

第 23 章 似不相关回归

Seemingly Unrelated Reg.

23.1 单一方程估计与系统估计 (SUR)

如果多方程间有联系, 同时估计这些方程可提高估计效率, 称为“系统估计”(system estimation)。

有时多个方程从同一个最大化问题推导而来(比如, 从企业的利润最大化问题导出对资本与劳动力的需求), 故存在“跨方程的参数约束”(cross-equation restrictions)。

多方程联合估计可检验这些跨方程约束。也可加上这些约束条件后再进行系统估计。

多方程联合估计的缺点是，如果某方程误差较大，将污染整个方程系统。选择单一方程估计或系统估计，也是“有效性”与“稳健性”的权衡。

多方程系统主要分为两类。一类为“联立方程组”(simultaneous equations)，即不同方程间存在内在联系，一个方程的解释变量是另一方程的被解释变量。

另一类为“似不相关回归”(Seemingly Unrelated Regression Estimation, 简记 SUR 或 SURE)，即各方程的变量之间没有内在联系，但各方程的扰动项之间存在相关性。

$$y_1 = x_1' \beta_1 + \varepsilon_1 \quad x_1: \text{kww, } s, \text{ med.}$$
$$y_2 = x_2' \beta_2 + \varepsilon_2 \quad x_2: \exp^2, \text{ if.}$$

例(似不相关回归) 以研一学生的计量成绩与英语成绩作为两个被解释变量。两个方程所包含的解释变量可以不同。由于同一学生的不可观测因素同时对计量成绩与英语成绩造成影响，故两个方程的扰动项相关。如进行联合估计，可提高估计效率。

23.2 似不相关回归的假定

假设有 n 个方程(n 个被解释变量)，每个方程有 T 个观测值，
 $T > n$ 。在第 i 个方程中，共有 K_i 个解释变量。

第 i 个方程可以写为

$$\underbrace{\mathbf{y}_i}_{T \times 1} = \underbrace{\mathbf{X}_i}_{T \times K_i} \underbrace{\boldsymbol{\beta}_i}_{K_i \times 1} + \underbrace{\boldsymbol{\varepsilon}_i}_{T \times 1} \quad (i = 1, 2, \dots, n)$$

将所有的方程叠放在一起可得

$$\mathbf{y} \equiv \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}}_{nT \times 1} = \underbrace{\begin{pmatrix} K_1 & K_2 & \cdots & K_n \\ X_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & X_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & X_n \end{pmatrix}}_{nT \times \sum_{i=1}^n K_i} \underbrace{\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix}}_{\sum_{i=1}^n K_i \times 1} + \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix}}_{nT \times 1} \equiv X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

考察“大”扰动项 $\boldsymbol{\varepsilon}$ 之协方差矩阵

$$\boldsymbol{\Omega} \equiv \text{Var} \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix} = E \left[\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix} (\boldsymbol{\varepsilon}'_1 \boldsymbol{\varepsilon}'_2 \cdots \boldsymbol{\varepsilon}'_n) \right] = E \begin{pmatrix} \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}'_1 & \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}'_2 & \cdots & \boldsymbol{\varepsilon}_1 \boldsymbol{\varepsilon}'_n \\ \boldsymbol{\varepsilon}_2 \boldsymbol{\varepsilon}'_1 & \boldsymbol{\varepsilon}_2 \boldsymbol{\varepsilon}'_2 & \cdots & \boldsymbol{\varepsilon}_2 \boldsymbol{\varepsilon}'_n \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}'_1 & \boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}'_2 & \cdots & \boldsymbol{\varepsilon}_n \boldsymbol{\varepsilon}'_n \end{pmatrix}_{nT \times nT}$$

假设同一方程不同期的扰动项不存在自相关，且方差也相同，记第 i 个方程的方差为 σ_{ii} 。

协方差阵 $\boldsymbol{\Omega}$ 中主对角线上的第 (i, i) 个矩阵为

$$E(\boldsymbol{\varepsilon}_i \boldsymbol{\varepsilon}'_i) = \boxed{\sigma_{ii} \mathbf{I}_T}$$

假设不同方程的扰动项之间存在同期相关，即

$$E(\varepsilon_{it}\varepsilon_{js}) = \begin{cases} \sigma_{ij}, & t = s \\ 0, & t \neq s \end{cases}$$

协方差阵 Ω 中的第 (i, j) 个矩阵 $(i \neq j)$ 为

$$\underbrace{E(\varepsilon_i \varepsilon_j')}_{\sigma_{ij}} \mathbf{I}_T$$

综合以上结果可知

$$\Omega = \begin{pmatrix} \sigma_{11} \mathbf{I}_T & \sigma_{12} \mathbf{I}_T & \cdots & \sigma_{1n} \mathbf{I}_T \\ \sigma_{21} \mathbf{I}_T & \sigma_{22} \mathbf{I}_T & \cdots & \sigma_{2n} \mathbf{I}_T \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} \mathbf{I}_T & \sigma_{n2} \mathbf{I}_T & \cdots & \sigma_{nn} \mathbf{I}_T \end{pmatrix}$$

可否把 \mathbf{I}_T 从右边提取出来?

