



# 高级计量经济学答案

## Solutions for Advanced Econometrics

版本：洪永淼. 高级计量经济学 [M]. 北京：高等教育出版社, 2011.

谁家今夜扁舟子？何处相思明月楼？

## 一般回归分析和模型设定

**2.1** 证明  $\text{var}(Y) = \text{var}(\mathbb{E}[Y|X]) + \text{var}(Y - \mathbb{E}[Y|X])$  并解释这一结果.

**解答** 根据  $\text{var}(Y) = \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2$  及 LIE 可得

$$\begin{aligned}\text{var}(\mathbb{E}[Y|X]) &= \mathbb{E}[(\mathbb{E}[Y|X])^2] - (\mathbb{E}[\mathbb{E}[Y|X]])^2 \\ &= \mathbb{E}[(\mathbb{E}[Y|X])^2] - (\mathbb{E}[Y])^2\end{aligned}\tag{2.1}$$

又根据  $\text{var}(Y|X) = \mathbb{E}[Y^2|X] - (\mathbb{E}[Y|X])^2$  可得

$$\begin{aligned}\mathbb{E}[\text{var}(Y|X)] &= \mathbb{E}[\mathbb{E}[Y^2|X]] - \mathbb{E}[(\mathbb{E}[Y|X])^2] \\ &= \mathbb{E}[Y^2] - \mathbb{E}[(\mathbb{E}[Y|X])^2]\end{aligned}\tag{2.2}$$

将式 (2.1) 和 (2.2) 相加可得 Law of total variance

$$\text{var}(Y) = \text{var}(\mathbb{E}[Y|X]) + \mathbb{E}[\text{var}(Y|X)]$$

因此只需证明  $\text{var}(Y - \mathbb{E}[Y|X]) = \mathbb{E}[\text{var}(Y|X)]$  即可.

根据上述已经证明的结论, 可以得到

$$\text{var}(Y - \mathbb{E}[Y|X]) = \text{var}(\mathbb{E}[Y - \mathbb{E}[Y|X]|X]) + \mathbb{E}[\text{var}(Y - \mathbb{E}[Y|X]|X)]$$

由于  $\mathbb{E}[Y - \mathbb{E}[Y|X]|X] = \mathbb{E}[Y|X] - \mathbb{E}[Y|X] = 0$ , 因此有

$$\text{var}(Y - \mathbb{E}[Y|X]) = \mathbb{E}[\text{var}(Y - \mathbb{E}[Y|X]|X)]$$

现在只需证明  $\text{var}(Y|X) = \text{var}(Y - \mathbb{E}[Y|X]|X)$ , 根据条件方差定义

$$\begin{aligned}\text{var}(Y - \mathbb{E}[Y|X]|X) &= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2|X] - (\mathbb{E}[Y - \mathbb{E}[Y|X]|X])^2 \\ &= \mathbb{E}[(Y - \mathbb{E}[Y|X])^2|X] = \text{var}(Y|X)\end{aligned}$$

即  $\text{var}(Y) = \text{var}(\mathbb{E}[Y|X]) + \text{var}(Y - \mathbb{E}[Y|X])$ .

□

**2.2** 假设  $(X, Y)$  服从二元正态分布, 其联合密度函数为

$$\begin{aligned}f_{XY}(x, y) &= \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \\ &\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]\right\}\end{aligned}$$

其中,  $-1 < \rho < 1$ ,  $-\infty < \mu_1, \mu_2 < \infty$ ,  $0 < \sigma_1, \sigma_2 < \infty$ . 求:

(1)  $\mathbb{E}[Y|X]$ .

(2)  $\text{var}(Y|X)$ .

**解答** (1) 首先定义

$$\begin{aligned}X &= \mu_1 + \sigma_1 Z_1 \\ Y &= \mu_2 + \sigma_2(\rho Z_1 + \sqrt{1-\rho^2} Z_2)\end{aligned}$$



由 Method of Transformations<sup>1</sup>可以得到

$$\begin{aligned} f_{Z_1 Z_2}(z_1, z_2) &= f_{XY}[\mu_1 + \sigma_1 Z_1, \mu_2 + \sigma_2(\rho Z_1 + \sqrt{1 - \rho^2} Z_2)]|J| \\ &= \frac{1}{2\pi} \exp\left[-\frac{1}{2}(z_1^2 + z_2^2)\right] \end{aligned}$$

其中,  $|J|$  为 Jacobian 行列式

$$|J| = \begin{vmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2\sqrt{1 - \rho^2} \end{vmatrix} = \sigma_1\sigma_2\sqrt{1 - \rho^2}$$

因此,  $(Z_1, Z_2)$  服从二元标准正态分布, 故  $Z_1$  和  $Z_2$  独立. 进一步可以得到

$$\begin{aligned} \mathbb{E}[Y|X] &= \mathbb{E}[\mu_2 + \sigma_2(\rho Z_1 + \sqrt{1 - \rho^2} Z_2)|X] \\ &= \mu_2 + \sigma_2(\rho Z_1 + \sqrt{1 - \rho^2} \mathbb{E}[Z_2|X]) \\ &= \mu_2 + \rho\sigma_2 \frac{X - \mu_1}{\sigma_1} \end{aligned}$$

即  $\mathbb{E}[Y|X] = \mu_2 + \rho\sigma_2 \frac{X - \mu_1}{\sigma_1}$ . 这里  $\mathbb{E}[Z_2|X] = \mathbb{E}[Z_2] = 0$  是根据  $Z_1$  和  $Z_2$  的独立性, 以及  $X$  线性于  $Z_1$  得到的.

(2) 利用  $X = \mu_1 + \sigma_1 Z_1$  及  $Y = \mu_2 + \sigma_2(\rho Z_1 + \sqrt{1 - \rho^2} Z_2)$ , 并且  $Z_1$  与  $Z_2$  独立, 同时注意到  $\text{var}(Z_1|X) = 0$  (这是因为一旦给定  $X$ ,  $Z$  也就完全确定了), 因此有

$$\begin{aligned} \text{var}(Y|X) &= \text{var}(\mu_2 + \sigma_2(\rho Z_1 + \sqrt{1 - \rho^2} Z_2)|X) \\ &= \text{var}(\sigma_2\sqrt{1 - \rho^2} Z_2|X) \\ &= \sigma_2^2(1 - \rho^2) \end{aligned}$$

即  $\text{var}(Y|X) = \sigma_2^2(1 - \rho^2)$ .

□

**2.3** 假设  $(Y, X)'$  是一个随机向量, 且  $g_o(X) \equiv \mathbb{E}[Y|X]$  存在, 其中  $X$  是一个  $(k + 1) \times 1$  随机向量. 假设用一个模型  $g(X)$  预测  $Y$ . 一个评价模型  $g(X)$  预测好坏的常用准则是均方误差  $\text{MSE}(g) \equiv \mathbb{E}[Y - g(X)]^2$ .

(1) 证明使均方误差  $\text{MSE}(g) \equiv \mathbb{E}[Y - g(X)]^2$  最小化的最优预测值  $g^*(X)$  是条件均值  $g_o(X)$ , 即  $g^*(X) = g_o(X)$ .

(2) 令  $\varepsilon \equiv Y - g_o(X)$ , 证明  $\mathbb{E}[\varepsilon|X] = 0$  并解释这一结果.

<sup>1</sup> 设  $(X, Y)$  为二元连续随机变量,  $(Z, W) = g(X, Y) = [g_1(X, Y), g_2(X, Y)]$ , 其中  $g: \mathbb{R}^2 \rightarrow \mathbb{R}^2$  是具有连续偏导数的 1-1 映射, 再设  $h = g^{-1}$ , 那么  $(Z, W)$  也是连续随机变量, 其联合概率密度为

$$f_{ZW}(z, w) = f_{XY}[h_1(z, w), h_2(z, w)]|J|$$

其中  $|J|$  为  $h$  的 Jacobian 行列式, 等于

$$|J| = \begin{vmatrix} \frac{\partial h_1}{\partial z} & \frac{\partial h_1}{\partial w} \\ \frac{\partial h_2}{\partial z} & \frac{\partial h_2}{\partial w} \end{vmatrix} = \frac{\partial h_1}{\partial z} \frac{\partial h_2}{\partial w} - \frac{\partial h_2}{\partial z} \frac{\partial h_1}{\partial w}$$

**解答** 先证明 (2), 根据 LIE 易得  $\mathbb{E}[\varepsilon|X] = \mathbb{E}[Y] - \mathbb{E}[g_o(X)|X] = \mathbb{E}[Y] - \mathbb{E}[\mathbb{E}[Y|X]] = 0$ .

进一步有

$$\begin{aligned}\text{MSE}(g) &\equiv \mathbb{E}[Y - g(X)]^2 \\ &= \mathbb{E}[\varepsilon + g_o(X) - g(X)]^2 \\ &= \mathbb{E}[\varepsilon^2] + \mathbb{E}[g_o(X) - g(X)]^2 + 2\mathbb{E}[\varepsilon(g_o(X) - g(X))] \\ &= \mathbb{E}[\varepsilon^2] + \mathbb{E}[g_o(X) - g(X)]^2 \geq \mathbb{E}[Y - g(X)]^2\end{aligned}$$

当且仅当  $g_o(X) = g(X)$  时取等号. 其中, 根据 (2) 的结果和 LIE 可以得到交叉项

$$\begin{aligned}\mathbb{E}[\varepsilon(g_o(X) - g(X))] &= \mathbb{E}[\mathbb{E}[\varepsilon(g_o(X) - g(X))|X]] \\ &= \mathbb{E}[(g_o(X) - g(X))\mathbb{E}[\varepsilon|X]] \\ &= 0\end{aligned}$$

因此, 使  $\text{MSE}(g)$  最小化的最优预测值  $g^*(X) = g_o(X)$ .

□

**2.4 问题 2.3** 中模型  $g(X)$  的集合是所有可测且平方可积的函数. 假设我们现在将  $g(X)$  的选择限制为线性 (或仿射) 模型  $\mathbb{A} = \{g : \mathbb{R}^{k+1} \rightarrow \mathbb{R} \mid g(X) = X'\beta\}$ , 其中  $\beta$  是  $(k+1) \times 1$  参数向量. 通过选择参数  $\beta$  的值可以选定线性函数  $g(X) = X'\beta$ , 不同的  $\beta$  值给出不同的线性函数. 使均方误最小化的最优线性函数定义为  $g^*(X) \equiv X'\beta^*$ , 其中

$$\beta^* \equiv \arg \min_{\beta \in \mathbb{R}^{k+1}} \mathbb{E}[Y - X'\beta]^2$$

称为最优线性近似系数.

(1) 设  $\mathbb{E}[Y^2] < \infty$ , 矩阵  $\mathbb{E}[XX']$  非奇异, 证明

$$\beta^* = (\mathbb{E}[XX'])^{-1}\mathbb{E}[XY]$$

(2) 定义  $u^* \equiv Y - X'\beta^*$ . 证明  $\mathbb{E}[Xu^*] = \mathbf{0}$ , 其中  $\mathbf{0}$  是  $(k+1) \times 1$  向量.

(3) 假设存在某一给定的  $\beta^o$ , 条件均值  $g_o(X) = X'\beta^o$ , 则说线性模型  $g_{\mathbb{A}}(X)$  是条件均值  $g_o(X)$  的正确设定, 且  $\beta^o$  是数据生成过程的真实参数值. 在此条件下, 证明  $\beta^* = \beta^o$ , 并且  $\mathbb{E}[u^*|X] = 0$ .

(4) 假设对于任意  $\beta$  值,  $g_o(X) \neq X'\beta$ , 则称线性模型  $g_{\mathbb{A}}(X)$  是条件均值  $g_o(X)$  的错误设定. 检查  $\mathbb{E}[u^*|X] = 0$  是否成立, 并讨论它的意义.

**解答** (1) 定义函数

$$L(\beta) = \mathbb{E}[Y - X'\beta]^2 = \mathbb{E}[Y^2] - 2\beta'\mathbb{E}[XY] + \beta'\mathbb{E}[XX']\beta$$

根据矩阵微分, 得到使得  $L(\beta)$  最小化的一阶条件为

$$\frac{\partial L(\beta)}{\partial \beta} = -2\mathbb{E}[XY] + 2\mathbb{E}[XX']\beta = 0$$

由于  $\mathbb{E}[XX']$  非奇异, 解得

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

检查二阶条件

$$\frac{\partial^2 L(\beta)}{\partial \beta \partial \beta'} = 2\mathbb{E}[XX']$$

由于  $\mathbb{E}[XX']$  满列秩<sup>2</sup>, 该Hessian矩阵正定, 因此

$$\beta^* = (\mathbb{E}[XX'])^{-1}\mathbb{E}[XY]$$

是使得仿射函数  $g(X) = X'\beta$  的 MSE 最小化的全局最优解.

(2) 利用  $\beta^*$  容易得到

$$\begin{aligned}\mathbb{E}[Xu^*] &= \mathbb{E}[X(Y - X'\beta^*)] \\ &= \mathbb{E}[XY] - \mathbb{E}[XX'](\mathbb{E}[XX'])^{-1}\mathbb{E}[XY] \\ &= \mathbf{0}\end{aligned}$$

即  $\mathbb{E}[Xu^*] = \mathbf{0}$ .

(3) 定义  $Y = \mathbb{E}[Y|X] + \varepsilon$ , 由于线性模型  $g_A(X)$  是条件均值  $g_o(X)$  的正确设定, 故存在  $\beta^o$  使得  $\mathbb{E}[Y|X] = X'\beta^o$ , 由于  $\mathbb{E}[\varepsilon|X] = 0$ , 因此  $\mathbb{E}[Y - X'\beta^o|X] = \mathbb{E}[u^*|X] = 0$ . 进一步,  $\mathbb{E}[X\varepsilon] = \mathbb{E}[X\mathbb{E}[\varepsilon|X]] = 0$ , 正交条件成立, 因此  $\beta^* = \beta^o$ .

(4) 由于线性模型  $g_A(X)$  是条件均值  $g_o(X)$  的错误设定, 对于任意  $\beta \in \mathbb{R}^{k+1}$ ,  $\varepsilon = Y - \mathbb{E}[Y|X] \neq Y - X'\beta$ , 因此  $\mathbb{E}[\varepsilon|X] = 0 \neq \mathbb{E}[u^*|X]$ .

□

**2.5** 假设  $Y = \beta_0^* + \beta_1^*X_1 + u_1$ , 其中  $Y$  和  $X_1$  是随机变量,  $\beta^* = (\beta_0^*, \beta_1^*)'$  最优线性最小二乘近似系数.

(1) 证明  $\beta_1^* = \text{cov}(Y, X_1)/\sigma_{X_1}^2$ ,  $\beta_0^* = \mathbb{E}[Y] - \beta_1^*\mathbb{E}[X_1]$ , 并且均方误为

$$\text{MSE}(\beta^*) = \mathbb{E}[Y - (\beta_0^* + \beta_1^*X_1)]^2 = \sigma_Y^2(1 - \rho_{X_1Y}^2)$$

其中,  $\sigma_Y^2 = \text{var}(Y)$ ,  $\rho_{X_1Y}$  是  $Y$  和  $X_1$  之间的相关系数.

(2) 进一步假设  $Y$  和  $X_1$  服从二元正态分布. 证明  $\mathbb{E}[Y|X_1] = \beta_0^* + \beta_1^*X_1$ ,  $\text{var}(Y|X_1) = \sigma_Y^2(1 - \rho_{X_1Y}^2)$ , 即  $Y$  的条件均值等于最优线性最小二乘预测值, 而  $Y$  的条件方差等于最优线性最小二乘预测的均方误.

**解答** (1) 根据问题 2.4 第 (2) 问  $\mathbb{E}[Xu^*] = \mathbf{0}$ , 当随机向量  $X$  中包含常数项时可得  $\mathbb{E}[u^*] = 0$ .

对式  $Y = \beta_0^* + \beta_1^*X_1 + u_1$  取数学期望, 注意到  $\mathbb{E}[u_1] = 0$ , 得到  $\mathbb{E}[Y] = \beta_0^* + \beta_1^*\mathbb{E}[X_1]$ . 设  $\mu_Y = \mathbb{E}[Y]$ ,  $\mu_{X_1} = \mathbb{E}[X_1]$ , 有  $\mu_Y = \beta_0^* + \beta_1^*\mathbb{E}[X_1]$ , 进一步得到

$$Y - \mu_Y = \beta_1^*(X_1 - \mu_{X_1}) + u_1$$

根据最优线性最小二乘近似系数易知

$$\begin{aligned}\beta_1 &= (\mathbb{E}[(X_1 - \mu_{X_1})(X_1 - \mu_{X_1})'])^{-1} \mathbb{E}[(X_1 - \mu_{X_1})(Y - \mu_Y)] \\ &= (\mathbb{E}[X_1 - \mu_{X_1}]^2)^{-1} \mathbb{E}[(X_1 - \mu_{X_1})(Y - \mu_Y)] \\ &= \text{cov}(Y, X_1)/\sigma_{X_1}^2\end{aligned}$$

<sup>2</sup>对于任意  $D \in \mathbb{R}^{m \times n}$ ,  $D'D$  半正定, 当且仅当  $D$  满列秩时正定.

因此  $\beta_1^* = \text{cov}(Y, X_1)/\sigma_{X_1}^2$ ,  $\beta_0^* = \mathbb{E}[Y] - \beta_1^*\mathbb{E}[X_1]$ .

进一步,  $\text{MSE}(\beta^*) = \mathbb{E}[u_1]^2 = \text{var}(u_1)$ , 将  $\beta_1^*$  代入展开后得

$$\begin{aligned}\text{var}(u_1) &= \text{var}(Y) + (\beta_1^*)^2 \text{var}(X_1) - 2\beta_1^* \text{cov}(Y, X_1) \\ &= \text{var}(Y) - (\text{cov}(Y, X_1))^2 / \sigma_{X_1}^2 \\ &= \sigma_Y^2 (1 - \rho_{X_1 Y}^2)\end{aligned}$$

即  $\text{MSE}(\beta^*) = \mathbb{E}[Y - (\beta_0^* + \beta_1^* X_1)]^2 = \sigma_Y^2 (1 - \rho_{X_1 Y}^2)$  成立.

(2) 根据问题 2.2, 当  $(Y, X_1)$  服从二元正态分布时有

$$\begin{aligned}\mathbb{E}[Y|X] &= \mu_Y + \rho\sigma_Y \frac{X_1 - \mu_{X_1}}{\sigma_X} \\ &= \beta_0^* + \beta_1^* \mu_{X_1} + \frac{\text{cov}(Y, X_1)}{\sigma_{X_1}^2} (X_1 - \mu_{X_1}) \\ &= \beta_0^* + \beta_1^* X_1\end{aligned}$$

同样地, 条件方差  $\text{var}(Y|X) = \sigma_Y^2 (1 - \rho_{X_1 Y}^2)$  并不依赖于  $Y$  和  $X$  的结构.

□

## 2.6 假设

$$Y = \beta_0^o + \beta_1^o X_1 + |X_1| \varepsilon$$

其中  $\mathbb{E}[X_1] = 0$ ,  $\text{var}(X_1) = \sigma_{X_1}^2 > 0$ ,  $\mathbb{E}[\varepsilon] = 0$ ,  $\text{var}(\varepsilon) = \sigma_\varepsilon^2 > 0$ ,  $\varepsilon$  和  $X_1$  是相互独立的,  $\beta_0^o$  和  $\beta_1^o$  均是常数.

(1) 求  $\mathbb{E}[Y|X_1]$ .

(2) 求  $\text{var}(Y|X_1)$ .

(3) 证明  $\beta_1^o = 0$  当且仅当  $\text{cov}(X_1, Y) = 0$ .

**解答** (1) 由于随机变量  $\varepsilon$  和  $X_1$  相互独立, 并且对于可测函数  $f: \mathbb{R} \rightarrow \mathbb{R}$ ,  $\varepsilon$  和  $f(X_1)$  也相互独立, 得到

$$\mathbb{E}[|X_1| \varepsilon | X_1] = |X_1| \mathbb{E}[\varepsilon] = 0$$

因此  $\mathbb{E}[Y|X_1] = \beta_0^o + \beta_1^o X_1$ .

(2) 根据条件方差定义易知

$$\begin{aligned}\text{var}(Y|X_1) &= \mathbb{E}[(Y - \mathbb{E}[Y|X_1])^2 | X_1] \\ &= \mathbb{E}[X_1^2 \varepsilon^2 | X_1] \\ &= X_1^2 \sigma_\varepsilon^2\end{aligned}$$

因此  $\text{var}(Y|X_1) = X_1^2 \sigma_\varepsilon^2$ .

