系统估计与面板数据

司继春

上海对外经贸大学统计与信息学院

在之前一章中，我们学习了在外生性条件下单方程的最小二乘估计问题。在这一节中，我们将主要介绍最小二乘法在多方程条件下的推广。而在这其中，整群抽样以及面板数据是最经常使用的两个例子，我们将重点介绍这些模型。

1 系统最小二乘

在此之前我们学习了单方程的线性回归问题，如果记 y_i 为被解释变量，x_i 为解释变量向量，假设两者存在如下线性关系：

$$y_i = x_i'\beta + u_i$$

那么在外生性条件 $\mathbb{E}(u_i|x_i) = 0$ 的条件下，可以使用普通最小二乘（OLS）对参数 β 进行估计。

然而在应用中，我们可能会碰到对于每个个体，同时存在不止一个方程需要估计的情况。其中，**似不相关回归**（**seemingly unrelated regression**）和面板数据模型是最经常遇到的两类需要同时估计多个方程的模型。

例 1. （似不相关回归）如果我们关心家庭的消费支出问题，对于每一个家庭 i，我们可以观察到其不同种类的支出占收入的比例，比如我们考虑食品支出比例（y_i^F）、医疗支出比例（y_i^M）以及教育支出比例（y_i^E），每一类支出都是由不同的价格水平及收入决定的：

$$y_i^j = \beta_j + \beta_j^F p_F + \beta_j^M p_M + \beta_j^E p_E + \beta_j^I I + u_i^j = x_i'\beta^j + u_i^j, j = F, M, E \quad (1)$$

其中 x_i 为 $K \times 1$ 维的由价格水平 p_j 和收入 I 所组成的解释向量。我们可以将以上模型用向量的形式表达，记：

$$Y_i = \begin{bmatrix} y_i^F \\ y_i^M \\ y_i^E \end{bmatrix}$$

以及：

$$X_i = \begin{bmatrix} x_i' & 0 & 0 \\ 0 & x_i' & 0 \\ 0 & 0 & x_i' \end{bmatrix}_{3\times5}, \beta = \begin{bmatrix} \beta^F \\ \beta^M \\ \beta^E \end{bmatrix}_{15\times1}, U_i = \begin{bmatrix} u_i^F \\ u_i^M \\ u_i^E \end{bmatrix}$$

从而我们可以将式 (1) 联合写为：

$$Y_i = X_i\beta + U_i = \begin{bmatrix} x_i' & 0 & 0 \\ 0 & x_i' & 0 \\ 0 & 0 & x_i' \end{bmatrix} \begin{bmatrix} \beta^F \\ \beta^M \\ \beta^E \end{bmatrix} + \begin{bmatrix} u_i^F \\ u_i^M \\ u_i^E \end{bmatrix}$$

在这里，对于每一个个体 i，我们都有三个方程需要同时进行估计。

例 2. （面板数据）如果我们可以对同一个个体在不同时间点上重复观测多次，我们就得到了一个面板数据。令 $t = 1, 2, ..., T$ 代表时间，且每期都有回归方程：

$$y_{it} = x_{it}'\beta + u_{it} \tag{2}$$

我们同样可以将以上方程使用向量表示。令：

$$Y_i = \begin{bmatrix} y_{i1} \\ y_{i2} \\ \vdots \\ y_{iT} \end{bmatrix}_{T\times1}$$

以及：

$$X_i = \begin{bmatrix} x_{i1}' \\ x_{i2}' \\ \vdots \\ x_{iT}' \end{bmatrix}_{T\times K}, U_i = \begin{bmatrix} u_{i1}' \\ u_{i2}' \\ \vdots \\ u_{iT}' \end{bmatrix}_{T\times1}$$