定义 对于任意两个矩阵 $\underline{\mathbf{A}}_{m \times n} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ 与 $\underline{\mathbf{B}}_{p \times q}$ (矩阵 \mathbf{A}, \mathbf{B} 的维度可以完全不同), 克罗内克尔乘积(Kronecker product)为

$$\mathbf{A} \otimes \mathbf{B} \equiv \begin{pmatrix} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{pmatrix}_{mp \times nq}$$

对于任意矩阵 \mathbf{A}, \mathbf{B} , 克罗内克尔乘积 $\mathbf{A} \otimes \mathbf{B}$ 总有定义。

可以证明，克罗内克尔乘积具有以下性质：

$$(1) \quad (\mathbf{A} \otimes \mathbf{B})(\mathbf{C} \otimes \mathbf{D}) = (\mathbf{AC}) \otimes (\mathbf{BD})$$

$$(2) \quad (\mathbf{A} \otimes \mathbf{B})' = \mathbf{A}' \otimes \mathbf{B}'$$

$$(3) \quad (\mathbf{A} \otimes \mathbf{B})^{-1} = \mathbf{A}^{-1} \otimes \mathbf{B}^{-1}$$



将扰动项 $\boldsymbol{\varepsilon}$ 的协方差矩阵简化为

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix} \otimes \mathbf{I}_T = \boxed{\boldsymbol{\Sigma} \otimes \mathbf{I}_T}$$

其中， $\Sigma \equiv \begin{pmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{pmatrix}$ 为同期协方差矩阵。

$$\frac{n(n-1)}{2}$$

WT

Ω 的逆矩阵可以写为

$$\Omega^{-1} = \boxed{\Sigma^{-1}} \otimes \boxed{I_T}$$

23.3 SUR 的 FGLS 估计

由于 Ω 不是单位矩阵，故用 OLS 不是最有效率。

假设 Ω 已知，则 GLS 最有效率：

$$\hat{\beta}_{\text{GLS}} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} y = [X' (\Sigma^{-1} \otimes I_T) X]^{-1} X' (\Sigma^{-1} \otimes I_T) y$$

此 GLS 估计量与单一方程 OLS 估计量不同。

如果出现以下两种情形之一，则 GLS 与单一方程 OLS 完全相同。

- (1) 各方程的扰动项互不相关。在 SUR 模型中，各方程间唯一的联系是扰动项间的相关性。如果扰动项互不相关， Ω 是单位矩阵，则系统估计与单一方程估计无区别。
- (2) 每个方程包含的解释变量完全相同。比如，VAR 的每个方

程包含完全相同的解释变量，故使用单一方程 OLS 估计 VAR。

除了以上两种特殊情形外，各方程的扰动项之间的相关性越大，则 GLS 所能带来的效率改进就越大；

各方程的数据矩阵 $\{X_1, X_2, \dots, X_n\}$ 之间的相关性越小，则 GLS 所能带来的效率改进也越大。

现实中，一般首先需要估计 $\hat{\Omega}$ ，然后进行 FGLS 估计。

首先，使用单一方程 OLS 的残差来一致地估计 σ_{ij} 。

假设第 i 个方程的 OLS 残差向量为 \mathbf{e}_i , 则 σ_{ij} 的一致估计量为

$$\hat{\sigma}_{ij} = \frac{1}{T} \mathbf{e}'_i \mathbf{e}_j = \frac{1}{T} \sum_{t=1}^T e_{it} e_{jt}$$

$$\hat{\Sigma}^{-1}$$

每个做 OLS
 $\hat{\beta}_1 \rightarrow e_1$
 $\hat{\beta}_2 \rightarrow e_2$
 \vdots
 $\hat{\beta}_n \rightarrow e_n$

因此, $\hat{\Omega} = \begin{pmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} & \cdots & \hat{\sigma}_{1n} \\ \hat{\sigma}_{21} & \hat{\sigma}_{22} & \cdots & \hat{\sigma}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\sigma}_{n1} & \hat{\sigma}_{n2} & \cdots & \hat{\sigma}_{nn} \end{pmatrix} \otimes \mathbf{I}_T$ 。将 $\hat{\Omega}$ 代入方程可得

$$\hat{\beta}_{\text{SUR}} = \underbrace{(\mathbf{X}' \hat{\Omega}^{-1} \mathbf{X})^{-1}}_{\downarrow} \mathbf{X}' \hat{\Omega}^{-1} \mathbf{y}$$

$\hat{\beta}_{1,\text{SUR}}$

$\hat{\beta}_{2,\text{SUR}}$

这就是“似不相关估计量”(Zellner, 1962), 记为 $\hat{\beta}_{\text{SUR}}$ 。

使用 FGLS 后得到新的残差，可再计算 $\hat{\boldsymbol{\Omega}}$ ，不断迭代直至系数估计值 $\hat{\boldsymbol{\beta}}_{\text{SUR}}$ 收敛为止。

23.4 SUR 的假设检验

对多方程系统进行 SUR 估计后，对线性假设 “ $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ ” 的检验可以照常进行。

由于 $\boldsymbol{\beta}$ 包含了所有方程的参数，故可检验跨方程的参数约束。

如果接受 “ $H_0 : \mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ ”，则可把 “ $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ ” 作为约束条件，进行有约束的 FGLS 估计。

即使各方程的解释变量完全相同，有时也使用 SUR 而不使用单一方程 OLS，以便检验跨方程的参数约束。

如果存在跨方程的参数约束，则即使各方程的解释变量完全相同，SUR 估计与单一方程 OLS 也不再等价。

SUR 模型的基本假设是，各方程的扰动项之间存在同期相关。

需要检验原假设“ H_0 :各方程的扰动项无同期相关”，即“ $H_0:\Sigma$ 为对角矩阵”。

Breusch and Pagan(1980)建议使用以下 LM 统计量：

$$\lambda_{\text{LM}} = T \sum_{i=2}^n \sum_{j=1}^{i-1} r_{ij}^2 \xrightarrow{d} \chi^2(n(n-1)/2)$$