(3) 充分性: 由于随机变量  $\varepsilon$  和  $X_1$  相互独立, 当  $\beta_1^o = 0$  时, 利用协方差的定义有

$$\begin{aligned}\text{cov}(X_1, Y) &= \text{cov}(|X_1| \varepsilon, X_1) \\ &= \mathbb{E}[|X_1| X_1 \varepsilon] - \mathbb{E}[|X_1| \varepsilon] \mathbb{E}[X_1] \\ &= \mathbb{E}[|X_1| X_1 \varepsilon] = 0\end{aligned}$$

其中,  $\mathbb{E}[X_1|\varepsilon] = \mathbb{E}[X_1]\mathbb{E}[\varepsilon] = 0$ ,  $\mathbb{E}[X_1X_1\varepsilon] = \mathbb{E}[X_1X_1]\mathbb{E}[\varepsilon] = 0$ .

必要性: 利用协方差的性质易知

$$\begin{aligned}\text{cov}(X_1, Y) &= \text{cov}(\beta_1^o X_1, X_1) + \text{cov}(X_1|\varepsilon, X_1) \\ &= (\beta_1^o)^2 \sigma_{X_1}^2 = 0\end{aligned}$$

由于  $\sigma_{X_1}^2 > 0$ , 因此必有  $\beta_1^o = 0$ .

□

**2.7** 假设一消费函数为

$$Y = 1 + 0.5X_1 + \frac{1}{4}(X_1^2 - 1) + \varepsilon$$

其中  $X_1 \sim N(0, 1)$ ,  $\varepsilon \sim N(0, 1)$ , 且  $X_1$  与  $\varepsilon$  独立.

(1) 求条件均值  $g_o(X) \equiv \mathbb{E}[Y|X]$ , 其中  $X \equiv (1, X_1)'$ .

(2) 求边际消费倾向 (MPC)  $\frac{\partial}{\partial X_1} g_o(X)$ .

(3) 假设用以下线性回归模型预测  $Y$

$$Y = X'\beta + u = \beta_0 + \beta_1 X_1 + u$$

其中  $\beta \equiv (\beta_0, \beta_1)'$ . 求最佳线性近似系数  $\beta^*$  和最佳线性预测值  $g_{\mathbb{A}}^*(X) \equiv X'\beta^*$ .

(4) 计算线性回归模型的偏导数  $\frac{\partial}{\partial X_1} g_{\mathbb{A}}^*(X)$ , 并与 (2) 部分的真实边际消费倾向相比较. 讨论所得的结果.

**解答** (1) 由于随机变量  $\varepsilon$  和  $X_1$  相互独立且  $\varepsilon \sim N(0, 1)$ , 故  $\mathbb{E}[\varepsilon|X] = \mathbb{E}[\varepsilon] = 0$ , 因此条件均值  $g_o(X) = 1 + 0.5X_1 + \frac{1}{4}(X_1^2 - 1)$ .

(2) 边际消费倾向为  $\frac{\partial}{\partial X_1} g_o(X) = \frac{1}{2} + \frac{1}{2}X_1$ .

(3) 根据最优线性最小二乘近似系数  $\beta^* = (\mathbb{E}[XX'])^{-1}\mathbb{E}[XY]$ ,  $X_1 \sim N(0, 1)$ , 并且

$$\begin{aligned}\mathbb{E}[XX'] &= \begin{bmatrix} 1 & \mathbb{E}[X_1] \\ \mathbb{E}[X_1] & \mathbb{E}[X_1^2] \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \mathbb{E}[XY] &= \begin{bmatrix} \mathbb{E}[Y] \\ \mathbb{E}[X_1 Y] \end{bmatrix} = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}\end{aligned}$$

代入  $\beta^*$  解得  $\beta^* = (1, \frac{1}{2})'$ , 并且  $g_{\mathbb{A}}^*(X) = 1 + \frac{1}{2}X_1$ .

(4)  $g_{\mathbb{A}}^*(X)$  的偏导数  $\frac{\partial}{\partial X_1} g_{\mathbb{A}}^*(X) = \frac{1}{2} \neq \frac{1}{2} + \frac{1}{2}X_1$ .

□

**2.8** 令  $g_o(X_1) = \mathbb{E}[Y|X_1]$ , 这里  $Y$  和  $X_1$  都是随机变量. 则

$$Y = g_o(X_1) + \varepsilon$$

其中  $\mathbb{E}[\varepsilon|X_1] = 0$ .

考虑  $g_o(X_1)$  在  $\mu_1 = \mathbb{E}[X_1]$  处的一阶 Taylor 展开

$$\begin{aligned} g_o(X_1) &\approx g_o(\mu_1) + g'_o(\mu_1)(X_1 - \mu_1) \\ &= [g_o(\mu_1) - g'_o(\mu_1)\mu_1] + g'_o(\mu_1)X_1 \end{aligned}$$

这里  $g'_o(\mu_1)$  是函数  $g_o(X_1)$  在  $\mu_1$  的一阶导数. 令  $\beta = (\beta_0^*, \beta_1^*)'$  为最优线性最小二乘近似系数. 问  $\beta_1^* = g'_o(\mu_1)$  吗? 请给出理由.

**解答** 将  $g_o(X_1)$  在  $\mu_1$  处 Taylor 展开得到

$$\begin{aligned} g_o(X_1) &= \sum_{j=0}^{\infty} \frac{g_o^{(j)}(X_1)}{j!} (X_1 - \mu_1)^j \\ &= [g_o(\mu_1) - g'_o(\mu_1)\mu_1] + g'_o(\mu_1)X_1 + \sum_{j=2}^{\infty} \frac{g_o^{(j)}(X_1)}{j!} (X_1 - \mu_1)^j \\ &= [g_o(\mu_1) - g'_o(\mu_1)\mu_1] + g'_o(\mu_1)X_1 + R(X_1) \end{aligned}$$

其中,  $R(X_1)$  为二阶及以上的 Taylor 余项. 令  $u = \varepsilon + R(X_1)$ , 得到

$$Y = [g_o(\mu_1) - g'_o(\mu_1)\mu_1] + g'_o(\mu_1)X_1 + u$$

由于  $u$  是关于  $X_1$  的函数, 显然  $\mathbb{E}[X_1 u] \neq 0$ , 因此  $\beta_1^* \neq g'_o(\mu_1)$ .

□

**2.9** 假设数据生成过程为

$$Y = 0.8X_1X_2 + \varepsilon$$

其中  $X_1 \sim N(0, 1)$ ,  $X_2 \sim N(0, 1)$ ,  $\varepsilon \sim N(0, 1)$ , 并且  $X_1$ ,  $X_2$  及  $\varepsilon$  是相互独立的. 令  $X = (1, X_1, X_2)'$ .

(1) 利用  $X$  的信息可预测  $Y$  的条件均值  $\mathbb{E}[Y|X]$  吗?

(2) 假设使用以下线性回归模型预测  $Y$

$$g_{\mathbb{A}}(X) = X'\beta + u = \beta_0 + \beta_1X_1 + \beta_2X_2 + u$$

这一线性回归模型对  $Y$  有预测能力吗? 请解释.

**解答** (1) 由于  $X_1$ ,  $X_2$  及  $\varepsilon$ ,  $\mathbb{E}[\varepsilon|X] = \mathbb{E}[\varepsilon] = 0$ , 因此  $\mathbb{E}[Y|X] = 0.8X_1X_2$ , 故利用  $X$  的信息可以预测  $Y$ .

(2) 如果使用仿射模型  $g_{\mathbb{A}}(X)$  预测  $Y$ , 则最佳线性最小二乘近似系数  $\beta^*$  可以使得 MSE 最小化.

由于  $X_1$ ,  $X_2$  和  $\varepsilon$  独立且均服从  $N(0, 1)$  分布, 故

$$\mathbb{E}[XY] = \begin{bmatrix} \mathbb{E}[Y] \\ \mathbb{E}[X_1Y] \\ \mathbb{E}[X_2Y] \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

代入  $\beta^* = (\mathbb{E}[XX'])^{-1}\mathbb{E}[XY]$  解得  $\beta_0^* = \beta_1^* = \beta_2^* = 0$ , 因此该线性回归模型对  $Y$  没有预测能力.



□

**2.10** 假设  $\mathbb{E}[\varepsilon|X]$  存在,  $X$  是有界的随机变量,  $h(X)$  是任意可测函数. 令  $g(X) = \mathbb{E}[\varepsilon|X]$ , 并假设  $\mathbb{E}[g^2(X)] < \infty$ .

(1) 证明: 如果  $g(X) = 0$ , 则  $\mathbb{E}[\varepsilon h(X)] = 0$ .

(2) 证明: 如果  $\mathbb{E}[\varepsilon h(X)] = 0$ , 则  $g(X) = 0$ .

**解答** (1) 由于  $g(X) = \mathbb{E}[\varepsilon|X] = 0$ , 根据 LIE 可得

$$\mathbb{E}[\varepsilon h(X)] = \mathbb{E}[\mathbb{E}[\varepsilon h(X)|X]] = \mathbb{E}[h(X)\mathbb{E}[\varepsilon|X]] = \mathbb{E}[h(X) \cdot 0] = 0$$

即  $g(X) = 0 \Rightarrow \mathbb{E}[\varepsilon h(X)] = 0$ .

(2) 考虑  $h(X) = e^{tX}$ , 其中  $t \in \mathbb{N}(0, \delta)$ ,  $\mathbb{N}(0, \delta)$  表示以原点 0 为中心, 半径为  $\delta$  的邻域,  $\delta$  为一个很小的正数. 易知  $\mathbb{E}[\varepsilon h(X)] = \mathbb{E}[e^{tX} \mathbb{E}[\varepsilon|X]] = \mathbb{E}[e^{tX} g(X)]$ . 根据幂级数

$$e^{tX} = \sum_{j=0}^{\infty} \frac{t^j}{j!} X^j$$

可以得到

$$\mathbb{E}[e^{tX} g(X)] = \sum_{j=0}^{\infty} \frac{t^j}{j!} \mathbb{E}[X^j g(X)] = \sum_{j=0}^{\infty} \frac{t^j}{j!} \beta_j = 0$$

设随机变量  $X$  的概率密度为  $f_X(x)$ , 于是 Fourier 系数  $\beta_j = \int_{-\infty}^{\infty} X^j g(x) f_X(x) dx$ .

因此, 对于任意  $t \in \mathbb{N}(0, \delta)$ , 所有的  $\beta_j = 0$ , 可以得到

$$g(X) = \mathbb{E}[\varepsilon|X] = \sum_{j=0}^{\infty} \beta_j X^j = 0$$

即  $\mathbb{E}[\varepsilon h(X)] = 0 \Rightarrow g(X) = 0$ .

□

**2.11** 考虑下面非线性最小二乘问题

$$\min_{\beta \in \mathbb{R}^{k+1}} \mathbb{E}[Y - g(X, \beta)]^2$$

其中  $g(X, \beta)$  是  $\beta$  的非线性函数. 对于任意  $\beta \in \mathbb{R}^{k+1}$ , 假设  $(k+1) \times (k+1)$  矩阵

$$\mathbb{E} \left[ \frac{\partial}{\partial \beta} g(X, \beta) \frac{\partial}{\partial \beta'} g(X, \beta) \right]$$

以及  $\frac{\partial^2 g(X, \beta)}{\partial \beta \partial \beta'}$  都是有界的和非奇异的, 其中  $\frac{\partial}{\partial \beta'} g(X, \beta)$  是  $(k+1) \times 1$  列向量  $\frac{\partial}{\partial \beta} g(X, \beta)$  的转置.

(1) 推导最优非线性最小二乘近似系数  $\beta^*$  的一阶条件.

(2) 令  $Y = g(X, \beta) + u$ . 证明  $\beta = \beta^*$  当且仅当  $\mathbb{E} \left[ u \frac{\partial g(X, \beta)}{\partial \beta} \right] = 0$ . 当  $g(X, \beta)$  是  $\beta$  的非线性函数时, 是否仍有  $\mathbb{E}[Xu] = 0$ ? 请解释.

(3) 如果存在某一参数值  $\beta^o$ , 使  $\mathbb{E}[Y|X] = g(X, \beta^o)$ , 则  $g(X, \beta)$  被称为  $\mathbb{E}[Y|X]$  的正确设定. 证明  $\beta^* = \beta^o$  当且仅当  $g(X, \beta)$  被称为  $\mathbb{E}[Y|X]$  的正确设定.

(4) 当模型  $g(X, \beta)$  设定正确时, 是否有  $\mathbb{E}[u|X] = 0$ ? 其中  $u = Y - g(X, \beta)$ .

(5) 如果存在某一  $\beta^o$  值, 使得  $\mathbb{E}[u|X] = 0$ , 这里  $u = Y - g(X, \beta)$ , 那么, 非线性回归模型  $g(X, \beta)$  是否为  $\mathbb{E}[Y|X]$  的正确设定?

**解答** (1) 将  $\mathbb{E}[Y - g(X, \beta)]^2$  展开后得  $\mathbb{E}[Y^2 - 2Yg(X, \beta) + g^2(X, \beta)]$ , 一阶条件为

$$\begin{aligned}\frac{\partial \mathbb{E}[Y - g(X, \beta)]^2}{\partial \beta} &= \mathbb{E}\left[-2Y \frac{\partial g(X, \beta)}{\partial \beta} + 2g(X, \beta) \frac{\partial g(X, \beta)}{\partial \beta}\right] \\ &= -2\mathbb{E}\left[(Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}\right] = 0\end{aligned}$$

由于  $\beta^*$  使上式成立, 因此一阶条件为  $\mathbb{E}\left[(Y - g(X, \beta^*)) \frac{\partial g(X, \beta^*)}{\partial \beta}\right] = 0$ .

(2) 充分性: 由于  $u = Y - g(X, \beta)$ , 并根据第 (1) 问, 当  $\beta = \beta^*$  时一阶条件成立, 故  $\mathbb{E}\left[u \frac{\partial g(X, \beta)}{\partial \beta}\right] = 0$ .

必要性: 当  $\mathbb{E}\left[u \frac{\partial g(X, \beta)}{\partial \beta}\right] = 0$  成立时, 显然

$$\mathbb{E}\left[u \frac{\partial g(X, \beta)}{\partial \beta}\right] = \mathbb{E}\left[(Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta}\right] = 0$$

当  $\beta = \beta^*$  时上式成立.

$\mathbb{E}[Xu] = 0$  仍成立, 理由如下.

设  $g(X, \beta) = X'\varphi(\beta)$ , 其中  $\varphi(\beta)$  是  $\beta$  的非线性函数. 当  $\beta = \beta^*$  时

$$\mathbb{E}\left[uX' \frac{\partial \varphi(\beta^*)}{\partial \beta}\right] = 0$$

由于  $\frac{\partial \varphi(\beta^*)}{\partial \beta}$  是一个  $k+1$  常向量, 因此  $\mathbb{E}[Xu] = 0$ .

(3) 充分性: 当  $\beta^* = \beta^o$  时, 要证  $g(X, \beta^*) = \mathbb{E}[Y|X]$ , 只需证明  $\mathbb{E}[u|X] = 0$ , 其中  $u = Y - g(X, \beta^*)$ .

根据第 (2) 问有  $\mathbb{E}\left[u \frac{\partial g(X, \beta^*)}{\partial \beta}\right] = 0$ , 考虑

$$\frac{\partial g(X, \beta^*)}{\partial \beta} = e^{tX} \varphi(\beta^*)$$

其中  $\varphi: \mathbb{R}^{k+1} \rightarrow \mathbb{R}$  为  $\beta$  的非线性函数,  $\varphi(\beta^*) \neq 0$ ,  $t \in \mathbb{N}(0, \delta)$ ,  $\delta$  为很小的正数. 进一步

$$\mathbb{E}\left[u \frac{\partial g(X, \beta^*)}{\partial \beta}\right] = \varphi(\beta^*) \mathbb{E}[ue^{tX}] = 0$$

因此  $\mathbb{E}[ue^{tX}] = \mathbb{E}[e^{tX} \mathbb{E}[u|X]] = 0$ . 设  $u(X) = \mathbb{E}[u|X]$ , 随机变量  $X$  的概率密度为  $f_X(x)$ , 根据问题 2.10 的结论

$$\mathbb{E}[e^{tX} \mathbb{E}[u|X]] = \sum_{j=0}^{\infty} \frac{t^j}{j!} \mathbb{E}[X^j u(X)] = \sum_{j=0}^{\infty} \frac{t^j}{j!} \gamma_j = 0$$

其中, Fourier 系数  $\gamma_j = \int_{-\infty}^{\infty} X^j u(x) f_X(x) dx$ . 因此,  $\mathbb{E}[u|X] = \sum_{j=1}^{\infty} \gamma_j X^j = 0$ .

必要性: 如果对于某个  $\beta^o \in \mathbb{R}^{k+1}$ , 使  $g(X, \beta^o)$  是  $\mathbb{E}[Y|X]$  的正确设定, 那么有  $\mathbb{E}[Y -$

$g(X, \beta^o)|X] = 0$ . 根据 LIE 有

$$\mathbb{E} \left[ (Y - g(X, \beta^o)) \frac{\partial g(X, \beta^o)}{\partial \beta} \right] = \mathbb{E} \left[ \frac{\partial g(X, \beta^o)}{\partial \beta} \mathbb{E}[(Y - g(X, \beta^o))|X] \right] = 0$$

根据第 (2) 问的结论  $\mathbb{E} \left[ (Y - g(X, \beta^*)) \frac{\partial g(X, \beta^*)}{\partial \beta} \right] = 0$ , 还需证明  $\beta^* = \beta^o$  是唯一的全局最优解.

考虑最优非线性最小二乘近似系数  $\beta^*$  的二阶条件

$$\begin{aligned} \frac{\partial^2 \mathbb{E}[Y - g(X, \beta)]^2}{\partial \beta \partial \beta'} &= -2 \frac{\partial}{\partial \beta'} \mathbb{E} \left[ (Y - g(X, \beta)) \frac{\partial g(X, \beta)}{\partial \beta} \right] \\ &= -2 \mathbb{E} \left[ \frac{\partial^2 g(X, \beta)}{\partial \beta \partial \beta'} Y - \frac{\partial g(X, \beta)}{\partial \beta} \frac{\partial g(X, \beta)}{\partial \beta'} - g(X, \beta) \frac{\partial^2 g(X, \beta)}{\partial \beta \partial \beta'} \right] \\ &= 2 \mathbb{E} \left[ \frac{\partial g(X, \beta)}{\partial \beta} \frac{\partial g(X, \beta)}{\partial \beta'} \right] \end{aligned}$$

该矩阵有界非奇异且正定, 因此  $\beta^* = \beta^o$ .

(4) 当模型设定正确时, 存在某个  $\beta^o \in \mathbb{R}^{k+1}$ , 使  $g(X, \beta^o) = \mathbb{E}[Y|X]$ , 因此  $\mathbb{E}[u|X] = \mathbb{E}[Y - \mathbb{E}[Y|X]|X] = 0$ .

(5) 非线性回归模型  $g(X, \beta)$  是  $\mathbb{E}[Y|X]$  的正确设定, 根据

$$\mathbb{E}[u|X] = \mathbb{E}[Y|X] - \mathbb{E}[g(X, \beta)|X] = \mathbb{E}[Y|X] - g(X, \beta) = 0$$

因此  $g(X, \beta) = \mathbb{E}[Y|X]$ .

□

## 经典线性回归模型

**3.1** 假设对于任意的  $t \geq m$ ,  $X_t = Q$ , 其中  $m$  是一个固定整数 (即不随样本容量  $n$  的增加而增加),  $Q$  是一个  $K \times 1$  常向量. 当  $n \rightarrow \infty$  时, 是否有  $\lambda_{\min}(\mathbf{X}'\mathbf{X}) \rightarrow \infty$ ? 请解释.

**解答** 利用  $\mathbf{X}'\mathbf{X} = \sum_{t=1}^n X_t X_t'$ , 易知

$$\begin{aligned}\sum_{t=1}^n X_t X_t' &= \sum_{t=1}^{m-1} X_t X_t' + \sum_{t=m}^n X_t X_t' \\ &= \sum_{t=1}^{m-1} X_t X_t' + (n-m+1)QQ'\end{aligned}$$

由于  $QQ'$  为实对称矩阵, 根据 Weyl 不等式<sup>3</sup>得到

$$\lambda_{\min}(\mathbf{X}'\mathbf{X}) \leq \lambda_{\max}\left(\sum_{t=1}^{m-1} X_t X_t'\right) + (n-m+1)\lambda_{\min}(QQ')$$

由于  $\text{rank}(QQ') = 1$  且  $QQ'$  半正定, 半正定矩阵特征值非负, 因此  $\lambda_{\min}(QQ') = 0$ , 又因为  $\lambda_{\max}\left(\sum_{t=1}^{m-1} X_t X_t'\right)$  为某个有限的数, 因此当  $n \rightarrow \infty$  时,  $\lambda_{\min}(\mathbf{X}'\mathbf{X}) \rightarrow \infty$  不成立.  $\square$

### 3.2 考虑双变量线性回归模型

$$Y_t = X_t' \beta^0 + \varepsilon_t, \quad t = 1, 2, \dots, n$$

其中  $X_t = (X_{0t}, X_{1t})' = (1, X_{1t})'$ ,  $\varepsilon_t$  是随机扰动项.

