目　录

经济计量模型表现为一个或多个数值方程，方程中每一个数值系数及常数，需要用回归分析方法来求。回归分析有线性的，有非线性的①。

一、线性回归法

设有一个随机变量Y，其与另一变量X之间存在的线性关系为：

$$Y=\alpha+\beta X \tag{1}$$

α为回归常数，β为回归系数。我们观察到N个X值和N个Y值。设X_i与Y_i相联系$i=1, 2, \cdots N$。设求出的经验回归方程为：

$$Y'=a+bX \tag{2}$$

Y'是Y的估算值，其与观察值之间的误差用ε_i表示

$$Y_i-Y'_i=Y_i-a-bX_i=\varepsilon_i \tag{3}$$

设误差是正态分布的，则误差ε_i的几率密度$P(\varepsilon_i)$是

$$P(\varepsilon_i)=\frac{1}{(\sqrt{2\pi})\sigma}\exp\left(-\frac{1}{2\sigma^2}\varepsilon_i^2\right) \tag{4}$$

同时发生误差ε_i，$i=1, 2, \cdots N$，的几率P'是

120

$$P'=P_1\ P_2\cdots P_N=\frac{1}{(2\pi)^{N/2}\sigma^N}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\varepsilon_i^2\right) \tag{5}$$

---

①笔者于解放前曾用非线性回归方法分析通货恶性膨胀下物价的趋势，见《现代知识》（1948）中《物价的趋势》一文。

我们用 Q 代表 $\sum e^2_i$

$$Q=\sum_{i=1}^{N} e^2_i=\sum_{i=1}^{N}(Y_i-a-bX_i)^2 \quad (6)$$

将几率密度重新写为

$$P'=\frac{1}{(2\pi)^{N/2}\sigma^N \exp\frac{1}{2\sigma^2}Q} \quad (7)$$

Q愈小，则几率密度 P′愈大，问题变为使Q为最小，Q是平方之和，故有最小二乘法之称。 因此才能发现问题

求Q的最小值，须求它对 a 和 b 的偏微商，使它们等于零。

求偏

$$Na^*+\left(\sum_{i=1}^{N} X_i\right)b^*=\sum_{i=1}^{N} Y_i$$

$$\left(\sum_{i=1}^{N} X_i\right)a^*+\left(\sum_{i=1}^{N} X^2_i\right)b^*=\sum_{i=1}^{N} X_iY_i \quad (8)$$

a^*和 b^*是方程(8)的解，它们是 α 和 β 的最小二乘估算值。

二、一般经济计量模型

设有 g 个内生变量（因变量）y_1，y_2，…，y_g 和 k 个预定变量（外生变量或前期内生变量）x_1，x_2，…，x_k，一般经济计量模型可以写成

$$y_1\gamma_{11}+y_2\gamma_{21}+\cdots+y_g\gamma_{g1}+x_1\beta_{11}+x_2\beta_{21}+\cdots+x_k\beta_{k1}=\epsilon_1$$

$$y_1\gamma_{12}+y_2\gamma_{22}+\cdots+y_g\gamma_{g2}+x_1\beta_{12}+x_2\beta_{22}+\cdots+x_k\beta_{k2}=\epsilon_2$$

$$\vdots$$

$$\vdots$$

$$\vdots$$

$$y_1 \gamma_{1g}+y_2 \gamma_{2g}+\cdots+y_g \gamma_{gg}+x_1 \beta_{1g}+x_2 \beta_{2g}+\cdots x_k \beta_{kg}=\epsilon_g$$

其中ϵ_1，ϵ_2，…，ϵ_g是g个随机干扰项（随机变量），γ是内在变量的系数，β是预定变量的系数。只要独立方程的数目和内生变量的数目一样多，方程组是完全的，可以求出内生变量的值。每个方程称为一个结构方程，方程组称为结构组。