其中, $r_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}}$ 为根据残差计算的扰动项 ε_i 与 ε_j 之间的同期相关系数, 而 $\sum_{i=2}^n \sum_{j=1}^{i-1} r_{ij}^2$ 为同期相关系数矩阵

$$\begin{pmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{pmatrix}$$

主对角线以下各项之平方和(该矩阵为对称矩阵)。

第 24 章 联立方程模型

24.1 联立方程模型的结构式与简化式

经济理论常常推导出一组相互联系的方程, 其中一个方程的解释变量是另一方程的被解释变量, 这就是联立方程组。

例 农产品市场均衡模型, 由需求函数、供给函数及市场均衡条件组成, 参见第 10 章。

例 简单的宏观经济模型, 参见第 10 章。

即使我们只关心单个方程，但如果该方程包含内生解释变量，则完整的模型仍然是联立方程组。

由 \underline{M} 个方程构成的联立方程模型的“结构式”(structural form):

$$\left\{ \begin{array}{l} \gamma_{11}y_{t1} + \gamma_{21}y_{t2} + \cdots + \gamma_{M1}y_{tM} + \beta_{11}x_{t1} + \cdots + \beta_{K1}x_{tK} = \varepsilon_{t1} \\ \gamma_{12}y_{t1} + \gamma_{22}y_{t2} + \cdots + \gamma_{M2}y_{tM} + \beta_{12}x_{t1} + \cdots + \beta_{K2}x_{tK} = \varepsilon_{t2} \\ \cdots \cdots \cdots \\ \gamma_{1M}y_{t1} + \gamma_{2M}y_{t2} + \cdots + \gamma_{MM}y_{tM} + \beta_{1M}x_{t1} + \cdots + \beta_{KM}x_{tK} = \varepsilon_{tM} \end{array} \right.$$

$\{y_{ti}\}$ 为内生变量， $\{x_{tj}\}$ 为外生变量，第一个下标表示第 t 个观测值($t = 1, \dots, T$)，第二个下标表示第*i*个内生变量($i = 1, \dots, M$)，或第*j*个外生变量($j = 1, \dots, K$)。

内生变量的系数为 $\{\gamma_{ik}\}_{m \times M}$, 其第一个下标表示它是第*i*个内生变量的系数, 而第二个下标表示它在第*k*个方程中($k = 1, \dots, M$)。

外生变量的系数为 $\{\beta_{jk}\}_{m \times M}$, 其第一个下标表示它是第*j*个外生变量的系数, 而第二个下标表示它在第*k*个方程中。

结构方程的扰动项为 $\{\varepsilon_{tk}\}_{m \times M}$, 其第一个下标表示第*t*个观测值($t = 1, \dots, T$), 而第二个下标表示它在第*k*个方程中。

“完整的方程系统”(complete system of equations)要求, 内生变量个数等于方程个数*M*。

将上述方程组写成更简洁的“横排”矩阵形式

$$\begin{aligned}
 & \text{Red Text: } \Gamma : M \times M \\
 & \text{Red Text: } y_t' \\
 & \left(y_{t1} \quad y_{t2} \quad \cdots \quad y_{tM} \right) \begin{pmatrix} \gamma_{11} & \gamma_{12} & \cdots & \gamma_{1M} \\ \gamma_{21} & \gamma_{22} & \cdots & \gamma_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \gamma_{M1} & \gamma_{M2} & \cdots & \gamma_{MM} \end{pmatrix} \\
 & + \left(x_{t1} \quad x_{t2} \quad \cdots \quad x_{tK} \right) \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1M} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{K1} & \beta_{K2} & \cdots & \beta_{KM} \end{pmatrix} = \left(\varepsilon_{t1} \quad \varepsilon_{t2} \quad \cdots \quad \varepsilon_{tM} \right) \\
 & \text{Red Text: } B \quad K \times M
 \end{aligned}$$

用矩阵来表示即

$$y_t' \boldsymbol{\Gamma} + x_t' \boldsymbol{B} = \boldsymbol{\varepsilon}_t'$$

其中，系数矩阵 $\Gamma_{M \times M}$ 与 $B_{K \times M}$ 的每一列对应于一个方程。

比如，第一个方程为

$$(y_{t1} \quad y_{t2} \quad \cdots \quad y_{tM}) \begin{pmatrix} \gamma_{11} \\ \gamma_{21} \\ \vdots \\ \gamma_{M1} \end{pmatrix} + (x_{t1} \quad x_{t2} \quad \cdots \quad x_{tK}) \begin{pmatrix} \beta_{11} \\ \beta_{21} \\ \vdots \\ \beta_{K1} \end{pmatrix} = \varepsilon_{t1}$$

扰动项 ε_t 由第 t 期各方程的扰动项所构成。

假设扰动项 ε_t 满足 $E(\varepsilon_t | \mathbf{x}_t) = \mathbf{0}$ (\mathbf{x}_t 外生)，记其协方差矩阵为，

$$\underline{\Sigma} \equiv E(\varepsilon_t \varepsilon_t' | \mathbf{x}_t)$$

由于存在内生变量，如果直接用 OLS 估计每一方程，将导致内生性偏差或联立方程偏差，得不到一致估计。

求解联立方程组：

True, Structural

$$\mathbf{y}'_t \boldsymbol{\Gamma} = -\mathbf{x}'_t \mathbf{B} + \boldsymbol{\varepsilon}'_t$$