(1) 令  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1)'$  为 OLS 估计量. 证明  $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1$ , 且

$$\hat{\beta}_1 = \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1)(Y_t - \bar{Y})}{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2} = \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1)Y_t}{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2} = \sum_{t=1}^n C_t Y_t$$

其中  $C_t = (X_{1t} - \bar{X}_1) / \sum_{t=1}^n (X_{1t} - \bar{X}_1)^2$ .

(2) 假设  $X_1 = (X_{11}, \dots, X_{1n})'$  和  $\varepsilon = (\varepsilon_1, \dots, \varepsilon_n)'$  是相互独立的, 证明

$$\text{var}(\hat{\beta}_1 | X_1) = \sigma_\varepsilon^2 / [(n-1)S_{X_1}^2]$$

其中  $S_{X_1}^2$  是  $\{X_{1t}\}_{t=1}^n$  的样本方差. 这个结果表明,  $\{X_{1t}\}$  的方差越大,  $\beta_1^0$  的 OLS 估计越准确.

(3) 令  $\hat{\rho}$  是  $Y_t$  和  $X_{1t}$  之间的样本相关系数, 即

<sup>3</sup> 设  $A, B$  为  $n \times n$  Hermitian 矩阵, 并且  $A, B$  和  $A+B$  的特征值各自按  $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n$  排序, 则对于任意的  $j \in \{1, 2, \dots, n\}$  有  $\lambda_j(A) + \lambda_1(B) \leq \lambda_j(A+B) \leq \lambda_j(A) + \lambda_n(B)$ .

$$\hat{\rho} = \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1)(Y_t - \bar{Y})}{\sqrt{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2 \sum_{t=1}^n (Y_t - \bar{Y})^2}}$$

证明  $R^2 = \hat{\rho}^2$ . 这个结果表明,  $Y$  和  $X_1$  之间的样本相关系数的平方是  $Y$  的样本方差中可以被  $X_1$  的线性模型的方差所预测的比例.

**解答** (1) 根据定理 3.1 的 OLS 估计量  $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'Y$  易知

$$\begin{aligned}\hat{\beta} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix} &= \begin{bmatrix} n & \sum X_{1t} \\ \sum X_{1t} & \sum X_{1t}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum Y_t \\ \sum X_{1t} Y_t \end{bmatrix} \\ &= \frac{1}{n \sum X_{1t}^2 - (\sum X_{1t})^2} \begin{bmatrix} \sum X_{1t}^2 \sum Y_t - \sum X_{1t} \sum X_{1t} Y_t \\ n \sum X_{1t} Y_t - \sum X_{1t} \sum Y_t \end{bmatrix}\end{aligned}$$

由此先得出  $\hat{\beta}_1$  的表达式

$$\begin{aligned}\hat{\beta}_1 &= \frac{n \sum_{t=1}^n X_{1t} Y_t - \sum_{t=1}^n X_{1t} \sum_{t=1}^n Y_t}{n \sum_{t=1}^n X_{1t}^2 - \left( \sum_{t=1}^n X_{1t} \right)^2} = \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1)(Y_t - \bar{Y})}{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2} \\ &= \frac{\sum_{t=1}^n (X_{1t} - \bar{X}_1) Y_t}{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2} = \sum_{t=1}^n C_t Y_t\end{aligned}$$

容易验证  $\hat{\beta}_0$  的表达式满足  $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1$ .

(2) 由于  $X_1$  和  $\varepsilon$  相互独立, 因此条件同方差假设成立, 根据定理 3.5 可以得到  $\hat{\beta}$  的协方差矩阵

$$\begin{aligned}\text{var}(\hat{\beta}|\mathbf{X}) &= \sigma_\varepsilon^2 (\mathbf{X}'\mathbf{X})^{-1} \\ &= \frac{\sigma_\varepsilon^2}{n \sum X_{1t}^2 - (\sum X_{1t})^2} \begin{bmatrix} \sum X_{1t}^2 & -\sum X_{1t} \\ -\sum X_{1t} & n \end{bmatrix} \\ &= \begin{bmatrix} \text{var}(\hat{\beta}_0|X_1) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1|X_1) \\ \text{cov}(\hat{\beta}_1, \hat{\beta}_0|X_1) & \text{var}(\hat{\beta}_1|X_1) \end{bmatrix}\end{aligned}$$

也即  $\text{var}(\hat{\beta}_1|X_1) = \sigma_\varepsilon^2 / [(n-1)S_{X_1}^2]$ .

(3) 要证  $R^2 = \hat{\rho}^2$ , 只需证明

$$\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2 \sum_{t=1}^n (X_{1t} - \bar{X}_1)^2 = \left[ \sum_{t=1}^n (X_{1t} - \bar{X}_1)(Y_t - \bar{Y}) \right]^2$$

其中  $\hat{Y}_t = \hat{\beta}_0 + \hat{\beta}_1 X_{1t}$ .



设  $M_0 = I - \iota(\iota'\iota)^{-1}\iota'$ , 其中  $I$  为  $n \times n$  单位矩阵,  $\iota$  为元素全为 1 的  $n \times 1$  向量, 满足  $M_0\iota = \mathbf{0}$  及  $\iota'M_0 = \mathbf{0}'$ . 易证

$$\begin{aligned}\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2 &= X_1' M_0 X_1 \\ \sum_{t=1}^n (X_{1t} - \bar{X}_1)(Y_t - \bar{Y}) &= X_1' M_0 Y\end{aligned}$$

要证明的等式立刻变为

$$\hat{Y}' M_0 \hat{Y} X_1' M_0 X_1 = X_1' M_0 Y X_1' M_0 Y \quad (3.1)$$

其中  $\hat{Y} = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)'$ , 式 (3.1) 左端

$$\begin{aligned}\text{LHS} &= \hat{Y}' M_0 \hat{Y} X_1' M_0 X_1 \\ &= (\hat{\beta}_0 \iota + \hat{\beta}_1 X_1)' M_0 (\hat{\beta}_0 \iota + \hat{\beta}_1 X_1) X_1' M_0 X_1 \\ &= (\hat{\beta}_1)^2 X_1' M_0 X_1 X_1' M_0 X_1\end{aligned}$$

定义残差向量  $e = (e_1, e_2, \dots, e_n)'$ , 满足  $X_1'e = 0$  及  $M_0e = e$ , 式 (3.1) 右端

$$\begin{aligned}\text{RHS} &= X_1' M_0 Y X_1' M_0 Y \\ &= X_1' M_0 (\hat{\beta}_0 \iota + \hat{\beta}_1 X_1 + e) X_1' M_0 (\hat{\beta}_0 \iota + \hat{\beta}_1 X_1 + e) \\ &= (\hat{\beta}_1)^2 X_1' M_0 X_1 X_1' M_0 X_1\end{aligned}$$

因此有  $\text{LHS} = \text{RHS}$ , 故  $R^2 = \hat{\rho}^2$ .

□

### 3.3 考虑线性回归模型

$$Y_t = X_t' \beta^o + \varepsilon_t$$

其中  $X_t$  和  $\beta^o$  是  $K \times 1$  向量. 令  $\hat{Y}_t = X_t' \hat{\beta}$ , 其中  $\hat{\beta}$  是 OLS 估计量. 证明:

$$R^2 = \hat{\rho}_{Y\hat{Y}}^2$$

这里  $\hat{\rho}_{Y\hat{Y}}$  是  $Y$  和  $\hat{Y}$  之间的样本相关系数.

**解答** 要证  $R^2 = \hat{\rho}_{Y\hat{Y}}^2$ , 只需证明

$$\sum_{t=1}^n (\hat{Y}_t - \bar{Y})^2 = \sum_{t=1}^n (\hat{Y}_t - \bar{Y})(Y_t - \bar{Y})$$

采用问题 3.3 中的符号, 设  $M_0 = I - \iota(\iota'\iota)^{-1}\iota'$ , 即

$$\hat{Y}' M_0 \hat{Y} = \hat{Y}' M_0 Y$$

根据  $Y = \hat{Y} + e$  以及  $\hat{Y}'e = \mathbf{0}$ , 得到

$$\begin{aligned}\text{RHS} &= \hat{Y}' M_0 Y \\ &= \hat{Y}' M_0 (\hat{Y} + e) \\ &= \hat{Y}' M_0 \hat{Y} = \text{LHS}\end{aligned}$$

因此  $R^2 = \hat{\rho}_{Y\hat{Y}}^2$ .

□

**3.4** 在线性回归模型中, 调整的  $R^2$ , 记为  $\bar{R}^2$ , 定义如下

$$\bar{R}^2 = 1 - \frac{e'e/(n-K)}{(Y - \bar{Y}\iota)'(Y - \bar{Y}\iota)/(n-1)}$$

这里  $\iota = (1, \dots, 1)'$  为  $n \times 1$  向量, 其中每个元素均为 1, 证明

$$\bar{R}^2 = 1 - \frac{n-1}{n-K}(1-R^2)$$

**解答** 设  $M_0 = I - \iota(\iota'\iota)^{-1}\iota'$ , 则

$$R^2 = 1 - \frac{e'e}{Y'M_0Y}$$

根据  $\bar{R}^2$  定义有

$$\bar{R}^2 = 1 - \frac{e'e/(n-K)}{Y'M_0Y/(n-1)}$$

显然  $\bar{R}^2 = 1 - \frac{n-1}{n-K}(1-R^2)$ .

□

**3.5 [多重共线性的影响]** 考虑回归模型

$$Y_t = \beta_0^o + \beta_1^o X_{1t} + \beta_2^o X_{2t} + \varepsilon_t$$

假设 3.1, 3.2, 3.3(1) 与 3.4 成立. 令  $\hat{\beta} = (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2)'$  为 OLS 估计量. 证明

$$\text{var}(\hat{\beta}_1|\mathbf{X}) = \frac{\sigma^2}{(1-\hat{r}^2) \sum_{t=1}^n (X_{1t} - \bar{X}_1)^2}$$

$$\text{var}(\hat{\beta}_2|\mathbf{X}) = \frac{\sigma^2}{(1-\hat{r}^2) \sum_{t=1}^n (X_{2t} - \bar{X}_2)^2}$$

其中  $\bar{X}_1 = n^{-1} \sum_{t=1}^n X_{1t}$ ,  $\bar{X}_2 = n^{-1} \sum_{t=1}^n X_{2t}$ , 并且

$$\hat{r}^2 = \frac{\left[ \sum_{t=1}^n (X_{1t} - \bar{X}_1)(X_{2t} - \bar{X}_2) \right]^2}{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2 \sum_{t=1}^n (X_{2t} - \bar{X}_2)^2}$$

**解答**  $\hat{r}^2$  可视为解释变量  $X_1$  和  $X_2$  之间相关系数的平方, 根据问题 3.2 第 (3) 问, 可以得到

$$\hat{r}_j^2 = R^2$$

其中  $R^2$  表示  $X_1$  对  $X_2$  回归的决定系数. 根据问题 3.6 的结论立刻证得

$$\text{var}(\hat{\beta}_1|\mathbf{X}) = \frac{\sigma^2}{(1-\hat{r}^2) \sum_{t=1}^n (X_{1t} - \bar{X}_1)^2}$$

同理可证  $\text{var}(\hat{\beta}_2|\mathbf{X})$  的表达式.

□

### 3.6 考虑线性回归模型

$$Y_t = X_t' \beta^o + \varepsilon_t$$

其中  $X_t = (1, X_{1t}, \dots, X_{kt})'$ . 假设 3.1, 3.2, 3.3(1) 与 3.4 成立. 令  $R_j^2$  是变量  $X_{jt}$  对所有其他解释变量  $\{X_{it}, 0 \leq i \leq k, i \neq j\}$  回归的决定系数. 证明

$$\text{var}(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{(1 - R_j^2) \sum_{t=1}^n (X_{jt} - \bar{X}_j)^2}$$

其中  $\bar{X}_j = n^{-1} \sum_{t=1}^n X_{jt}$ . 因子  $1/(1 - R_j^2)$  被称为方差膨胀因子 (variance inflation factor, VIF), 它用来衡量解释变量  $X_t$  之间的多重共线性的程度.

**解答** 不失一般性, 设  $\hat{\beta}_j$  为  $\beta$  的第一个分量, 根据定理 3.5 有  $\text{var}(\hat{\beta}_j | \mathbf{X}) = \sigma^2 (\mathbf{X}'\mathbf{X})_{jj}^{-1}$ .

将矩阵  $(\mathbf{X}'\mathbf{X})^{-1}$  改写为

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} X_j' X_j & X_j' X_{-j} \\ X_{-j}' X_j & X_{-j}' X_{-j} \end{bmatrix}^{-1}$$

其中  $X_j$  为  $n \times 1$  列向量,  $X_{-j}$  为  $X_j$  以外的其它解释变量组成的  $n \times (K-1)$  矩阵.

根据分块矩阵求逆公式, 以上矩阵的 (1, 1) 元素为

$$[X_j' X_j - X_j' X_{-j} (X_{-j}' X_{-j})^{-1} X_{-j}' X_j]^{-1}$$

可以得到

$$\begin{aligned} \text{var}(\hat{\beta}_j | \mathbf{X}) &= \sigma^2 (\mathbf{X}'\mathbf{X})_{jj}^{-1} \\ &= \sigma^2 [X_j' X_j - X_j' X_{-j} (X_{-j}' X_{-j})^{-1} X_{-j}' X_j]^{-1} \\ &= \sigma^2 [X_j' X_j - X_j' X_{-j} (X_{-j}' X_{-j})^{-1} (X_{-j}' X_{-j}) (X_{-j}' X_{-j})^{-1} X_{-j}' X_j]^{-1} \\ &= \sigma^2 [X_j' X_j - \hat{\gamma}' (X_{-j}' X_{-j}) \hat{\gamma}]^{-1} \end{aligned}$$

其中  $\hat{\gamma} = (X_{-j}' X_{-j})^{-1} X_{-j}' X_j$ , 代表  $X_j$  关于  $X_{-j}$  的 OLS 回归系数, 即

$$X_j = X_{-j} \hat{\gamma} + e$$

其中  $e$  为  $n \times 1$  残差向量. 并且残差平方和  $\text{SSR}_j$

$$\begin{aligned} e'e &= (X_j - X_{-j} \hat{\gamma})' (X_j - X_{-j} \hat{\gamma}) \\ &= X_j' X_j - 2X_j' X_{-j} \hat{\gamma} + \hat{\gamma}' X_{-j}' X_{-j} \hat{\gamma} \\ &= X_j' X_j - \hat{\gamma}' X_{-j}' X_{-j} \hat{\gamma} \end{aligned}$$

由此得到  $\text{var}(\hat{\beta}_j | \mathbf{X}) = \sigma^2 / \text{SSR}_j$ . 根据  $R_j^2 = 1 - e'e / \sum_{t=1}^n (X_{jt} - \bar{X}_j)^2$ , 最终得到

$$\text{var}(\hat{\beta}_j | \mathbf{X}) = \frac{\sigma^2}{(1 - R_j^2) \sum_{t=1}^n (X_{jt} - \bar{X}_j)^2}$$

表明  $X_j$  与  $X_{-j}$  的线性关系越强时,  $R_j^2$  越接近于 1, OLS 回归的多重共线性越强.

□

### 3.7 考虑以下线性回归模型

$$Y_t = X_t' \beta^o + u_t, \quad t = 1, 2, \dots, n$$

其中

$$u_t = \sigma(X_t) \varepsilon_t$$

这里  $\{X_t\}_{t=1}^n$  是一个非随机序列, 并且  $\sigma(X_t)$  是  $X_t$  的一个正函数, 使得

$$\begin{bmatrix} \sigma^2(X_1) & 0 & \cdots & 0 \\ 0 & \sigma^2(X_2) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma^2(X_n) \end{bmatrix} = \Omega^{\frac{1}{2}} \Omega^{\frac{1}{2}}$$

其中

$$\Omega^{\frac{1}{2}} = \begin{bmatrix} \sigma(X_1) & 0 & \cdots & 0 \\ 0 & \sigma(X_2) & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma(X_n) \end{bmatrix}$$

假设  $\{\varepsilon_t\} \sim \text{IID. } N(0, 1)$ , 则  $\{u_t\} \sim N(0, \sigma^2(X_t))$ . 这与经典线性回归分析的假设 3.5 不同, 因为  $\{u_t\}$  存在有条件异方差. 令  $\hat{\beta}$  表示  $\beta^o$  的 OLS 估计量.

(1)  $\hat{\beta}$  是  $\beta^o$  的无偏估计量吗?

(2) 证明:  $\text{var}(\hat{\beta}|\mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ .

考虑另一个估计量

$$\begin{aligned} \tilde{\beta} &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-1}Y \\ &= \left[ \sum_{t=1}^n \sigma^{-2}(X_t) X_t X_t' \right]^{-1} \sum_{t=1}^n \sigma^{-2}(X_t) X_t Y_t \end{aligned}$$

(3)  $\tilde{\beta}$  是  $\beta^o$  的无偏估计量吗?

(4) 证明:  $\text{var}(\tilde{\beta}|\mathbf{X}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$ .

(5)  $\text{var}(\hat{\beta}|\mathbf{X}) - \text{var}(\tilde{\beta}|\mathbf{X})$  是半正定 (PSD) 的吗? 估计量  $\hat{\beta}$  和  $\tilde{\beta}$ , 哪一个更有效?

(6)  $\tilde{\beta}$  是  $\beta^o$  的最优线性无偏估计量 (BLUE) 吗?

(7) 构造两个关于原假设  $\mathbb{H}_0: \beta_2^o = 0$  的检验统计量. 一个检验是基于  $\hat{\beta}$ , 另一个检验是基于  $\tilde{\beta}$ . 当  $\mathbb{H}_0: \beta_2^o = 0$  成立时, 所构造的检验统计量的有限样本分布分别是什么? 在有限样本条件下哪一个检验更有效?

(8) 为检验原假设  $\mathbb{H}_0: R\beta^o = r$ , 构造两个检验统计量, 其中  $R$  是  $J \times K$  满秩矩阵,  $r$  是  $J \times 1$  向量, 且  $J \leq K$ . 一个检验是基于  $\hat{\beta}$ , 另一个检验是基于  $\tilde{\beta}$ . 当  $\mathbb{H}_0: R\beta^o = r$  成立时, 所构造的检验统计量的有限样本分布分别是什么?

解答 (1) 根据  $\hat{\beta} - \beta^o = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}\varepsilon$ , 可以得到

$$\mathbb{E}[\hat{\beta} - \beta^o | \mathbf{X}] = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}\mathbb{E}[\varepsilon | \mathbf{X}] = \mathbf{0}$$

因此  $\hat{\beta}$  是  $\beta^o$  的无偏估计量.

(2) 根据  $\hat{\beta} - \beta^o = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}\varepsilon$ , 由于  $\{\varepsilon_t\} \sim \text{i.i.d. } N(0, 1)$ , 因此  $\mathbb{E}[\varepsilon\varepsilon' | \mathbf{X}] = I$ , 可以得到

$$\begin{aligned}\text{var}(\hat{\beta} - \beta^o | \mathbf{X}) &= \mathbb{E}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}\varepsilon\varepsilon'\Omega^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} | \mathbf{X}] \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}\mathbb{E}[\varepsilon\varepsilon' | \mathbf{X}]\Omega^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\end{aligned}$$

即  $\text{var}(\hat{\beta} | \mathbf{X}) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$ .

(3) 根据  $\tilde{\beta} - \beta^o = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}\varepsilon$ , 可以得到

$$\mathbb{E}[\tilde{\beta} - \beta^o | \mathbf{X}] = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}\mathbb{E}[\varepsilon | \mathbf{X}] = \mathbf{0}$$

因此  $\tilde{\beta}$  是  $\beta^o$  的无偏估计量.

(4) 根据  $\tilde{\beta} - \beta^o = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}\varepsilon$ , 可以得到

$$\begin{aligned}\text{var}(\tilde{\beta} - \beta^o | \mathbf{X}) &= \mathbb{E}[(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}\varepsilon\varepsilon'\Omega^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}] \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}\mathbb{E}[\varepsilon\varepsilon' | \mathbf{X}]\Omega^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\end{aligned}$$

因此  $\text{var}(\tilde{\beta} | \mathbf{X}) = (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}$ .