用矢量——矩阵符号，结构组可写成

$$\underset{1\times g}{y}\ \Gamma + \underset{1\times k}{x}\ \mathrm{B} = \underset{1\times g}{\varepsilon} \tag{9}$$

y和x是行矢量

$$y=(y_1\ y_2\ \cdots\ y_g) \tag{10}$$

$$x=(x_1\ x_2\cdots\ x_k) \tag{11}$$

ε也是一个行矢量

$$\varepsilon=(\epsilon_1\ \epsilon_2\ \cdots\ \epsilon_g) \tag{10}$$

矩阵Γ和B分别是g^2和gk个结构常数的矩阵

$$\underset{g\times g}{\Gamma}=(\gamma_{jh})=\begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1g} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2g} \\ \vdots & \vdots & & \\ \gamma_{g1} & \gamma_{g2} & \cdots & \gamma_{gg} \end{pmatrix} \tag{13}$$

$$\mathrm{B}_{k\times g}=(\beta_{jh})=\begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1g} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2g} \\ \vdots & \vdots & & \\ \beta_{k1} & \beta_{k2} & \cdots & \beta_{kg} \end{pmatrix} \tag{14}$$

结构方程有一个不确定性。用一个非零常数乘某个方程的所有各项不改变这个方程。用一个归一规则可以消除不确定性，给每个方程的一个非零结构系数选一个具体数值。例如用 $-\gamma_{hh}$ 除方程 h 的一切系数，所以

$$\gamma_{hh} = -1, \qquad h = 1, 2, \cdots, g. \tag{15}$$

这样 Γ 矩阵的主对角线上所有元素都是 -1。

如果 Γ 矩阵有一个非零行列式，因而有逆矩阵，则用 Γ 的逆矩阵右乘方程(9)得

$$y\Gamma\Gamma^{-1} + xB\Gamma^{-1} = \varepsilon\Gamma^{-1} \tag{16}$$

求解 y，

$$y = -xB\Gamma^{-1} + \varepsilon\Gamma^{-1} \tag{17}$$

引入以下符号

$$\underset{k\times g}{\Pi} \equiv -\underset{k\times g}{B}\,\underset{g\times g}{\Gamma^{-1}} \tag{18}$$

不需引伸
V.级核论

$$\underset{1\times g}{u} \equiv \underset{1\times g}{\varepsilon}\,\underset{g\times g}{\Gamma^{-1}} \tag{19}$$

方程(17)可以写成

$$\underset{1\times g}{y} = \underset{1\times k}{x}\,\underset{k\times g}{\Pi} + \underset{1\times g}{u} \tag{20}$$

称为约化式。也可用求和符号表示为

$$y_h = \sum_{j=1}^{k} x_j \Pi_{jh} + u_h, \qquad h = 1, 2, \cdots, g \tag{21}$$

约化式系数矩阵的元素可以看成是比较静力学结果，说明任何预定变量变化时，每个内生变量的变化程度。对 (21) 求导

$$\frac{\partial y_h}{\partial x_j} = \Gamma_{jh}, \quad j=1, 2, \cdots, k, \quad h=1, 2, \cdots, y \tag{22}$$

譬如x_j为农产品价格，则通过（22）可以求出它的变动对国民经济所有其他内生变量的影响。

三、再谈线性回归法

为简化计，将方程（21）中h下标去掉，各个参数写成β而不写成Γ：

$$y = \sum_{j=1}^{k} x_j \beta_j + u \tag{22}$$

这个方程也可表示为矢量形式

$$y = x\beta + u = (x_1 \ x_2 \cdots x_k)\begin{pmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_k\end{pmatrix} + u \tag{23}$$

方程（22），（23）代表基本线性回归模型。如果只有一个自变量，$k=2$，称为简单线性回归法。更广义地说，如果$k\geq 2$，称为多元线性回归法。如果$\hat{\beta}_j$是（23）中第j个回归系数的估算值，则

$$\hat{\beta}_j = \frac{\partial y}{\partial x_j}, \quad j=1, 2, \cdots, k \tag{24}$$