假设 $\boldsymbol{\Gamma}$ 非退化，两边同时右乘 $\boldsymbol{\Gamma}^{-1}$ ，

$$\mathbf{y}'_t = -\mathbf{x}'_t \mathbf{B} \boldsymbol{\Gamma}^{-1} + \boldsymbol{\varepsilon}'_t \boldsymbol{\Gamma}^{-1}$$

$$\mathbf{y}'_t = \mathbf{x}'_t \boldsymbol{\Pi} + \mathbf{v}'_t$$

此方程称为“简化式”(reduced form)。

其系数矩阵为 $\underbrace{\boldsymbol{\Pi}}_{K \times M} \equiv -\underbrace{\mathbf{B}}_{K \times M} \underbrace{\boldsymbol{\Gamma}^{-1}}_{M \times M}$ ，扰动项为 $\mathbf{v}'_t \equiv \boldsymbol{\varepsilon}'_t \boldsymbol{\Gamma}^{-1}$ ，故 $\mathbf{v}_t \equiv \boldsymbol{\Gamma}^{-1'} \boldsymbol{\varepsilon}_t$ 。

简化式扰动项 v_t 仍与外生变量 x_t 不相关，因为

$$E(v_t | x_t) = E(\Gamma^{-1}' \varepsilon_t | x_t) = \Gamma^{-1}' E(\varepsilon_t | x_t) = \mathbf{0}$$

v_t 的协方差矩阵为

$$\Omega \equiv E(v_t v_t' | x_t) = E(\Gamma^{-1}' \varepsilon_t \varepsilon_t' \Gamma^{-1} | x_t) = \Gamma^{-1}' E(\varepsilon_t \varepsilon_t' | x_t) \Gamma^{-1} = \boxed{\Gamma^{-1}' \Sigma \Gamma^{-1}}$$

简化式方程的解释变量全部为外生变量 x_t ，故可用 OLS 得到简化式参数 Π 与 Ω 的一致估计。

但通常我们最终关心的是结构式参数。

在什么情况下，才能从简化式参数(Π, Ω)反推出结构式参数(Γ, B, Σ)呢？

这涉及联立方程模型的“识别问题”(problem of identification)。

24.2 联立方程模型的识别

在对模型的总体参数进行估计之前，其参数必须“可识别”(identified)。

如果一个总体参数可识别，则该参数的任意两个不同取值，都会在随机样本中显示出系统差异，即如果样本容量足够大，则应该能够在统计意义上区分这两个不同的参数值。

反之，如果无论多大的样本都区分不开，即由不同参数值的总体产生的观测数据在统计意义上是一样的，则该参数“不可识别”(unidentified)。

例 考虑以下回归模型：

$$y_i = \alpha_1 + \alpha_2 + \beta x_i + \varepsilon_i$$

仅通过样本数据 $\{y_i, x_i\}_{i=1}^n$ 是无法对 α_1 与 α_2 分别进行识别的，但可以识别二者之和 $(\alpha_1 + \alpha_2)$ 。

回到联立方程模型的情形，“可识别”意味着，可以从简化式参数 (Π, Ω) 求出结构式参数 (Γ, B, Σ) 的唯一解(unique solution)。

这两组参数之间的关系如下：

$$\boldsymbol{\Pi} \equiv -\mathbf{B}\boldsymbol{\Gamma}^{-1}$$

$$\boldsymbol{\Omega} \equiv \boldsymbol{\Gamma}^{-1'} \boldsymbol{\Sigma} \boldsymbol{\Gamma}^{-1}$$

如果 $\boldsymbol{\Gamma}$ 已知，则可通过 $\boldsymbol{\Pi}$ 与 $\boldsymbol{\Omega}$ 求得 \mathbf{B} 与 $\boldsymbol{\Sigma}$ 。但 $\boldsymbol{\Gamma}$ 一般是由未知参数组成的矩阵。

事实上，结构式的参数个数比简化式的参数个数多出 M^2 个。

简化式参数 $(\boldsymbol{\Pi}, \boldsymbol{\Omega})$ 的总个数为 $[K \times M + M(M+1)/2]$ (其中， $\boldsymbol{\Pi}_{K \times M}$ 含 $K \times M$ 个参数，而对称矩阵 $\boldsymbol{\Omega}_{M \times M}$ 含 $M(M+1)/2$ 个参数)；

结构式参数 $(\Gamma, \mathbf{B}, \Sigma)$ 的总个数为 $\left[M^2 + K \times M + M(M+1)/2\right]$ (其中, $\Gamma_{M \times M}$ 含 M^2 个参数, $\mathbf{B}_{K \times M}$ 含 $K \times M$ 个参数, 对称矩阵 Σ 含 $M(M+1)/2$ 个参数)。

一般地, 不可能从 (Π, Ω) 求出 $(\Gamma, \mathbf{B}, \Sigma)$ 的唯一解。

如不对结构式参数进行约束, 将不可能从简化式参数得到结构式参数的唯一解。

为识别结构方程, 常对结构参数施加如下约束。

(1) 标准化(normalization): 在每个结构方程中, 可以将一个内生变量视为被解释变量, 并将其系数标准化为 1。

(2) 恒等式(identity): 比如, 供需相等的均衡条件、会计恒等式、定义式。恒等式中每个变量的系数均为已知, 不需要识别或估计。

(3) 排斥约束(exclusion restrictions): 在结构方程中排斥某些内生或外生变量, 这相当于对结构矩阵(Γ, \mathbf{B})施以“零约束”(zero restrictions), 即让(Γ, \mathbf{B})中的某些元素为 0。

