

第六章 联立方程 计量经济模型理论方法

The Theory and Methodology of the Simultaneous-Equations Model(SEM)

本章内容

- § 6.1 联立方程计量经济学模型的提出
- § 6.2 联立方程计量经济学模型的基本概念*
- § 6.3 联立方程计量经济学模型的识别***
- § 6.4 联立方程计量经济学模型的估计**
- § 6.5 联立方程计量经济学模型的讨论

§ 6.1 问题的提出

- 一、经济研究中的联立方程计量经济学问题
- 二、计量经济学方法中的联立方程问题

一、经济研究中的 联立方程计量经济学问题

1. 研究对象

- 经济系统，而不是单个经济活动

“系统”的相对性

- 相互依存、互为因果，而不是单向因果关系
- 必须用一组方程才能描述清楚

2. 一个简单的宏观经济系统

- 由国内生产总值Y、居民消费总额C、投资总额I和政府消费额G等变量构成简单的宏观经济系统。
- 将政府消费额G由系统外部给定，其他内生。

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

- 在消费方程、投资方程中，国内生产总值 Y 决定居民消费总额 C 和投资总额 I ；
- 在国内生产总值方程中， Y 又由居民消费总额 C 和投资总额 I 所决定。

二、计量经济学方法中的联立方程问题

1.随机解释变量问题

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

- 解释变量中出现随机变量，而且与误差项相关。
- 为什么_(y与c的随机性)？

2.损失变量信息问题

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

- 如果用单方程模型的方法估计某个方程，将损失变量信息。
- 为什么？

3.损失方程之间的相关性信息问题

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

- 联立方程模型系统中每个随机方程之间往往存在某种相关性。
- 表现于“不同方程随机误差项之间”。
- 如果用单方程模型方法估计某一个方程，将损失不同方程之间相关性信息。

4.结论

- 如果采用单方程OLS估计联立方程计量经济学模型，会产生**联立性偏误** (simultaneity bias)。
- 必须发展新的估计方法估计联立方程计量经济学模型，以尽可能避免出现这些问题。
- 这就从计量经济学理论方法上提出了联立方程问题。

§ 6. 2联立方程计量经济学模型的 若干基本概念*

一、变量**

二、结构式模型*

三、简化式模型

四、参数关系体系

一、变量

1.内生变量（Endogenous Variables）

- 对联立方程模型系统而言，已经不能用被解释变量与解释变量来划分变量，而将变量分为内生变量和外生变量两大类。
- 内生变量是具有“某种概率分布”的随机变量，它的参数是联立方程系统估计的元素。
- 内生变量是由模型系统决定的，同时也对模型系统产生影响。
- 内生变量一般都是经济变量。

- 一般情况下，内生变量与随机项相关，即

$$\begin{aligned}Cov(Y_i, \mu_i) &= E((Y_i - E(Y_i))(\mu_i - E(\mu_i))) \\&= E((Y_i - E(Y_i))\mu_i) \\&= E(Y_i\mu_i) - E(Y_i)E(\mu_i) \\&= E(Y_i\mu_i) \\&\neq 0\end{aligned}$$

- 在联立方程模型中，内生变量既作为被解释变量，又可在不同方程中作为解释变量。

2.外生变量 (Exogenous Variables)

- 外生变量一般是确定性变量，或是具有“临界/已知概率分布”的随机变量，其参数“不是模型系统研究的元素”。
- 外生变量影响系统，但本身不受系统的影响。
- 外生变量一般是经济变量、条件变量、政策变量、虚变量。
- 一般情况下，外生变量与随机项不相关。

3. 先决变量 (Predetermined Variables)

- 外生变量与滞后内生变量(Lagged Endogenous Variables), 统称为先决变量。
- 滞后内生变量 y_{t-s} 是联立方程计量经济学模型中重要的不可缺少的一部分变量, 用以反映经济系统的动态性与连续性。
- 先决变量只能作为解释变量。

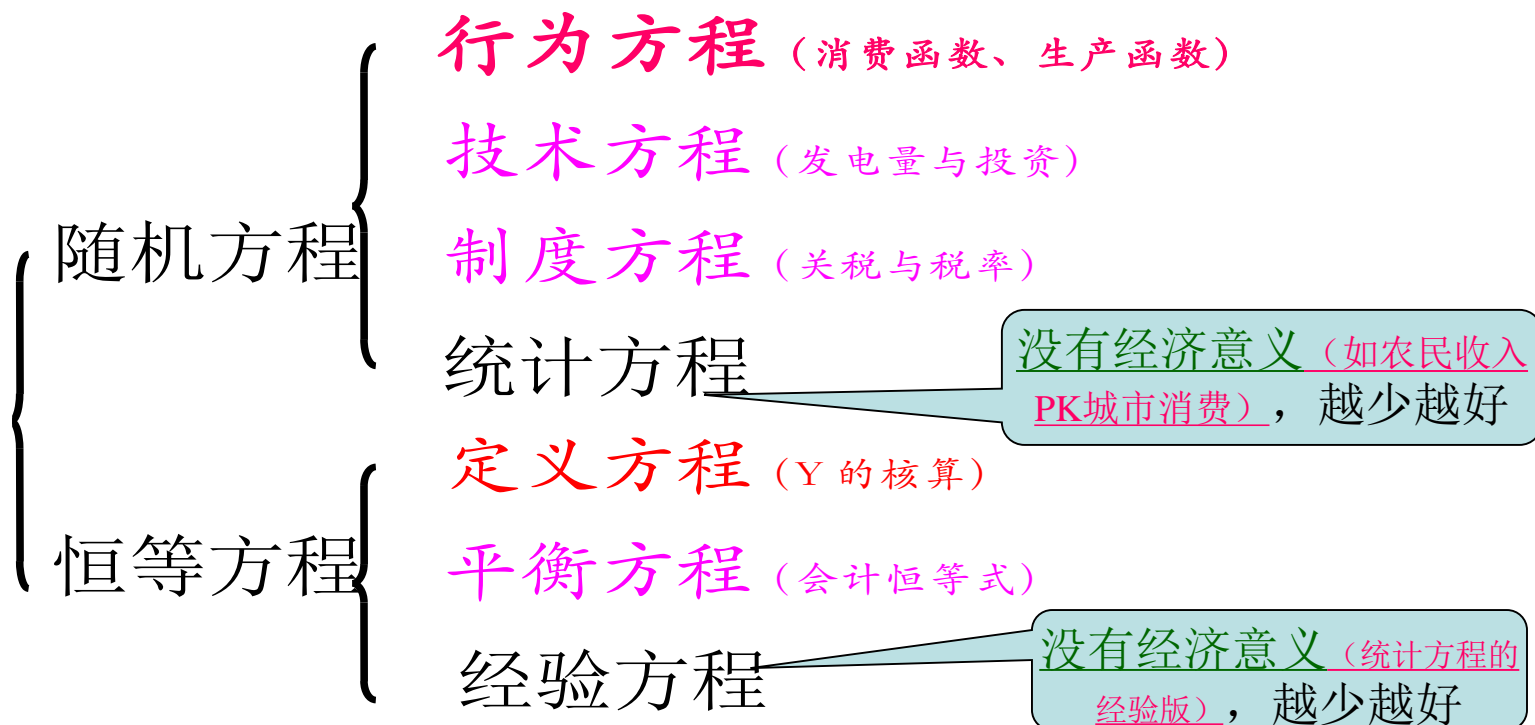
二、结构式模型

Structural Model

1.定义

- 根据经济理论和行为规律、建立“描述经济变量之间直接结构关系”的计量经济学方程系统，称为结构式模型。
- 结构式模型中的每一个方程都是结构方程（ **Structural Equations** ）。
- 各个结构方程的参数被称为结构参数（ **Structural Parameters or Coefficients** ）。
- 将一个内生变量表示为其它内生变量、先决变量和随机误差项的函数形式，称为结构方程正规形式。

2.结构方程的方程类型



3.完备的结构式模型

- 具有 g 个内生变量、 k 个先决变量、 g 个结构方程的模型，被称为完备结构式模型。
- 完备结构式模型中，独立结构方程数目、等于内生变量数目；
每个内生变量，都分别由一个方程来描述。

4.完备结构式模型的矩阵表示

- 习惯上用 Y 表示内生变量， X 表示先决变量， μ 表示随机项； β 表示内生变量的结构参数， γ 表示先决变量的结构参数；
- 如果模型中有常数项，可看成为一个外生的虚变量，它的观测值始终取1。

$$BY + \Gamma X = N$$

$$(B\Gamma) \begin{pmatrix} Y \\ X \end{pmatrix} = N$$

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{pmatrix} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ y_{21} & y_{22} & \cdots & y_{2n} \\ \vdots & & & \\ y_{g1} & y_{g2} & \cdots & y_{gn} \end{bmatrix}_{g \times n}$$

$$\mathbf{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_k \end{pmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & & & \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix}_{k \times n}$$

$$\mathbf{N} = \begin{pmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ \vdots \\ \mathbf{N}_g \end{pmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ \vdots & & & \\ \mu_{g1} & \mu_{g2} & \cdots & \mu_{gn} \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1g} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2g} \\ \vdots & & & \\ \beta_{g1} & \beta_{g2} & \cdots & \beta_{gg} \end{bmatrix} \quad \Gamma = \begin{bmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1k} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2k} \\ \vdots & & & \\ \gamma_{g1} & \gamma_{g2} & \cdots & \gamma_{gk} \end{bmatrix}_{g \times k}$$