是第j个自变量对y的影响的估计。

经济计量方法除一个经济计量模型外，还需要变量的数据。这些

数据能用来估算模型的参数。如系（23）那样的一般线性模型，一宗数据样本包括因变量（内在变量）y的n个观察值的矢量和自变量（外在变量）X的矢量的n个观察值的矩阵，写成

$$\underset{n\times 1}{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix},\quad \underset{n\times k}{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k} \\ x_{21} & x_{22} & \cdots & x_{2k} \\ \vdots & \vdots & & \\ x_{n1} & x_{n2} & \cdots & x_{nk} \end{pmatrix} \quad (25)$$

y的元素是数字y_i，其中y_i是变量y在观察i的数值，i是一个观察指数。这里$i=1, 2, \cdots, n$，其中n是样本大小。X的元素是数字x_{ij}，x_{ij}是变量x_j在观察i的数值。X矩阵的每一行代表所有自变量在某一次观察的观察值，而每一列代表一个自变量的所有观察值。一般地数据点比要估算的参数多，所以表示样本大小的n比k大，k是β中参数的数目。n和k的差额，也是X的行数与列数之差，称为问题的自由度d、f

并且一般地假设X矩阵中k个自变量是线性独立的，所以各列是线性独立矢量。也就是要求X矩阵有一个$k\times k$的子矩阵，其行列式不等于零，称为阶数假设，用符号表示为$\rho(X)=k$。

任何回归分析涉及三部分，一个模型，一组数据和一种估算方法。这种方法把y和X代表的数据变换为模型参数的估计量，$\hat{\beta}$

最简单的方法是用眼和手画一个最佳曲线。不但不准确，自变量多了无法绘图。现代经济计量技术用两种方法。最小二乘法或最大似然法

简单线性回归法只有一个自变量，模型表示为

$$y=\beta_1 x+\beta_2+u \quad (26)$$

其中y是因变量；x是自变量，假设不是随机的（固定的）；u是随

机干扰项；β_1 和 β_2 是斜率和截距参数。样本中的变量用 i 标志，并且对随机干扰项加一些假设，得到简单线性回归的基本线性回归模型：

$$
\begin{aligned}
& y_i = \beta_1 x_i + \beta_2 + u_i \quad i = 1, 2, \cdots, n \\
& E(u_i) = 0 \\
& V_{ar}(u_i) = \sigma^2 < \infty \qquad (27) \\
& C_{ov}(u_i u_j) = 0 \qquad i \neq j \\
& x_i
\end{aligned}
$$

$E(u_i)$是不能观察的随机变量 u_i 的预期值，$E(u_i) = 0$ 是说每个随机干扰项的平均值为零。V_{ar}是方差；C_{ov}是协方差。

最小二乘法产生 β_1 和 β_2 的估计量，不产生方差 σ^2 的估计量。最大似然法则产生所有三种估计量。本文对最大似然法不作介绍。

多元线性回归的基本线性回归模型可以写成

$$
\begin{aligned}
& y_i = \sum_{j=1}^{k} x_{ij} \beta_j + u_i = x_i \beta + u_i \\
& \qquad\qquad i = 1, 2, \cdots, n \\
& E(u_i) = 0 \qquad (28) \\
& V_{ar}(u_i) = \sigma^2 < \infty \\
& C_{ov}(u_i u_j) = 0 \qquad i \neq j \\
& x_{ij}
\end{aligned}
$$

y_i 代表因变量 y 在第 i 次观察所取的值，x_{ij}是第 j 个自变量在第 i 次观察中的数值，x_i 是在第 i 次观察中 k 个自变量的矢量，β是要估算的参量的 $k \times 1$ 矢量，u_i 是第 i 次观察中随机干扰项。求了一个参量 β_k 是截距，所以对所有 i 来说 $x_{ik} \equiv 1$。问题是求 $k+1$ 未知数的估计量，即 $k-1$ 个斜率系数 β_1，β_2，…，

β_{k-1}，一个截距 β_k，和方差 σ^2。

基本线性模型用矩阵表示最为方便，令

$$\underset{n\times 1}{y}=\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix} \qquad \underset{n\times k}{X}=\begin{pmatrix} X_{11} & X_{12} & \cdots & X_{1\,k-1} & 1 \\ X_{21} & X_{22} & \cdots & X_{2\,k-1} & 1 \\ \vdots & & & & \\ X_{n1} & X_{n2} & \cdots & X_{n\,k-1} & 1 \end{pmatrix}$$

$$\underset{n\times 1}{u}=\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \qquad \underset{k\times 1}{\beta}=\begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} \tag{29}$$