(5) 定义矩阵  $P_1 = \Omega^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}$ , 易证  $P_1$  为幂等矩阵和对称矩阵. 定义矩阵  $M_1 = I - P_1$ ,  $M_1$  也是幂等矩阵和对称矩阵. 再定义矩阵  $A = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}$ , 可以得到

$$\begin{aligned}\text{var}(\hat{\beta} | \mathbf{X}) - \text{var}(\tilde{\beta} | \mathbf{X}) &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} - (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'[\Omega - \mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}']\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega^{\frac{1}{2}}[I - \Omega^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}]\Omega^{\frac{1}{2}}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &= (AM_1)(AM_1)' \sim \text{p.s.d.}\end{aligned}$$

因此  $\text{var}(\hat{\beta} | \mathbf{X}) - \text{var}(\tilde{\beta} | \mathbf{X})$  是半正定, 估计量  $\tilde{\beta}$  更有效.

(6) 设  $\hat{b} = C'Y$  为  $\beta^o$  的任一线性无偏估计量, 其中  $C = C(\mathbf{X})$  为  $n \times K$  矩阵. 由于  $\hat{b}$  是无偏估计量, 可以得到

$$\mathbb{E}[\hat{b} | \mathbf{X}] = C'\mathbf{X}\beta^o + C\mathbb{E}[u | \mathbf{X}] = \beta^o$$

因此  $C'\mathbf{X} = I$ . 进一步得到  $\hat{b}$  的条件方差为

$$\text{var}(\hat{b} | \mathbf{X}) = C\mathbb{E}[uu' | \mathbf{X}]C' = C'\Omega C$$

故有

$$\begin{aligned}\text{var}(\hat{b} | \mathbf{X}) - \text{var}(\tilde{\beta} | \mathbf{X}) &= C'\Omega C - (\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1} \\ &= C'\Omega^{\frac{1}{2}}[I - \Omega^{-\frac{1}{2}}\mathbf{X}(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}\mathbf{X}'\Omega^{-\frac{1}{2}}]\Omega^{\frac{1}{2}}C \\ &= (C'\Omega^{\frac{1}{2}}M_1)(C'\Omega^{\frac{1}{2}}M_1)' \sim \text{p.s.d.}\end{aligned}$$



因此  $\tilde{\beta}$  是  $\beta^o$  的 BLUE.

(7) 由于  $\varepsilon_t \sim \text{IID } N(0, 1)$ , 根据第 (2) 问的结果易知

$$\hat{\beta}_2 - \beta_2^o | \mathbf{X} \sim N(0, [(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]_{22})$$

由于协方差矩阵  $\Omega$  已知, 故对  $\hat{\beta}_2$  可以构建统计量

$$Z = \frac{\hat{\beta}_2}{\sqrt{[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]_{22}}}$$

使得当  $\mathbb{H}_0 : \beta_2^o = 0$  成立时有  $Z \sim N(0, 1)$ .

对于  $\tilde{\beta}_2$  可以得到

$$\tilde{\beta}_2 - \beta | \mathbf{X} \sim N(0, [(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}]_{22})$$

因此可以构建统计量

$$\tilde{Z} = \frac{\tilde{\beta}_{22}}{\sqrt{[(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}]_{22}}}$$

使得当  $\mathbb{H}_0 : \beta_2^o = 0$  成立时有  $\tilde{Z} \sim N(0, 1)$ .

由于  $\text{var}(\hat{\beta}|\mathbf{X}) - \text{var}(\tilde{\beta}|\mathbf{X})$  半正定, 因此  $\tilde{Z}$  统计量更有效.

(8) 对于  $\hat{\beta}$ , 当假设  $\mathbb{H}_0$  成立时有

$$R\hat{\beta} - r | \mathbf{X} \sim N(0, R(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}R')$$

根据引理 3.11 可构建统计量

$$F = (R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Omega\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r) \sim \chi_J^2$$

对于  $\tilde{\beta}$ , 当假设  $\mathbb{H}_0$  成立时有

$$R\tilde{\beta} - r | \mathbf{X} \sim N(0, R(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}R')$$

则可构建统计量

$$\tilde{F} = (R\tilde{\beta} - r)'[R(\mathbf{X}'\Omega^{-1}\mathbf{X})^{-1}R']^{-1}(R\tilde{\beta} - r) \sim \chi_J^2$$

二者的有限样本分布均为  $\chi_J^2$  分布.

□

### 3.8 考虑经典回归模型

$$Y_t = X_t'\beta^o + \varepsilon_t$$

假定假设 3.1, 3.3(1) 成立. 目的是检验原假设

$$\mathbb{H}_0 : R\beta^o = r$$

$F$  检验统计量定义为

$$F = \frac{(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)/J}{s^2}$$

证明:

$$F = \frac{(\tilde{e}'\tilde{e} - e'e)/J}{e'e/(n - K)}$$

其中  $e'e$  是无约束回归模型的残差平方和,  $\tilde{e}'\tilde{e}$  是有约束回归模型的残差平方和, 其中约束条件是  $R\beta^o = r$ .

**解答** 由于  $s^2 = \frac{e'e}{n-K}$ , 因此只需证明

$$(R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r) = \tilde{e}'\tilde{e} - e'e$$

其中  $\tilde{e}$  是约束条件下进行 OLS 回归得到的残差向量.

设  $\tilde{\beta}$  是约束条件下的 OLS 回归估计量, 也即

$$\begin{aligned}\tilde{\beta} &= \arg \min_{\beta \in \mathbb{R}^K} (Y - \mathbf{X}\beta)'(Y - \mathbf{X}\beta) \\ \text{s.t. } R\beta &= r\end{aligned}$$

定义 Lagrange 函数

$$L(\beta, \lambda) = (Y - \mathbf{X}\beta)'(Y - \mathbf{X}\beta) - 2\lambda'(R\beta - r)$$

其中  $\lambda$  为  $J \times 1$  Lagrange 乘子向量. 得到一阶条件为

$$\begin{aligned}\frac{\partial L(\beta, \lambda)}{\partial \beta} &= -2\mathbf{X}'Y + 2\mathbf{X}'\mathbf{X}\tilde{\beta} - 2R'\tilde{\lambda} = \mathbf{0} \\ \frac{\partial L(\beta, \lambda)}{\partial \lambda} &= -2(R\tilde{\beta} - r) = \mathbf{0}\end{aligned}\tag{3.2}$$

其中  $\tilde{\beta}$  和  $\tilde{\lambda}$  为一阶条件的最优解. 对式 (3.2) 左乘  $(\mathbf{X}'\mathbf{X})^{-1}$  得到

$$\hat{\beta} - \tilde{\beta} = -(\mathbf{X}'\mathbf{X})^{-1}R'\tilde{\lambda}\tag{3.3}$$

对式 (3.3) 左乘  $R$  得到

$$R\hat{\beta} - r = -R(\mathbf{X}'\mathbf{X})^{-1}R'\tilde{\lambda}$$

进一步有

$$\tilde{\lambda} = -R'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)\tag{3.4}$$

将式 (3.4) 代入到式 (3.3) 中得到

$$\hat{\beta} - \tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1}R'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)\tag{3.5}$$

另一方面, 注意到

$$\begin{aligned}\tilde{e}'\tilde{e} &= (Y - \mathbf{X}\tilde{\beta})'(Y - \mathbf{X}\tilde{\beta}) \\ &= Y'Y - Y'\mathbf{X}\tilde{\beta} - \tilde{\beta}'\mathbf{X}'Y + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta} \\ &= (\mathbf{X}\hat{\beta} + e)'(\mathbf{X}\hat{\beta} + e) - (\mathbf{X}\hat{\beta} + e)'\mathbf{X}\tilde{\beta} - \tilde{\beta}'\mathbf{X}'(\mathbf{X}\hat{\beta} + e) + \tilde{\beta}'\mathbf{X}'\mathbf{X}\tilde{\beta} \\ &= (\hat{\beta} - \tilde{\beta})'\mathbf{X}'\mathbf{X}(\hat{\beta} - \tilde{\beta}) + e'e\end{aligned}$$

也即  $\tilde{e}'\tilde{e} - e'e = (\hat{\beta} - \tilde{\beta})'\mathbf{X}'\mathbf{X}(\hat{\beta} - \tilde{\beta})$ . 将式 (3.5) 直接代入即得

$$\tilde{e}'\tilde{e} - e'e = (R\hat{\beta} - r)'[R(\mathbf{X}'\mathbf{X})^{-1}R']^{-1}(R\hat{\beta} - r)$$

因此原式成立.

□

3.9 考虑问题 3.8 的检验问题. 证明:

$$F = \frac{\sum_{t=1}^n (\hat{Y}_t - \bar{Y}_t)^2 / J}{s^2} = \frac{(\hat{\beta} - \tilde{\beta})' \mathbf{X}' \mathbf{X} (\hat{\beta} - \tilde{\beta}) / J}{s^2}$$

其中  $\hat{Y}_t = X_t' \hat{\beta}$ ,  $\tilde{Y}_t = X_t' \tilde{\beta}$ , 且  $\hat{\beta}$ ,  $\tilde{\beta}$  分别是无约束回归模型和有约束回归模型的 OLS 估计量.

**解答** 根据问题 3.8 有  $\tilde{e}'\tilde{e} - e'e = (\hat{\beta} - \tilde{\beta})' \mathbf{X}' \mathbf{X} (\hat{\beta} - \tilde{\beta})$ , 因此只需证明

$$\tilde{e}'\tilde{e} - e'e = \sum_{t=1}^n (\hat{Y}_t - \tilde{Y}_t)^2$$

设  $\hat{Y} = (\hat{Y}_1, \dots, \hat{Y}_n)'$ ,  $\tilde{Y} = (\tilde{Y}_1, \dots, \tilde{Y}_n)'$ , 可以得到

$$\begin{aligned} \sum_{t=1}^n (\hat{Y}_t - \tilde{Y}_t)^2 &= (\hat{Y} - \tilde{Y})'(\hat{Y} - \tilde{Y}) \\ &= (\mathbf{X}\hat{\beta} - \mathbf{X}\tilde{\beta})'(\mathbf{X}\hat{\beta} - \mathbf{X}\tilde{\beta}) \\ &= (\hat{\beta} - \tilde{\beta})' \mathbf{X}' \mathbf{X} (\hat{\beta} - \tilde{\beta}) \end{aligned}$$

因此原式成立.

□

3.10 考虑经典线性回归模型

$$\begin{aligned} Y_t &= X_t' \beta^o + \varepsilon_t \\ &= \beta_0^o + \sum_{j=1}^k \beta_j X_{jt} + \varepsilon_t, \quad t = 1, \dots, n \end{aligned} \quad (3.6)$$

目的是检验原假设

$$\mathbb{H}_0: \beta_1^o = \dots = \beta_k^o = 0$$

考虑  $F$  检验统计量

$$F = \frac{(\tilde{e}'\tilde{e} - e'e) / J}{e'e / (n - K)}$$

其中  $e'e$  为无约束回归模型 3.6 的残差平方和,  $\tilde{e}'\tilde{e}$  是下面有约束回归模型

$$Y_t = \beta_0^o + \varepsilon_t \quad (3.7)$$

的残差平方和.

(1) 证明: 给定假设 3.1 和 3.3(1)

$$F = \frac{R^2 / (K - 1)}{(1 - R^2) / (n - K)}$$

其中  $R^2$  是无约束模型 3.6 的决定系数.

(2) 再给定假设 3.5, 证明: 当原假设  $\mathbb{H}_0: \beta_1^o = \dots = \beta_k^o = 0$  成立及  $n \rightarrow \infty$  时

$$(n - K)R^2 \xrightarrow{d} \chi_{K-1}^2$$

**解答** (1) 原假设  $\mathbb{H}_0$  表明共  $J = K - 1$  个约束条件, 要证明原式成立, 只需证明

$$\frac{R^2}{1 - R^2} = \frac{\tilde{e}'\tilde{e} - e'e}{e'e}$$

对线性模型 3.7 进行 OLS 回归可得估计量  $\tilde{\beta} = (\bar{Y}, 0, \dots, 0)'$ , 进一步有  $\mathbf{X}\tilde{\beta} = \bar{Y}\iota$ , 其中  $\iota = (1, \dots, 1)'$ . 定义  $M_0 = I - \iota(\iota'\iota)^{-1}\iota'$ , 有  $M_0\iota = 0$  及  $M_0\tilde{e} = \tilde{e}$ . 易证

$$\begin{aligned} Y'M_0Y &= (\mathbf{X}\tilde{\beta} + \tilde{e})'M_0(\mathbf{X}\tilde{\beta} + \tilde{e}) \\ &= (\bar{Y}\iota + \tilde{e})'M_0(\bar{Y}\iota + \tilde{e}) = \tilde{e}'\tilde{e} \end{aligned}$$

根据  $R^2$  的定义有

$$R^2 = 1 - \frac{e'e}{Y'M_0Y} = 1 - \frac{e'e}{\tilde{e}'\tilde{e}}$$

进一步得到

$$\frac{R^2}{1 - R^2} = \frac{1 - \frac{e'e}{\tilde{e}'\tilde{e}}}{\frac{e'e}{\tilde{e}'\tilde{e}}} = \frac{\tilde{e}'\tilde{e} - e'e}{e'e}$$

因此原式成立.

(2) 由于  $F \sim F_{K-1, n-K}$ , 因此当  $n \rightarrow \infty$  时,  $(K-1)F \xrightarrow{d} \chi_{K-1}^2$ , 也即

$$(n-K) \frac{R^2}{1-R^2} \xrightarrow{d} \chi_{K-1}^2$$

由于  $K-1$  是有限的, 故而  $(K-1)F = O_p(1)$ . 根据  $O_p(1) \cdot o_p(1) = o_p(1)$  可得

$$\frac{R^2}{1-R^2} = \frac{(K-1)F}{n-K} = o_p(1)$$

也即当  $n \rightarrow \infty$  时,  $R^2 \xrightarrow{p} 0$ . 因此

$$(n-K)R^2 = (K-1)F \cdot (1-R^2) \xrightarrow{d} \chi_{K-1}^2$$

□

**3.11 [结构变化]** 在假设 3.1 和 3.3(1) 成立的条件下, 考虑对整个样本建立如下模型

$$Y_t = X_t'\beta^o + (D_t X_t)'\alpha^o + \varepsilon_t, \quad t = 1, \dots, n$$

其中  $D_t$  为时间虚拟变量. 当  $t \leq n_1$  时,  $D_t = 0$ , 当  $t > n_1$  时,  $D_t = 1$ . 该模型也可写成两个单独模型

$$Y_t = X_t'\beta^o + \varepsilon_t, \quad t = 1, \dots, n_1$$

和

$$Y_t = X_t'(\beta^o + \alpha^o) + \varepsilon_t, \quad t = n_1 + 1, \dots, n$$

令  $SSR_u$ ,  $SSR_1$ ,  $SSR_2$  分别代表三个 OLS 回归方程的残差平方和, 证明:

$$SSR_u = SSR_1 + SSR_2$$

该等式意味着对第一个包含时间虚拟变量的全样本回归方程进行 OLS 估计得到的残差平方和, 等价于对两个子样本回归方程分别估计而得到的残差平方和的加总.

**解答** 设  $\mathbf{X}_1$  为包含了前  $n_1$  个观测的  $n_1 \times K$  矩阵,  $\mathbf{X}_2$  为包含了后  $n - n_1$  个观测的  $(n - n_1) \times K$

矩阵, 因此全样本 OLS 回归方程记为

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{bmatrix} \mathbf{X}_1 & \mathbf{O} \\ \mathbf{O} & \mathbf{X}_2 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \tilde{\beta} \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

其中

$$\begin{aligned} Y_1 &= (Y_1, \dots, Y_{n_1})' \\ Y_2 &= (Y_{n_1+1}, \dots, Y_n)' \\ \hat{\beta} &= (\hat{\beta}_1, \dots, \hat{\beta}_K)' \\ \tilde{\beta} &= (\tilde{\beta}_1 + \tilde{\alpha}_1, \dots, \tilde{\beta}_K + \tilde{\alpha}_K)' \\ e_1 &= (e_1, \dots, e_{n_1})' \\ e_2 &= (e_{n_1+1}, \dots, e_n)' \end{aligned}$$

可以得到

$$\begin{aligned} \text{SSR}_u &= [(Y_1 - \mathbf{X}_1 \hat{\beta})', (Y_2 - \mathbf{X}_2 \tilde{\beta})'] \begin{bmatrix} Y_1 - \mathbf{X}_1 \hat{\beta} \\ Y_2 - \mathbf{X}_2 \tilde{\beta} \end{bmatrix} \\ &= (Y_1 - \mathbf{X}_1 \hat{\beta})'(Y_1 - \mathbf{X}_1 \hat{\beta}) + (Y_2 - \mathbf{X}_2 \tilde{\beta})'(Y_2 - \mathbf{X}_2 \tilde{\beta}) \\ &= \text{SSR}_1 + \text{SSR}_2 \end{aligned}$$

因此原式成立.

□

**3.12** 假设  $\mathbf{X}'\mathbf{X}$  是一个  $K \times K$  矩阵,  $V$  是一个  $n \times n$  矩阵,  $\mathbf{X}'\mathbf{X}$  和  $V$  均是对称和非奇异的, 并且当  $n \rightarrow \infty$  时, 最小的特征值  $\lambda_{\min}(\mathbf{X}'\mathbf{X}) \rightarrow \infty$ . 此外,  $0 < c \leq \lambda_{\max}(V) \leq C < \infty$ . 证明: 对于任意的  $\tau \in \mathbb{R}^K$ , 满足  $\tau'\tau = 1$ , 当  $n \rightarrow \infty$  时

$$\tau' \text{var}(\hat{\beta}|\mathbf{X}) \tau = \sigma^2 \tau' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' V \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \tau \rightarrow 0$$

因此, 在条件异方差情形下, 当  $n \rightarrow \infty$  时,  $\text{var}(\hat{\beta}|\mathbf{X})$  缩小至零.

**解答** 由于  $\text{var}(\hat{\beta}|\mathbf{X})$  是对称矩阵, 因此它的谱范数<sup>4</sup>为  $\|\text{var}(\hat{\beta}|\mathbf{X})\| = \lambda_{\max}[\text{var}(\hat{\beta}|\mathbf{X})]$ .

由于  $V$  为实对称矩阵, 存在正交矩阵  $Q$  使得  $V = Q \Lambda Q'$ , 其中  $\Lambda = \text{diag}\{\lambda_1, \dots, \lambda_n\}$  为矩阵  $V$  的全部特征值构成的对角矩阵. 再因为  $\text{var}(\hat{\beta}|\mathbf{X})$  也是实对称矩阵, 根据矩阵二次不等式<sup>5</sup>, 当  $n \rightarrow \infty$  时有

<sup>4</sup>对于  $m \times k$  维实矩阵  $A$ , 它的谱范数被定义为

$$\|A\| = [\lambda_{\max}(A'A)]^{\frac{1}{2}}$$

特别地, 当  $A$  为  $m \times m$  对称矩阵时,  $\|A\| = \max_{j \leq m} |\lambda_j(A)|$ .

<sup>5</sup>对任意列向量  $b \in \mathbb{R}^m$  以及对称矩阵  $A \in \mathbb{R}^{m \times m}$ , 若  $\|A\|$  是  $A$  的谱范数, 那么  $b'Ab \leq \|A\|b'b$ .



$$\begin{aligned}
\tau' \text{var}(\hat{\beta}|\mathbf{X}) \tau &= \sigma^2 \tau' (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' V \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \tau \\
&\leq \sigma^2 \lambda_{\max} [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' V \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}] \\
&\leq \sigma^2 C \lambda_{\max} [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' Q Q' \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1}] \\
&= \sigma^2 C \lambda_{\min}^{-1} (\mathbf{X}'\mathbf{X}) \rightarrow 0
\end{aligned}$$

其中  $C$  是矩阵  $V$  的最大特征值.

□

**3.13** 假设本章第 9 节中的条件成立, 可以证明 OLS 估计量  $\hat{\beta}$  和 GLS 估计量  $\hat{\beta}^*$  的方差分别为

$$\begin{aligned}
\text{var}(\hat{\beta}|\mathbf{X}) &= \sigma^2 (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' V \mathbf{X} (\mathbf{X}'\mathbf{X})^{-1} \\
\text{var}(\hat{\beta}^*|\mathbf{X}) &= \sigma^2 (\mathbf{X}' V^{-1} \mathbf{X})^{-1}
\end{aligned}$$

证明:  $\text{var}(\hat{\beta}|\mathbf{X}) - \text{var}(\hat{\beta}^*|\mathbf{X})$  是半正定的.

