收敛。

4.6 广义矩法估计的 MATLAB 程序①

```
function [para,stats]=gmm(moment,para0,Y,X,Z,number,L)
```

① MATLAB 程序主要参考曹志广编著的《金融计算与编程——基于 MATLAB 的应用》一书[7]。

```
    % input:
    % moment:moment conditions function defined by users
    % para0:initial values for estimated parameters
    % number:maximum convergence number when choosing optimal weighting matrix
    % L:number of moment conditions
    % Y:data for dependent variable
    % X:data for independent variables
    % Z:data for instrument variables
    % output:
    % para:estimated parameters
    % stats:a structure variable,its fields represent as follows:
    % std:standard errors for each estimated parameters
    % t_stat:T statistics for each estimated parameters
    % V:covariance matrix for estimated parameters
    % it:number of iterations
    % chi_stat:chi2 statistics for over-identifying test,null hypothesis is moment conditions
    % are feasible
    % p:P value for chi2 statistics
%%%%%% main function
    nlag=round(size(Y,1)^(1/3));
    W(:,:,1)=eye(L);
    [para(:,1),fv(:,1)]=fminsearch(moment,para0,[],1,Y,X,Z,W(:,:,1));
    for i=2:number
        mom=feval(moment,para(:,i-1),2,Y,X,Z,W(:,:,i-1));
        W(:,:,i)=gmm_weight(mom,nlag);
        [para(:,i),fv(i)]=fminsearch(moment,para(:,i-1),[],1,Y,X,Z,W(:,:,i));
        if abs(fv(i)-fv(i-1))/abs(fv(:,i-1))<1e-4|fv(i)<=1e-10
          break
        end
    end
    it=i;
    if it==number
      error('number of iteration exceeds defined maximum number')
    else
        para=para(:,it);
        f0=feval(moment,para,3,Y,X,Z,W(:,:,it));
        for j=1:length(para)
          a=zeros(length(para),1);
          eps=max(para(j)*1e-6,1e-5);
          a(j)=eps;
          M(:,j)=(feval(moment,para+a,3,Y,X,Z,W(:,:,it))-f0)/eps;
        end
    end
    stats.V=pinv(M'*W(:,:,it)*M)/size(Y,1);
    stats.std=sqrt(diag(stats.V));
```

```
stats. t_stat=para. /(stats. std);
stats. chi_stat=size(Y,1) * fv(it);
stats. p=1-chi2cdf(stats. chi_stat,L-length(para0));
stats. it=it;
%%%%%%%%%% auxiliary function to find the weight matrix using Newey and West method
function W= gmm_weight(mom,nlag)
q=size(mom,2);T=size(mom,1);a2=zeros(q,q);a3=zeros(q,q);
for j=1:nlag
    a1=zeros(q,q);
    for i=j+1:T
        a1=mom(i,:)' * mom(i-j,:)+a1;
    end
    S(:,:,j)=a1/T;a2=(1-j/(nlag+1)) * S(:,:,j)+a2;a3=(1-j/(nlag+1)) * S(:,:,j)+a3;
end
b1=zeros(q,q);
for i=1:T
    b1= mom(i,:)' * mom(i,:)+b1;
end
if nlag==0   newS=b1/T;else   newS=a2+a3+b1/T;
end
W=pinv(newS);
```

下面以一元线性回归为例，使用广义矩法估计参数：

$$Y=\beta_0+\beta_1 X+\eta$$

考虑最简单的两个矩条件：$E(\eta)=E(Y-\beta_0-\beta_1 X)=0$，$E(X\eta)=E(X(Y-\beta_0-\beta_1 X))=0$。首先定义矩条件函数：

```
function f=linearmodel(para,num,Y,X,Z,W)
    [T,q]=size(Y);beta_0=para(1);beta_1=para(2);eta=[Y-(beta_0+beta_1 * X)];
    for i=1:T
        m_t(i,:)=kron(eta(i,:),Z(i,:));
    end
    m=mean(m_t)';obj=m' * W * m;
    if num ==1 f=obj;elseif num == 2 f=m_t;elseif num == 3 f=m;
end
```