那么对于每一个个体 i，我们都有 T 个方程，此时方程 (2) 可以表示为：

$$Y_i = X_i\beta + U_i = \begin{bmatrix} x_{i1}' \\ x_{i2}' \\ \vdots \\ x_{iT}' \end{bmatrix} \beta + \begin{bmatrix} u_{i1} \\ u_{i2} \\ \vdots \\ u_{iT} \end{bmatrix}$$

注意在这里我们假设了在每一期，x_{it} 对 y_{it} 的影响都是相同的，其系数都是 β。

在以上两个例子中，我们所面临的问题都是相似的，对于每一个个体，我们都有 J 个方程需要同时进行估计：在例 (1) 中，我们需要同时估计三个产品的

需求，而在例 (2) 中，我们需要同时估计同一个个体 T 期的方程。在每种情况下，我们都可以将问题转化为以下形式：

$$Y_i = X_i\beta + U_i \tag{3}$$

与一元线性回归不同的是，这里对于每一个个体，Y_i 和 U_i 都是 $J\times 1$ 的向量，而 X_i 为 $J\times K$ 的矩阵。

实际上，如果我们假设 $\mathbb{E}(u_{ij}|x_{ij})=0$，那么我们可以对第 j 个方程做普通最小二乘估计，就可以得到对于 β_j 的一致估计。然而很多时候，同时估计 J 个方程可以带来额外的好处。比如，某些时候我们希望对跨方程的约束进行检验，例如在例 (1) 中，我们可能需要对 $H_0:\beta_E^F=\beta_F^E$ 进行假设检验，如果将三个方程单独估计，那么我们无法获得两个系数 β_E^F,β_F^E 的协方差，因而无法完成假设检验。或者，在例 (2) 中，我们假设了每一期 x_{it} 对 y_{it} 的影响都是相同的，如果单独每一期进行回归，那么我们会得到每一期都有不同的 β_t，而如果联合估计多期的方程，可以潜在的提高估计的有效性。

对于方程 (3)，在 $\mathbb{E}(u_{ij}|x_{ij})=0$ 满足的条件下，我们有：$\mathbb{E}(u_{ij}x_i)=0$。如果对于每一个 $j=1,...,J$，该条件都成立，那么我们可以将以上条件写为：

$$\mathbb{E}(X_i'U_i)=\mathbb{E}[X_i\text{'}(Y_i-X_i\beta)]=0$$

从而得到：

$$\beta=[\mathbb{E}(X_i'X_i)]^{-1}[\mathbb{E}(X_i'Y_i)]$$

使用样本矩代替总体矩，我们便得到了 β 的矩估计：

$$\hat{\beta}=\left[\frac{1}{N}\sum_{i=1}^{N}(X_i'X_i)\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}(X_i'Y_i)\right]=\left[\sum_{i=1}^{N}(X_i'X_i)\right]^{-1}\left[\sum_{i=1}^{N}(X_i'Y_i)\right]$$

以上估计称为**系统普通最小二乘估计**(**system ordinary least squares**, SOLS)。

例 3. 在例 (1) 中，带入以上公式，我们得到：

$$\sum_{i=1}^{N}(X_i'X_i)=\sum_{i=1}^{N}\begin{bmatrix} x_i & 0 & 0\\ 0 & x_i & 0\\ 0 & 0 & x_i\end{bmatrix}\begin{bmatrix} x_i' & 0 & 0\\ 0 & x_i' & 0\\ 0 & 0 & x_i'\end{bmatrix}=\sum_{i=1}^{N}\begin{bmatrix} x_ix_i' & 0 & 0\\ 0 & x_ix_i' & 0\\ 0 & 0 & x_ix_i'\end{bmatrix}$$

同时：

$$\sum_{i=1}^{N}(X_i'Y_i)=\sum_{i=1}^{N}\begin{bmatrix} x_i & 0 & 0\\ 0 & x_i & 0\\ 0 & 0 & x_i\end{bmatrix}\begin{bmatrix} y_i^F\\ y_i^M\\ y_i^E\end{bmatrix}=\sum_{i=1}^{N}\begin{bmatrix} x_iy_i^F & 0 & 0\\ 0 & x_iy_i^M & 0\\ 0 & 0 & x_iy_i^E\end{bmatrix}$$