**解答** 与问题 3.7 的条件不同,  $V$  是一个  $n \times n$  有限且对称的正定矩阵, 表明可能同时存在条件异方差和条件自相关, 而  $\Omega$  仅包括了条件异方差一种情况.

由 Cholesky 分解, 存在一个非奇异的  $n \times n$  矩阵使得  $V^{-1} = C'C$ . 定义矩阵  $M^* = I - C\mathbf{X}(\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'C'$ , 容易验证  $M^*$  是对称且幂等的. 类似问题 3.7 第 (5) 问的做法, 容易写出

$$\text{var}(\hat{\beta}|\mathbf{X}) - \text{var}(\hat{\beta}^*|\mathbf{X}) = \sigma^2 [(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' C^{-1} M^*][(\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}' C^{-1} M^*]'$$

因此  $\text{var}(\hat{\beta}|\mathbf{X}) - \text{var}(\hat{\beta}^*|\mathbf{X})$  是半正定的.

□

**3.14** 假设数据生成过程为

$$Y_t = X_t' \beta^o + \varepsilon_t = \beta_1^o X_{1t} + \beta_2^o X_{2t} + \varepsilon_t$$

其中  $X_t = (X_{1t}, X_{2t})'$ ,  $\mathbb{E}[X_t X_t']$  是非奇异的, 并且  $\mathbb{E}[\varepsilon_t | X_t] = 0$ . 简单起见, 进一步假设  $\mathbb{E}[X_{2t}] = 0$ ,  $\mathbb{E}[X_{1t} X_{2t}] \neq 0$ , 且  $X_{2t}$  不是  $X_{1t}$  的一个确定性函数, 即不存在一个可测函数  $g(\cdot)$ , 使得  $X_{2t} = g(X_{1t})$ . 此外, 假设  $\beta_2^o \neq 0$ .

考虑以下双变量线性回归模型

$$Y_t = \beta_1 X_{1t} + u_t$$

(1) 证明:  $\mathbb{E}[Y_t | X_t] = X_t \beta^o \neq \mathbb{E}[Y_t | X_{1t}]$ . 即双变量回归模型中存在遗漏变量  $X_{2t}$ .

(2) 证明: 对于所有的  $\beta_1 \in \mathbb{R}$ ,  $\mathbb{E}[Y_t | X_{1t}] \neq \beta_1 X_{1t}$ . 即双变量线性回归模型是  $\mathbb{E}[Y_t | X_{1t}]$  的错误设定.

(3) 双变量线性回归模型的最优最小二乘近似系数  $\beta_1^*$  等于  $\beta_1^o$  吗? 请解释.

**解答** (1) 由  $\mathbb{E}[\varepsilon_t | X_t] = 0$  易证  $\mathbb{E}[Y_t | X_t] = X_t' \beta^o$ . 而  $\mathbb{E}[Y_t | X_{1t}] = \beta_1^o X_{1t} + \beta_2^o \mathbb{E}[X_{2t} | X_{1t}]$ , 因为  $\forall g: \mathbb{R} \rightarrow \mathbb{R}$  都有  $X_{2t} \neq g(X_{1t})$ , 故  $\mathbb{E}[X_{2t} | X_{1t}] \neq X_{2t}$ , 也即  $\mathbb{E}[Y_t | X_{1t}] \neq X_t' \beta^o$ .

(2) 由于  $\mathbb{E}[X_{1t} X_{2t}] = \mathbb{E}[X_{1t} \mathbb{E}[X_{2t} | X_{1t}]] \neq 0$ , 故  $\mathbb{E}[X_{2t} | X_{1t}] \neq 0$ . 因此对任意  $\beta_1 \in \mathbb{R}$ , 都有  $\mathbb{E}[Y_t | X_{1t}] = \beta_1^o X_{1t} + \beta_2^o \mathbb{E}[X_{2t} | X_{1t}] \neq \beta_1 X_{1t}$ .

(3) 由于  $\mathbb{E}[X_{1t}u] = \mathbb{E}[X_{1t}(\beta_2^o X_{2t} + \varepsilon_t)] = \beta_2^o \mathbb{E}[X_{1t}X_{2t}] \neq 0$ , 因此最优最小二乘近似系数  $\beta_1^* \neq \beta_1^o$ .

□

**3.15** 假设数据生成过程为

$$Y_t = X_t' \beta^o + \varepsilon_t = \beta_1^o X_{1t} + \beta_2^o X_{2t} + \varepsilon_t$$

其中  $X_t = (X_{1t}, X_{2t})'$ , 并且假设 3.1, 3.2, 3.3(1) 与 3.4 成立. OLS 估计量记为  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$ .

如果已知  $\beta_2^o = 0$ , 考虑以下线性回归模型

$$Y_t = \beta_1^o X_{1t} + \varepsilon_t$$

记该双变量回归模型的 OLS 估计量为  $\tilde{\beta}_1$ .

请比较  $\hat{\beta}_1$  与  $\tilde{\beta}_1$  之间的相对效率, 即哪一个  $\beta_1^o$  的估计量更精确, 并给出理由.

**解答** 由 OLS 估计量可得

$$\tilde{\beta}_1 = \frac{\text{cov}(X_1, Y)}{\text{var}(X_1)} = \hat{\beta}_1 + \hat{\beta}_2 \frac{\text{cov}(X_1, X_2)}{\text{var}(X_1)}$$

因为  $\beta_2^o = 0$ , 故  $\mathbb{E}[\tilde{\beta}_1] = \mathbb{E}[\hat{\beta}_1] = \beta_1^o$ .

在给定的假设下,  $\tilde{\beta}_1$  的条件方差为

$$\text{var}(\tilde{\beta}_1 | \mathbf{X}) = \frac{\sigma^2}{\sum_{t=1}^n (X_{1t} - \bar{X}_1)^2}$$

再根据问题 3.5 的结论,  $\hat{\beta}_1$  的条件方差为

$$\text{var}(\hat{\beta}_1 | \mathbf{X}) = \frac{\sigma^2}{(1 - \hat{r}^2) \sum_{t=1}^n (X_{1t} - \bar{X}_1)^2}$$

其中  $\hat{r}^2$  是  $X_1$  和  $X_2$  之间的相关系数的平方, 在假设 3.3(1) 的条件下不存在完全多重共线性, 故  $0 \leq \hat{r}^2 < 1$ . 因此

$$\text{var}(\tilde{\beta}_1 | \mathbf{X}) \leq \text{var}(\hat{\beta}_1 | \mathbf{X})$$

当且仅当  $X_1$  和  $X_2$  不相关时取等号. 因此  $\tilde{\beta}_1$  比  $\hat{\beta}_1$  更有效.

□

**3.16** 假定假设 3.1–3.3(1) 及以下假设 3.6' 成立.

**假设 3.6'**:  $\varepsilon | \mathbf{X} \sim N(\mathbf{0}, V)$ , 其中  $V = V(\mathbf{X})$  是一个已知的  $n \times n$  对称, 有限与正定的矩阵.

与假设 3.6 相比, 假设 3.6' 下  $\text{var}(\varepsilon | \mathbf{X}) = V$  完全已知, 没有未知常数  $\sigma^2$ . 定义 GLS 估计量为  $\hat{\beta}^* = (\mathbf{X}'V^{-1}\mathbf{X})^{-1}\mathbf{X}'V^{-1}Y$ .

(1)  $\hat{\beta}^*$  是 BLUE 吗?

(2) 令  $\mathbf{X}^* = C\mathbf{X}$ ,  $s^{*2} = e^{*'}e^*/(n - K)$ , 其中  $e^* = Y - \mathbf{X}^*\hat{\beta}^*$ ,  $C'C = V^{-1}$ . 通常的  $t$  检验和  $F$  检验定义如下

$$T^* = \frac{R\hat{\beta}^* - r}{\sqrt{s^{*2}R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R'}}, J = 1$$

$$F^* = \frac{(R\hat{\beta}^* - r)'[R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R']^{-1}(R\hat{\beta}^* - r)/J}{s^{*2}}, J > 1$$

在原假设  $\mathbb{H}_0: R\beta^o = r$  下, 它们分别服从  $t_{n-K}$  和  $F_{J, n-K}$  分布吗? 请解释.

(3) 构造两个新的检验统计量

$$\tilde{T}^* = \frac{R\hat{\beta}^* - r}{\sqrt{R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R'}}, J = 1$$

$$\tilde{W}^* = (R\hat{\beta}^* - r)'[R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R']^{-1}(R\hat{\beta}^* - r), J > 1$$

在原假设  $\mathbb{H}_0: R\beta^o = r$  下, 这两个检验统计量分别服从什么分布? 请解释.

(4) 在相同的显著性水平下,  $(T^*, F^*)$  和  $(\tilde{T}^*, \tilde{W}^*)$  这两个检验, 哪个更为有效, 也就是说, 哪个检验有更大的概率拒绝错误的原假设  $\mathbb{H}_0: R\beta^o = r$ ?

**解答** (1) 容易  $\hat{\beta}^*$  是  $Y$  的线性无偏估计量, 再根据问题 3.13 的结论,  $\hat{\beta}^*$  是 BLUE.

(2) 易证  $\text{var}(\hat{\beta}^*|\mathbf{X}) = (\mathbf{X}'V^{-1}\mathbf{X})^{-1} = (\mathbf{X}^{*'}\mathbf{X}^*)^{-1}$ , 可以得到

$$\hat{\beta}^* - \beta^o|\mathbf{X} \sim N(\mathbf{0}, (\mathbf{X}^{*'}\mathbf{X}^*)^{-1})$$

在原假设  $\mathbb{H}_0: R\beta^o = r$  下

$$R\hat{\beta}^* - r|\mathbf{X} \sim N(\mathbf{0}, R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R') \quad (3.8)$$

因此

$$\frac{R\hat{\beta}^* - r}{\sqrt{R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R'}} \sim N(\mathbf{0}, I)$$

其中  $I$  为  $J \times J$  单位矩阵. 特别地, 当  $J = 1$  时, 上式服从标准正态分布.

由定理 3.9 与定理 3.17,  $(n-K)s^{*2} \sim \chi_{n-K}^2$  且  $\text{cov}(\hat{\beta}^*, e^*) = \mathbf{0}$ . 当  $J = 1$  时, 根据  $t$  分布的定义容易得到

$$T^* = \frac{N(0, 1)}{\sqrt{\chi_{n-K}^2/(n-K)}} \sim t_{n-K}$$

再根据式 (3.8) 易证

$$(R\hat{\beta}^* - r)'[R(\mathbf{X}^{*'}\mathbf{X}^*)^{-1}R']^{-1}(R\hat{\beta}^* - r) \sim \chi_J^2$$

故而

$$F^* = \frac{\chi_J^2/J}{\chi_{n-K}^2/(n-K)} \sim F_{J, n-K}$$

因此在原假设  $\mathbb{H}_0: R\beta^o = r$  下,  $T^*$  和  $F^*$  分别服从  $t_{n-K}$  和  $F_{J, n-K}$  分布.

(3) 第 (2) 问已证当  $J = 1$  时  $\tilde{T}^*$  服从标准正态分布. 根据引理 3.11 和式 (3.8), 易证  $\tilde{W}^*$  服从  $\chi_J^2$  分布.

(4) 由于  $t$  分布相比标准正态分布具有厚尾的特点, 在相同显著性水平上, 标准正态分布的临界值小于  $t$  分布的临界值, 因此  $\tilde{T}^*$  比  $T^*$  有效.

---

并且当  $n \rightarrow \infty$  时,  $J \cdot F^* \xrightarrow{d} \chi_J^2$ , 在有限样本和相同显著性水平下,  $\chi_J^2$  分布的临界值小于渐近  $\chi_J^2$  的临界值, 因此  $\tilde{W}^*$  比  $F^*$  有效.

□

## 独立同分布随机样本的线性回归模型

4.1 给定假设 3.1, 3.3 和 3.5, 证明:

(1)  $s^2 \xrightarrow{p} \sigma^2$ .

(2)  $s \xrightarrow{p} \sigma$ .

**解答** (1) 在给定假设成立下,  $(n-K) \frac{s^2}{\sigma^2} \Big| \mathbf{X} \sim \chi_{n-K}^2$ . 进而有

$$\mathbb{E}[s^2 | \mathbf{X}] = \sigma^2$$

$$\text{var}(s^2 | \mathbf{X}) = \frac{2\sigma^4}{n-K}$$

当  $n \rightarrow \infty$  时,  $\mathbb{E}||s^2 - \sigma^2||^2 = \frac{2\sigma^4}{n-K} \rightarrow 0$  并不依赖于  $\mathbf{X}$ . 因此  $s^2 \xrightarrow{q.m.} \sigma^2$ , 由引理 4.3 可知  $s^2 \xrightarrow{p} \sigma^2$ .

(2) 由于  $g(\cdot) = \sqrt{\cdot}$  连续, 根据连续映射定理有  $s \xrightarrow{p} \sigma$ .

□

4.2 令  $\{Z_t\}_{t=1}^n$  为来自均值为  $\mu$ , 方差为  $\sigma^2$  的总体的一个独立同分布随机样本, 并定义  $\bar{Z}_n = n^{-1} \sum_{t=1}^n Z_t$ . 证明:

$$\mathbb{E} \left[ \frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma} \right] = 0$$
$$\text{var} \left[ \frac{\sqrt{n}(\bar{Z}_n - \mu)}{\sigma} \right] = 1$$

**解答** 易证  $\mathbb{E}[\bar{Z}_n] = \mu$ ,  $\text{var}(\bar{Z}_n) = \sigma^2/n$ . 原式成立是显然的.

□

4.3 假设一个随机变量序列  $\{Z_n, n = 1, 2, \dots\}$ , 其中  $Z_n$  的概率分布为

$$\begin{array}{cc} Z_n & \begin{array}{cc} \frac{1}{n} & n \end{array} \\ P_{Z_n} & \begin{array}{cc} 1 - \frac{1}{n} & \frac{1}{n} \end{array} \end{array}$$

(1)  $\{Z_n\}$  是否依均方收敛于 0? 给出理由.

(2)  $\{Z_n\}$  是否依概率收敛于 0? 给出理由.

**解答** (1) 当  $n \rightarrow \infty$  时

$$\mathbb{E}||Z_n - 0||^2 = n + \frac{1}{n^2} - \frac{1}{n^3} \rightarrow \infty$$

因此  $\{Z_n\}$  不依均方收敛于 0.

(2)  $\forall \epsilon > 0$ , 当  $n \rightarrow \infty$  时

$$\mathbb{P}[||Z_n - 0|| > \epsilon] = \mathbb{P}[Z_n = n] = \frac{1}{n} \rightarrow 0$$



因此  $\{Z_n\}$  依概率收敛于 0.

□

**4.4** 设样本空间  $\Omega$  是闭区间  $[0, 1]$ , 且其基本元素  $\omega \in \Omega$  服从均匀分布. 定义随机变量  $Z(\omega) = \omega, \omega \in [0, 1]$ . 另外, 对于  $n = 1, 2, \dots$ , 定义随机变量序列

$$Z_n(\omega) = \begin{cases} \omega + \omega^n, & \omega \in [0, 1 - n^{-1}] \\ \omega + 1, & \omega \in (1 - n^{-1}, 1] \end{cases}$$

(1)  $\{Z_n\}$  是否依均方收敛于  $Z$ ?

(2)  $\{Z_n\}$  是否依概率收敛于  $Z$ ?

(3)  $\{Z_n\}$  是否几乎必然收敛于  $Z$ ?

**解答** (1) 由于当  $n \rightarrow \infty$  时

$$\begin{aligned} \mathbb{E}\|Z_n - Z\|^2 &= \int_0^{1-n^{-1}} \omega^{2n} d\omega + \int_{1-n^{-1}}^1 d\omega \\ &= \frac{1}{2n+1} \left(1 - \frac{1}{n}\right)^{2n+1} + \frac{1}{n} \rightarrow 0 \end{aligned}$$

因此  $\{Z_n\}$  依均方收敛于  $Z$ .

(2) 首先  $\omega \in [0, 1], \forall 0 < \epsilon < 1$ , 当  $n \rightarrow \infty$  时

$$\mathbb{P}[\|Z_n - Z\| \leq \epsilon] = \mathbb{P}[Z_n = \omega + \omega^n] = 1 - n^{-1} \rightarrow 1$$

因此  $\{Z_n\}$  依概率收敛于  $Z$ .

(3) 令  $A = \left\{\omega \in \Omega : \lim_{n \rightarrow \infty} Z_n = Z\right\}$ , 证明  $Z_n \xrightarrow{a.s.} Z$  等价于证明  $\mathbb{P}[A] = 1$ .

对  $\omega \in [0, 1)$  都有

$$\begin{aligned} \mathbb{P}\left[\lim_{n \rightarrow \infty} \|Z_n - Z\| = 0\right] &= (1 - n^{-1})\mathbb{P}\left[\lim_{n \rightarrow \infty} \omega^n = 0\right] \\ &\quad + n^{-1}\mathbb{P}\left[\lim_{n \rightarrow \infty} 1 = 0\right] = 1 \end{aligned}$$

也即  $A = \Omega \setminus \{1\}$ , 又因为  $\{1\}$  是零测集, 故  $\mathbb{P}[A] = 1$ . 因此  $\{Z_n\}$  几乎必然收敛于  $Z$ .

□

**4.5** 假设  $g(\cdot)$  是一个实值连续函数,  $\{Z_n, n = 1, 2, \dots\}$  是一个实值随机变量序列, 并依概率收敛于随机变量  $Z$ . 证明:  $g(Z_n) \xrightarrow{P} g(Z)$ .

**解答** 由于  $g(\cdot)$  在  $Z$  处连续, 故  $\forall \epsilon > 0, \exists \delta > 0$ , 当  $\|Z_n - Z\| < \delta$  时有  $\|g(Z_n) - g(Z)\| < \epsilon$ . 可以得到

$$\mathbb{P}[\|Z_n - Z\| < \delta] \leq \mathbb{P}[\|g(Z_n) - g(Z)\| < \epsilon]$$

又因为  $Z_n \xrightarrow{P} Z$ , 即  $\forall \epsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}[\|Z_n - Z\| < \epsilon] = 1$ , 进而上式左端在  $n \rightarrow \infty$  时也收敛于 1, 因此

$$\lim_{n \rightarrow \infty} \mathbb{P}[\|g(Z_n) - g(Z)\| < \epsilon] = 1$$

也即  $g(Z_n) \xrightarrow{P} g(Z)$ .

□

**4.6** 假设随机过程  $\{Y_t, X_t'\}$  满足下列假设:

假设 4.6.1  $\{Y_t, X_t'\}_{t=1}^n$  是一个可观测的独立同分布随机样本, 并且

$$Y_t = X_t' \beta^o + \varepsilon_t, \quad t = 1, \dots, n$$

其中  $\beta^o$  是  $K \times 1$  的为未知参数向量,  $\varepsilon_t$  是不可观测扰动项.

假设 4.6.2  $K \times K$  矩阵  $\mathbb{E}[X_t X_t'] = Q$  是有限, 对称和非奇异的.

假设 4.6.3 (i)  $\mathbb{E}[X_t \varepsilon_t] = 0$ ; (ii)  $\mathbb{E}[\varepsilon_t^2 | X_t] \neq \sigma^2$ ; (iii) 对所有  $j \in \{0, \dots, k\}$ ,  $\mathbb{E}[X_{jt}^4] \leq C$ ,  $\mathbb{E}[\varepsilon_t^4] \leq C$ , 这里  $C$  是一个常数.

(1) 证明:  $\hat{\beta} \xrightarrow{p} \beta^o$ .

(2) 证明:  $\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(0, Q^{-1} V Q^{-1})$ , 其中  $V = \mathbb{E}[X_t X_t' \varepsilon_t^2]$ .

(3) 证明:  $\sqrt{n} \hat{\beta}$  的渐近方差估计量

$$\hat{Q}^{-1} \hat{V} \hat{Q}^{-1} \xrightarrow{p} Q^{-1} V Q^{-1}$$

其中  $\hat{Q} = n^{-1} \sum_{t=1}^n X_t X_t'$ ,  $\hat{V} = n^{-1} \sum_{t=1}^n X_t X_t' \varepsilon_t^2$ . 这称为 White 异方差一致性方差 - 协方差矩阵估计量.

(4) 考虑检验原假设  $H_0: R\beta^o = r$ , 这里  $R$  是一个  $J \times K$  满秩矩阵,  $r$  是  $J \times 1$  向量,  $J \leq K$ . 是否有  $J \cdot F \xrightarrow{d} \chi_J^2$ ? 这里

$$F = \frac{(R\hat{\beta} - r)' [R(\mathbf{X}'\mathbf{X})^{-1} R']^{-1} (R\hat{\beta} - r) / J}{s^2}$$

是经典  $F$  检验统计量. 如果成立, 给出理由; 如果不成立, 给出一个新的依分布收敛于  $\chi_J^2$  的检验统计量.