↓ (4) 线性约束(linear restriction): 比如, 在理论上可以假设生产函数为规模报酬不变(constant returns to scale), 则资本的产出弹性与劳动力的产出弹性之和为 1。

(5) 对扰动项协方差矩阵的约束(restrictions on the disturbance covariance matrix): 比如, 在某些情况下, 可以假设不同方程的扰动项之间不相关。

实践中最重要的约束方法是“排斥变量”(即零约束)。

对于线性约束，可通过重新定义变量转化为“排斥变量”约束。

究竟需要多少零约束才可以保证结构方程可识别呢？

不失一般性，考虑第一个结构方程。

假设在第一个方程中，内生变量 y_1 的系数已被标准化为 1，另有 M_1 个内生变量也包括在此方程中，而其余 M_1^* 个内生变量则被排斥在此方程之外，故 $1 + M_1 + M_1^* = M$ 。

假设第一个方程包含 K_1 个外生变量，而其余 K_1^* 个外生变量则被排斥在此方程之外，故 $K_1 + K_1^* = K$ 。

可识别的必要条件为

$$K_1^* \geq M_1$$

称为“阶条件”(order condition)，即结构方程所排斥的外生变量的个数(K_1^*)应大于或等于该方程所包含的内生解释变量的个数(M_1)。

从工具变量法的角度，被第一个结构方程排斥的所有外生变量都是有效工具变量，因为根据外生变量的定义，它们与扰动项不相关(外生性)；而根据简化式，内生变量可以表示为外生变量的函数，故它们与内生解释变量相关(相关性)。

在可识别(即秩条件满足)的情况下，如果恰好 $K_1^* = M_1$ ，则称该结构方程“恰好识别”(just identified)，即工具变量个数正好相等内生解释变量的个数。

如果 $K_1^* > M_1$, 则称该结构方程“过度识别”(overidentified), 即工具变量个数大于内生解释变量的个数。

24.3 单一方程估计法

估计联立方程组的方法可以分为两类:

“单一方程估计法”(single equation estimation), 也称“有限信息估计法”(limited information estimation);

“系统估计法”, 也称“全信息估计法”(full information estimation)。

1. 普通最小二乘法

对于一种特殊的递归模型(recursive model)，即 Γ 为下三角矩阵(lower triangular matrix)而协方差矩阵 Σ 为对角矩阵(不同方程之间的扰动项不相关)的情形，OLS 依然是一致的。

以一个三方程的系统为例：

$$\begin{cases} y_1 = \mathbf{x}'\boldsymbol{\beta}_1 + \varepsilon_1 \\ y_2 = \mathbf{x}'\boldsymbol{\beta}_2 + \gamma_{12}y_1 + \varepsilon_2 \\ y_3 = \mathbf{x}'\boldsymbol{\beta}_3 + \gamma_{13}y_1 + \gamma_{23}y_2 + \varepsilon_3 \end{cases}$$

第一个方程不含内生解释变量，可用 OLS 得到一致估计。

在第二个方程中，唯一的内生解释变量为 y_1 ，且与扰动项不相关：

$$\text{Cov}(y_1, \varepsilon_2) = \text{Cov}(\mathbf{x}'\boldsymbol{\beta}_1 + \varepsilon_1, \varepsilon_2) = \underbrace{\text{Cov}(\mathbf{x}'\boldsymbol{\beta}_1, \varepsilon_2)}_{=0} + \underbrace{\text{Cov}(\varepsilon_1, \varepsilon_2)}_{=0} = 0$$

故可用 OLS 来估计第二个方程。

在第三个方程中，内生解释变量为 (y_1, y_2) ，而且 $\text{Cov}(y_1, \varepsilon_3) = \text{Cov}(y_2, \varepsilon_3) = 0$ ，故也可用 OLS 来估计。

2. 间接最小二乘法

在恰好识别的情况下，可先用 OLS 来一致地估计简化式参数，然后通过结构式参数与简化式参数的关系来求解结构式参数，称为“间接最小二乘法”(Indirect Least Square, 简记 ILS)。

在恰好识别的情况下，ILS 是一致的，但却不是最有效率的。

在过度识别的情况下，无法使用 ILS。

3. 二阶段最小二乘法

在结构方程可识别的情况下，其排斥的外生变量个数大于或等于包含的内生解释变量个数，而所有排斥的外生变量都是有效工具变量，故可以用工具变量法来估计。

如果结构方程的扰动项满足同方差、无自相关的古典假定，则(2SLS)是最有效率的工具变量法，也是最常见的单一方程估计法。

4. 广义矩估计法

在过度识别的情况下，如果结构方程的扰动项存在异方差或自相关，则 GMM 比 2SLS 更有效率。

5. 有限信息最大似然估计法

假定结构方程的扰动项服从正态分布，可使用 MLE 对单一方程进行估计，称为“有限信息最大似然估计法”(Limited Information Maximum Likelihood Estimation，简称 LIML)。