从而该模型的 SOLS 估计量为：

$$\begin{bmatrix} \hat{\beta}^F \\ \hat{\beta}^M \\ \hat{\beta}^E \end{bmatrix} = \left(\sum_{i=1}^{N} \begin{bmatrix} x_i x_i' & 0 & 0 \\ 0 & x_i x_i' & 0 \\ 0 & 0 & x_i x_i' \end{bmatrix}\right)^{-1} \left(\sum_{i=1}^{N} \begin{bmatrix} x_i y_i^F & 0 & 0 \\ 0 & x_i y_i^M & 0 \\ 0 & 0 & x_i y_i^E \end{bmatrix}\right)$$

$$= \begin{bmatrix} \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i^F \\ \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i^M \\ \left(\sum_{i=1}^{N} x_i x_i'\right)^{-1} \sum_{i=1}^{N} x_i y_i^E \end{bmatrix} = \begin{bmatrix} \beta_{OLS}^F \\ \beta_{OLS}^M \\ \beta_{OLS}^E \end{bmatrix}$$

因而在该模型中，SOLS 估计与单方程分别做最小二乘估计是等价的。

例 4. 在例 (2) 中，可以计算，SOLS 估计量为：

$$\hat{\beta} = \left[\sum_{i=1}^{N} (X_i' X_i)\right]^{-1} \left[\sum_{i=1}^{N} (X_i' Y_i)\right] = \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} x_{it}')\right]^{-1} \left[\sum_{i=1}^{N} \sum_{t=1}^{T} (x_{it} y_{it})\right]$$

实际上，以上估计等价于将所有的数据放在一起做 OLS，因而以上估计也被成为混合 OLS 估计（pooled OLS）。

我们可以进一步建立 SOLS 的大样本性质。首先，将式 (3) 带入，得到：

$$\hat{\beta} = \left[\frac{1}{N} \sum_{i=1}^{N} (X_i' X_i)\right]^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} (X_i' Y_i)\right] = \beta + \left[\frac{1}{N} \sum_{i=1}^{N} (X_i' X_i)\right]^{-1} \left[\frac{1}{N} \sum_{i=1}^{N} (X_i' U_i)\right]$$

对于一致性，根据大数定律，在一定条件下，有：

$$\frac{1}{N} \sum_{i=1}^{N} (X_i' X_i) \xrightarrow{p} \mathbb{E}(X_i' X_i) \triangleq A$$

以及：

$$\frac{1}{N} \sum_{i=1}^{N} (X_i' U_i) = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{J} x_{ij} u_{ij} \xrightarrow{p} \mathbb{E} \sum_{j=1}^{J} x_{ij} u_{ij} = 0$$

从而得到：$\hat{\beta} \xrightarrow{p} \beta$，即 SOLS 是一致估计量。

而对于渐进分布，由于：

$$\mathrm{Var}(X_i' U_i) = \mathbb{E}(X_i' U_i U_i' X_i) \triangleq B$$

根据中心极限定理，在一定的条件下，有：

$$\frac{1}{\sqrt{N}} \sum_{i=1}^{N} (X_i' U_i) \overset{a}{\sim} N(0, B)$$

因而：

$$\sqrt{N}\left(\hat{\beta}-\beta\right)\overset{a}{\sim}N\left(0,A^{-1}BA^{-1}\right)$$