解答 (1) 由 OLS 估计量  $\hat{\beta} = \hat{Q}^{-1} n^{-1} \sum_{t=1}^n X_t Y_t$ , 可以得到

$$\hat{\beta} = \beta^o + \hat{Q}^{-1} n^{-1} \sum_{t=1}^n X_t \varepsilon_t$$

由于对  $\forall j \in \{0, \dots, k\}$ , 存在充分大的实数  $C$ , 使得  $\mathbb{E}[X_{jt}^4] \leq C$ ,  $\mathbb{E}[\varepsilon_t^4] \leq C$  成立, 根据 Jensen 不等式<sup>6</sup>

$$\begin{aligned} (\mathbb{E}[X_{jt}^2])^2 &\leq \mathbb{E}[X_{jt}^4] \leq C \\ (\mathbb{E}[\varepsilon_t^2])^2 &\leq \mathbb{E}[\varepsilon_t^4] \leq C \end{aligned}$$

因此二阶矩  $\mathbb{E}[X_{jt}^2]$  和  $\mathbb{E}[\varepsilon_t^2]$  均有限, 假定它们也都小于常数  $C$ , 再根据 Cauchy-Schwarz 不等式<sup>7</sup>可得

$$\mathbb{E}|X_{jt} \varepsilon_t| \leq (\mathbb{E}[X_{jt}^2])^{\frac{1}{2}} (\mathbb{E}[\varepsilon_t^2])^{\frac{1}{2}} \leq C$$

<sup>6</sup> 设  $(\Omega, \mathcal{F}, \mu)$  为概率空间,  $\mu(\Omega) = 1$ ,  $f: \Omega \rightarrow \mathbb{R} \in L^1(\Omega, \mu)$ , 且  $\varphi: \mathbb{R} \rightarrow \mathbb{R}$  为凸函数, 那么

$$\varphi\left(\int_{\Omega} f d\mu\right) \leq \int_{\Omega} \varphi \circ f d\mu$$

<sup>7</sup> 对于任意随机矩阵  $X, Y \in \mathbb{R}^{m \times n}$ ,  $\mathbb{E}\|X'Y\| \leq (\mathbb{E}\|X\|^2)^{\frac{1}{2}} (\mathbb{E}\|Y\|^2)^{\frac{1}{2}}$ .

故 WLLN 的使用条件成立, 进一步可得

$$n^{-1} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{p} \mathbb{E}[X_t \varepsilon_t] = \mathbf{0}$$

同理可证对任意  $j, l \in \{0, \dots, k\}$ , 均有  $\mathbb{E}[X_{jt} X_{lt}] \leq C$ , 再次利用 WLLN 可得

$$\hat{Q} = n^{-1} \sum_{t=1}^n X_t X_t' \xrightarrow{p} Q = \mathbb{E}[X_t X_t']$$

根据连续映射定理有  $\hat{Q}^{-1} \xrightarrow{p} Q^{-1}$ , 因此

$$\begin{aligned} \hat{\beta} - \beta^o &= \hat{Q}^{-1} n^{-1} \sum_{t=1}^n X_t \varepsilon_t \\ &\xrightarrow{p} Q^{-1} \cdot \mathbb{E}[X_t \varepsilon_t] = \mathbf{0} \end{aligned}$$

也即  $\hat{\beta} \xrightarrow{p} \beta^o$ .

(2) 定义  $Z_t = X_t \varepsilon_t$ , 并且  $\bar{Z}_n = n^{-1} \sum_{t=1}^n X_t \varepsilon_t$ , 由独立同分布随机样本的中心极限定理可得

$$\sqrt{n} \bar{Z}_n = n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{d} Z \sim N(\mathbf{0}, V)$$

其中  $V = \mathbb{E}[X_t X_t' \varepsilon_t^2]$ . 再根据  $\hat{Q}^{-1} \xrightarrow{p} Q^{-1}$  以及 Slutsky 定理, 可以得到

$$\begin{aligned} \sqrt{n}(\hat{\beta} - \beta^o) &= \hat{Q}^{-1} n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \\ &\xrightarrow{d} Q^{-1} Z \sim N(\mathbf{0}, Q^{-1} V Q^{-1}) \end{aligned}$$

因此  $\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, Q^{-1} V Q^{-1})$ , 其中  $V = \mathbb{E}[X_t X_t' \varepsilon_t^2]$ ,  $Q = \mathbb{E}[X_t X_t']$ .

(3) 因为  $e_t = \varepsilon_t - X_t'(\hat{\beta} - \beta^o)$ , 可以得到

$$\begin{aligned} \hat{V} &= n^{-1} \sum_{t=1}^n X_t X_t' \varepsilon_t^2 \\ &\quad - 2n^{-1}(\hat{\beta} - \beta^o)' \sum_{t=1}^n X_t X_t' X_t \varepsilon_t \\ &\quad + n^{-1}(\hat{\beta} - \beta^o)' \left( \sum_{t=1}^n X_t X_t' X_t X_t' \right) (\hat{\beta} - \beta^o) \end{aligned} \tag{4.1}$$

需要对这三项进行渐近分析.

对任意  $i, j, l, m \in \{0, \dots, k\}$ , 由 Cauchy-Schwarz 不等式可得

$$\mathbb{E}[X_{it} X_{jt} \varepsilon_t^2] \leq (\mathbb{E}[X_{it}^2 X_{jt}^2])^{\frac{1}{2}} (\mathbb{E}[\varepsilon_t^4])^{\frac{1}{2}} \leq C$$

当  $n \rightarrow \infty$  时, 根据 WLLN 可得式 (4.1) 的第一项为

$$n^{-1} \sum_{t=1}^n X_t X_t' \varepsilon_t^2 \xrightarrow{p} \mathbb{E}[X_t X_t' \varepsilon_t^2] = V$$

由假设 4.6.3 以及 WLLN, 可以得到

$$n^{-1} \sum_{t=1}^n X_{it} X_{jt} X_{lt} X_{mt} \xrightarrow{p} \mathbb{E}[X_{it} X_{jt} X_{lt} X_{mt}] = O(1)$$

再根据 Hölder 不等式<sup>8</sup>, 以下矩条件成立

$$\mathbb{E}|X_{it} X_{jt} X_{lt} \varepsilon_t| \leq (\mathbb{E}|X_{it}^{\frac{4}{3}} X_{jt}^{\frac{4}{3}} X_{lt}^{\frac{4}{3}}|)^{\frac{3}{4}} \mathbb{E}[\varepsilon_t^4]^{\frac{1}{4}} \leq C$$

故而由 WLLN 可得

$$n^{-1} \sum_{t=1}^n X_{it} X_{jt} X_{lt} \varepsilon_t \xrightarrow{p} \mathbb{E}[X_{it} X_{jt} X_{lt} \varepsilon_t] = O(1)$$

当  $n \rightarrow \infty$  时,  $\hat{\beta} \xrightarrow{p} \beta^o$ , 故式 (4.1) 的第二项和第三项为  $o_p(1)$ , 因此  $\hat{V} \xrightarrow{p} V$ , 由连续映射定理, 即可得到  $\hat{Q}^{-1} \hat{V} \hat{Q}^{-1} \xrightarrow{p} Q^{-1} V Q^{-1}$ .

(4) 由于存在条件异方差, 在原假设  $\mathbb{H}_0: R\beta^o = r$  成立时,  $J \cdot F$  统计量不再服从渐近  $\chi_J^2$  分布.

在原假设  $\mathbb{H}_0: R\beta^o = r$  下, 可以得到

$$\sqrt{n}(R\hat{\beta} - r) | \mathbf{X} \xrightarrow{d} N(\mathbf{0}, RQ^{-1}VQ^{-1}R')$$

由引理 3.11 可得

$$W = n(R\hat{\beta} - r)'[R\hat{Q}^{-1}\hat{V}\hat{Q}^{-1}R']^{-1}(R\hat{\beta} - r) \xrightarrow{d} \chi_J^2$$

因此上述稳健 Wald 统计量即为所求.

□

#### 4.7 假定下列假设成立

**假设 4.7.1**  $\{Y_t, X_t'\}_{t=1}^n$  是一个可观测的独立同分布随机样本, 并且

$$Y_t = X_t' \beta^o + \varepsilon_t, \quad t = 1, \dots, n$$

其中  $\beta^o$  是  $K \times 1$  的为未知参数向量,  $\varepsilon_t$  是不可观测扰动项.

**假设 4.7.2**  $\mathbb{E}[\varepsilon_t | X_t] = 0$ .

**假设 4.7.3** (i)  $W_t = W(X_t)$  是  $X_t$  的一个非负函数;  $K \times K$  阶矩阵  $\mathbb{E}[X_t W_t X_t'] = Q_W$  是有限, 对称与非奇异的; (iii) 对所有的  $0 \leq j \leq k$ ,  $\mathbb{E}[\varepsilon_t^4] \leq C$ ,  $\mathbb{E}[W_t^8] \leq C$ ,  $\mathbb{E}[X_{jt}^8] \leq C$ , 这里  $C$  是一有限常数.

**假设 4.7.4**  $V_W = \mathbb{E}[W_t^2 X_t X_t' \varepsilon_t^2]$  是一个  $K \times K$  有限, 对称与非奇异的矩阵.

考虑  $\beta^o$  的加权最小二乘 (weighed least square, WLS) 估计量

$$\hat{\beta}_W = \left( n^{-1} \sum_{t=1}^n X_t W_t X_t' \right)^{-1} n^{-1} \sum_{t=1}^n X_t W_t Y_t$$

(1) 证明:  $\hat{\beta}_W$  是下面加权最小二乘问题的最优解

$$\min_{\beta \in \mathbb{R}^K} \sum_{t=1}^n W_t (Y_t - X_t' \beta)^2$$

<sup>8</sup>如果  $p > 1, q > 1$ , 且  $1/p + 1/q = 1$ , 则对于任意随机矩阵  $X, Y \in \mathbb{R}^{m \times n}$ ,  $\mathbb{E}||X'Y|| \leq (\mathbb{E}||X||^p)^{\frac{1}{p}} (\mathbb{E}||Y||^q)^{\frac{1}{q}}$ .

(2) 证明:  $\hat{\beta}_W$  是  $\beta^o$  的一致估计.

(3) 证明:  $\sqrt{n}(\hat{\beta}_W - \beta^o) \xrightarrow{d} N(\mathbf{0}, \Omega_W)$ , 其中  $\Omega_W$  是一个  $K \times K$  有限正定矩阵. 分别在条件同方差 ( $\mathbb{E}[\varepsilon_t^2|X_t] = \sigma^2$ ) 和条件异方差 ( $\mathbb{E}[\varepsilon_t^2|X_t] \neq \sigma^2$ ) 情形下推导出  $\Omega_W$  的表达式.

(4) 分别在条件同方差和条件异方差的情形下构建  $\Omega_W$  的估计量  $\hat{\Omega}_W$ , 并证明  $\hat{\Omega}_W$  是  $\Omega_W$  的一致估计量.

(5) 分别在条件同方差和条件异方差的情形下, 构造一个统计量以检验原假设  $\mathbb{H}_0: R\beta^o = r$ , 其中  $R$  是  $J \times K$  矩阵,  $r$  是一个  $J \times 1$  向量, 且  $J \leq K$ . 推导出在每种情形下, 检验统计量在原假设  $\mathbb{H}_0: R\beta^o = r$  成立时的渐近分布.

(6) 假设存在条件异方差且  $\mathbb{E}[\varepsilon_t^2|X_t] = \sigma^2(X_t)$  已知, 并选择  $W_t = \sigma^{-1}(X_t)$ . 构造一个统计量以检验原假设  $\mathbb{H}_0: R\beta^o = r$ , 并推导出检验统计量在原假设  $\mathbb{H}_0: R\beta^o = r$  成立时的渐近分布.

**解答** (1) 定义残差平方和

$$\text{SSR} = \sum_{t=1}^n W_t(Y_t - X_t'\beta)^2$$

考虑一阶条件

$$\frac{\partial \text{SSR}}{\partial \beta} = -2 \sum_{t=1}^n X_t W_t(Y_t - X_t'\beta) = 0$$

解得 WLS 估计量

$$\hat{\beta}_W = \left( n^{-1} \sum_{t=1}^n X_t W_t X_t' \right)^{-1} n^{-1} \sum_{t=1}^n X_t W_t Y_t$$

继续考虑二阶条件

$$\frac{\partial^2 \text{SSR}}{\partial \beta \partial \beta'} = \sum_{t=1}^n W_t X_t X_t' \sim \text{p.s.d.}$$

因此  $\hat{\beta}_W$  即加权最小二乘问题的最优解.

(2) 定义  $\hat{Q}_W = n^{-1} \sum_{t=1}^n X_t W_t X_t'$ , 由 WLS 估计量易证

$$\hat{\beta}_W - \beta^o = \hat{Q}_W^{-1} n^{-1} \sum_{t=1}^n X_t W_t \varepsilon_t$$

又因为对任意  $i, j \in \{0, \dots, k\}$ ,  $\mathbb{E}[\varepsilon_t^4] \leq C$ ,  $\mathbb{E}[W_t^8] \leq C$ ,  $\mathbb{E}[X_{jt}^8] \leq C$ , 根据 Cauchy-Schwarz 不等式可得

$$\begin{aligned} \mathbb{E}|X_{it} W_t X_{jt}| &\leq (\mathbb{E}[X_{it}^2 W_t^2])^{\frac{1}{2}} (\mathbb{E}[X_{jt}^2])^{\frac{1}{2}} \\ &\leq (\mathbb{E}[X_{it}^4])^{\frac{1}{4}} (\mathbb{E}[W_t^4])^{\frac{1}{4}} (\mathbb{E}[X_{jt}^2])^{\frac{1}{2}} = O(1) \end{aligned}$$

以及

$$\begin{aligned} \mathbb{E}|X_{jt} W_t \varepsilon_t| &\leq (\mathbb{E}[X_{jt}^2 W_t^2])^{\frac{1}{2}} (\mathbb{E}[\varepsilon_t^2])^{\frac{1}{2}} \\ &\leq (\mathbb{E}[X_{jt}^4])^{\frac{1}{4}} (\mathbb{E}[W_t^4])^{\frac{1}{4}} (\mathbb{E}[\varepsilon_t^2])^{\frac{1}{2}} = O(1) \end{aligned}$$

再使用 WLLN 和连续映射定理可得  $\hat{Q}_W^{-1} \xrightarrow{p} Q_W^{-1}$ , 以及

$$n^{-1} \sum_{t=1}^n X_t W_t \varepsilon_t \xrightarrow{p} \mathbb{E}[X_t W_t \varepsilon_t] = \mathbf{0}$$

其中  $\mathbb{E}[X_t W_t \varepsilon_t] = \mathbb{E}[X_t W_t \mathbb{E}[\varepsilon_t | X_t]] = \mathbf{0}$ . 当  $n \rightarrow \infty$  时

$$\begin{aligned} \hat{\beta}_W - \beta^o &= \hat{Q}_W^{-1} n^{-1} \sum_{t=1}^n X_t W_t \varepsilon_t \\ &\xrightarrow{p} Q_W^{-1} \cdot \mathbb{E}[X_t W_t \varepsilon_t] = \mathbf{0} \end{aligned}$$

因此  $\hat{\beta}_W$  是  $\beta^o$  的一致估计.

(3) 定义  $Z_t = X_t W_t \varepsilon_t$ , 并且  $\bar{Z}_n = n^{-1} \sum_{t=1}^n X_t W_t \varepsilon_t$ , 由独立同分布随机样本中心极限定理可得

$$\sqrt{n} \bar{Z}_n = n^{-\frac{1}{2}} \sum_{t=1}^n X_t W_t \varepsilon_t \xrightarrow{p} Z_W \sim N(\mathbf{0}, V_W)$$

其中  $V_W = \mathbb{E}[W_t^2 X_t X_t' \varepsilon_t^2]$ . 再根据  $\hat{Q}_W^{-1} \xrightarrow{p} Q_W^{-1}$  以及 Slutsky 定理, 可以得到

$$\begin{aligned} \sqrt{n}(\hat{\beta}_W - \beta^o) &= \hat{Q}_W^{-1} n^{-\frac{1}{2}} \sum_{t=1}^n X_t W_t \varepsilon_t \\ &\xrightarrow{d} Q_W^{-1} Z_W \sim N(\mathbf{0}, \Omega_W) \end{aligned}$$

其中  $\Omega_W = Q_W^{-1} V_W Q_W^{-1}$ ,  $Q_W = \mathbb{E}[X_t W_t X_t']$ .

假定条件同方差成立, 即  $\mathbb{E}[\varepsilon_t^2 | X_t] = \sigma^2$ , 则

$$V_W = \mathbb{E}[W_t^2 X_t X_t' \mathbb{E}[\varepsilon_t^2 | X_t]] = \sigma^2 \mathbb{E}[W_t^2 X_t X_t']$$

记  $V_{W_1} = \mathbb{E}[W_t^2 X_t X_t']$ , 此时  $\Omega_W = \sigma^2 Q_W^{-1} V_{W_1} Q_W^{-1}$ . 当存在条件异方差时

$$\Omega_W = Q_W^{-1} V_W Q_W^{-1}$$

不能继续化简.

(4.1) 条件同方差成立时, 考虑估计量  $\hat{\Omega}_W = s^2 \hat{Q}_W^{-1} \hat{V}_{W_1} \hat{Q}_W^{-1}$ . 其中

$$\begin{aligned} \hat{V}_{W_1} &= n^{-1} \sum_{t=1}^n W_t^2 X_t X_t' \\ s^2 &= \frac{1}{n-K} \sum_{t=1}^n e_t^2 \end{aligned}$$

由于  $\hat{Q}_W^{-1} \xrightarrow{p} Q_W^{-1}$  已在第 (2) 问中证明, 还需要证明  $\hat{V}_{W_1} \xrightarrow{p} V_{W_1}$  以及  $s^2 \xrightarrow{p} \sigma^2$ .

由 Cauchy-Schwarz 不等式, 对任意  $0 \leq j, l \leq k$  以下矩条件成立

$$\begin{aligned} \mathbb{E}[W_t^2 X_{jt} X_{lt}] &\leq (\mathbb{E}[W_t^4 X_{jt}^2])^{\frac{1}{2}} (\mathbb{E}[X_{lt}^2])^{\frac{1}{2}} \\ &\leq (\mathbb{E}[W_t^8])^{\frac{1}{4}} (\mathbb{E}[X_{jt}^4])^{\frac{1}{4}} (\mathbb{E}[X_{lt}^4])^{\frac{1}{2}} = O(1) \end{aligned}$$

再根据 WLLN 可得  $\hat{V}_{W_1} \xrightarrow{p} V_{W_1}$ .

根据  $e_t = \varepsilon_t - X_t'(\hat{\beta}_W - \beta^o)$  可得

$$\begin{aligned} s^2 &= \frac{n}{n-K} \left( n^{-1} \sum_{t=1}^n \varepsilon_t^2 \right) \\ &\quad - \frac{2n}{n-K} (\hat{\beta}_W - \beta^o)' \left( n^{-1} \sum_{t=1}^n X_t \varepsilon_t \right) \\ &\quad + \frac{n}{n-K} (\hat{\beta}_W - \beta^o)' \left( n^{-1} \sum_{t=1}^n X_t X_t' \right) (\hat{\beta}_W - \beta^o) \end{aligned}$$

当  $n \rightarrow \infty$  时,  $\frac{n}{n-K} \rightarrow 1$ ,  $\hat{\beta}_W \xrightarrow{p} \beta^o$ ,  $n^{-1} \sum_{t=1}^n \varepsilon_t^2 \xrightarrow{p} \mathbb{E}[\varepsilon^2] = \sigma^2$ , 并且  $\mathbb{E}|X_t \varepsilon_t|$  和  $\mathbb{E}|X_t X_t'|$  均有限, 最终有

$$s^2 \xrightarrow{p} \sigma^2$$

因此在条件同方差下, 由连续映射定理可得  $\hat{\Omega}_W = s^2 \hat{Q}_W^{-1} \hat{V}_W \hat{Q}_W^{-1} \xrightarrow{p} \Omega_W$ .

(4.2) 条件异方差成立时, 考虑估计量  $\hat{\Omega}_W = \hat{Q}_W^{-1} \hat{V}_W \hat{Q}_W^{-1}$ , 由于  $\hat{Q}_W^{-1} \xrightarrow{p} Q_W^{-1}$  已在第 (2) 问中证明, 还需要证明  $\hat{V}_W \xrightarrow{p} V_W$ , 其中  $\hat{V}_W = n^{-1} \sum_{t=1}^n W_t^2 X_t X_t' e_t^2$ .

