

第 23 章 似不相关回归

23.1 单一方程估计与系统估计

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。

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

$$\mathbf{y} \equiv \underbrace{\begin{pmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_n \end{pmatrix}}_{nT \times 1} = \underbrace{\begin{pmatrix} \mathbf{X}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{X}_n \end{pmatrix}}_{nT \times \sum_{i=1}^n K_i} \underbrace{\begin{pmatrix} \boldsymbol{\beta}_1 \\ \boldsymbol{\beta}_2 \\ \vdots \\ \boldsymbol{\beta}_n \end{pmatrix}}_{\sum_{i=1}^n K_i \times 1} + \underbrace{\begin{pmatrix} \boldsymbol{\varepsilon}_1 \\ \boldsymbol{\varepsilon}_2 \\ \vdots \\ \boldsymbol{\varepsilon}_n \end{pmatrix}}_{nT \times 1} \equiv \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

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

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

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

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

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 $\boldsymbol{\Omega}$ 中的第 (i, j) 个矩阵($i \neq j$)为

$$E(\boldsymbol{\varepsilon}_i\boldsymbol{\varepsilon}_j') = \sigma_{ij}\boldsymbol{I}_T$$

综合以上结果可知

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

可否把 \mathbf{I}_T 从右边提取出来？

定义 对于任意两个矩阵 $\mathbf{A}_{m \times n} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ 与 $\mathbf{B}_{p \times q}$ (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

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

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。

Ω 的逆矩阵可以写为

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

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：

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

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 $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n\}$ 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

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

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

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

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

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

23.4 SUR 的假设检验

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

由于 β 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“ $H_0: R\beta = r$ ”，则可将“ $R\beta = r$ ”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)。