LIML 与 2SLS 在大样本下是渐近等价的。

如果存在弱工具变量，LIML 比 2SLS 更稳健。

24.4 三阶段最小二乘法

最常见的系统估计法为“三阶段最小二乘法”(Three Stage Least Square, 简记 3SLS)。

在某种意义上, 3SLS 将 2SLS 与 SUR 相结合。

3SLS 的基本步骤如下。

前两步: 对每个方程进行 2SLS 估计。

第三步: 根据前两步的估计, 得到对整个系统的扰动项之协方差矩阵的估计。然后, 据此对整个系统进行 GLS 估计(类似于 SUR 的做法)。具体操作如下。

联立方程模型的第 j 个方程可写为(忽略不在方程中的内生变量与外生变量):

$$\underbrace{\mathbf{y}_j}_{T \times 1} = \underbrace{\mathbf{Y}_j}_{T \times M_j} \underbrace{\boldsymbol{\gamma}_j}_{M_j \times 1} + \underbrace{\mathbf{X}_j}_{T \times K_j} \underbrace{\boldsymbol{\beta}_j}_{K_j \times 1} + \underbrace{\boldsymbol{\varepsilon}_j}_{T \times 1} \equiv \underbrace{\mathbf{Z}_j}_{T \times (M_j + K_j)} \underbrace{\boldsymbol{\delta}_j}_{(M_j + K_j) \times 1} + \underbrace{\boldsymbol{\varepsilon}_j}_{T \times 1} \quad (j = 1, \dots, M)$$

其中, $\mathbf{Z}_j \equiv (\mathbf{Y}_j \ \mathbf{X}_j)$, $\boldsymbol{\delta}_j \equiv \begin{pmatrix} \boldsymbol{\gamma}_j \\ \boldsymbol{\beta}_j \end{pmatrix}$

将所有 M 个方程叠放在一起可得

$$\mathbf{y} \equiv \begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_M \end{pmatrix}_{MT \times 1} = \begin{pmatrix} \mathbf{Z}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Z}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Z}_M \end{pmatrix} \begin{pmatrix} \boldsymbol{\delta}_1 \\ \boldsymbol{\delta}_2 \\ \vdots \\ \boldsymbol{\delta}_M \end{pmatrix} + \begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_M \end{pmatrix}_{MT \times 1} \equiv \mathbf{Z}\boldsymbol{\delta} + \boldsymbol{\varepsilon}$$

假设 $E(\boldsymbol{\varepsilon} | X) = \mathbf{0}$, $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' | X) = \boldsymbol{\Sigma} \otimes \mathbf{I}$, 其中 X 包含整个方程系统中所有的外生变量(都可作为工具变量)。

记 $\hat{\mathbf{Z}}_j \equiv \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}_j$ 为第 j 个方程解释变量 \mathbf{Z}_j 对所有外生变量(工具变量) X 进行回归的拟合值(第一阶段回归), 则第 j 个方程的 2SLS 估计量为

$$\hat{\boldsymbol{\delta}}_{j, 2SLS} \equiv (\hat{\mathbf{Z}}_j' \hat{\mathbf{Z}}_j)^{-1} \hat{\mathbf{Z}}_j' \mathbf{y}_j$$

$$\text{定义 } \hat{\mathbf{Z}} \equiv \begin{pmatrix} \hat{\mathbf{Z}}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{Z}}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \hat{\mathbf{Z}}_M \end{pmatrix},$$

可将所有方程的单一方程 2SLS 估计量写在一起:

$$\hat{\boldsymbol{\delta}}_{2\text{SLS}} \equiv \begin{pmatrix} \hat{\boldsymbol{\delta}}_{1, 2\text{SLS}} \\ \hat{\boldsymbol{\delta}}_{2, 2\text{SLS}} \\ \vdots \\ \hat{\boldsymbol{\delta}}_{M, 2\text{SLS}} \end{pmatrix} = (\hat{\mathbf{Z}}' \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}' \mathbf{y}$$

为进行 3SLS 估计，须先得到对协方差矩阵 Σ 的估计值 $\hat{\Sigma}$ 。

记矩阵 $\hat{\Sigma}$ 的 (i, j) 元素为 $\hat{\sigma}_{ij}$ ，利用单一方程 2SLS 估计的残差可得

$$\hat{\sigma}_{ij} = \frac{1}{T} (\mathbf{y}_i - \mathbf{Z}_i \hat{\boldsymbol{\delta}}_{i, 2\text{SLS}})' (\mathbf{y}_j - \mathbf{Z}_j \hat{\boldsymbol{\delta}}_{j, 2\text{SLS}})$$

类比 SUR，可定义 3SLS 估计量为

$$\hat{\boldsymbol{\delta}}_{3\text{SLS}} = [\hat{\mathbf{Z}}' (\hat{\Sigma}^{-1} \otimes \mathbf{I}) \hat{\mathbf{Z}}]^{-1} \hat{\mathbf{Z}}' (\hat{\Sigma}^{-1} \otimes \mathbf{I}) \mathbf{y}$$

对于 3SLS，也可进行迭代，即用 3SLS 的残差重新估计协方差矩阵 Σ ，然后再使用 GLS，如此反复，直至收敛。

24.5 三阶段最小二乘法的 Stata 实例

24.6 结构 VAR

Vector Autoregression

Sims (1980)提出 VAR 模型，但简化式 VAR 的脉冲响应函数依赖于变量次序，而且无法揭示经济结构(变量之间没有当期影响)。

经济学家又试图将结构重新纳入 VAR 模型中，允许变量之间存在当期影响，形成“结构 VAR”的方法。

考虑如下二元动态联立方程组(忽略常数项):

$$\begin{cases} y_{1t} = -a_{12}y_{2t} + \gamma_{11}y_{1,t-1} + \gamma_{12}y_{2,t-1} + \varepsilon_{1t} \\ y_{2t} = -a_{21}y_{1t} + \gamma_{21}y_{1,t-1} + \gamma_{22}y_{2,t-1} + \varepsilon_{2t} \end{cases}$$