5.简单宏观经济模型的矩阵表示

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t} \\ Y_t = C_t + I_t + G_t \end{cases}$$

$$\mathbf{Y} = \begin{bmatrix} C_t \\ I_t \\ Y_t \end{bmatrix} = \begin{bmatrix} C_1 & C_2 & \cdots & C_n \\ I_1 & I_2 & \cdots & I_n \\ Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} 1 \\ Y_{t-1} \\ G_t \end{bmatrix} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ Y_0 & Y_1 & \cdots & Y_{n-1} \\ G_1 & G_2 & \cdots & G_n \end{bmatrix}$$

$$\mathbf{N} = \begin{bmatrix} \mathbf{N}_1 \\ \mathbf{N}_2 \\ 0 \end{bmatrix} = \begin{bmatrix} \mu_{11} & \mu_{12} & \cdots & \mu_{1n} \\ \mu_{21} & \mu_{22} & \cdots & \mu_{2n} \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

$$(\mathbf{B}\Gamma) = \begin{bmatrix} \overset{C_t}{1} & \overset{I_t}{0} & \overset{Y_t}{-\alpha_1} & \overset{1}{-\alpha_0} & \overset{Y_{t-1}}{0} & \overset{G_t}{0} \\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2 & 0 \\ -1 & -1 & 1 & 0 & 0 & -1 \end{bmatrix}$$

三、简化式模型

Reduced-Form Model

1.定义

- 用所有先决变量作为每个内生变量的解释变量，所形成的模型称为简化式模型。
- 简化式模型并不反映经济系统中变量之间的直接关系，并不是经济系统的客观描述。
- 由于简化式模型中作为解释变量的变量中没有内生变量，可以采用普通最小二乘法估计每个方程的参数，所以它在联立方程模型研究中具有重要的作用。
- 简化式模型中每个方程称为简化式方程 (Reduced-Form Equations)，方程的参数称为简化式参数 (Reduced-Form Coefficients)。

2.简化式模型的矩阵形式

$$\mathbf{Y} = \mathbf{\Pi X} + \mathbf{E}$$

$$\mathbf{\Pi} = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1k} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2k} \\ \vdots & & & \\ \pi_{g1} & \pi_{g2} & \cdots & \pi_{gk} \end{bmatrix} \quad \mathbf{E} = \begin{bmatrix} \mathbf{E}_1 \\ \mathbf{E}_2 \\ \vdots \\ \mathbf{E}_g \end{bmatrix} = \begin{bmatrix} \varepsilon_{11} & \varepsilon_{12} & \cdots & \varepsilon_{1n} \\ \varepsilon_{21} & \varepsilon_{22} & \cdots & \varepsilon_{2n} \\ \vdots & & & \\ \varepsilon_{g1} & \varepsilon_{g2} & \cdots & \varepsilon_{gn} \end{bmatrix}$$

3.简单宏观经济模型的简化式模型

$$\begin{cases} C_t = \pi_{10} + \pi_{11}Y_{t-1} + \pi_{12}G_t + \varepsilon_t \\ I_t = \pi_{20} + \pi_{21}Y_{t-1} + \pi_{22}G_t + \varepsilon_t \\ Y_t = \pi_{30} + \pi_{31}Y_{t-1} + \pi_{32}G_t + \varepsilon_t \end{cases}$$

四、参数关系体系

1.定义

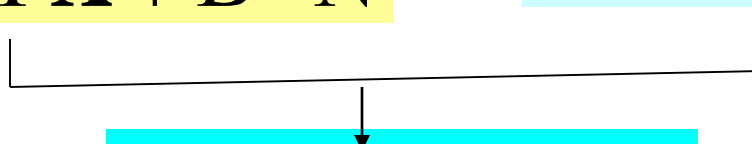
$$\mathbf{B}\mathbf{Y} + \mathbf{\Gamma}\mathbf{X} = \mathbf{N}$$



$$\mathbf{B}\mathbf{Y} = -\mathbf{\Gamma}\mathbf{X} + \mathbf{N}$$

$$\mathbf{Y} = -\mathbf{B}^{-1}\mathbf{\Gamma}\mathbf{X} + \mathbf{B}^{-1}\mathbf{N}$$

$$\mathbf{Y} = \mathbf{\Pi}\mathbf{X} + \mathbf{E}$$



$$\mathbf{\Pi}_{g \times k} = -\mathbf{B}^{-1}\mathbf{\Gamma}$$

- 该式描述了简化式参数与结构式参数之间的关系，称为参数关系体系。

2.作用

- 利用参数关系体系：首先估计简化式参数，然后可计算得到结构式参数。
- 从参数关系体系还可以看出，简化式参数反映了先决变量对内生变量的直接与间接影响之和，这是简化式模型的另一个重要作用。

例如，在上述模型中存在如下关系：

$$\pi_{21} = \frac{\beta_2 - \alpha_1 \beta_2}{1 - \alpha_1 - \beta_1} = \beta_2 + \frac{\beta_1 \beta_2}{1 - \alpha_1 - \beta_1}$$

Π_{21} 反映 Y_{t-1} 对 I_t 的**直接与间接影响之和**；而其中的 β_2 正是结构方程中 Y_{t-1} 对 I_t 的结构参数，显然，它只反映 Y_{t-1} 对 I_t 的**直接影响**。

- 在这里， β_2 是 Y_{t-1} 对 I_t 的部分乘数， Π_{21} 反映 Y_{t-1} 对 I_t 的完全乘数。

注意：简化式参数与结构式参数之间的区别与联系。

§ 6. 3联立方程计量经济学模型的识别***

The Identification of SEMs

- 一、识别的概念*
- 二、从定义出发识别模型
- 三、结构式识别条件***
- 四、简化式识别条件
- 五、实际应用中的经验方法

一、识别的概念

1.为什么要对模型进行识别？

- 从一个例子看：

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \mu_{2t} \\ Y_t = C_t + I_t \end{cases}$$

- 消费方程是包含C、Y和常数项的直接线性方程。
- 投资方程和国内生产总值方程的某种线性组合
(消去I) 所构成的新方程，也是包含C、Y和常数项的直接线性方程。

- 如果利用C、Y的样本观测值并进行参数估计后，很难判断得到的是“消费方程的参数估计量”还是“新组合方程的参数估计量”。

这二个方程被认为是“观测上无区别”
(**observationally indistinguishable**)。

- 只能认为原模型中的消费方程，是不可估计的。
- 这种情况被称为**不可识别(unidentified)**。
- 只有可以识别的方程，才是可以估计的。

2.识别的定义

- 3种定义:

“如果联立方程模型中某个结构方程不具有确定的统计形式，则称该方程为不可识别。”

“如果联立方程模型中某些方程的线性组合可构成与某一个方程相同的统计形式，则称该方程为不可识别。”

“根据参数关系体系，在已知简化式参数估计值时，如果不能得到联立方程模型中某个结构方程的确定的结构参数估计值，则称该方程为不可识别。”

- 以是否具有确定统计形式，作为识别的基本定义。
- 什么是“统计形式”？
- 什么是“具有确定的统计形式”（一组变量，只对应一组确定的待估参数）？

3.模型的识别

- 上述识别的定义，是针对结构方程而言的。
- 模型中每个需要估计其参数的随机方程都存在识别问题。
- 如果一个模型中的所有随机方程都是可以识别的，则认为该联立方程模型系统是可以识别的；
反过来，如果一个模型系统中存在一个不可识别的随机方程，则认为该联立方程模型系统是不可以识别的。
- 恒等方程由于不存在参数估计问题，故也不存在识别问题。但是，判断随机方程的识别性问题时，应将恒等方程考虑在内。

4.恰好识别与过度识别

- 如果某一个随机方程具有一组参数估计量，称其为恰好识别(Just-Identified) ；
- 如果某一个随机方程具有多组参数估计量，称其为过度识别(Over-identified) 。

二、从定义出发识别模型

1.例题1

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \mu_{2t} \\ Y_t = C_t + I_t \end{cases}$$

- 第2与第3个方程的线性组合得到的新方程具有与消费方程相同的统计形式，所以消费方程也是不可识别的。

- 第1与第3个方程的线性组合得到的新方程、具有与投资方程相同的统计形式，所以投资方程也是不可识别的。
- 于是，该模型系统不可识别。
- 参数关系体系由3个方程组成，剔除一个矛盾方程
(笨方法)：2个方程，不能求得4个结构参数确定值。
也证明消费方程与投资方程都是不可识别的。

2.例题2

$$C_t = \alpha_0 + \alpha_1 Y_t + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

- 消费方程是可以识别的，因为任何方程的线性组合都不能构成与它相同的统计形式。
- 投资方程仍然是不可识别的，因为第1、第2与第3个方程的线性组合（消去C，但保留Y与I）构成与它相同的统计形式。
- 于是，该模型系统仍然不可识别。

- 参数关系体系由6个方程($g \times k$)组成，剔除2个矛盾方程：由4个方程，是不能求得所有5个结构参数的确定估计值。
- 可得到消费方程参数确定值，证明消费方程可识别；因为只能得到它的一组确定值，所以消费方程是恰好识别的方程。
- 投资方程，都是不可识别的。
- 注意：与例题1相比，在投资方程中增加了1个变量，消费方程变成可以识别。

3.例题3

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

- 消费方程仍是可识别的，因为任何方程的线性组合、都不能构成与它相同的统计形式。
- 投资方程也是可识别的，因为任何方程的线性组合、都不能构成与它相同的统计形式。
- 于是，该模型系统是可以识别的。

- 参数关系体系由9 ($g \times k$) 个方程组成，剔除3个矛盾方程，在已知简化式参数估计值时：由6个方程，可求得所有6个结构参数的确定估计值。
- 所以也证明消费方程和投资方程，都是可识别的。
- 而且，只能得到所有6个结构参数的一组确定值，所以消费方程和投资方程都是恰好识别的方程。
- 注意：与例题2相比，在消费方程中增加了1个变量，投资方程变成可以识别。