以上即 SOLS 的大样本分布。

在实践中，我们需要对 A 和 B 两个矩阵进行估计。对于矩阵 A，可以使用 $\frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)$ 直接进行估计，而对于矩阵 B，可以令残差向量：

$$\hat{U}_i=Y_i-X_i'\hat{\beta}$$

从而可以矩阵 B 可以使用 $\frac{1}{N}\sum_{i=1}^{N}\left(X_i'\hat{U}_i\hat{U}_i'X_i\right)$ 进行估计。最终，$\hat{\beta}$ 的渐进方差可以使用：

$$\begin{aligned}\widehat{\mathrm{Var}\left(\hat{\beta}\right)}&=\frac{1}{N}\left[\frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)\right]^{-1}\left[\frac{1}{N}\sum_{i=1}^{N}\left(X_i'\hat{U}_i\hat{U}_i'X_i\right)\right]\left[\frac{1}{N}\sum_{i=1}^{N}\left(X_i'X_i\right)\right]^{-1}\\&=\left[\sum_{i=1}^{N}\left(X_i'X_i\right)\right]^{-1}\left[\sum_{i=1}^{N}\left(X_i'\hat{U}_i\hat{U}_i'X_i\right)\right]\left[\sum_{i=1}^{N}\left(X_i'X_i\right)\right]^{-1}\end{aligned}$$

进行估计。

在有了渐进正态性之后，对于任意的关于 β 的原假设，都可以使用 Wald 检验以及 delta 方法构建检验统计量。

2 广义最小二乘

以上我们构建了多方程条件下的系统最小二乘估计。然而注意在 SOLS 估计的方差中，存在着 $\mathbb{E}\left(X_i'U_iU_i'X_i\right)$ 这一项，即方程之间可能存在着相关性，导致我们所得到的 SOLS 估计并不是最有效的估计。实际上，如果我们加强一些假定，可以得到更加有效的估计量。

记 U_i 的协方差矩阵为 $\Omega=\mathbb{E}\left(U_iU_i'\right)$，该矩阵度量了不同方程之间 u_{ij} 的相关性。比如在例 (1) 中，不同之处项目的误差项之间可能存在着相关性，而在例 (2) 中，不同时间的误差项 u_{it} 也可能存在着相关性（这种时间上的相关性我们称之为**自相关**，**autocorrelation**）。实际上，这种相关性的存在使得 SOLS 的估计不是最有效的估计。

为了消除这种相关性，我们可以使用 $\Omega^{-1/2}$ 左乘 U_i，即 $U_i^*=\Omega^{-1/2}U_i$，现在我们有：

$$\mathrm{Var}\left(U_i^*\right)=\mathbb{E}\left(\Omega^{-1/2}U_iU_i'\Omega^{-1/2}\right)=\Omega^{-1/2}\mathbb{E}\left(U_iU_i'\right)\Omega^{-1/2}=\Omega^{-1/2}\Omega\Omega^{-1/2}=I$$

从而得到的新的 U_i^* 的分量之间不再存在着相关性。

进一步，我们可以使用 $\Omega^{-1/2}$ 左乘式 (3)，得到：

$$\Omega^{-1/2}Y_i=\Omega^{-1/2}X_i'\beta+\Omega^{-1/2}U_i$$

令 $Y_i^* = \Omega^{-1/2} Y_i$，$X_i^* = \Omega^{-1/2} X_i$，我们得到了新的回归方程：

$$Y_i^* = X_i^{*'} \beta + U_i^*$$

在该方程中，误差项 U_i^* 不再存在着自相关性，因而使用以上回归方程做 OLS，就可以得到有效的估计，即：

$$\hat{\beta}^* = \left[\sum_{i=1}^{N} \left(X_i^{*'} X_i^*\right)\right]^{-1} \left[\sum_{i=1}^{N} \left(X_i^{*'} Y_i^*\right)\right] = \left[\sum_{i=1}^{N} \left(X_i' \Omega^{-1} X_i\right)\right]^{-1} \left[\sum_{i=1}^{N} \left(X_i' \Omega^{-1} Y_i\right)\right]$$