其中，扰动项的分布满足

$$\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix} \sim i.i.d. \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_1^2 & 0 \\ 0 & \sigma_2^2 \end{pmatrix} \right]$$

此方程组的显著特征是在方程右边的解释变量中包含了当期变量，即 y_{1t} 的解释变量包括 y_{2t} ，而 y_{2t} 的解释变量也包括 y_{1t} 。

一般认为，方程组来自于经济理论对于经济结构的建模，故称为“结构 VAR”(Structural VAR，简记 SVAR)。

假设结构方程的扰动项 ε_{1t} 与 ε_{2t} 相互独立，称为“结构新息”(structural innovation)。

例 y_{1t} 为去势(detrended)的实际 GDP 对数， y_{2t} 为去势的名义货币供给对数；则结构新息的假设意味着，对产出的意外冲击(unexpected shocks to output)与对货币供给的意外冲击不相关。

例 y_{1t} 为实际 GDP 增长率， y_{2t} 为失业率；则 ε_{1t} 与 ε_{2t} 可分别解释为需求冲击(demand shock)与供给冲击(supply shock)，而需求冲击(例如消费者偏好变化)与供给冲击(例如石油价格波动)不相关(Blanchard and Quah, 1989)。

将方程组写为矩阵形式：

$$\underbrace{\begin{pmatrix} 1 & a_{12} \\ a_{21} & 1 \end{pmatrix}}_A \underbrace{\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} \gamma_{11} & \gamma_{12} \\ \gamma_{21} & \gamma_{22} \end{pmatrix}}_{\boldsymbol{\Gamma}_1} \underbrace{\begin{pmatrix} y_{1,t-1} \\ y_{2,t-1} \end{pmatrix}}_{\boldsymbol{y}_{t-1}} + \underbrace{\begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{pmatrix}}_{\boldsymbol{\varepsilon}_t}$$

上式可写为

$$\boldsymbol{A}\boldsymbol{y}_t = \boldsymbol{\Gamma}_1\boldsymbol{y}_{t-1} + \boldsymbol{\varepsilon}_t$$

其中，矩阵 \boldsymbol{A} 反映了 y_{1t} 与 y_{2t} 的当期互动，即内生性。

假设矩阵 \boldsymbol{A} 非退化，在方程两边同时左乘 \boldsymbol{A}^{-1} ，可得简化式 VAR(reduced-form VAR)：

$$\mathbf{y}_t = \mathbf{A}^{-1} \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \mathbf{A}^{-1} \boldsymbol{\varepsilon}_t$$

其中，简化式 VAR 的扰动项 $\mathbf{u}_t \equiv \mathbf{A}^{-1} \boldsymbol{\varepsilon}_t$ 的协方差矩阵为

$$\text{Var}(\mathbf{u}_t) = \text{Var}(\mathbf{A}^{-1} \boldsymbol{\varepsilon}_t) = \mathbf{A}^{-1} \text{Var}(\boldsymbol{\varepsilon}_t) \mathbf{A}^{-1'}$$

其中， $\text{Var}(\boldsymbol{\varepsilon}_t)$ 为对角矩阵； $\text{Var}(\mathbf{u}_t)$ 不是对角矩阵，包含 3 个参数。

方程可识别的必要条件(阶条件)是，结构 VAR 的待估参数个数小于或等于简化 VAR 的待估参数个数。

在本例中, SVAR 的待估参数为 8 个(6 个系数, 2 个方差), 而 VAR 的待估参数为 7 个(4 个系数, 3 个协方差)。

为了识别此 SVAR, 至少需要对方程施加一个约束, 比如 $a_{12} = 0$ (意味着 y_{2t} 对 y_{1t} 无直接影响)。

考虑一般形式的 SVAR。从 p 阶简化 VAR 出发:

$$\mathbf{y}_t = \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \mathbf{y}_{t-p} + \mathbf{u}_t$$

其中, \mathbf{y}_t 为 $M \times 1$ 向量; \mathbf{u}_t 为简化式扰动项, 允许存在同期相关 (contemporaneous correlation)。

在方程两边同时左乘某非退化矩阵 A :

$$A\mathbf{y}_t = A\Gamma_1\mathbf{y}_{t-1} + \cdots + A\Gamma_p\mathbf{y}_{t-p} + A\mathbf{u}_t$$

经移项整理可得:

$$A(\mathbf{I} - \Gamma_1 L - \cdots - \Gamma_p L^p)\mathbf{y}_t = A\mathbf{u}_t$$

我们希望 SVAR 的扰动项正交。

一种简单的作法为令 $A\mathbf{u}_t = \boldsymbol{\varepsilon}_t$ ，其中 $\boldsymbol{\varepsilon}_t$ 为 SVAR 的结构扰动项，不存在同期相关。

但此假定可能过强(矩阵 A 来自经济理论对经济结构的建模，未必能使 $A\mathbf{u}_t$ 同期不相关)。

一般地，假设 $A\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$ ，其中 \mathbf{B} 为 $M \times M$ 矩阵；则方程可写为

$$A(\mathbf{I} - \Gamma_1 L - \cdots - \Gamma_p L^p) \mathbf{y}_t = A\mathbf{u}_t = \mathbf{B}\boldsymbol{\varepsilon}_t$$

结构扰动项 $\boldsymbol{\varepsilon}_t$ 的协方差矩阵被标准化为单位矩阵 \mathbf{I}_M 。

此方程称为 SVAR 的“AB 模型”(AB-Model)(Amisano and Giannini, 1997)。

对于传统的联立方程模型，分析的重点在于解释变量的边际效应，故一般不要求结构扰动项正交。

对于 AB 模型，分析的重点在于正文化冲击的效应，故一般假设结构扰动项 $\boldsymbol{\varepsilon}_t$ 正交。

如果令 $A = \mathbf{I}_M$ ，则为 B 模型。如果令 $B = \mathbf{I}_M$ ，则为 A 模型。A 模型与 B 模型都是 AB 模型的特例。

