

高级计量经济学

Lecture 9: Nonlinear Regression, NLS, and GLS

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室	粤海校区汇文楼1510
E-mail	huangjp@szu.edu.cn
Website	https://huangjp.com

非线性回归模型和 MM估计量

线性模型与非线性模型

Linear and Nonlinear Models

线性回归模型可以表达为

$$y = X\beta + u, \quad u \mid X \sim \text{IID}(\mathbf{0}, \sigma^2 I)$$

或

$$y_t = X_t\beta + u_t, \quad u_t \mid X \sim \text{IID}(0, \sigma^2), \quad t = 1, \dots, n$$

定义信息集 (information set) Ω_t 为所有可能对 y_t 产生影响的变量的集合。一般情况下我们要求 Ω_t 中的变量为外生或者前定变量。

基于信息集 Ω_t ，我们可以将线性模型拓展为下面的非线性模型

$$y_t = x_t(\beta) + u_t, \quad u_t \mid \Omega_t \sim \text{IID}(0, \sigma^2), \quad t = 1, \dots, n \quad \text{因此 } E[u_t \mid \Omega_t] = 0$$

这里 $x_t(\beta)$ 是参数向量 β 的非线性函数。非线性模型也可以写成

$$y = x(\beta) + u, \quad u \sim \text{IID}(\mathbf{0}, \sigma^2 I)$$

注： $x_t(\beta)$ 也可以写成 $f(x_t, \beta)$ 。为了使符号和线性模型相对应，这里采用 $x_t(\beta)$ 的写法。

矩条件

Moment Conditions

在线性模型下，如果假设前定性 $E[u_t | X_t] = 0$ ，我们可以推出 $E[X_t^\top u_t] = \mathbf{0}$ (Lecture 5)，而这恰好是求 MM 估计量的条件 (Lecture 2)。此条件对应的样本条件是

$$X^\top (y - X\beta) = \mathbf{0} \quad \text{左式称为 moment conditions}$$

在非线性模型下，我们可以从信息集 Ω_t 选择 k 个变量并写成 $1 \times k$ 向量 W_t 。此时可得 $E[u_t | W_t] = 0$ ，因此 $E[W_t^\top u_t] = \mathbf{0}$ ，对应的矩条件为

$$W^\top (y - x(\beta)) = \mathbf{0}, \quad W \text{ 的第 } t \text{ 行是 } W_t$$

非线性模型的矩条件中包含 k 个关于系数的非线性方程，其解（若存在）就是 β 的 MM 估计量，可记作 $\hat{\beta}_{\text{MM}}$ 。

关于 MM 估计量

- MM 估计量是 $W^{\top}(y - x(\beta)) = \mathbf{0}$ 的解。因为条件是非线形的，我们无法获得解的一般表达式。
- 矩条件基于 W_t 的前定性 $E[u_t | W_t] = 0$ 而非外生性，因此 MM 估计量可能有偏。
- MM 估计量的性质会随 W 中变量的选择而变化。 W 的选择一般不会影响 MM 估计量的一致性，但会影响它的渐进协方差矩阵。因此我们需要找到使 MM 估计量满足渐进有效性的 W 。

回归系数的渐进识别

Asymptotic Identification

识别 (Identification) 指在确立样本和估计方法后, 可以求出唯一的参数值的情况。如果一种估计方法在 $n \rightarrow \infty$ 时可以求出唯一的参数值, 我们称其为渐进可识别。

在线性模型的 OLS 估计中, 如果 X 列满秩, 则 β 是可识别的 (标准方程有唯一解)。在非线性模型的 MM 估计中, 如果 $W^\top(y - x(\beta)) = \mathbf{0}$ 存在唯一解, 则 β 是可识别的。

MM 估计量的渐进可识别性取决于 $\frac{1}{n}W^\top(y - x(\beta))$ 的概率极限是否可以求出唯一的参数值。令

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n}W^\top(y - x(\beta)) = \alpha(\beta), \quad \alpha(\beta) \text{ 为确定性函数。}$$

已知 $\frac{1}{n}W^\top(y - x(\beta_0)) = \frac{1}{n} \sum_{t=1}^n W_t^\top u_t$, $E[W_t^\top u_t] = \mathbf{0}$, 根据 LLN 可得

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n}W^\top(y - x(\beta_0)) = \mathbf{0}$$