4.例题4

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \alpha_3 P_{t-1} + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

- 消费方程和投资方程仍然是可识别的，因为任何方程的线性组合、都不能构成与它们相同的统计形式。
- 于是，该模型系统是可以识别的。

- 参数关系体系由12个方程组成，剔除4个矛盾方程，在已知简化式参数估计值时：8个方程，能够求得所有7个结构参数的确定估计值。
- 所以也证明消费方程和投资方程都是可以识别的。
- 但是，求解结果表明：对消费方程参数，只能得到一组确定值，所以消费方程是恰好识别的方程；
而对投资方程的参数，能够得到多组确定值，所以投资方程是过度识别的方程。

- **注意：**

- 在求解线性代数方程组时，如果方程数目大于未知数数目，被认为无解；如果方程数目小于未知数数目，被认为有无穷多解。
- 但在这里，无穷多解、意味着没有确定值——所以，如果参数关系体系中有效方程数目、小于未知结构参数估计量数目，被认为不可识别。
- 如果参数关系体系中有效方程数目、大于未知结构参数估计量数目：那么每次从中选择与未知结构参数估计量数目相等的方程数，可以解得一组结构参数估计值；换一组方程，又可以解得一组结构参数估计值。

这样就可得到多组结构参数估计值，被认为可识别，但不是恰好识别、而是过度识别。

5.如何修改模型：使不可识别的方程、变成可识别

- 在其它方程中、增加变量；
- 在该不可识别方程中、减少变量。
- 必须保持经济意义的合理性。

三、结构式识别条件 (Structural Condition for Identification)

1.结构式识别条件

- 直接从结构模型出发
- 一种规范的判断方法
- 每次、用于1个随机方程
- 具体描述为：

联立方程计量经济学模型的结构式

$$\mathbf{B}\mathbf{Y} + \mathbf{\Gamma}\mathbf{X} = \mathbf{N}$$

中的第 i 个方程中包含 g_i 个内生变量（含被解释变量）和 k_i 个先决变量（含常数项），模型系统中内生变量和先决变量的数目仍用 g 和 k 表示，矩阵 $(\mathbf{B}_0\mathbf{\Gamma}_0)$ 表示第 i 个方程中未包含的变量（包括内生变量和先决变量）在其它 $g-1$ 个方程中对应系数所组成的矩阵。于是，判断第 i 个结构方程识别状态的结构式条件为：

如果 $R(\mathbf{B}_0\mathbf{\Gamma}_0) < g-1$ ，则第 i 个结构方程不可识别；

如果 $R(\mathbf{B}_0\mathbf{\Gamma}_0) = g-1$ ，则第 i 个结构方程可以识别，并且

如果 $k - k_i = g_i - 1$ ，则第 i 个结构方程恰好识别，

如果 $k - k_i > g_i - 1$ ，则第 i 个结构方程过度识别。

$$(k - k_i) + \bar{g}_i < (g_i + \bar{g}_i) - 1$$

- 一般将该条件的前一部分称为秩条件（**Rank Condition**），用以判断结构方程是否识别；
- 将后一部分称为阶条件（**Order Condition**），用以判断结构方程恰好识别或者过度识别。

为什么在秩条件中没有讨论 $R(B_0\Gamma_0) > g - 1$ 的情况？

为什么在阶条中没有讨论 $k - k_i < g_i - 1$ 的情况？

2.例题

$$C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \alpha_3 P_{t-1} + \mu_{1t}$$

$$I_t = \beta_0 + \beta_1 Y_t + \beta_2 Y_{t-1} + \mu_{2t}$$

$$Y_t = C_t + I_t$$

$$[B\Gamma] = \begin{bmatrix} \overset{C_t}{1} & \overset{I_t}{0} & \overset{Y_t}{-\alpha_1} & \overset{1}{-\alpha_0} & \overset{Y_{t-1}}{0} & \overset{C_{t-1}}{-\alpha_2} & \overset{P_{t-1}}{-\alpha_3} \\ 0 & 1 & -\beta_1 & -\beta_0 & -\beta_2 & 0 & 0 \\ -1 & -1 & 1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

- 判断第1个结构方程的识别状态

$$[B_0\Gamma_0] = \begin{bmatrix} 1 & -\beta_2 \\ -1 & 0 \end{bmatrix} \longrightarrow R(B_0\Gamma_0) = 2 = g - 1$$

所以，该方程可以识别。

因为

$$k - k_1 = 1 = g_1 - 1$$

所以，第1个结构方程为恰好识别的结构方程。

- 判断第2个结构方程的识别状态

$$[B_0\Gamma_0] = \begin{bmatrix} 1 & -\alpha_2 & -\alpha_3 \\ -1 & 0 & 0 \end{bmatrix} \longrightarrow R(B_0\Gamma_0) = 2 = g - 1$$

所以，该方程可以识别。

因为

$$k - k_2 = 2 > g_2 - 1$$

所以，第2个结构方程为过度识别的结构方程。

- 第3个方程是平衡方程，不存在识别问题。
- 综合以上结果，该联立方程模型是可以识别的。
- 与从定义出发识别的结论一致。

四、简化式识别条件☆ (略)
(Reduced Form Condition for
Identification)

1.简化式识别条件

- 如果已经知道联立方程模型的简化式模型参数，那么可以通过对简化式模型的研究达到判断结构式模型是否识别的目的。
- 由于需要首先估计简化式模型参数，所以很少实际应用。

2.例题

- 需要识别的结构式模型

$$\begin{cases} y_{1i} = \alpha_1 y_{2i} + \alpha_2 x_{1i} + \alpha_3 x_{2i} + \mu_{1i} \\ y_{2i} = \beta_1 y_{3i} + \beta_2 x_{3i} + \mu_{2i} \\ y_{3i} = \gamma_1 y_{1i} + \gamma_2 y_{2i} + \gamma_3 x_{3i} + \mu_{3i} \end{cases}$$

- 已知其简化式模型参数矩阵为

$$\Pi = \begin{bmatrix} 4 & -2 & 3 \\ 2 & -1 & 1 \\ 2 & -1 & 0 \end{bmatrix}$$

- 判断第1个结构方程的识别状态

$$\Pi_2 = \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

$$R(\Pi_2) = 1 = g_1 - 1$$

所以该方程是可以识别的。又因为

$$k - k_1 = 1 = g_1 - 1$$

所以该方程是恰好识别的。

- 判断第2个结构方程的识别状态

$$\Pi_2 = \begin{bmatrix} 2 & -1 \\ 2 & -1 \end{bmatrix}$$

$$R(\Pi_2) = 1 = g_2 - 1$$

所以该方程是可以识别的。又因为

$$k - k_2 = 2 > g_2 - 1$$

所以该方程是过度识别的。

- 判断第3个结构方程的识别状态

$$\Pi_2 = \begin{bmatrix} 4 & -2 \\ 2 & -1 \\ 2 & -1 \end{bmatrix} \quad R(\Pi_2) = 1 < g_3 - 1$$

所以该方程是不可识别的。

- 所以该模型是不可识别的。

- 可从数学上严格证明，简化式识别条件、和结构式识别条件是等价的。

《计量经济学学习指南与练习》（潘文卿、李子奈编著，高等教育出版社，2010年6月）第127—129页。

五、实际应用中的经验方法

(要求太高，不太实用：略)

- 当一个联立方程计量经济学模型系统中的方程数目比较多时，无论是从识别的概念出发，还是利用规范的结构式或简化式识别条件——对模型进行识别，困难都是很大的，或者说不可能的。
- 理论上很严格的方法，实际中可能是无法应用的；实际中应用的，有时是一些经验方法。
- 关于联立方程计量经济学模型的识别问题，实际上不是等到理论模型已经建立了之后再进行识别，而是在建立模型的过程中设法保证模型的可识别性。

- “在建立某个结构方程时，要使该方程包含前面每一个方程中都不包含的至少1个变量（内生或先决变量）；同时使前面每一个方程中都包含至少1个该方程所未包含的变量，并且互不相同（要求太高，不太实用）。”
- 该原则的前一句话是保证该方程的引入不破坏前面已有方程的可识别性。只要新引入方程包含前面每一个方程中都不包含的至少1个变量，那么它与前面方程的任意线性组合都不能构成与前面方程相同的统计形式，原来可以识别的方程仍然是可以识别的。
- 该原则的后一句话是保证该新引入方程本身是可以识别的。只要前面每个方程都包含至少1个该方程所未包含的变量，并且互不相同。那么所有方程的任意线性组合都不能构成与该方程相同的统计形式。

- 在实际建模时，将每个方程所包含的变量记录在如下表所示的表式中，将是有帮助的。

	变量 1	变量 2	变量 3	变量 4	变量 5	变量 6	...
方程 1	×	×		×			
方程 2		×	×	×	×		
方程 3	×		×	×		×	
方程 4		×	×				×
...							