在方程两边同时左乘 A^{-1} ，可得简化 VAR:

$$\mathbf{y}_t = \boldsymbol{\Gamma}_1 \mathbf{y}_{t-1} + \cdots + \boldsymbol{\Gamma}_p \mathbf{y}_{t-p} + \underbrace{\mathbf{A}^{-1} \mathbf{B} \boldsymbol{\varepsilon}_t}_{\mathbf{u}_t}$$

由于 $\mathbf{u}_t = \mathbf{A}^{-1} \mathbf{B} \boldsymbol{\varepsilon}_t$, 故简化式扰动项 \mathbf{u}_t 的协方差矩阵为

$$\text{Var}(\mathbf{u}_t) = \mathbf{A}^{-1} \mathbf{B} \mathbf{B}' \mathbf{A}^{-1'}$$

结构 VAR 模型的待估参数总数为“ M^2 (\mathbf{A} 的参数个数) + M^2 (\mathbf{B} 的参数个数) + pM^2 ($\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_p$ 的参数个数)”, 即 $2M^2 + pM^2$ 。

简化 VAR 模型的待估参数总数为“ $M(M+1)/2$ ($\text{Var}(\mathbf{u}_t)$ 的参数个数) + pM^2 ($\boldsymbol{\Gamma}_1, \dots, \boldsymbol{\Gamma}_p$ 的参数个数)”, 即 $[M(M+1)/2] + pM^2$ 。

一般地, SVAR 的参数比 VAR 的参数多 $[2M^2 - M(M+1)/2]$ 个。

为识别 AB 模型, 需对矩阵 \mathbf{A} 与 \mathbf{B} 施加 $[2M^2 - M(M+1)/2]$ 个约束。

即使将矩阵 A 的主对角线元素都标准化为 1，还需附加 $[2M^2 - M - M(M + 1)/2]$ 个约束条件。

如果正好施加如此多约束，为恰好识别；

如施加更多约束，为过度识别。

此阶条件(order condition)为识别 AB 模型的必要条件。

为估计 SVAR 模型，一般假设结构扰动项 ε_t 服从多维正态分布，即 $\varepsilon_t \sim N(\boldsymbol{\theta}, \mathbf{I}_M)$ ，然后进行带约束条件的 MLE。

虽然此 MLE 估计量在多维正态的假设下导出，但在更弱的条件下，QMLE 估计量依然一致。

一般来说，应从经济理论或对简化式 VAR 的估计结果出发，来设置约束条件。

较常用的方法沿用乔利斯基分解的思路，将矩阵 A 设为下三角矩阵且主对角线元素全部为 1，并将矩阵 B 设为对角矩阵，称为“乔利斯基约束”(Cholesky restrictions)。

以 $M = 3$ 为例，约束条件可写为

$$A = \begin{pmatrix} 1 & 0 & 0 \\ \cdot & 1 & 0 \\ \cdot & \cdot & 1 \end{pmatrix}, \quad B = \begin{pmatrix} \cdot & 0 & 0 \\ 0 & \cdot & 0 \\ 0 & 0 & \cdot \end{pmatrix}$$

其中，缺失值 “.” 表示自由参数(即没有约束)。

从矩阵 A 的第一行可以看出， y_{2t} 与 y_{3t} 对 y_{1t} 无直接影响。

从矩阵 A 的第二行可看出， y_{1t} 对 y_{2t} 有直接影响，但 y_{3t} 对 y_{2t} 无直接影响。

从矩阵 A 的第三行可看出， y_{1t} 与 y_{2t} 对 y_{3t} 都有直接影响。

使用乔利斯基约束来识别 SVAR，其估计结果依赖于变量次序。

对于所选择的特定变量次序，需要从理论上进行说明；或进行敏感度分析，即变换变量次序，并对比结果。

针对矩阵 A 与 B 所施加的约束也称为“短期约束”(short-run restrictions)，其 SVAR 模型称为“短期 SVAR”(short-run SVAR)。

另一类约束为“长期约束”(long-run restrictions)，即对结构冲击 ε_t 对于 y_t 的长期效应进行约束，由 Blanchard and Quah (1989) 所首倡；其 SVAR 模型称为“长期 SVAR”(long-run SVAR)。

例 根据货币中性假说(money neutrality hypothesis)，货币在长期内是中性的，即货币供给在长期内对于实际产出的累积影响为零。

从简化式 VAR 出发，可推导出 SVAR 模型的脉冲响应函数。

根据与第 20 章类似的推导，结构冲击 ε_t 对 y_t 的长期效应为

$$\mathbf{C} \equiv (\mathbf{I} - \boldsymbol{\Gamma}_1 - \cdots - \boldsymbol{\Gamma}_p)^{-1} \mathbf{A}^{-1} \mathbf{B}$$

在长期内，可将 SVAR 模型简洁地写为

$$y_t = \mathbf{C} \varepsilon_t$$

例 假设 $M = 2$ ，第一个变量为实际 GDP，第二个变量为货币供给；则可约束长期效应矩阵 \mathbf{C} 的 $(1, 2)$ 元素为 0，即对第二个变量(货币供给)的结构冲击在长期内对第一个变量(实际 GDP)无作用。