因此, 如果 β_0 是 $\alpha(\beta) = \mathbf{0}$ 的唯一解, MM 估计量是渐进可识别的。

需要注意的是, 可识别性无法推出渐进可识别性, 反之亦然 (pp.217-218)。

MM 估计量的一致性

Consistency of MM Estimators

假设 MM 估计量是渐进可识别的，则

$$\text{plim}_{n \rightarrow \infty} \frac{1}{n} \mathbf{W}^\top (\mathbf{y} - \mathbf{x}(\hat{\boldsymbol{\beta}}_{\text{MM}})) = \boldsymbol{\alpha}(\hat{\boldsymbol{\beta}}_{\text{MM}}) = \mathbf{0}$$

存在唯一解。如果我们假设 $\text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}}_{\text{MM}} = \boldsymbol{\beta}_\infty$ (非随机)，则根据上式可得 $\boldsymbol{\alpha}(\boldsymbol{\beta}_\infty) = \mathbf{0}$ ，但是渐进可识别性告诉我们只有 $\boldsymbol{\beta}_0$ 满足 $\boldsymbol{\alpha}(\boldsymbol{\beta}) = \mathbf{0}$ ，因此

$$\text{plim}_{n \rightarrow \infty} \hat{\boldsymbol{\beta}}_{\text{MM}} = \boldsymbol{\beta}_0$$

即 MM 估计量满足一致性。

MM 估计量的渐进可识别性取决于如何选取 \mathbf{W} 里的变量。

MM 估计量的其他渐进性质

Other Asymptotic Properties of MM Estimators

非线性模型的 MM 估计量还具有以下渐进性质：

- 渐进正态性： $\sqrt{n}(\hat{\beta}_{\text{MM}} - \beta_0)$ 服从均值为 **0** 的渐进正态分布 (pp.220-222)
- 渐进有效性： $\sqrt{n}(\hat{\beta}_{\text{MM}} - \beta_0)$ 的渐进协方差矩阵是

$$\sigma_0^2 \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} X_0^\top W (W^\top W)^{-1} W^\top X_0 \right)^{-1} = \sigma_0^2 \text{plim}_{n \rightarrow \infty} \left(\frac{1}{n} X_0^\top P_W X_0 \right)^{-1}$$

其中 $X_0 = X(\beta_0) = [\partial x_t(\beta) / \partial \beta_i]_{\beta=\beta_0}$ 。

当 $W = X_0$ 时， $\sqrt{n}(\hat{\beta}_{\text{MM}} - \beta_0)$ 在 MM 估计量当中最有效。 (pp.222-223)

需要注意的是， X_0 是未知量 β_0 的函数，因此这样定义的有效估计量不可行 (infeasible) 。

线性模型与 MM 估计量

线性模型 $y = X\beta + u$ 是非线形模型的一个特例，因此非线性模型的 MM 估计量也适用于线性模型。

在线性模型下，MM 估计量是

$$W^{\top}(y - X\beta) = \mathbf{0}$$

的解。根据定义 $X(\beta) = [\partial x_t(\beta) / \partial \beta_i] = X$ ，因此 $X_0 = X$ ，最有效的 MM 估计量满足

$$X^{\top}(y - X\hat{\beta}_{\text{MM}}) = \mathbf{0} \quad \Rightarrow \quad \hat{\beta}_{\text{MM}} = \hat{\beta}_{\text{OLS}}$$

非线性最小二乘法

Nonlinear Least Squares

因为 $X_0 = X(\beta_0) = [\partial x_t(\beta) / \partial \beta_i]_{\beta=\beta_0}$ 是未知参数 β_0 的函数，我们无法通过它求得最有效的 MM 估计量。

假设估计量 $\hat{\beta}$ 是下面 MM 条件的解

$$X^\top(\beta)(y - x(\beta)) = 0$$

同时我们也可以考虑非线性最小二乘法

$$\min_{\beta} (y - x(\beta))^\top (y - x(\beta))$$

上面的 MM 条件等价于非线性最小二乘法的一阶条件。因此，我们将 $\hat{\beta}$ 称为非线性最小二乘估计量 (NLS)，并记为 $\hat{\beta}_{\text{NLS}}$ 。