§ 6.4 联立方程模型的估计**

The Estimation of the Simultaneous-Equations Econometric Model

本节内容

- 一、概述
- 二、狭义的工具变量法（IV）*
- 三、间接最小二乘法（ILS）
- 四、二阶段最小二乘法（2SLS）**
- 五、三种方法的等价性证明
- 六、简单宏观经济模型实例演示
- *七、主分量法的应用
- *八、 k 级估计式

一、概述

1、单方程估计方法与系统估计方法

- 联立方程计量经济学模型的估计方法分为两大类：单方程估计方法与系统估计方法。

所谓单方程估计方法，指每次只估计模型系统中的一个方程，依次逐个估计。也将单方程估计方法称为有限信息估计方法。

所谓系统估计方法，指同时对全部方程进行估计，同时得到所有方程的参数估计量。也将系统估计方法称为完全信息估计方法。

- 注意：联立方程模型的单方程估计方法，不同于单方程模型的估计方法。

2、单方程估计方法按其方法原理分类

- 一类仍以最小二乘为原理，称其为经典方法。例如：
 - 间接最小二乘法（ILS, Indirect Least Square）
 - 两阶段最小二乘法(2SLS, Two Stage Least Squares)
 - 工具变量法（IV, Instrumental Variables: M估计）
- 一类不以最小二乘为原理，或不直接从最小二乘原理出发。例如：
 - 以最大或然为原理的有限信息最大或然法(LIML, Limited Information Maximum Likelihood)
 - 仍然应用最小二乘原理、但并不以残差平方和最小为判断标准的最小方差比方法(LVR, Least Variance Ratio)

3、系统估计方法

- 系统估计方法主要包括：
 - 三阶段最小二乘法(3SLS, Three Stage Least Squares)
 - 完全信息最大或然法(FIML, Full Information Maximum Likelihood)
- 本节只介绍几种简单的、常用的单方程估计方法。
- 在大量的联立方程模型的应用研究中，仍然广泛应用普通最小二乘OLS法进行模型的估计。

二、狭义的工具变量法

(IV, Instrumental Variables)

1.方法思路

- “狭义的工具变量法” 与 “广义的工具变量法”
- 解决结构方程中与随机误差项相关的内生解释变量问题。
- 方法原理与单方程模型的IV方法相同。
- “模型系统中”、提供了可供选择的工具变量，使得IV方法的应用成为可能。

2.工具变量的选取

- 对联立方程模型的每个结构方程，例如第1个方程，可写成如下形式：

$$Y_1 = \beta_{12}Y_2 + \beta_{13}Y_3 + \cdots + \beta_{1g_1}Y_{g_1} + \gamma_{11}X_1 + \gamma_{12}X_2 + \cdots + \gamma_{1k_1}X_{k_1} + N_1$$

- 内生解释变量 (g_1-1) 个，先决解释变量 k_1 个。
- 如果方程是恰好识别的，有 $(g_1-1) = (k - k_1)$ 。
- 可选择 $(k - k_1)$ 个 方程 i 中没有包含的先决变量，作为 (g_1-1) 个内生解释变量的工具变量。

3. IV参数估计量

- 方程的矩阵表示为:

$$Y_1 = (Y_0, X_0) \begin{pmatrix} B_0 \\ \Gamma_0 \end{pmatrix} + N_1$$

- 选择方程中“**没有包含的先决变量 X_0^*** ”、作为“**包含的内生解释变量 Y_0** ”的工具变量，得到参数估计量为:

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{IV} = \left(\begin{pmatrix} X_0^* & X_0 \end{pmatrix}' \begin{pmatrix} Y_0 & X_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} X_0^* & X_0 \end{pmatrix}' Y_1$$

4.讨论

- 该估计量与OLS估计量的区别是什么？
- 该估计量具有什么统计特性？
- $(k - k_1)$ 工具变量与 $(g_1 - 1)$ 个内生解释变量的对应关系，是否影响参数估计结果？为什么（同一组正规方程组）？
- IV是否利用了模型系统中方程之间相关性信息？
- 对于过度识别的方程，可否应用IV？为什么？
- 对于过度识别的方程，可否应用GMM？为什么？

三、间接最小二乘法

(ILS, Indirect Least Squares)

1.方法思路

- 联立方程模型的结构方程中包含有内生解释变量，不能直接采用OLS估计其参数；但是对于简化式方程，可采用OLS直接估计其参数。
- 间接最小二乘法：先对关于内生解释变量的简化式方程、采用OLS估计简化式参数，得到简化式参数估计量；**然后通过参数关系体系**，计算得到结构式参数的估计量。
- **间接最小二乘法、只适用于恰好识别的结构方程的参数估计**，因为只有恰好识别的结构方程，才能从参数关系体系中得到**唯一一组结构参数的估计量**。

2.一般间接最小二乘法的估计过程

$$Y_1 = (\mathbf{Y}_0, \mathbf{X}_0) \begin{pmatrix} \mathbf{B}_0 \\ \Gamma_0 \end{pmatrix} + \mathbf{N}_1$$



$$Y_1 - \mathbf{B}_0 \mathbf{Y}_0 - \Gamma_0 \mathbf{X}_0 = \mathbf{N}_1$$



$$\begin{pmatrix} 1 & -\mathbf{B}_0 & -\Gamma_0 \end{pmatrix} \begin{pmatrix} Y_1 \\ \mathbf{Y}_0 \\ \mathbf{X}_0 \end{pmatrix} = \mathbf{N}_1$$



$$\begin{pmatrix} \mathbf{B}_{00} & \Gamma_{00} \end{pmatrix} \begin{pmatrix} \mathbf{Y}_{00} \\ \mathbf{X}_0 \end{pmatrix} = \mathbf{N}_1$$

$$\mathbf{Y}_{00} = \Pi_{00} \mathbf{X} + \mathbf{E}$$

$$\mathbf{B}_{00} \Pi_{00} \mathbf{X} + \Gamma_{00} \mathbf{X}_0 = 0$$

补充: $\mathbf{B}_{00} \mathbf{E} = \mathbf{N}_1$

$$\mathbf{B}_{00} \Pi_{00} \begin{pmatrix} \mathbf{X}_0 \\ \mathbf{X}_0^* \end{pmatrix} + \Gamma_{00} \mathbf{X}_0 = 0$$

$$\Pi_{00} = \begin{pmatrix} \Pi_{00}^1 & \Pi_{00}^2 \end{pmatrix}$$

$$\begin{cases} B_{00}\Pi_{00}^1 = -\Gamma_{00} = \Gamma_0 \\ B_{00}\Pi_{00}^2 = 0 \end{cases}$$

- 先用OLS估计简化式模型，得到简化式参数估计量 Π_{00} ；
再由上述该参数关系体系：1) 先由第2组方程，计算得到内生解释变量的参数 B_{00} ；
2) 再代入第1组方程，计算得到先决解释变量的参数 Γ_0 。
于是，得到了结构方程的所有结构参数估计量。

3.间接最小二乘法也是一种工具变量方法

- **ILS等价于一种工具变量方法：依次选择 \mathbf{X} 作为 $(\mathbf{Y}_0, \mathbf{X}_0)$ 的工具变量。**
- 数学证明见《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年3月）第126—128页。
- 估计结果为：

$$\begin{bmatrix} \hat{\mathbf{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{ILS} = \left(\mathbf{X}' \begin{pmatrix} \mathbf{Y}_0 & \mathbf{X}_0 \end{pmatrix} \right)^{-1} \mathbf{X}' \mathbf{Y}_1$$

四、二阶段最小二乘法**

(2SLS, Two Stage Least Squares)

1. 2SLS是应用最多的单方程估计方法

- IV和ILS一般只适用于联立方程模型中恰好识别的结构方程的估计。
- 实际的联立方程模型中，恰好识别的结构方程很少出现；一般情况下，结构方程都是过度识别的。为什么？
- 2SLS是一种既适用于恰好识别的结构方程，又适用于过度识别的结构方程的单方程估计方法。

2.2SLS的方法步骤

- 第一阶段：对内生解释变量向量的简化式方程使用OLS（每个分量方程，单独使用OLS）。得到：

$$\hat{\mathbf{Y}}_0 = \mathbf{X}\hat{\boldsymbol{\Pi}}_0 = \mathbf{X}((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_0)$$

- 用估计量代替结构方程中的内生解释变量，得到新模型：

$$Y_1 = (\hat{\mathbf{Y}}_0, \mathbf{X}_0) \begin{pmatrix} \mathbf{B}_0 \\ \boldsymbol{\Gamma}_0 \end{pmatrix} + \mathbf{N}_1$$

- 第二阶段：对该模型应用OLS估计，得到参数估计量、即为原结构方程参数的TOLS估计量。

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{2SLS} = \left(\begin{pmatrix} \hat{Y}_0 & X_0 \end{pmatrix}' \begin{pmatrix} \hat{Y}_0 & X_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \hat{Y}_0 & X_0 \end{pmatrix}' Y_1$$

注意：模型中的内生解释变量，
已被“它们的估计量”替换。

3.二阶段最小二乘法，也是一种工具变量方法

- 如果用“ Y_0 的估计量 \hat{Y}_0 ”、作为工具变量，按照工具变量方法估计过程，应得到如下结构参数估计量：

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix} = \left(\begin{pmatrix} \hat{Y}_0 & X_0 \end{pmatrix}' \begin{pmatrix} Y_0 & X_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \hat{Y}_0 & X_0 \end{pmatrix}' Y_1$$