以上方法我们称之为**广义最小二乘**（**generalized least squares**, GLS）。

然而注意到，由于：

$$\hat{\beta}^* = \left[\sum_{i=1}^{N} \left(X_i' \Omega^{-1} X_i\right)\right]^{-1} \left[\sum_{i=1}^{N} \left(X_i' \Omega^{-1} Y_i\right)\right] = \beta + \left[\sum_{i=1}^{N} \left(X_i' \Omega^{-1} X_i\right)\right]^{-1} \left[\sum_{i=1}^{N} \left(X_i' \Omega^{-1} U_i\right)\right]$$

因而为了得到一致性，我们必须要假设 $\mathbb{E}\left(X_i' \Omega^{-1} U_i\right) = 0$，由于 Ω^{-1} 结构未知，因而实际上我们必须假设

$$\mathbb{E}\left(u_{ij} x_{ih}\right) = 0, h = 1, ..., J$$

即 u_{ij} 必须对所有方程的控制变量 x_{ih} 都是不相关的，而 SOLS 则只需要 u_{ij} 对该方程的 x_{ij} 是不相关的，即只需要假设 $\mathbb{E}\left(u_{ij} x_{ij}\right) = 0$。实际上，该条件的一个充分条件是：

$$\mathbb{E}\left(u_{ij} | x_i\right) = 0$$

即 u_{ij} 对于所有的控制变量 $x_i = \left(x_{i1}', ..., x_{iJ}'\right)'$ 都是外生的。因而 GLS 虽然可以提高有效性，但是代价是必须做更加严格的外生性假设。

然而现实中，由于协方差矩阵 Ω 的结构是未知的，因而我们必须对该协方差矩阵进行估计。由于我们实际上假设了对于不同个体，其误差项 U_i 的协方差矩阵都是相同的，即所有个体的误差项的协方差矩阵都为 $\Omega = \mathbb{E}\left(U_i U_i'\right)$，因而我们使用矩估计的方法对该协方差矩阵进行估计。由于 Ω 是 $J \times J$ 维的实对称矩阵，因而为了估计 Ω，我们必须估计 $J\left(J+1\right)/2$ 个未知参数。如果 J 不随着 N 的增大而增大，那么我们可以获得 Ω 的一致估计[1]。

实际上，我们可以使用：

$$\hat{\Omega} = \frac{1}{N} \sum_{i=1}^{N} \left(\hat{U}_i \hat{U}_i'\right)$$

[1]该方法与单方程的加权最小二乘估计（WLS）是不一样的，在 WLS 中，每个个体的方差都是未知的，因而我们会遇到所谓的伴生参数问题（incidental parameters problem），即随着样本量 N 的增大，参数的维数与 N 同阶或者更高阶的增加，在这种情况下不可能获得这些伴生变量的一致估计。但是在这里，由于 J 不随着 N 的增大而增加，因而并不存在伴生参数问题。

进行估计，而为了获得 $\hat{U}_i$，我们必须首先获得一个参数 β 的一致估计。因而对于方程 (3)，可以首先使用 SOLS 获得一个 β 的一致估计 $\hat{\beta}_{SOLS}$，然后计算残差向量：

$$\hat{\hat{U}}_i = Y_i - X_i'\hat{\beta}_{SOLS}$$

最终得到 Ω 的估计：

$$\hat{\Omega} = \frac{1}{N}\sum_{i=1}^{N}\left(\hat{\hat{U}}_i\hat{\hat{U}}_i'\right)$$

最后使用 $\hat{\Omega}$ 代替 Ω，使用公式：

$$\hat{\beta}_{FGLS} = \left[\sum_{i=1}^{N}\left(X_i'\hat{\Omega}^{-1}X_i\right)\right]^{-1}\left[\sum_{i=1}^{N}\left(X_i'\hat{\Omega}^{-1}Y_i\right)\right]$$

计算广义最小二乘估计量，以上估计两被称为**可行的广义最小二乘**（**feasible generalized least squares**, FGLS）估计量。

习题