当 $n \rightarrow \infty$ 时， $\hat{\beta}_{\text{NLS}}$ 收敛于有效 MM 估计量 ($W = X_0$)。

广义最小二乘法

线性模型与广义最小二乘法

OLS 和 NLS 估计量都需要误差项满足同方差性。下面我们放松这个假设，考虑线性模型

$$y = X\beta + u, \quad E[uu^\top] = \Omega$$

一般情况下 $\Omega \neq \sigma^2 I$ ，因此不满足 Gauss-Markov 定理的条件。

为了获得有效估计量，我们可以将原模型变换为满足条件的新模型，并将新模型的 OLS 估计量作为 β 的估计量。这种方法被称为广义最小二乘法（generalized least squares, GLS）。

广义最小二乘估计量

GLS Estimator

已知 Ω 是正定矩阵且可逆，因此存在 $n \times n$ 非奇异矩阵 Ψ 满足

$$\Omega^{-1} = \Psi\Psi^{\top}$$

将回归方程 $y = X\beta + u$ 从左侧乘以 Ψ^{\top} 可得 $\Psi^{\top}y = \Psi^{\top}X\beta + \Psi^{\top}u$ 。

GLS 估计量 $\hat{\beta}_{\text{GLS}}$ 是变换后模型的 OLS 估计量，即

$$\hat{\beta}_{\text{GLS}} = (X^{\top}\Psi\Psi^{\top}X)^{-1}X^{\top}\Psi\Psi^{\top}y = (X^{\top}\Omega^{-1}X)^{-1}X^{\top}\Omega^{-1}y$$

变换后的误差项的协方差矩阵是

$$E[\Psi^{\top}uu^{\top}\Psi] = \Psi^{\top}E[uu^{\top}]\Psi = \Psi^{\top}(\Psi\Psi^{\top})^{-1}\Psi = I$$

由此可得到 GLS 估计量的协方差矩阵

$$\text{Var}(\hat{\beta}_{\text{GLS}}) = 1 \cdot (X^{\top}\Psi\Psi^{\top}X)^{-1} = (X^{\top}\Omega^{-1}X)^{-1}$$

广义最小二乘法

变换后模型 $\Psi^T y = \Psi^T X\beta + \Psi^T u$ 的 OLS 估计量是 SSR 最小化问题的解，该问题可以写成

$$\begin{aligned} & \min_{\beta} (\Psi^T y - \Psi^T X\beta)^T (\Psi^T y - \Psi^T X\beta) \\ &= \min_{\beta} (\Psi^T (y - X\beta))^T (\Psi^T (y - X\beta)) \\ &= \min_{\beta} (y - X\beta)^T \Psi \Psi^T (y - X\beta) \\ &= \min_{\beta} (y - X\beta)^T \Omega^{-1} (y - X\beta) \end{aligned}$$

我们将目标函数 $(y - X\beta)^T \Omega^{-1} (y - X\beta)$ 称为 GLS 准则函数 (GLS criterion function)。

GLS 估计量是 MM 估计量

最小化问题 $\min_{\beta} (y - X\beta)^\top \Omega^{-1} (y - X\beta)$ 的一阶条件是

$$X^\top \Omega^{-1} (y - X\beta) = \mathbf{0}$$

因此, $\hat{\beta}_{\text{GLS}}$ 也可以看作满足矩条件 $X^\top \Omega^{-1} (y - X\hat{\beta}_{\text{GLS}}) = \mathbf{0}$ 的 MM 估计量。一般情况下, 矩条件 $W^\top (y - X\beta) = \mathbf{0}$ 的解可以写成

$$\hat{\beta}_W = (W^\top X)^{-1} W^\top y$$

$\hat{\beta}_{\text{GLS}}$ 是 $W = X\Omega^{-1}$ 时的 $\hat{\beta}_W$ 。

GLS 估计量的统计学性质

GLS 估计量的统计学性质类似于 MM 估计量的统计学性质。

- 当 X 和 W 外生时, 即 $E[u \mid X, W] = \mathbf{0}$ 时, $\hat{\beta}_W$ 非偏。
- 当 W 满足前定性, 即 $E[u_t \mid W_t] = 0$ 时, $\hat{\beta}_W$ 满足一致性。
- $\hat{\beta}_W$ 的协方差矩阵是

$$\begin{aligned}\text{Var}(\hat{\beta}_W) &= E[(\hat{\beta}_W - \beta_0)(\hat{\beta}_W - \beta_0)^\top] \\ &= E[(W^\top X)^{-1} W^\top u u W^\top (X^\top W)^{-1}] \\ &= (W^\top X)^{-1} W^\top \Omega W^\top (X^\top W)^{-1}\end{aligned}$$