- 可严格证明，两组参数估计量是完全等价的；所以，可把2SLS也看成为一种工具变量方法。

证明过程见《计量经济学—方法与应用》（李子奈编著，清华大学出版社，1992年3月）第130—131页。

五、三种方法的等价性证明

1.三种单方程估计方法得到的参数估计量

$$\begin{bmatrix} \hat{\mathbf{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{IV} = \left(\begin{pmatrix} \mathbf{X}_0^* & \mathbf{X}_0 \end{pmatrix}' \begin{pmatrix} \mathbf{Y}_0 & \mathbf{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \mathbf{X}_0^* & \mathbf{X}_0 \end{pmatrix}' \mathbf{Y}_1$$

$$\begin{bmatrix} \hat{\mathbf{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{ILS} = \left(\mathbf{X}' \begin{pmatrix} \mathbf{Y}_0 & \mathbf{X}_0 \end{pmatrix} \right)^{-1} \mathbf{X}' \mathbf{Y}_1$$

$$\begin{bmatrix} \hat{\mathbf{B}}_0 \\ \hat{\Gamma}_0 \end{bmatrix}_{2SLS} = \left(\begin{pmatrix} \hat{\mathbf{Y}}_0 & \mathbf{X}_0 \end{pmatrix}' \begin{pmatrix} \mathbf{Y}_0 & \mathbf{X}_0 \end{pmatrix} \right)^{-1} \begin{pmatrix} \hat{\mathbf{Y}}_0 & \mathbf{X}_0 \end{pmatrix}' \mathbf{Y}_1$$

2.IV与ILS估计量的等价性

- 在恰好识别情况下：

工具变量集合相同，只是次序不同；

工具变量次序不同、不影响正规方程组的解。

2.2SLS与ILS估计量的等价性

- 恰好识别情况下：
 - ILS的工具变量是全体先决变量；
 - 2SLS每个工具变量 \hat{Y}_0 ，都是全体先决变量的线性组合。
 - 2SLS的正规方程组，相当于ILS的正规方程组经过一系列的初等变换的结果。
 - 线性代数方程组经过初等变换，不影响方程组的解。

六、简单宏观经济模型实例演示

1.模型

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \mu_{2t} \\ Y_t = I_t + C_t + G_t \end{cases}$$

消费方程是恰好识别的；
投资方程是过度识别的；
模型是可以识别的。

- 下列演示中采用了1978-2007年的数据。
- 作业：采用1978-2009年的数据，独立完成下列演示的全过程。

2.数据

obs	Y	I	X	G
1978	3605.600	1377.900	1759.100	468.6000
1979	4092.600	1478.900	2011.500	602.2000
1980	4592.900	1599.700	2331.200	662.0000
1981	5008.800	1630.200	2627.900	750.7000
1982	5590.000	1784.200	2902.900	902.9000
1983	6216.200	2039.000	3231.100	946.1000
1984	7362.700	2515.100	3742.000	1105.600
1985	9076.700	3457.500	4687.400	931.8000
1986	10508.50	3941.900	5302.100	1264.500
1987	12277.40	4462.000	6126.100	1689.300
1988	15388.60	5700.200	7868.100	1820.300
1989	17311.30	6332.700	8812.600	2166.000
1990	19347.80	6747.000	9450.900	3149.900
1991	22577.40	7868.000	10730.60	3978.800
1992	27565.20	10086.30	13000.10	4478.800
1993	36938.10	15717.70	16412.10	4808.300
1994	50217.40	20341.10	21844.20	8032.100
1995	63216.90	25470.10	28369.70	9377.100
1996	74163.60	28784.90	33955.90	11422.80
1997	81658.50	29968.00	36921.50	14769.00
1998	86531.60	31314.20	39229.30	15988.10
1999	91125.00	32951.50	41920.40	16253.10
2000	98749.00	34842.80	45854.60	18051.60
2001	108972.4	39769.40	49213.20	19989.80
2002	120350.3	45565.00	52571.30	22214.00
2003	136398.8	55963.00	56834.40	23601.40
2004	160280.4	69168.40	63833.50	27278.50
2005	188692.1	80646.30	71217.50	36828.30
2006	221651.3	94402.00	80476.90	46772.40
2007	263242.5	111417.4	93317.20	58507.90

G的数据采用
(Y-I-C)计
算得到，不是
实际数据

3.用狭义的工具变量法估计消费方程

← 用 G_t 作为 Y_t 的工具变量

$$\hat{\alpha}_0 = 886.5725$$

$$\hat{\alpha}_1 = 0.085101$$

$$\hat{\alpha}_2 = 0.861701$$

- 估计结果显示 (借用2SLS, 完成IV估计)

Dependent Variable: X
 Method: Two-Stage Least Squares
 Date: 09/28/10 Time: 17:49
 Sample (adjusted): 1979 2007
 Included observations: 29 after adjustments
 Instrument list: C G X(-1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	886.5725	326.2414	2.717535	0.0115
Y	0.085101	0.020009	4.253073	0.0002
X(-1)	0.861701	0.059276	14.53708	0.0000
R-squared	0.998347	Mean dependent var	28096.42	
Adjusted R-squared	0.998219	S.D. dependent var	26706.49	
S.E. of regression	1126.948	Sum squared resid	33020308	
F-statistic	7842.749	Durbin-Watson stat	0.584922	
Prob(F-statistic)	0.000000			

4.用间接最小二乘法估计消费方程

$$\begin{cases} C_t = \pi_{10} + \pi_{11}C_{t-1} + \pi_{12}G_t + \varepsilon_{1t} \\ Y_t = \pi_{20} + \pi_{21}C_{t-1} + \pi_{22}G_t + \varepsilon_{2t} \end{cases}$$

$$\begin{aligned} \hat{\pi}_{10} &= 996.4507 \\ \hat{\pi}_{11} &= 0.968812 \\ \hat{\pi}_{12} &= 0.237809 \end{aligned}$$

$$\begin{aligned} \hat{\pi}_{20} &= 1291.152 \\ \hat{\pi}_{21} &= 1.258635 \\ \hat{\pi}_{22} &= 2.794436 \end{aligned}$$

$$\begin{aligned} \hat{\alpha}_1 &= \hat{\pi}_{12} / \hat{\pi}_{22} = 0.085101 \\ \hat{\alpha}_2 &= \hat{\pi}_{11} - \hat{\alpha}_1 \hat{\pi}_{21} = 0.861701 \\ \hat{\alpha}_0 &= \hat{\pi}_{10} - \hat{\alpha}_1 \hat{\pi}_{20} = 886.5724 \end{aligned}$$

- C简化式模型估计结果 (简单OLS)

Dependent Variable: X
Method: Least Squares
Date: 09/29/10 Time: 12:02
Sample (adjusted): 2 30
Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	996.4507	413.3076	2.410918	0.0233
X(-1)	0.968812	0.042493	22.79913	0.0000
G	0.237809	0.068707	3.461226	0.0019
R-squared	0.997503	Mean dependent var	28096.42	
Adjusted R-squared	0.997311	S.D. dependent var	26706.49	
S.E. of regression	1384.767	Akaike info criterion	17.40215	
Sum squared resid	49857072	Schwarz criterion	17.54359	
Log likelihood	-249.3312	F-statistic	5194.248	
Durbin-Watson stat	0.610763	Prob(F-statistic)	0.000000	

- Y简化式模型估计结果

Dependent Variable: Y
 Method: Least Squares
 Date: 09/29/10 Time: 12:04
 Sample (adjusted): 2 30
 Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	1291.152	1317.612	0.979919	0.3362
X(-1)	1.258635	0.135468	9.291045	0.0000
G	2.794436	0.219034	12.75798	0.0000
R-squared	0.996431	Mean dependent var	67210.48	
Adjusted R-squared	0.996157	S.D. dependent var	71211.15	
S.E. of regression	4414.594	Akaike info criterion	19.72092	
Sum squared resid	5.07E+08	Schwarz criterion	19.86236	
Log likelihood	-282.9533	F-statistic	3629.861	
Durbin-Watson stat	0.698679	Prob(F-statistic)	0.000000	

5.用两阶段最小二乘法估计消费方程（2S，手工完成）

$$\hat{Y}_t = 1291.152 + 1.258635C_{t-1} + 2.794436G_t$$



← 代替原消费方程中的 Y_t ，应用OLS估计

$$\hat{\alpha}_0 = 886.7525$$

$$\hat{\alpha}_1 = 0.085101$$

$$\hat{\alpha}_2 = 0.861701$$

- 比较上述消费方程的3种估计结果，证明这3种方法对于恰好识别的结构方程是等价的。
估计量的差别，只是很小的计算误差。

- 第2阶段估计结果

Dependent Variable: X
Method: Least Squares
Date: 09/29/10 Time: 12:07
Sample (adjusted): 2 30
Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	886.5725	400.8777	2.211579	0.0360
YF	0.085101	0.024587	3.461226	0.0019
X(-1)	0.861701	0.072837	11.83053	0.0000
R-squared	0.997503	Mean dependent var	28096.42	
Adjusted R-squared	0.997311	S.D. dependent var	26706.49	
S.E. of regression	1384.767	Akaike info criterion	17.40215	
Sum squared resid	49857072	Schwarz criterion	17.54359	
Log likelihood	-249.3312	F-statistic	5194.248	
Durbin-Watson stat	0.610763	Prob(F-statistic)	0.000000	

6.用两阶段最小二乘法估计投资方程

- 投资方程是过度识别结构方程，只能用2SLS估计。

估计过程，与上述2SLS估计消费方程的过程相同。
得到投资方程的参数估计量为：

$$\hat{\beta}_0 = -1455.814$$

$$\hat{\beta}_1 = 0.419774$$

- 至此，完成了该模型系统的估计。

- 2SLS第2阶段估计结果 (先得到YF; 亦可直接用2SLS内置程序)

Dependent Variable: I
Method: Least Squares
Date: 09/29/10 Time: 12:09
Sample (adjusted): 2 30
Included observations: 29 after adjustments

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1455.814	962.6275	-1.512334	0.1421
YF	0.419774	0.009931	42.26953	0.0000
R-squared	0.985113	Mean dependent var	26757.40	
Adjusted R-squared	0.984562	S.D. dependent var	30063.82	
S.E. of regression	3735.415	Akaike info criterion	19.35558	
Sum squared resid	3.77E+08	Schwarz criterion	19.44987	
Log likelihood	-278.6559	F-statistic	1786.713	
Durbin-Watson stat	0.476678	Prob(F-statistic)	0.000000	