基本线性模型（28）于是可写成

$$\begin{aligned} &\underset{n\times 1}{y}=\underset{n\times k}{X}\ \underset{k\times 1}{\beta}+\underset{n\times 1}{u} \\ &E(u)=0 \\ &C_{ov}(u)=\sigma^2\ I \\ &X\text{是一个固定矩阵，}\ \rho(X)=k<n \end{aligned} \tag{30}$$

其中 I 是 $n\times n$ 单位矩阵。

（30）中第一个统计假设，写成矢量方程

$$E(u)=0 \tag{31}$$

是说所有随机干扰项的平均值是零。它与把 u 解释为随机干扰项的矢量是一致的。而且平均来说，模型正确，因为

$$E(y)=X\beta+E(u)=X\beta \tag{32}$$

第二个统计假设可以表示为矩阵方程

$$
\begin{aligned}
Cov(u) &= E\{(u-E(u))(u-E(u))'\} \\
&= E(uu') = \sigma^2 I
\end{aligned}
\tag{33}
$$

其中u'是u的转置矩阵，都写出来是

$$
\begin{pmatrix}
Var(u_1) & Cov(u_1 u_2) & \cdots & Cov(u_1 u_n) \\
Cov(u_2 u_1) & Var(u_2) & \cdots & Cov(u_2 u_r) \\
\vdots & \vdots & & \vdots \\
Cov(u_n u_1) & Cov(u_n u_2) & \cdots & Var(u_n)
\end{pmatrix}
$$

$$
= \begin{pmatrix}
\sigma^2 & 0 & \cdots & 0 \\
0 & \sigma^2 & \cdots & 0 \\
\vdots & \vdots & & \vdots \\
0 & 0 & & \sigma^2
\end{pmatrix}
\tag{34}
$$

方差都是σ^2，称为同方差性。协方差都是零，说明没有序列相关。同方差性和没有序列相关两个假设合起来称为球状干扰假设。

四、高斯——马尔柯夫定理和最小二乘法一致性定理

用最小二乘法或最大似然法（后者假设正态分布的随机项）在多元线性回归中得到的系数矢量的估计量是

$$
\hat{\beta} = (X'X)^{-1} X'y \tag{35}
$$

高斯——马尔柯夫定理和最小二乘一致性定理是说（35）中的最小二乘估计量的最佳性质，它们在一定条件下也是最大似然估计量。

按照高斯——马尔柯夫定理，在上述基本线性回归模型的假设下，这些估计量是线性的和无偏的估计量，它们是所有线性无偏估计量中

最佳者；也就是，它们是在线性无偏估计量的一类中方差最小的估计量。这个定理有时称为BLUE定理。BLUE是最佳线性无偏估计量的缩写语。

如果$\hat{\theta}$是θ的估计量，令$|\hat{\theta}-\theta|$为估计量与真实量之间距离的标量量度，如果对任何$\epsilon>0$，

$$\lim_{n\to\infty} P(|\hat{\theta}-\theta|<\epsilon)=1 \tag{36}$$

这里P(1)是距离小于规定距离ϵ的几率。只有样本无限增大时，这个几率逼近一，估计量才称为具备一致性。最小二乘一致性定理是说最小二乘估计量是具备一致性的。

五、最小二乘法的新发展

估算联立方程组的参数的简单方法是普通最小二乘法(OLS)。采用这种方法的宏观动态经济计量模型如丁伯根模型(1939)，克来因两次大战间模型(1950)，Duesenberry 等人的季度模型(1960)，Suits 年度模型(1962)，美国商业经济局和经济分析局季度模型(1974)，美国联邦准备局／麻省理工学院——宾州大学——社会科学研究委员会季度模型(1974)，圣路易季度模型(1974)，密歇根季度模型(1974)，大通银行经济计量季度模型(1974)。这种方法将最小二乘法分别用于模型的每一个方程，忽视内生自变量与外生变量之间的区别。这种方法导致有偏的和不一致的估计量。不过由于计算简单，一个国家或一个单位在开始搞宏观动态经济计量模型时还是可以采用的。