已知, $\text{Var}(\hat{\beta}_{\text{GLS}}) = (X^\top \Omega^{-1} X)^{-1}$, 可以通过证明两者之差是半正定矩阵证明 GLS 估计量的有效性。

OLS 估计量也是 MM 估计量, 因此 GLS 估计量至少和 OLS 估计量同样有效, 且在大多数情况下比 OLS 估计量更有效。

GLS 估计量的计算

虽然 GLS 估计量有很好的性质，但是并不容易计算。

$\hat{\beta}_{\text{GLS}} = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y$ ，即使 Ω 已知，我们需要求它的逆矩阵。当 n 很大时， Ω^{-1} 的计算需要占用大量的计算和储存资源（例如 $n = 10,000$ 时，储存 Ω 和 Ω^{-1} 需要 1600Mb 空间）。

如果 Ψ 已知，且我们可以将数据变换为 $\Psi^T y$ 和 $\Psi^T X$ 而不需要储存 Ψ ，就可以通过 OLS 估计 $\hat{\beta}_{\text{GLS}}$ 。

OLS 估计不需要用到误差项的方差，因此我们可以利用这一特性。如果 $\Omega = \sigma^2 \Delta$ ，且 Δ 为已知，则可以用 Δ 替代 Ω 求 $\hat{\beta}_{\text{GLS}}$ 。具体做法是找到满足 $\Delta^{-1} = \Psi \Psi^T$ 的 Ψ ，并对原模型进行变换，再用变换后模型进行 OLS 估计，所获得的估计量就是原模型的 GLS 估计量：

$$(X^T \Delta^{-1} X)^{-1} X^T \Delta^{-1} y = (X^T \Omega^{-1} X)^{-1} X^T \Omega^{-1} y = \hat{\beta}_{\text{GLS}}$$

这时 GLS 估计量的协方差矩阵可以写成 $\sigma^2 (X^T \Delta^{-1} X)^{-1}$ 。如果 σ^2 未知，我们依然要对其进行估计，即 $\widehat{\text{Var}}(\hat{\beta}_{\text{GLS}}) = s^2 (X^T \Delta^{-1} X)^{-1}$ ， s^2 是变换后模型误差项的估计量。

加权最小二乘估计量

Weighted Least Square Estimator

如果误差项是**异方差**但是**无自相关**，那么就可以很容易地计算 GLS 估计量。这时的 $\mathbf{\Omega}$ 是对角矩阵，我们将其第 t 对角要素写为 ω_t^2 ，即 $\mathbf{\Omega} = \text{diag}(\omega_1^2, \dots, \omega_n^2)$ ，则可得

$$\mathbf{\Omega}^{-1} = \text{diag}(\omega_1^{-2}, \dots, \omega_n^{-2}), \quad \mathbf{\Psi} = \text{diag}(\omega_1^{-1}, \dots, \omega_n^{-1})$$

变换后的模型可以写成

$$\omega_t^{-1}y_t = \omega_t^{-1}\mathbf{X}_t\boldsymbol{\beta} + \omega_t^{-1}u_t, \quad t = 1, \dots, n$$

通过对这个模型进行 OLS 估计得到的估计量被称为加权最小二乘估计量（WLS 估计量），因为这可以看作是用权重 ω_t^{-1} 给样本中第 t 观测值进行加权。

因为变换后模型的误差项 $\omega_t^{-1}u_t$ 满足同方差性，WLS 估计可以作为应对异方差性的一种方法。

如何决定 ω_t

WLS 估计的难点是如何决定 ω_t 。这可以分为以下几种情况：

- 通过理论或者对样本数据的检验，我们相信 $E[u_t^2]$ 和某个可观测变量 z_t^2 成正比。此时可以用 z_t^{-1} 作为权重。
- 样本变量是针对不同大小的集合获得的统计数据。
 - 如果变量是集合中的均值，例如不同城市的人均可支配收入，则 u_t 的方差和集合的要素数 N_t 呈反比。此时可用 $N_t^{1/2}$ 作为权重。
 - 如果变量是总和而不是均值，则 u_t 的方差和 N_t 成正比，此时需用 $N_t^{-1/2}$ 作为权重。