*七、主分量法的应用（略）

1.方法的提出

- 主分量方法本身并不是联立方程模型的估计方法，而是配合其它方法，例如**2SLS**使用于模型的估计过程之中。
- 数学上的主分量方法早就成熟，**Kloek**和**Mennes**于**1960**年提出将它用于计量经济学模型的估计。
- **2SLS**是一种普遍适用的联立方程模型的单方程估计方法，但是当它在实际模型估计中被应用时，立刻就会遇到不可逾越的困难。其第一阶段—用**OLS**估计简化式方程，是难以实现的。
为什么？

2.方法的原理

- 所谓主分量方法，就是用较少数目的新变量重新表示原模型中较多数目的先决变量的方法。
- 例如，如果能够找到5个左右的新变量表示宏观经济模型中的30个先决变量，那么只需要15组以上的样本，就可以进行2SLS第一阶段的估计。
- 对充当主分量的变量是有严格要求：一是它必须是先决变量的线性组合，二是它们之间必须是正交的。前一条是保证主分量对先决变量的代表性；后一条是保证主分量之间不出现共线性。

3.主分量的选取

- 用两个主分量表示两个原变量

$$Z_1 = a_{11}X_1 + a_{12}X_2$$

$$Z_2 = a_{21}X_1 + a_{22}X_2$$

$$A = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

可以证明， \mathbf{a}_1 、 \mathbf{a}_2 分别是 $\mathbf{X}'\mathbf{X}$ 的2个特征值对应的特征向量。

- 用**k**个主分量表示**k**个原变量

$$\mathbf{Z} = \mathbf{XA}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_k \end{pmatrix}$$

同样可以证明， \mathbf{a}_1 、 \mathbf{a}_2 、...、 \mathbf{a}_k 分别是 $\mathbf{X}'\mathbf{X}$ 的**k**个特征值对应的特征向量。

- 用**f**个主分量表示**k**个原变量

$$\mathbf{Z} = \mathbf{XA}$$

$$\mathbf{A} = \begin{pmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_f \end{pmatrix}$$

选择 \mathbf{a}_1 、 \mathbf{a}_2 、...、 \mathbf{a}_f 分别是 $\mathbf{X}'\mathbf{X}$ 的**f**个最大特征值对应的特征向量。

- 在**2SLS**中主分量的选取

对于简化式方程

$$\mathbf{Y}_0 = \mathbf{X}\Pi_0 + \mathbf{E}_0 = \begin{pmatrix} \mathbf{X}_0 & \mathbf{X}_0^* \end{pmatrix} \Pi_0 + \mathbf{E}_0$$

一般情况下，结构方程包含的先决解释变量 \mathbf{x}_0 中变量的数目很有限，变量主要集中在结构方程未包含的先决变量 \mathbf{x}_0^* 中。所以只需要选择主分量重新表示 \mathbf{x}_0^* ，就可以有效地减少简化式方程中解释变量的数目，使得在有限样本的支持下模型得到估计。

4.主分量法在ILS中的应用

- 对于**2SLS**，直接利用主分量完成第一阶段的估计，得到内生解释变量的估计量。
- 对于**ILS**，必须求得到简化式参数，进而计算结构式参数。
- 首先估计 $Y=Z\Delta+E$ ，然后将 $Z=XA$ 代入，得到 $Y=X\Pi$ 中 Π 的估计量。

*八、k级估计式 (略)

1.k级估计式

- 本身不是一种估计方法，而是对上述几种方法得到的估计式的概括。
- 对于联立方程模型中的第1个结构方程：

$$Y_1 = (Y_0, X_0) \begin{pmatrix} B_0 \\ \Gamma_0 \end{pmatrix} + N_1$$

- **k级估计式** 为：

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix} = ((Y_0 + k(\hat{Y}_0 - Y_0), X_0)'(Y_0, X_0))^{-1} (Y_0 + k(\hat{Y}_0 - Y_0), X_0)' Y_1$$

$$\begin{bmatrix} \hat{B}_0 \\ \hat{\Gamma}_0 \end{bmatrix} = ((Y_0 + k(\hat{Y}_0 - Y_0), X_0)'(Y_0, X_0))^{-1} (Y_0 + k(\hat{Y}_0 - Y_0), X_0)' Y_1$$

- **k=0**时，即为**OLS**估计式；
- **k=1**时，即为**2SLS**估计式；
- **k**等于有限信息估计方法中的 λ 时，即为有限信息估计式。

2.k级估计式的性质

- 假设工具变量与随机误差项不相关，即

$$P\lim \frac{1}{n} (\mathbf{Y}_0 + k(\hat{\mathbf{Y}}_0 - \mathbf{Y}_0))\mathbf{N}_1 = 0$$

且先决变量与随机误差项不相关，即

$$P\lim \frac{1}{n} (\mathbf{X}_0'\mathbf{N}_1) = 0$$

那么，容易证明k级估计式是一致性估计式。

- 工具变量与随机误差项不相关，对 k 是有限制的，必须有（证明见教科书）：

$$Plim(1 - k) = 0$$

- 这就是说，只有在**2SLS**或有限信息估计方法中， k 级估计式是一致性估计式，而在**OLS**方法中，不具有有一致性。

§ 6. 5联立方程计量经济学模型若干问题的讨论

- 一、模型估计方法的比较
- 二、为什么普通最小二乘法被普遍采用
- 三、联立方程模型的检验*

一、模型估计方法的比较 (粗略结果)

1.大样本估计特性的比较

- 在大样本情况下，各种参数估计方法的统计特性可依渐近理论从数学上进行严格的证明，因而也可将各种方法按照各个性质比较优劣。
- 按渐近无偏性/一致性比较优劣

除OLS方法外，所有方法的参数估计量都具有大样本下渐近无偏性；

因而，除OLS方法最差外，其它方法无法比较优劣。

- 按渐近有效性比较优劣

OLS 非一致性估计，未利用任何单方程外的信息；

IV 利用了模型系统部分先决变量的数据信息；

2SLS、LIML 利用了模型系统全部先决变量的数据信息；

3SLS、FIML 利用了模型系统全部先决变量的数据信息和结构方程相关性信息。

2.小样本估计特性的Monte Carlo试验

- 参数估计量的大样本特性只是理论上的，实际上并没有“大样本”；所以，对小样本估计特性进行比较更有实际意义。
- 在小样本情况下，各种参数估计方法的统计特性、无法从数学上进行严格证明，因而提出了一种Monte Carlo试验方法。
- Monte Carlo试验方法，在经济实验中被广泛采用。

- 小样本估计特性的Monte Carlo试验过程 (略)

第一步： 利用随机数发生器产生随机项分布的一组样本；

第二步： 代入已经知道结构参数和先决变量观测值的结构模型中；

第三步： 计算内生变量的样本观测值；

第四步： 选用各种估计方法估计模型的结构参数。

上述步骤反复进行数百次，得到每一种估计方法的参数估计值的序列。

第五步： 对每种估计方法的参数估计值序列进行统计分析；

第六步： 与真实参数（即试验前已经知道的结构参数）进行比较，以判断各种估计方法的优劣。

- 小样本估计特性实验结果比较

(1)无偏性

OLS 2SLS 3SLS (LIML, FIML)



(2)最小方差性

LIML 2SLS FIML OLS



(3)最小均方差性

OLS LIML 2SLS 3SLS (FIML)



为什么OLS具有最好的最小方差性？

方差的计算公式：

$$V = \frac{1}{N} \sum_{i=1}^N (\hat{\beta}_i - \bar{\hat{\beta}})^2$$

均方差的计算公式：

$$MSE = E(\hat{\beta} - \beta)^2 = \frac{1}{n} \sum_{i=1}^N (\hat{\beta}_i - \beta)^2$$

前者反映估计量偏离实验均值的程度；
后者反映估计量偏离真实值的程度。

所以尽管OLS具有最小方差性，但是由于它是有偏的，偏离真实值最为严重，所以它的最小均方差性仍然是最差的。

二、为什么普通最小二乘法被普遍采用 (略)

1. 小样本特性

- 从理论上讲，在小样本情况下，各种估计方法的估计量都是有偏的。

2. 充分利用样本数据信息

- 除OLS之外的其它估计方法可以部分地或者全部地利用某个结构方程中未包含的先决变量的数据信息，从而提高参数估计量的统计性质；但是其前提是，所有变量具有相同的样本容量。
- 但实际上，变量经常不具有相同的样本容量。
- 采用先进估计方法所付出的代价，经常是牺牲了该方程所包含的变量的样本数据信息。

3. 确定性误差传递

- 确定性误差：外生变量的观测误差，以及结构方程的关系误差。
- 采用OLS方法，当估计某个结构方程时：方程中没有包含的外生变量的观测误差和其它结构方程的关系误差，对该方程的估计结果没有影响。
- 如果采用2SLS方法 ...
- 如果采用3SLS方法...