由 Cauchy-Schwarz 不等式, 对任意  $0 \leq j, l \leq k$  以下矩条件成立

$$\begin{aligned} \mathbb{E}|W_t^2 X_{jt} X_{lt} \varepsilon_t^2| &\leq (\mathbb{E}[W_t^4 X_{jt}^2 X_{lt}^2])^{\frac{1}{2}} (\mathbb{E}[\varepsilon_t^4])^{\frac{1}{2}} \\ &\leq (\mathbb{E}[W_t^8])^{\frac{1}{4}} (\mathbb{E}[X_{jt}^4 X_{lt}^4])^{\frac{1}{4}} (\mathbb{E}[\varepsilon_t^4])^{\frac{1}{2}} \\ &\leq (\mathbb{E}[W_t^8])^{\frac{1}{4}} (\mathbb{E}[X_{jt}^8])^{\frac{1}{8}} (\mathbb{E}[X_{lt}^8])^{\frac{1}{8}} (\mathbb{E}[\varepsilon_t^4])^{\frac{1}{2}} = O(1) \end{aligned}$$

故  $n^{-1} \sum_{t=1}^n W_t^2 X_t X_t' \varepsilon_t^2 \xrightarrow{p} V_W$ .

利用  $e_t = \varepsilon_t - X_t'(\hat{\beta}_W - \beta^o)$  将估计量  $\hat{V}_W$  展开后得

$$\begin{aligned} \hat{V}_W &= n^{-1} \sum_{t=1}^n W_t^2 X_t X_t' \varepsilon_t \\ &\quad - 2n^{-1} (\hat{\beta}_W - \beta^o)' \left( \sum_{t=1}^n W_t^2 X_t X_t' X_t \varepsilon_t \right) \\ &\quad + n^{-1} (\hat{\beta}_W - \beta^o)' \left( \sum_{t=1}^n W_t^2 X_t X_t' X_t X_t' \right) (\hat{\beta}_W - \beta^o) \end{aligned}$$

同理易证  $\mathbb{E}|W_t^2 X_t X_t' X_t \varepsilon_t|$  和  $\mathbb{E}|W_t^2 X_t X_t' X_t X_t'|$  均是有限的. 因此当  $n \rightarrow \infty$  时, 上式后两项是  $o_p(1)$ , 也即  $\hat{V}_W \xrightarrow{p} V_W$ , 再由连续映射定理可得  $\hat{\Omega}_W = \hat{Q}_W^{-1} \hat{V}_W \hat{Q}_W^{-1} \xrightarrow{p} \Omega_W$ .

(5) 在原假设  $\mathbb{H}_0: R\beta^o = r$  下, 如果条件同方差成立, 易知

$$\sqrt{n}(R\hat{\beta}_W - r) | \mathbf{X} \sim N(0, \sigma^2 R Q_W^{-1} V_W Q_W^{-1} R')$$

由 Slutsky 定理, 当  $J = 1$  时

$$T_W = \frac{\sqrt{n}(R\hat{\beta}_W - r)}{\sqrt{s^2 R \hat{Q}_W^{-1} \hat{V}_W Q_W^{-1} R'}} \xrightarrow{d} N(0, 1)$$



当  $J > 1$  时

$$J \cdot F_W = n(R\hat{\beta}_W - r)'[s^2 R \hat{Q}_W^{-1} \hat{V}_{W_1} \hat{Q}_W^{-1} R']^{-1}(R\hat{\beta}_W - r) \xrightarrow{d} \chi_J^2$$

如果存在条件异方差, 易知

$$\sqrt{n}(R\hat{\beta}_W - r)|\mathbf{X} \sim N(0, Q_W^{-1} V_W Q_W^{-1})$$

由 Slutsky 定理, 当  $J = 1$  时

$$T_W^* = \frac{\sqrt{n}(R\hat{\beta}_W - r)}{\sqrt{R \hat{Q}_W^{-1} \hat{V}_W \hat{Q}_W^{-1} R'}} \xrightarrow{d} N(0, 1)$$

当  $J > 1$  时

$$W_W^* = n(R\hat{\beta}_W - r)'[R \hat{Q}_W^{-1} \hat{V}_W \hat{Q}_W^{-1} R']^{-1}(R\hat{\beta}_W - r) \xrightarrow{d} \chi_J^2$$

(6) 由题设条件易知  $V = \mathbb{E}[W_t^2 X_t X_t' \varepsilon_t^2] = \mathbb{E}[X_t X_t']$ , 此时  $V$  的一致估计量为  $\frac{\mathbf{X}'\mathbf{X}}{n}$ . 根据第 (5) 问的结论可知, 当  $J = 1$  时

$$\tilde{T}_W = \frac{n(R\hat{\beta}_W - r)}{\sqrt{R \hat{Q}_W^{-1} \mathbf{X}'\mathbf{X} \hat{Q}_W^{-1} R'}} \xrightarrow{d} N(0, 1)$$

当  $J > 1$  时

$$\tilde{W}_W = n^2(R\hat{\beta}_W - r)'[R \hat{Q}_W^{-1} \mathbf{X}'\mathbf{X} \hat{Q}_W^{-1} R']^{-1}(R\hat{\beta}_W - r) \xrightarrow{d} \chi_J^2$$

□

**4.8** 考虑以下线性回归模型的条件同方差检验问题

$$Y_t = X_t' \beta^0 + \varepsilon_t$$

其中  $X_t$  是  $K \times 1$  随机变量, 包括一个截距项和经济解释变量. 为检验条件同方差, 考虑以下辅助回归

$$\begin{aligned} \varepsilon_t^2 &= \text{vech}(X_t, X_t')' \gamma + v_t \\ &= U_t' \gamma + v_t \end{aligned}$$

其中  $U_t = \text{vech}(X_t, X_t')$ .

证明: 当  $\mathbb{H}_0 : \mathbb{E}[\varepsilon_t^2 | X_t] = \sigma^2$  成立时, (1)  $\mathbb{E}[v_t | X_t] = 0$ ; (2)  $\mathbb{E}[v_t^2 | X_t] = \sigma_v^2$  当且仅当  $\mathbb{E}[\varepsilon_t^4 | X_t] = \mu_4$ , 这里  $\mu_4$  为一常数.

**解答** (1) 当  $\mathbb{H}_0 : \mathbb{E}[\varepsilon_t^2 | X_t] = \sigma^2$  成立时

$$\varepsilon_t^2 = \gamma_0^o + v_t$$

$\mathbb{E}[\varepsilon_t^2 | X_t] = \sigma^2 + \mathbb{E}[v_t | X_t]$ , 其中  $\gamma_0^o = \sigma^2$ , 因此  $\mathbb{E}[v_t | X_t] = 0$ .

(2) 进一步可得

$$\begin{aligned} \mathbb{E}[\varepsilon_t^4 | X_t] &= \mathbb{E}[(\gamma_0^o + v_t)^2 | X_t] \\ &= \sigma^4 + 2\sigma^2 \mathbb{E}[v_t | X_t] + \mathbb{E}[v_t^2 | X_t] \\ &= \sigma^4 + \mathbb{E}[v_t^2 | X_t] = \mu_4 \end{aligned}$$

因此  $\mathbb{E}[v_t^2|X_t] = \sigma_v^2$  当且仅当  $\mathbb{E}[\varepsilon_t^4|X_t] = \mu_4$ .

□

**4.9** 考虑以下线性回归模型的条件同方差检验问题

$$Y_t = X_t' \beta^o + \varepsilon_t$$

其中  $X_t$  是  $K \times 1$  随机变量, 包括一个截距项和经济解释变量. 为检验条件同方差, 考虑以下辅助回归

$$\begin{aligned} \varepsilon_t^2 &= \text{vech}(X_t, X_t')' \gamma + v_t \\ &= U_t' \gamma + v_t \end{aligned} \quad (4.2)$$

假定假设 4.1–4.4 及 4.7 成立, 且  $\mathbb{E}[\varepsilon_t^4|X_t] \neq \mu_4$ , 即  $\mathbb{E}[\varepsilon_t^4|X_t]$  是  $X_t$  的非负函数.

(1) 证明: 当  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t^2|X_t] = \sigma^2$  成立时,  $\text{var}(v_t|X_t) \neq \sigma_v^2$ . 即辅助回归模型的随机扰动项  $v_t$  存在条件异方差.

(2) 假设可以直接观察到  $\{\varepsilon_t\}_{t=1}^n$ . 构造一个检验  $\varepsilon_t$  是否存在条件同方差的渐近有效统计量, 并推导其在原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t^2|X_t] = \sigma^2$  成立时的渐近分布.

**解答** (1) 当  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t^2|X_t] = \sigma^2$  成立时, 根据上一题的结论有  $\mathbb{E}[v_t|X_t] = 0$ , 并且

$$\text{var}(v_t^2|X_t) = \mathbb{E}[v_t^2|X_t] = \mathbb{E}[\varepsilon_t^4|X_t] - \sigma^4$$

由于  $\mathbb{E}[\varepsilon_t^4|X_t]$  是  $X_t$  的非负函数, 故  $\text{var}(v_t|X_t)$  也是  $X_t$  的非负函数, 也即随机扰动项  $v_t$  存在条件异方差.

(2) 此时要检验的原假设为  $\mathbb{H}_0: R\gamma^o = \mathbf{0}$ , 其中  $R = (\mathbf{0}, I)$  为一个  $(J+1) \times J$  矩阵, 并且  $J = \frac{K(K+1)}{2} - 1$ . 由于  $v_t$  存在条件异方差, 故  $(n-K)R^2$  检验不再适用.

设  $\hat{\gamma}$  为  $\gamma^o$  的 OLS 估计量, 定义  $\hat{Q}_U = n^{-1} \sum_{t=1}^n U_t U_t'$ , 易证  $\hat{Q}_U^{-1} \xrightarrow{p} Q_U^{-1}$ , 以及

$$\sqrt{n}(\hat{\gamma} - \gamma^o) \xrightarrow{d} N(\mathbf{0}, Q_U^{-1} V_U Q_U^{-1})$$

其中  $Q_U = \mathbb{E}[U_t U_t']$ ,  $V_U = \mathbb{E}[U_t U_t' v_t^2]$ . 在  $\mathbb{H}_0: R\gamma^o = \mathbf{0}$  成立的条件下

$$\sqrt{n}R\hat{\gamma} \xrightarrow{d} N(\mathbf{0}, RQ_U^{-1}V_U Q_U^{-1}R')$$

再定义  $\hat{V}_U = n^{-1} \sum_{t=1}^n U_t U_t' \hat{v}_t^2$ , 其中  $\hat{v}_t$  为辅助回归 (4.2) 的残差, 类似地有  $\hat{V}_U \xrightarrow{p} V_U$ . 根据 Slutsky 定理, 可构造统计量

$$W = n(R\hat{\gamma})'[R\hat{Q}_U^{-1}\hat{V}_U\hat{Q}_U^{-1}R']^{-1}(R\hat{\gamma}) \xrightarrow{d} \chi_J^2$$

□

## 平稳时间序列的线性回归模型

5.1 假设时间序列  $\{Z_t\}$  服从 AR(1) 过程

$$Z_t = \alpha Z_{t-1} + \varepsilon_t, \quad |\alpha| < 1$$

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$$

(1) 求  $\mathbb{E}[Z_t]$ ,  $\text{var}(Z_t)$ ,  $\text{cov}(Z_t, Z_{t-j})$  及  $\text{corr}(Z_t, Z_{t-j})$ ,  $j = 0, \pm 1, \dots$ .

(2) 这是一个弱平稳过程吗?

(3) 如果  $|\alpha| \geq 1$ ,  $\{Z_t\}$  是弱平稳吗? 请解释.

**解答** (1) 定义滞后算子  $L$  使得  $L^j Z_t = Z_{t-j}$  及  $L^j \varepsilon_t = \varepsilon_{t-j}$ , 故而  $Z_t = (1 - \alpha L)^{-1} \varepsilon_t$ , 幂级数展开得

$$Z_t = (1 - \alpha L)^{-1} \varepsilon_t = \sum_{j=0}^{\infty} \alpha^j L^j \varepsilon_t = \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

因为  $\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$ , 故  $\mathbb{E}[Z_t] = 0$ . 同时易证

$$\text{var}(Z_t) = \text{var}\left(\sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}\right) = \sum_{j=0}^{\infty} \text{var}(\alpha^j \varepsilon_{t-j}) = \frac{\sigma^2}{1 - \alpha^2}$$

又因为  $\forall l \neq m, \text{cov}(\varepsilon_l, \varepsilon_m) = 0$ , 故而

$$\begin{aligned} \text{cov}(Z_t, Z_{t-j}) &= \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha^{l+m} \mathbb{E}[\varepsilon_{t-l} \varepsilon_{t-m-j}] \\ &= \sum_{m=0}^{\infty} \alpha^{j+2m} \mathbb{E}[\varepsilon_{t-m-j}^2] \\ &= \frac{\sigma^2}{1 - \alpha^2} \alpha^{|j|} \end{aligned}$$

显然也有  $\text{corr}(Z_t, Z_{t-j}) = \alpha^{|j|}$ .

(2) 由于  $\mathbb{E}[Z_t]$  和  $\text{var}(Z_t)$  均有限且  $\text{cov}(Z_t, Z_{t-j}) = \gamma(j)$  与时间点  $t$  无关, 因此 AR(1) 是弱平稳.

(3) 不是. 当  $|\alpha| \geq 1$  时,  $Z_t$  的方差不存在.

5.2 假设时间序列  $\{Z_t\}$  服从 MA(1) 过程

$$Z_t = \alpha \varepsilon_{t-1} + \varepsilon_t$$

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$$

(1) 求  $\mathbb{E}[Z_t]$ ,  $\text{var}(Z_t)$ ,  $\text{cov}(Z_t, Z_{t-j})$  及  $\text{corr}(Z_t, Z_{t-j})$ ,  $j = 0, \pm 1, \dots$ .

(2) 假设不对参数  $\alpha$  作任何限制.  $\{Z_t\}$  是一个弱平稳过程吗? 请解释.

(3) 对同一个自相关系数  $\text{corr}(Z_t, Z_{t-j})$ , 是否存在两个不同的  $\alpha$  值? 请解释.

**解答** (1) 不难发现  $\mathbb{E}[Z_t] = 0$ , 由于  $\forall l \neq m, \text{cov}(\varepsilon_l, \varepsilon_m) = 0$ , 易证  $\text{var}(Z_t) = (\alpha^2 + 1)\sigma^2$ .

当  $j = 0$  时,  $\text{cov}(Z_t, Z_{t-j}) = (\alpha^2 + 1)\sigma^2$ ,  $\text{corr}(Z_t, Z_{t-j}) = 1$ .

当  $j = \pm 1$  时,  $\text{cov}(Z_t, Z_{t-j}) = \alpha\sigma^2$ ,  $\text{corr}(Z_t, Z_{t-j}) = \frac{\alpha}{\alpha^2 + 1}$ .

当  $|j| > 1$  时,  $\text{cov}(Z_t, Z_{t-j}) = 0$ ,  $\text{corr}(Z_t, Z_{t-j}) = 0$ .

(2) 只要  $\alpha$  是有限的, 则  $\{Z_t\}$  是弱平稳过程.

(3) 当  $j = \pm 1$  或  $j = 0$  时, 同一个  $\text{corr}(Z_t, Z_{t-j}) = \frac{\alpha}{\alpha^2 + 1}$  对应唯一的  $\alpha$  值; 而当  $|j| > 1$  时, 由于  $\text{corr}(Z_t, Z_{t-j}) = 0$  与  $\alpha$  无关, 因此  $\alpha$  可以不同.

□

**5.3** 假设时间序列  $\{Z_t\}$  服从 ARMA(1, 1) 过程

$$Z_t = \alpha Z_{t-1} + \beta \varepsilon_{t-1} + \varepsilon_t, \quad |\alpha| < 1$$

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$$

(1) 求  $\mathbb{E}[Z_t]$ ,  $\text{var}(Z_t)$ ,  $\text{cov}(Z_t, Z_{t-j})$  及  $\text{corr}(Z_t, Z_{t-j})$ ,  $j = 0, \pm 1, \dots$ .

(2) 这是一个弱平稳过程吗? 请解释.

(3) 假设  $|\beta| < 1$ , 将  $Z_t$  表示为一个 AR( $\infty$ ) 过程.

(4) 将  $Z_t$  表示为一个 MA( $\infty$ ) 过程.

**解答** (1) 同样地, 利用滞后算子  $L$  可得  $Z_t = \alpha L Z_t + \beta L \varepsilon_t + \varepsilon_t$ , 也即

$$Z_t = \beta \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j-1} + \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j}$$

不难发现  $\mathbb{E}[Z_t] = 0$ . 进一步

$$\begin{aligned} \text{var}(Z_t) &= (1 + \beta^2)\sigma^2 \sum_{j=0}^{\infty} \alpha^{2j} + 2\beta \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \mathbb{E}[\alpha^{j+l} \varepsilon_{t-j-1} \varepsilon_{t-l}] \\ &= \frac{1 + \beta^2}{1 - \alpha^2} \sigma^2 + 2\beta \sum_{j=0}^{\infty} \alpha^{2j+1} \mathbb{E}[\varepsilon_{t-j-1}^2] \\ &= \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma^2 \end{aligned}$$

又因为  $\forall l \neq m, \text{cov}(\varepsilon_l, \varepsilon_m) = 0$ , 故而当  $|j| \geq 1$  时

$$\mathbb{E}[Z_t Z_{t-j}] = \beta^2 \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha^{l+m} \mathbb{E}[\varepsilon_{t-l-1} \varepsilon_{t-m-j-1}] \quad (A_1)$$

$$+ \beta \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha^{l+m} \mathbb{E}[\varepsilon_{t-l-1} \varepsilon_{t-m-j}] \quad (A_2)$$

$$+ \beta \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha^{l+m} \mathbb{E}[\varepsilon_{t-l} \varepsilon_{t-m-j-1}] \quad (A_3)$$

$$+ \sum_{l=0}^{\infty} \sum_{m=0}^{\infty} \alpha^{l+m} \mathbb{E}[\varepsilon_{t-l} \varepsilon_{t-m-j}] \quad (A_4)$$

其中

$$\begin{aligned}
A_1 &= \beta^2 \sum_{m=0}^{\infty} \alpha^{2m+j} \mathbb{E}[\varepsilon_{t-m-j-1}^2] = \frac{\alpha^{|j|} \beta^2 \sigma^2}{1 - \alpha^2} \\
A_2 &= \beta \sum_{m=0}^{\infty} \alpha^{2m+j-1} \mathbb{E}[\varepsilon_{t-m-j}^2] = \frac{\alpha^{|j|-1} \beta \sigma^2}{1 - \alpha^2} \\
A_3 &= \beta \sum_{m=0}^{\infty} \alpha^{2m+j+1} \mathbb{E}[\varepsilon_{t-m-j-1}^2] = \frac{\alpha^{|j|+1} \beta \sigma^2}{1 - \alpha^2} \\
A_4 &= \sum_{m=0}^{\infty} \alpha^{2m+j} \mathbb{E}[\varepsilon_{t-m-j}^2] = \frac{\alpha^{|j|} \sigma^2}{1 - \alpha^2}
\end{aligned}$$

因此对于  $\forall |j| \geq 1$  有

$$\text{cov}(Z_t, Z_{t-j}) = \frac{(\alpha^{|j|} \beta^2 + \alpha^{|j|-1} \beta + \alpha^{|j|+1} \beta + \alpha^{|j|}) \sigma^2}{1 - \alpha^2}$$

以及

$$\text{corr}(Z_t, Z_{t-j}) = \frac{\alpha^{|j|} \beta^2 + \alpha^{|j|-1} \beta + \alpha^{|j|+1} \beta + \alpha^{|j|}}{1 + \beta^2 + 2\alpha\beta}$$

而当  $j = 0$  时有

$$\text{cov}(Z_t, Z_{t-j}) = \frac{1 + \beta^2 + 2\alpha\beta}{1 - \alpha^2} \sigma^2$$

并且  $\text{corr}(Z_t, Z_{t-j}) = 1$ .

(2) 只要  $\beta$  是有限的, 则  $\{Z_t\}$  是弱平稳过程.

(3) 将  $\varepsilon_t$  改写为  $\varepsilon_t = (1 - \alpha L)(1 + \beta L)^{-1} Z_t$ , 在  $|\beta| < 1$  时进行幂级数展开得

$$\begin{aligned}
\varepsilon_t &= (1 - \alpha L) \sum_{j=0}^{\infty} (-\beta)^j Z_{t-j} \\
&= \sum_{j=0}^{\infty} (-\beta)^j Z_{t-j} - \alpha \sum_{j=0}^{\infty} (-\beta)^j Z_{t-j-1} \\
&= Z_t - \beta \sum_{j=1}^{\infty} (-\beta)^{j-1} Z_{t-j} - \alpha \sum_{j=1}^{\infty} (-\beta)^{j-1} Z_{t-j}
\end{aligned}$$

因此  $Z_t = (\alpha + \beta) \sum_{j=1}^{\infty} (-\beta)^{j-1} Z_{t-j} + \varepsilon_t$ .