设计经济计量模型有个重要的“识别”问题。估算结构方程从估算约化方程

$$y=X\Pi+u \tag{37}$$

开始，利用对g个内生变量和k个预定变量X的观察数据。所得估计量总结为$k \times g$矩阵Π。这个情报然后被用来估算结构式

$$y\Gamma + XB = \varepsilon \qquad (38)$$

中内生变量和预定变量的系数，它们分别是Γ和B。前面已介绍过

$$\Pi = -B\Gamma^{-1} \qquad (39)$$

但是关于Π的元素的估计量的情报显然不足以区别B和Γ对决定Π的影响。为了决定B和Γ，显然需要更多的情报。这种情报通常称为先验情报，面为它在估算之前。面(39)中包含的情报称为后验情报，因为它根据约化式方程的估算，特别是根据Π的估算。

识别问题主要是提供足够的先验情报，可以从约化式参数Π决定结构参数B和Γ。如果先验情报不足，不能从Π决定B和Γ，这个情况称为不足识别。在此情况下不可能将真实的结构参数与"虚假"参数区别。有了刚好足够的情报，可以唯一地（在归一之后）决定B和Γ的元素，所谓精确识别。如有更多的情报，则从Π和先验情报会有不止一种方式来决定B和Γ的元素，这是超额识别，经济计量模型往往如此。

现举一种农产品的微现模型为例，包括以下三个方根：

$$q^D = r_1 P + \beta_1 I + \delta_1 + \in^D \qquad (40)$$

$$q^S = r_2 P + \beta_2 r + \delta_2 + \in^S \qquad (41)$$

$$q^D = q^S \qquad (42)$$

其中q^D是需求量，q^S是供应量，P是价格，I是收入，r是降雨

量，$\in^{D}$是需求的随机干扰项，$\in^{S}$是供应的随机干扰项。令q为平衡数量，即q^{D}和q^{S}的共同值。则微观模型包括(40)和(41)两个结构方程。采用矢量——矩阵符号，结构式是下列结构方程组

$$
(q \quad P)\begin{pmatrix} 1 & 1 \\ -r_1 & -r_2 \end{pmatrix} + (I \quad r \quad 1)\begin{pmatrix} -\beta_1 & 0 \\ 0 & -\beta_2 \\ -\delta_1 & -\delta_2 \end{pmatrix} = (\in^{D} \quad \in^{S}) \tag{43}
$$

使(40)和(41)相等，可以求P和q

$$
P = \frac{\beta_1}{r_2 - r_1} I - \frac{\beta_2}{r_2 - r_1} r + \frac{\delta_1 - \delta_2}{r_2 - r_1} + \frac{\in^{D} - \in^{S}}{r_2 - r_1} \tag{44}
$$

$$
q = \frac{r_2 \beta_1}{r_2 - r_1} I - \frac{r_1 \beta_2}{r_2 - r_1} r + \frac{r_2 \delta_1 - r_1 \delta_2}{r_2 - r_1} + \frac{r_2 \in^{D} - r_1 \in^{S}}{r_2 - r_1} \tag{45}
$$

(44) 和 (45) 称为约化式方程，可以写成

$$
(q \quad P) = (I \quad r \quad 1)\begin{pmatrix} \dfrac{r_2 \beta_1}{r_2 - r_1} & \dfrac{\beta_1}{r_2 - r_1} \\[2ex] \dfrac{-r_1 \beta_1}{r_2 - r_1} & \dfrac{-\beta_2}{r_2 - r_1} \\[2ex] \dfrac{r_2 \delta_1 - r_1 \delta_2}{r_2 - r_1} & \dfrac{\delta_1 - \delta_2}{r_2 - r_1} \end{pmatrix}
$$

$$
+ \left(\frac{r_2 \in^{D} - r_1 \in^{S}}{r_2 - r_1} \quad \frac{\in^{D} - \in^{S}}{r_2 - r_1} \right) \tag{46}
$$