在 MATLAB 主窗口中输入：

```
kk=1000;X=randn(kk,1);Y=1+X+trnd(6,kk,1)/2;Z=[ones(kk,1),X];number=100;para0=[0;1];
[para,Stats]=gmm('linearmodel',para0,Y,X,Z,number,2);
```

可得截距项和斜率项分别为：1.0064 和 0.9953。而普通最小二乘结果分别为：0.9763 和 0.9883。显然该算例中，使用广义矩法估计比普通最小二乘结果更接近真实数据。然而，这样的优势并不是稳健的，当改变一元线性回归模型的参数真值或残差项的分布时，有时普通最小二乘结果更优。因此，针对不同研究问题，在方法的选择上还是要谨慎。

5 结论

广义矩法可以在非常宽松的条件下，得到参数的一致估计。因为在现实中很难满足线性回归模型的经典假设，因此广义矩法估计更符合实际应用；而且可以看到，广义矩法估计是涵盖普通最小二乘法、广义

最小二乘法、工具变量法以及极大似然方法的更一般的方法。尽管广义矩法有诸多优势，但我们仍需要注意几个问题：①大样本性。这是一致估计满足的前提条件。②工具变量的选择。并不是所有的问题都可以很容易地找到工具变量，当工具变量选择有误时，会（严重）影响参数估计的标准误差。③工具变量个数的确定。当工具变量的个数大于参数的个数时，需反复做关于样本矩条件的检验，且可能出现多个结果。因此学者可以根据自己的研究需要做出选择，若仅关心回归系数的显著性，采用 Newey-West 回归是很好的一个选择；若更关心系数的取值，则需选择更为精确的估计方法。

参考文献：

[1] Greene, W. Econometric analysis [M]. *Prentice Hall*, New Jersey, 2008.

[2] Sargan, T. The estimation of economic relationships using instrumental variables [J]. *Econometrica*, 1958, (26): 393-415.

[3] Hansen, L. Large sample properties of generalized method of moments estimators [J]. *Econometrica*, 1982, (50): 1029-1054.

[4] Hausman, J. Specification tests in economics [J]. *Econometrica*, 1978, (46): 1251-1271.

[5] White, H. Heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity [J]. *Econometrica*, 1980, (48): 817-838.

[6] Newey, W., K. West. A simple, positive semi-definite heteroskedasticity and auto-correlation consistent covariance matrix [J]. *Econometrica*, 1987, (55): 703-708.

[7] 曹志广. 金融计算与编程——基于 MATLAB 的应用 [M]. 上海：上海财经大学出版社，2013.
Cao, Z. Financial calculating and programming: applications based on MATLAB [M], *Shanghai University of Finance and Economics Press*, Shanghai, 2013.

Multiple Linear Regression: Generalized Method of Moments and MATLAB Implementation

Dong Chenyu[1], *Liu Weiqi*[2]

1. School of Mathematical Sciences, Shanxi University, Taiyuan 030006, China;
2. Institute of Management and Decision, Shanxi University, Taiyuan 030006, China

Abstract: In linear regression model, the excellent properties of ordinary least squares estimation depend heavily on various assumptions, when one or more assumptions does not hold, how to estimate the regression coefficients effectively? Generalized least squares, instrumental variable method and generalized method of moments can be used to solve these problem, and ordinary least squares, generalized least squares and instrumental variables method are special cases of the generalized method of moments. In this paper, I based on the principle of moment method, analysis generalized method of moments estimation of regression coefficients in the light of the classical hypothesis of linear regression models. The numerical simulation example is given to illustrate the implementation of generalized method of moment estimation in MATLAB.

Key words: Generalized least squares; Instrumental variable method; Generalized method of moments