4. 样本容量不支持

- 实际的联立方程模型中每个结构方程、往往是过度识别的，适宜采用2SLS或3SLS方法；

但是在其第一阶段要以所有先决变量作为解释变量，这就需要很大容量的样本。实际上是难以实现的。

- 采用主分量方法等可以克服这个矛盾，但又带来方法的复杂性和新的误差。

5. 实际模型的递推（Recurrent）结构

- 应用中的联立方程模型主要是宏观经济计量模型。
- 宏观经济计量模型有时具有递推结构。
- 具有递推结构的模型，可以采用OLS。

补充：递推模型（Recursive Model）

$$\mathbf{B}\mathbf{Y} + \mathbf{\Gamma}\mathbf{X} = \mathbf{N}$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\beta_{21} & 1 & 0 & \cdots & 0 \\ -\beta_{31} & -\beta_{32} & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ -\beta_{g1} & -\beta_{g2} & -\beta_{g3} & \cdots & 1 \end{bmatrix}$$

$$\Gamma = \begin{bmatrix} -\gamma_{11} & -\gamma_{12} & \cdots & -\gamma_{1k} \\ -\gamma_{21} & -\gamma_{22} & \cdots & -\gamma_{2k} \\ \vdots & & & \\ -\gamma_{g1} & -\gamma_{g2} & \cdots & -\gamma_{gk} \end{bmatrix}$$

- 可采用OLS依次估计每个结构方程；
- 在估计后面的结构方程时，认为其中的内生解释变量、是“先决”的。

三、联立方程模型的检验*

- 包括单方程检验、和方程系统的检验。

凡是在单方程模型中必须进行的各项检验：
对联立方程模型中的结构方程，以及应用2SLS或3SLS方法过程中的简化式方程——都是适用的和需要的。

- 模型系统的检验主要包括：

1.拟合效果检验

- 将样本期的先决变量观测值代入估计后的模型，求解该模型系统，得到内生变量的估计值；

将估计值与实际观测值进行比较，据此判断模型系统的拟合效果。

- 常用判断模型系统拟合效果的检验统计量、是“均方百分比误差”，用RMS表示。

$$RMS_i = \sqrt{\sum_{t=1}^T e_{it}^2 / T}$$

$$e_{it} = (y_{it} - \hat{y}_{it}) / y_{it}$$

- 当 $RMS_i=0$ ，表示第*i*个内生变量估计值与观测值完全拟合。
- 一般地，在g个内生变量中， $RMS<5\%$ 的变量数目占70%以上，并且每个变量的RMS不大于10%，则认为模型系统总体拟合效果较好。

2.预测性能检验

- 如果样本期之外的某个时间截面上的内生变量实际观测值已经知道，这就有条件对模型系统进行预测检验。
- 将该时间截面上的先决变量实际观测值代入模型，计算所有内生变量预测值，并计算其相对误差。

$$RE = (y_{i0} - \hat{y}_{i0}) / y_{i0}$$

- 一般认为， $RE < 5\%$ 的变量数目占70%以上，并且每个变量的相对误差不大于10%，则认为模型系统总体预测性能较好。

3.方程间误差传递检验

- 寻找模型中描述主要经济行为主体的经济活动过程的、方程之间存在明显递推关系的关键路径。
- 在关键路径上进行误差传递分析，可检验总体模型的模拟优度和预测精度。
- 例如，计算：

$$\left(\sum_{i=2}^N (e_i - e_{i-1})^2 / \sum_{i=1}^N e_i^2 \right) \frac{N}{N-1}$$

- 称为冯诺曼比，如果误差在方程之间没有传递，该比值为0。

4.样本点间误差传递检验

- 在联立方程模型系统中，由于经济系统的动态性，决定了有一定数量的滞后内生变量。
- 由于滞后内生变量的存在，使得模型预测误差不仅在方程之间传递，而且在不同的时间截面之间，即样本点之间传递。
- 必须对模型进行滚动预测检验。

Step 1:

- 给定 $t=1$ 时的所有先决变量的观测值，包括滞后内生变量，求解方程组，得到内生变量 Y_1 的预测值；
- 对于 $t=2$ ，只给定外生变量的观测值，滞后内生变量、则以“前一时期的预测值”代替，求解方程组，得到内生变量第二期的 Y_2 预测值；
- “逐年滚动预测”，直至得到 $t=n$ 时、内生变量 Y_n 的预测值；
- 求出该滚动预测值与实际观测值的相对误差。

Step 2:

- 将 $t=n$ 时的所有先决变量的观测值，包括滞后内生变量的实际观测值，代入模型，求解方程组，得到内生变量 Y_n 的“非滚动预测值”；
- 求出该非滚动预测值与实际观测值的相对误差。
- 比较两种结果，二者的差异、表明模型预测误差在不同时间截面之间的传递。

联立方程模型估计方法

(补充升级, 仅供参考)

- 一、过度识别方程的GMM估计
- 二、其它有限信息估计方法简介
- 三、系统估计方法

一、过度识别方程的GMM估计

(GMM, Generalized Method of Moments)

1、过度识别方程的GMM估计

$$\begin{cases} C_t = \alpha_0 + \alpha_1 Y_t + \alpha_2 C_{t-1} + \mu_{1t} \\ I_t = \beta_0 + \beta_1 Y_t + \mu_{2t} \\ Y_t = I_t + C_t + G_t \end{cases}$$

投资方程是过度识别的，
可以采用**2SLS**和**GMM**
估计

$$\begin{cases} \sum_t I_t - \sum_t (\beta_0 + \beta_1 Y_t) = 0 \\ \sum_t I_t G_t - \sum_t (\beta_0 + \beta_1 Y_t) G_t = 0 \\ \sum_t I_t C_{t-1} - \sum_t (\beta_0 + \beta_1 Y_t) C_{t-1} = 0 \end{cases}$$

投资方程
GMM估计的**矩条件**

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

EViews - [Workfile: E6.4 - (c:\program files\evIEWS5\example files\data\

File Edit Object View Proc Quick Options Window Help

View Proc Object Print Save Details+/- Show Fetch Store Delete Genr Sample

Range: 1 30 -- 30 obs
Sample: 1 30 -- 30 obs

-  c
-  g
-  i
-  resid
-  x
-  y

Equation Estimation

Specification Options

Equation specification
Dependent variable followed by list of AR and PDL terms, OR an explicit

I C Y

Instrument list
C G X(-1)

☒ include lagged regressors for linear eqns

Estimation settings
Method: GMM - Generalized Method of Moments
Sample: 1 30

Weighting matrix
☐ Cross section (White Cov)
☒ Time series ()

HAC options
☐ Prewhiten

Kernel
☒ Bartlett ☐ Quadratic

Bandwidth
☒ Fixed Number or NW for Newey
☐ Andrews
☐ Variable - Newey-W.

确定 取消

Dependent Variable: I

Method: Generalized Method of Moments

Date: 12/04/10 Time: 12:46

Sample (adjusted): 2 30

Included observations: 29 after adjustments

Kernel: Bartlett, Bandwidth: Fixed (3), No prewhitening

Simultaneous weighting matrix & coefficient iteration

Convergence achieved after: 6 weight matrices, 7 total coef iterations

Instrument list: C G X(-1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-897.3970	254.0871	-3.531848	0.0015
Y	<u>0.428524</u>	0.001566	273.5763	0.0000
R-squared	0.992582	Mean dependent var	26757.40	
Adjusted R-squared	0.992307	S.D. dependent var	30063.82	
S.E. of regression	2636.860	Sum squared resid	1.88E+08	
Durbin-Watson stat	0.195148	J-statistic	0.063942	

Dependent Variable: I

Method: Two-Stage Least Squares

Date: 12/04/10 Time: 12:52

Sample (adjusted): 2 30

Included observations: 29 after adjustments

Instrument list: C G X(-1)

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1455.814	590.5228	-2.465297	0.0203
Y	<u>0.419774</u>	0.006092	68.90472	0.0000

R-squared	0.994398	Mean dependent var	26757.40
Adjusted R-squared	0.994190	S.D. dependent var	30063.82
S.E. of regression	2291.486	Sum squared resid	1.42E+08
F-statistic	4747.860	Durbin-Watson stat	0.264423
Prob(F-statistic)	0.000000		

2、2SLS与GMM

- 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，从信息利用角度、是等价的。

- 2SLS估计

$$y_i = X_i' \beta + \mu_i \quad i = 1, \dots, n$$

$$X = Z\pi + \mu$$

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

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

第1阶段估计:
Z全部外生

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

第2阶段估计:
以 \hat{X} 作为X的工具

- **GMM估计**

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

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

大括号部分的
一阶极值条件

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

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

$$\hat{W} = \frac{\hat{\sigma}^2}{n^2} Z'Z$$

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

Dependent Variable: l

Method: Generalized Method of Moments

Date: 12/04/10 Time: 13:07

Sample (adjusted): 2 30

Included observations: 29 after adjustments

Kernel: Bartlett, Bandwidth: Fixed (2), Prewhitening

Simultaneous weighting matrix & coefficient iteration

Convergence achieved after: 55 weight matrices, 56 total coef iterations

Instrument list: C G X(-1)

改变权矩阵
设置，估计
结果改变

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-1317.781	1176.200	-1.120372	0.2724
Y	<u>0.418288</u>	0.009751	42.89795	0.0000
R-squared	0.994363	Mean dependent var	26757.40	
Adjusted R-squared	0.994155	S.D. dependent var	30063.82	
S.E. of regression	2298.518	Sum squared resid	1.43E+08	
Durbin-Watson stat	0.264405	J-statistic	1.02E-05	

- **GMM和2SLS比较，哪种方法更优？**

二、其它有限信息估计方法简介 (Limited Information Estimation Methods)

1.有限信息最大或然法 (LIML, Limited Information Maximum Likelihood)

- 以最大或然为准则，通过对简化式模型进行最大或然估计，以得到结构方程参数估计量的联立方程模型的单方程估计方法。
- 由Anderson和Rubin于1949年提出，早于两阶段最小二乘法。
- 适用于恰好识别和过度识别结构方程的估计。