在（43）中，B矩阵有两个零，反映需求不依赖降雨和供应不依赖收入。这类零值限制可以用来估算结构式。例如由（46）

$$\begin{pmatrix} \Pi_{11} & \Pi_{12} \\ \Pi_{21} & \Pi_{22} \\ \Pi_{31} & \Pi_{32} \end{pmatrix} = \frac{1}{\gamma_2-\gamma_1}\begin{pmatrix} \gamma_2\beta_1 & \beta_1 \\ -\gamma_1\beta_2 & -\beta_2 \\ \gamma_2\delta_1-\gamma_1\delta_2 & \delta_1-\delta_2 \end{pmatrix} \tag{47}$$

在此情况下结构常数——β，γ，δ——可以从约化式参数——Π——唯一地决定

$$\gamma_1=\frac{\Pi_{21}}{\Pi_{22}} \qquad \gamma_2=\frac{\Pi_{11}}{\Pi_{12}}$$

$$\beta_1=\Pi_{12}\left(\frac{\Pi_{11}}{\Pi_{12}}-\frac{\Pi_{21}}{\Pi_{22}}\right) \qquad \beta_2=-\Pi_{22}\left(\frac{\Pi_{11}}{\Pi_{12}}-\frac{\Pi_{21}}{\Pi_{22}}\right)$$

$$\delta_1=\Pi_{31}-\Pi_{32}\frac{\Pi_{21}}{\Pi_{22}} \qquad \delta_2=\Pi_{31}-\Pi_{32}\frac{\Pi_{11}}{\Pi_{12}} \tag{48}$$

这样成为精确识别的情况。

间接最小二乘法是一种有限情报技术，可用来得到精确识别的一致性估计量。在精确识别情况下，结构参数由约化式参数唯一地决定，所以估算的约化式参数可以用来间接推导结构参数的估计量，故称为“间接最小二乘法。”此法产生的估计量是一致性的，但是有偏的。它不能用于超额识别的情况。

1957年Basmann提出两级最小二乘法，可用于超额识别情况。它产生的估计量是有偏的，但是一致性的。用此法的经济计量

模型如刘大中季度模型（1963），布鲁金斯季度模型（1975），华通季度模型（1972），Fair季度模型（1974），数据公司季度模型（1976），华通年度和产业模型（1975）等，

设我们要估算的联立方程组是

$$\underset{1\times g}{y_i}\ \underset{g\times g}{\Gamma}+\underset{1\times k}{X_i}\ \underset{k\times g}{B}=\underset{1\times g}{\varepsilon_i},\quad i=1,2,\cdots,n \tag{49}$$

y_i 是 g 个内生变量的矢量，X_i 是 k 个预定变量的矢量，ε_i 是 g 个随机干扰项的矢量，共 g 个方程。下标 i 是观察番号，共观察 n 次。

对 g 个内生变量的每一个的 n 次观察可以总结为数据矩阵

$$\underset{n\times g}{Y}=\begin{pmatrix} y_1\\ y_2\\ \vdots\\ y_n\end{pmatrix}=\begin{pmatrix} y_{11} & y_{12} & \cdots & y_{1g}\\ y_{21} & y_{22} & \cdots & y_{2g}\\ \vdots & & & \\ y_{n1} & y_{n2} & \cdots & y_{ng}\end{pmatrix} \tag{50}$$

对 k 个预定变量的每一个的 n 次观察可以总结为数据矩阵

$$\underset{n\times k}{X}=\begin{pmatrix} X_1\\ X_2\\ \vdots\\ X_n\end{pmatrix}=\begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1k}\\ x_{21} & x_{22} & \cdots & x_{2k}\\ \vdots & & & \\ x_{n1} & x_{n2} & \cdots & x_{nk}\end{pmatrix} \tag{51}$$

利用这些数据矩阵，联立方程组可以写成

$$\underset{n\times g}{Y}\ \underset{g\times g}{\Gamma}+\underset{n\times k}{X}\ \underset{k\times g}{B}=\underset{n\times g}{E} \tag{52}$$

其中 E 是 $n\times g$ 矩阵，其每一行是（49）中随机干扰项的 ε_i 矢量。

联立方程式估算问题是用矩阵 Y 和 X 估算（52）的参数，即系数矩阵 Γ 和 B 以及 ε_i 的协方差的矩阵 Σ。

$$\mathrm{Cov}(\varepsilon_i) = E(\varepsilon_i' \varepsilon_i) = \underset{g\times g}{\Sigma} \tag{53}$$

每个方程可以归一，使Γ矩阵主对角线上n个元素都等于-1

$$\gamma_{hh} = -1, \quad h = 1,\ 2,\ \cdots,\ g \tag{54}$$

解方程h求y_{ih}

$$y_{ih} = \sum_{\substack{h'=1 \\ h'\neq h}}^{g} y_{ih'}\gamma_{h'h} + \sum_{j=1}^{k} x_{ij}\beta_{jh} - \varepsilon_{ih}, \quad h = 1,\ 2,\ \cdots,\ g \tag{55}$$

其中h是方程的指数，h'是内生变量的指数，j是外生变量的指数

现在考虑方程组的第一方程（$h=1$）并且假设对系数的先验限制都是零限制。必要时可将变量重新编号，所以只有前g_1个内生变量（$g_1<g$）和只有前k_1个外生变量（$k_1<k$）包括在方程中，其他（$g-g_1$）+（$k-k_1$）个变量的系数为零。第一方程于是可写成

$$y_{i1} = \sum_{h'=2}^{g_1} y_{ih'}\gamma_{h'1} + \sum_{j=1}^{k_1} x_{ij}\beta_{j1} - \epsilon_{i1} \tag{56}$$

引入自变量矢量

$$\begin{aligned} Y_{i1} &= (y_{i2} \quad y_{i3} \quad \cdots \quad y_{ig_1}), \\ X_{i1} &= (x_{i1} \quad x_{i2} \quad \cdots \quad x_{ik_1}) \end{aligned} \tag{57}$$

并且引入

$$\gamma_1 = (\gamma_{21} \quad \gamma_{31} \cdots \gamma_{g_1 1})', \quad \beta_1 = (\beta_{11} \quad \beta_{21} \quad \cdots \quad \beta_{k_1 1})' \tag{58}$$

代表非零系数，方程（56）可以写成

$$y_{i1} = \underset{1\times(g_1-1)}{Y_{i1}}\ \underset{(g_1-1)\times 1}{\gamma_1} + \underset{1\times k_1}{X_{i1}}\ \underset{k_1\times 1}{\beta_1} - \epsilon_{i1} \tag{59}$$

其中下标1表示“包括在第一方程中”。

在（50）（51）中总结为Y和X矩阵的数据，可以按（59）中指示的变量分块。

$$\underset{n\times g}{Y} = \left(\underset{n\times 1}{y_1} \;\middle|\; \underset{n\times(g_1-1)}{Y_1} \;\middle|\; \underset{n\times(g-g_1)}{Y_2}\right) \tag{60}$$

$$\underset{n\times k}{X} = \left(\underset{n\times k_1}{X_1} \;\middle|\; \underset{n\times(k-k_1)}{X_2}\right) \tag{61}$$

方程（59）可用数据矩阵表示为

$$\underset{n\times 1}{y_1} = \underset{n\times(g_1-1)}{Y_1}\ \underset{(g_1-1)\times 1}{\gamma_1} + \underset{n\times k_1}{X_1}\ \underset{k_1\times 1}{\beta_1} + \underset{n\times 1}{\varepsilon_1} \tag{62}$$

将最小二乘法直接用于估算这个方程的问题在于存在自变内生变量，Y_1，它们与随机干扰项ε_1有关。如果这些变量能用与随机干扰项无关的相应变量来代替，所得的估计量是一致性的。二阶最小二乘法利用估计的约化型达到这一点，用估计值代表内生自变量。例如在（62）中Y_1用$\hat{Y}_1$代替。所将方程的最小二乘估计量是二阶最小二乘估计量，