- 在该方法中，以下两个概念是重要的：
 - 这里的“有限信息”指的是每次估计只考虑一个结构方程的信息，而没有考虑模型系统中其它结构方程的信息；
 - 这里的“最大或然法”是针对结构方程中包含的内生变量的简化式模型的，即应用最大或然法求得的是简化式参数估计量，而不是结构式参数估计量。

$$Y_1 = (Y_0, X_0) \begin{pmatrix} B_\theta \\ \Gamma_\theta \end{pmatrix} + N_1$$

$$(Y_0^1, X_0) \begin{pmatrix} B_0^1 \\ \Gamma_\theta \end{pmatrix} + N_1 = 0$$

$$Y_0^1 = X \Pi_0^1 + E_0^1$$

$$LnL(Y_0^1) = c + \frac{n}{2} Ln|\Omega_0^{-1}| - \frac{1}{2} tr \Omega_0^{-1} (Y_0^1 - X \Pi_0^1)' (Y_0^1 - X \Pi_0^1)$$

简化式模型随机误差项
方差—协方差矩阵

2.有限信息最小方差比方法(LVR, **Least Variable Ratio**)

- 估计某一个结构方程参数时，仍然只利用关于该方程的信息，没有利用方程系统的信息，所以是一种有限信息估计方法。
- 参见教科书《计量经济学——方法与应用》**p143-145。**

三、联立方程计量经济学模型的系统 估计方法

the Systems Estimation Methods

1.联立模型随机误差项方差—协方差矩阵

- 随机误差项的同期相关性

- 随机误差项的相关性不仅存在于每个结构方程不同样本点之间，而且存在于不同结构方程之间。
- 对于不同结构方程的随机误差项之间，不同时期互不相关，只有同期的随机误差项之间才相关，称为具有同期相关性。

•具有同期相关性的方差—协方差矩阵 （略）

$$\mathbf{YB} + \mathbf{X}\Gamma = \mathbf{N}$$

$$\mathbf{Y} = \mathbf{Z}\Delta + \tilde{\mathbf{N}}$$



$$Y_i = \mathbf{Z}_i \Delta_i + \tilde{N}_i$$

$$\mathbf{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_g \end{bmatrix}$$

$$Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{in} \end{bmatrix}$$

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_0^i & \mathbf{X}_0^i \end{pmatrix}$$

$$\Delta_i = \begin{pmatrix} \mathbf{B}_0^i \\ \Gamma_0^i \end{pmatrix}$$

假设:

- 对于一个结构方程的随机误差项，在不同样本点之间，具有同方差性和序列不相关性。即

$$Cov(\tilde{N}_i) = \sigma_{ii}^2 \mathbf{I}$$

- 对于不同结构方程的随机误差项之间，具有且仅具有同期相关性。即

$$Cov(\tilde{N}_i, \tilde{N}_j) = \sigma_{ij} \mathbf{I}$$

联立模型系统随机误差项方差—协方差矩阵为

$$: \quad Cov(\tilde{\mathbf{N}}) = \begin{bmatrix} \sigma_{11}^2 \mathbf{I} & \sigma_{12} \mathbf{I} & \cdots & \sigma_{1g} \mathbf{I} \\ \sigma_{21} \mathbf{I} & \sigma_{22}^2 \mathbf{I} & \cdots & \sigma_{2g} \mathbf{I} \\ \vdots & & & \\ \sigma_{g1} \mathbf{I} & \sigma_{g2} \mathbf{I} & & \sigma_{gg}^2 \mathbf{I} \end{bmatrix}$$

$$= \Sigma \otimes \mathbf{I}$$

$$= \Omega$$

2.三阶段最小二乘法简介

(3SLS, Three Stages Least Squares)

- 概念

- **3SLS**是由**Zellner**和**Theil**于**1962**年提出的同时估计联立方程模型全部结构方程的系统估计方法。
- 其基本思路是 $3SLS = 2SLS + GLS$
- 即首先用**2SLS**估计模型系统中每一个结构方程，然后再用GLS估计模型系统。

- 三阶段最小二乘法的步骤（略）

(1) 用2SLS估计结构方程

$$Y_i = \mathbf{Z}_i \Delta_i + \tilde{N}_i$$

得到方程随机误差项的估计值。

$$\mathbf{Z}_i = \begin{pmatrix} \mathbf{Y}_0^i & \mathbf{X}_0^i \end{pmatrix} \longrightarrow \mathbf{Y}_0^i = \mathbf{X}\Pi_0^i + \mathbf{E}_0^i$$

OLS估计

$$\hat{\mathbf{Y}}_0^i = \mathbf{X}\hat{\Pi}_0^i = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Y}_0^i$$

$$\hat{\mathbf{Z}}_i = \begin{pmatrix} \hat{\mathbf{Y}}_0^i & \mathbf{X}_0^i \end{pmatrix}$$

OLS估计

$$\hat{\Delta}_i = (\hat{\mathbf{Z}}_i' \hat{\mathbf{Z}}_i)^{-1} \hat{\mathbf{Z}}_i' \mathbf{Y}_i$$

$$\hat{\hat{\mathbf{Y}}}_i = \hat{\mathbf{Z}}_i \hat{\Delta}_i \longrightarrow e_{il} = y_{il} - \hat{\hat{y}}_{il}$$

(2) 求随机误差项方差—协方差矩阵的估计量

$$\mathbf{e}_i = (e_{i1} \quad e_{i2} \quad \cdots \quad e_{in})'$$

$$\hat{\sigma}_{ij} = \frac{\mathbf{e}_i' \mathbf{e}_j}{\sqrt{(n - g_i + 1 - k_i)(n - g_j + 1 - k_j)}}$$

$$\hat{\Sigma} = (\hat{\sigma}_{ij})$$

$$\hat{\Sigma} \otimes \mathbf{I} = \hat{\Omega}$$

(3) 用**GLS**估计原模型系统

$$\mathbf{Y} = \mathbf{Z}\Delta + \tilde{\mathbf{N}}$$

得到结构参数的**3SLS**估计量为：

$$\begin{aligned}\hat{\hat{\Delta}} &= (\hat{\mathbf{Z}}' \hat{\mathbf{\Omega}}^{-1} \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}' \hat{\mathbf{\Omega}}^{-1} \mathbf{Y} \\ &= (\hat{\mathbf{Z}}' (\hat{\mathbf{\Sigma}} \otimes \mathbf{I})^{-1} \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}' (\hat{\mathbf{\Sigma}} \otimes \mathbf{I})^{-1} \mathbf{Y}\end{aligned}$$

- 三阶段最小二乘法估计量的统计性质

- 如果联立方程模型系统中所有结构方程都是可以识别的，并且非奇异，则**3SLS**估计量是一致估计量。
- **3SLS**估计量，比**2SLS**估计量更有效。
- 如果 Σ 是对角矩阵，即模型系统中不同结构方程的随机误差项之间无相关性，那么可以证明**3SLS**估计量与**2SLS**估计量是等价的。
- 反过来说明，**3SLS**方法主要优点是考虑了模型系统中不同结构方程的随机误差项之间的相关性。

3.完全信息最大似然法简介

(FIML, Full Information Maximum Likelihood)

- 概念

- 另一种已有实际应用的联立方程模型的系统估计方法。
- **Rothenberg**和**Leenders**于**1964**年提出一个线性化的**FIML**估计量。
- **FIML**是**ML**的直接推广，是在已经得到样本观测值的情况下：使整个联立方程模型系统的或然函数达到最大、以得到所有结构参数的估计量。

- 完全信息最大似然函数（略）

$$\mathbf{Y} = \mathbf{Z}\Delta + \tilde{\mathbf{N}}$$

$$\tilde{\mathbf{N}} \sim \text{正态}(\mathbf{0}, \Sigma \otimes \mathbf{I})$$

MLE的直接推广

$$L(\mathbf{Y}) = \frac{1}{(2\pi)^{gn/2} |\Sigma \otimes \mathbf{I}|^{-\frac{1}{2}}} \left| \frac{\partial \tilde{\mathbf{N}}}{\partial \mathbf{Y}} \right| e^{-\frac{1}{2}(\mathbf{Y} - \mathbf{Z}\Delta)'(\Sigma^{-1} \otimes \mathbf{I})(\mathbf{Y} - \mathbf{Z}\Delta)}$$

$$= \frac{1}{(2\pi)^{gn/2} |\Sigma \otimes \mathbf{I}|^{-\frac{1}{2}}} |\mathbf{B}|^n e^{-\frac{1}{2}(\mathbf{Y} - \mathbf{Z}\Delta)'(\Sigma^{-1} \otimes \mathbf{I})(\mathbf{Y} - \mathbf{Z}\Delta)}$$

$$L(\mathbf{Y}) = \frac{1}{(2\pi)^{gn/2} |\Sigma \otimes \mathbf{I}|^{-\frac{1}{2}}} \left| \frac{\partial \tilde{\mathbf{N}}}{\partial \mathbf{Y}} \right| e^{-\frac{1}{2}(\mathbf{Y} - \mathbf{Z}\Delta)'(\Sigma^{-1} \otimes \mathbf{I})(\mathbf{Y} - \mathbf{Z}\Delta)} \\ - \frac{1}{2}(\mathbf{Y} - \mathbf{Z}\Delta)'(\Sigma^{-1} \otimes \mathbf{I})(\mathbf{Y} - \mathbf{Z}\Delta)$$

- 对数或然函数对于协方差逆矩阵的元素取极大值的一阶条件，得到协方差矩阵的元素的FIML估计量；
- 对数或然函数对于待估计参数取极大值的一阶条件，求解该方程系统，即可得到结构参数的FIML估计量。
- 研究的重点是如何求解非线性方程系统。