(4) 将  $Z_t$  改写为  $Z_t = (1 + \beta L)(1 - \alpha L)^{-1} \varepsilon_t$ , 在  $|\alpha| < 1$  时进行幂级数展开得

$$\begin{aligned}
Z_t &= (1 + \beta L) \sum_{j=0}^{\infty} \alpha^j \varepsilon_{t-j} \\
&= \varepsilon_t + \alpha \sum_{j=1}^{\infty} \alpha^{j-1} \varepsilon_{t-j} + \beta \sum_{j=1}^{\infty} \alpha^{j-1} \varepsilon_{t-j} \\
&= (\alpha + \beta) \sum_{j=1}^{\infty} \alpha^{j-1} \varepsilon_{t-j} + \varepsilon_t
\end{aligned}$$

□

**5.4** 随机时间序列  $\{X_t\}$  是一个 Gaussian 过程, 如果对任何正整数  $t, k$

$$Z_t \equiv (X_t, X_{t-1}, \dots, X_{t-k})'$$

服从联合正态分布. 证明: 如果  $\{X_t\}$  是一个零均值平稳 Gaussian 过程, 则独立同分布, 鞅差分序列, 白噪声三者合一, 即从其中任何一个定义, 可推出另外两个定义.

**解答** (i) 如果  $\{X_t\}$  是 IID, 则  $\mathbb{E}[X_t] = 0$ ,  $\text{var}(X_t) = \sigma^2$ , 并且对于  $\forall |j| > 0$ ,  $\text{cov}(X_t, X_j) = 0$ , 因此  $\{X_t\}$  是 WN. 同时由 IID 可得  $\mathbb{E}[X_t|I_{t-1}] = \mathbb{E}[X_t] = 0$ , 说明  $\{X_t\}$  也是 MDS.

(ii) 如果  $\{X_t\}$  是 WN, 则  $\text{var}(X_t) = \sigma^2$ , 并且对于  $\forall |j| > 0$ ,  $\text{cov}(X_t, X_{t-j}) = 0$ , 由于  $\{X_t\}$  是 Gaussian 过程, 因此  $\{X_t\}$  是 IID, 由 (i) 可得  $\{X_t\}$  也是 MDS.

(iii) 如果  $\{X_t\}$  是 MDS, 由 LIE 首先可得  $\mathbb{E}[X_t] = 0$ , 并且对于  $\forall |j| > 0$ ,  $\mathbb{E}[X_t X_{t-j}] = 0$ , 即  $\text{cov}(X_t, X_{t-j}) = 0$ , 因此  $\{X_t\}$  是 WN.

设  $Z_t = (X_t, X_{t-1}, \dots, X_{t-k})'$  为  $K \times 1$  向量, 其中  $K = k + 1$ . 对于  $\forall t, k \in \mathbb{N}^*$  可得  $Z_t$  的概率密度

$$f_{Z_t}(z_t) = \frac{1}{(2\pi)^{\frac{K}{2}} |\Sigma|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} z_t' \Sigma^{-1} z_t\right)$$

由于已证得  $\{X_t\}$  是 WN, 故  $K \times K$  协方差矩阵为  $\Sigma = \text{diag}(\sigma^2, \dots, \sigma^2)$ . 因此

$$f_{Z_t}(z_t) = \prod_{k=0}^K f_{X_{t-k}}(x_{t-k})$$

也即  $\{X_t\}$  是 IID.

□

**5.5** (1) 假设利用某一检验, 发现  $\{\varepsilon_t\}$  存在序列相关. 那么, 是否可以确定  $\{\varepsilon_t\}$  不是一个鞅差分序列? 给出理由.

(2) 假设  $\{\varepsilon_t\}$  不存在序列相关, 是否能确定  $\{\varepsilon_t\}$  是一个鞅差分序列? 给出理由.

**解答** (1) 简单起见, 设  $\mathbb{E}[\varepsilon_t] = 0$ ,  $\text{var}(\varepsilon_t) = 1$ , 且  $\text{cov}(\varepsilon_t, \varepsilon_{t-1}) \neq 0$ , 可以得到

$$\mathbb{E}[\varepsilon_t \varepsilon_{t-1}] = \mathbb{E}[\varepsilon_{t-1} \mathbb{E}[\varepsilon_t | \varepsilon_{t-1}]] \neq 0 \Rightarrow \mathbb{E}[\varepsilon_t | \varepsilon_{t-1}] \neq 0$$

故  $\{\varepsilon_t\}$  不是 MDS.

(2) 设  $\varepsilon_t = z_{t-1} z_{t-2} + z_t$ , 其中  $z_t \sim \text{IID}(0, \sigma^2)$ . 易证  $\mathbb{E}[\varepsilon_t] = 0$ , 并且对于  $\forall |j| \geq 1$  有

$$\mathbb{E}[\varepsilon_t \varepsilon_{t-j}] = \mathbb{E}[(Z_{t-1} Z_{t-2} + Z_t)(Z_{t-j-1} Z_{t-j-2} + Z_{t-j})] = 0$$

也即  $\{\varepsilon_t\}$  不存在序列相关. 但显然  $\mathbb{E}[\varepsilon_t | I_{t-1}] \neq 0$ ,  $\{X_t\}$  不是 MDS.

□

**5.6** 假设  $\{Z_t\}$  是一个均值为零的弱平稳过程, 其谱密度函数为  $h(\omega)$ , 标准谱密度函数为  $f(\omega)$ . 证明:

(1)  $f(\omega)$ ,  $\omega \in [-\pi, \pi]$ , 是实值函数.

(2)  $f(\omega)$  是对称函数, 即  $f(-\omega) = f(\omega)$ ,  $\omega \in [-\pi, \pi]$ .

(3)  $\int_{-\pi}^{\pi} f(\omega) d\omega = 1$ .

(4)  $\forall \omega \in [-\pi, \pi]$ ,  $f(\omega) \geq 0$ .

**解答** (1) 在  $\sum_{j=-\infty}^{\infty} |\gamma(j)| < \infty$  成立的条件下, 标准谱密度函数  $f(\omega)$  的定义为

$$f(\omega) = \frac{h(\omega)}{\gamma(0)} = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j) e^{-ij\omega}$$

其中  $\gamma(j) = \text{cov}(Z_t, Z_{t-j})$  为  $\{Z_t\}$  的自协方差函数,  $\rho(j) = \gamma(j)/\gamma(0)$ ,  $i = \sqrt{-1}$ .

另一方面由 Euler 公式可得  $e^{-ij\omega} = \cos j\omega - i \sin j\omega$ . 注意到  $\rho(j) = \rho(-j)$  以及  $\sin x$  为  $\mathbb{R}$  上的奇函数, 故而  $\sum_{j=-\infty}^{\infty} \rho(j) i \sin j\omega = 0$ . 因此

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j) (\cos j\omega - i \sin j\omega) \\ &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \rho(j) \cos j\omega \end{aligned}$$

也即  $f(\omega)$  是实值函数.

(2) 根据第 (1) 问得出的  $f(\omega)$  的表达式易证  $f(-\omega) = f(\omega)$ ,  $\omega \in [-\pi, \pi]$ .

(3) 进一步将  $f(\omega)$  改写为

$$f(\omega) = \frac{1}{\pi} \sum_{j=1}^{\infty} \rho(j) \cos j\omega + \frac{1}{2\pi} \rho(0)$$

因此有

$$\begin{aligned} \int_{-\pi}^{\pi} f(\omega) d\omega &= \frac{1}{\pi} \sum_{j=1}^{\infty} \int_{-\pi}^{\pi} \rho(j) \cos j\omega d\omega + \frac{1}{2\pi} \int_{-\pi}^{\pi} \rho(0) d\omega \\ &= \frac{2}{\pi} \sum_{j=1}^{\infty} \rho(j) \frac{\sin j\pi}{j} + \rho(0) \\ &= 1 \end{aligned}$$

其中  $\rho(0) = \gamma(0)/\gamma(0) = 1$ .

(4) 设  $Z_n(\omega) = n^{-1/2} \sum_{t=1}^n Z_t e^{it\omega}$  为复随机变量<sup>9</sup>, 可以得到

$$\begin{aligned} \mathbb{E}|Z_n(\omega)|^2 &= \frac{1}{n} \sum_{t=1}^n \sum_{s=1}^n \mathbb{E}[Z_t Z_s] e^{i(s-t)\omega} \quad (\text{令 } s = t - j) \\ &= \frac{1}{n} \left[ n \mathbb{E}[Z_t^2] + \sum_{j=1-n}^{-1} (n+j) \gamma(j) e^{-ij\omega} + \sum_{j=1}^{n-1} (n-j) \gamma(j) e^{-ij\omega} \right] \\ &= \frac{1}{n} \sum_{j=1-n}^{n-1} (n - |j|) \gamma(j) e^{-ij\omega} \end{aligned}$$

<sup>9</sup> 设复随机变量  $Z, W \in \mathbb{C}$ , 则  $\mathbb{E}[Z] = \mathbb{E}[\Re(Z)] + i\mathbb{E}[\Im(Z)]$ ,  $\text{var}(Z) = \mathbb{E}|Z|^2 - |\mathbb{E}[Z]|^2$ ,  $\text{cov}(Z, W) = \mathbb{E}[Z\overline{W}] - \mathbb{E}[Z]\mathbb{E}[\overline{W}]$ , 其中  $\Re(Z)$  和  $\Im(Z)$  分别表示  $Z$  的实部和虚部.



当  $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=-n}^n |j| |\gamma(j)| = 0$  成立时

$$\begin{aligned} h(\omega) &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega} \\ &= \frac{1}{2\pi} \lim_{n \rightarrow \infty} \mathbb{E} |Z_n(\omega)|^2 \geq 0 \end{aligned}$$

又因为  $f(\omega) = h(\omega)/\gamma(0)$ ,  $\gamma(0) > 0$ , 因此  $f(\omega) \geq 0$ .

□

### 5.7 假设时间序列线性回归模型

$$Y_t = X_t' \beta^0 + \varepsilon_t$$

满足假设 5.1–5.3, 其中随机扰动项  $\varepsilon$  直接可观测. 这一模型分别包括静态时间序列回归模型和动态时间序列回归模型两种情形.

(1) 从模型正确设立条件  $\mathbb{E}[\varepsilon_t | X_t] = 0$  是否可推出  $\{\varepsilon_t\}$  是白噪声? 请解释.

(2) 如果  $\{\varepsilon_t\}$  是鞅差分序列, 是否可推出  $\mathbb{E}[\varepsilon_t | X_t] = 0$ ? 请解释.

(3) 如果  $\{\varepsilon_t\}$  存在序列相关, 是否必然意味着  $\mathbb{E}[\varepsilon_t | X_t] \neq 0$ , 即  $\mathbb{E}[Y_t | X_t]$  的线性回归模型存在模型误设? 请解释.

**解答** (1) 不能.  $\mathbb{E}[X_t | \varepsilon_t] = 0$  没有蕴含有关  $\{\varepsilon_t\}$  是否存在序列相关的信息.

(2) 不能.  $\{\varepsilon_t\}$  是 MDS 没有蕴含有关  $X_t$  的信息, 无法说明条件均值  $\mathbb{E}[\varepsilon_t | X_t] = 0$ .

(3) 如果模型是静态的,  $\{\varepsilon_t\}$  是否存在序列相关与模型是否正确设定没有必然关系; 如果模型是动态的, 说明存在模型误设. 以 ARMA(1, 1) 为例

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + \beta \varepsilon_{t-1} + \varepsilon_t, \quad |\alpha| < 1 \\ \{\varepsilon_t\} &\sim \text{WN}(0, \sigma^2) \end{aligned}$$

如果  $\{\varepsilon_t\}$  存在序列相关, 则  $\mathbb{E}[\varepsilon_t | X_t] \neq 0$ .

□

### 5.8 假设时间序列线性回归模型

$$Y_t = X_t' \beta^0 + \varepsilon_t$$

的扰动项  $\varepsilon_t$  直接可观测. 假定假设 5.1–5.5 成立, 且存在自回归条件同方差, 即  $\text{var}(\varepsilon_t | I_{t-1}) = \sigma^2$ , 这里  $I_t$  是由  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  生成的  $\sigma$ -域. 我们的目的是检验原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t | I_{t-1}] = 0$ .

(1) 考虑辅助回归

$$\varepsilon_t = \sum_{j=1}^p \alpha_j \varepsilon_{t-j} + v_t, \quad t = p+1, \dots, n$$

令  $\tilde{R}_{\text{uc}}^2$  是辅助回归的非中心化  $R^2$ . 证明: 当原假设  $\mathbb{H}_0$  成立, 且  $n \rightarrow \infty$  时,  $n \tilde{R}_{\text{uc}}^2 \xrightarrow{d} \chi_p^2$ .

(2) 现在考虑另一个有截距项的辅助回归

$$\varepsilon_t = \alpha_0 + \sum_{j=1}^p \alpha_j \varepsilon_{t-j} + u_t, \quad t = p+1, \dots, n$$

令  $\tilde{R}^2$  是这一辅助回归模型中心化  $R^2$ . 证明: 当原假设  $\mathbb{H}_0$  成立, 且  $n \rightarrow \infty$  时,  $n\tilde{R}^2 \xrightarrow{d} \chi_p^2$ .

(3) 在样本容量  $n$  较小的情形下,  $n\tilde{R}_{uc}^2$  和  $n\tilde{R}^2$  这两个检验统计量, 哪个表现更好? 给出直观理由.

**解答** (1) 根据问题 3.10 的第 (1) 问可知

$$F = \frac{\tilde{R}_{uc}^2/p}{(1 - \tilde{R}_{uc}^2)/(n - p - 1)} \sim F_{p, n-p-1}$$

在  $\text{var}(\varepsilon_t | I_{t-1}) = \sigma^2$  的条件下, 当  $n \rightarrow \infty$  时

$$pF = \frac{(n - p - 1)\tilde{R}_{uc}^2}{1 - \tilde{R}_{uc}^2} \xrightarrow{d} \chi_p^2$$

由于  $p$  是有限的, 故  $pF = O_p(1)$ , 进一步有

$$\frac{\tilde{R}_{uc}^2}{1 - \tilde{R}_{uc}^2} = \frac{pF}{n - p - 1} = o_p(1)$$

也即当  $n \rightarrow \infty$  时,  $\tilde{R}_{uc}^2 \xrightarrow{p} 0$ , 因此

$$n\tilde{R}_{uc}^2 = \frac{n}{n - p - 1} pF \cdot (1 - \tilde{R}_{uc}^2) \xrightarrow{d} \chi_p^2$$

(2) 证明与第 (1) 问相似.

(3) 令  $M_0 = I - \iota(\iota'\iota)^{-1}\iota'$ , 可以得到

$$Y'Y - Y'M_0Y = Y'\iota(\iota'\iota)^{-1}\iota'Y \sim \text{p.s.d.}$$

而根据  $R_{uc}^2$  和  $R^2$  的定义

$$R_{uc}^2 = 1 - \frac{e'e}{Y'Y}$$

$$R^2 = 1 - \frac{e'e}{Y'M_0Y}$$

故而  $\tilde{R}_{uc}^2 \geq \tilde{R}^2$ , 因此当样本容量  $n$  较小时, 在同一显著性水平下统计量  $n\tilde{R}_{uc}^2$  表现更好, 犯 I 类错误的概率更小.

□

## 5.9 假设时间序列线性回归模型

$$Y_t = X_t'\beta^o + \varepsilon_t$$

的扰动项  $\varepsilon_t$  直接可观测. 假定假设 5.1–5.5 成立, 且存在自回归条件异方差, 即  $\text{var}(\varepsilon_t | I_{t-1}) \neq \sigma^2$ , 这里  $I_t$  是由  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  生成的  $\sigma$ -域. 我们的目的是检验原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t | I_{t-1}] = 0$ . 考虑辅助回归

$$\varepsilon_t = \sum_{j=1}^p \alpha_j \varepsilon_{t-j} + v_t, \quad t = p+1, \dots, n$$

请构造一个渐近有效的检验统计量以检验原假设  $\mathbb{H}_0$ . 并推导出其渐近分布.

**解答** 设  $\hat{\alpha} = (\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p)'$  为  $\alpha_j$  的 OLS 估计量,  $\tilde{\varepsilon}_t = (\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-p})'$ ,  $\hat{Q}_\varepsilon = (n -$

$p)^{-1} \sum_{t=p+1}^n \tilde{\varepsilon}_t \tilde{\varepsilon}_t', \hat{V}_\varepsilon = (n-p)^{-1} \sum_{t=p+1}^n \tilde{\varepsilon}_t \tilde{\varepsilon}_t' \hat{v}_t^2$ , 其中  $j = 1, 2, \dots, p$ ,  $\hat{v}_t$  是残差. 于是

$$\sqrt{n-p}(\hat{\alpha} - \alpha) = \hat{Q}_\varepsilon^{-1}(n-p)^{-\frac{1}{2}} \sum_{t=p+1}^n \tilde{\varepsilon}_t v_t$$

根据遍历平稳 MDS 的中心极限定理与 Slutsky 定理可知

$$\sqrt{n-p}(\hat{\alpha} - \alpha) \xrightarrow{d} N(\mathbf{0}, Q_\varepsilon^{-1} V_\varepsilon Q_\varepsilon^{-1})$$

其中  $Q_\varepsilon = \mathbb{E}[\tilde{\varepsilon}_t \tilde{\varepsilon}_t']$ ,  $V_\varepsilon = \mathbb{E}[\tilde{\varepsilon}_t \tilde{\varepsilon}_t' v_t^2]$ . 在原假设在  $\mathbb{H}_0$  成立的情况下,  $R\alpha = \mathbf{0}$ , 于是

$$\sqrt{n-p}R\hat{\alpha}' \xrightarrow{d} N(\mathbf{0}, RQ_\varepsilon^{-1}V_\varepsilon Q_\varepsilon^{-1}R')$$

由  $\mathbb{H}_0$  可知  $R$  为单位矩阵, 因此由定理 5.15 可知

$$W = (n-p)\hat{\alpha}'[\hat{Q}_\varepsilon^{-1}\hat{V}_\varepsilon\hat{Q}_\varepsilon^{-1}]^{-1}\hat{\alpha} \xrightarrow{d} \chi_p^2$$

□

**5.10** 假设  $\varepsilon_t$  服从一个 ARCH(1) 过程

$$\begin{aligned}\varepsilon_t &= z_t \sigma_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \alpha_0 > 0, \quad 0 < \alpha_1 < 1 \\ \{z_t\} &\sim \text{i.i.d. } N(0, 1)\end{aligned}$$

(1) 证明:  $\mathbb{E}[\varepsilon_t | I_{t-1}] = 0$ ,  $\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = 0$ ,  $j > 0$ , 其中  $I_{t-1}$  是由  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  生成的  $\sigma$ -域.

(2) 证明:  $\text{corr}(\varepsilon_t^2, \varepsilon_{t-1}^2) = \alpha_1$ .

(3) 证明: 当  $0 < \alpha_1 < \frac{1}{\sqrt{3}}$  时,  $\varepsilon_t$  的峰度为

$$K = \frac{\mathbb{E}[\varepsilon_t^4]}{(\mathbb{E}[\varepsilon_t^2])^2} = \frac{3(1-\alpha_1^2)}{1-3\alpha_1^2} > 3$$

**解答** (1) 显然地

$$\mathbb{E}[\varepsilon_t | I_{t-1}] = \mathbb{E}[z_t \sigma_t | I_{t-1}] = \sigma_t \mathbb{E}[z_t | I_{t-1}] = 0$$

由 LIE 可知  $\mathbb{E}[\varepsilon_t] = 0$ . 于是

$$\text{cov}(\varepsilon_t, \varepsilon_{t-j}) = \mathbb{E}[\mathbb{E}[\varepsilon_t \varepsilon_{t-j} | I_{t-1}]] = \mathbb{E}[\varepsilon_{t-j} \mathbb{E}[\varepsilon_t | I_{t-1}]] = 0$$

(2) 首先有  $\mathbb{E}[\varepsilon_t^2] = \mathbb{E}[\sigma_t^2 \mathbb{E}[z_t^2 | I_{t-1}]] = \mathbb{E}[\sigma_t^2]$ , 故而

$$\mathbb{E}[\varepsilon_t^2] = \alpha_0 + \alpha_1 \mathbb{E}[\varepsilon_t^2] \Rightarrow \mathbb{E}[\varepsilon_t^2] = \frac{\alpha_0}{1-\alpha_1}$$

与下标  $t$  无关. 现在考虑

$$\begin{aligned}\text{cov}(\varepsilon_t^2, \varepsilon_{t-1}^2) &= \mathbb{E}[\varepsilon_t^2 \varepsilon_{t-1}^2] - \mathbb{E}[\varepsilon_t^2] \mathbb{E}[\varepsilon_{t-1}^2] \\ &= \alpha_1 \mathbb{E}[z_t^2] \mathbb{E}[\varepsilon_{t-1}^4] - \alpha_1 \mathbb{E}[z_t^2] \mathbb{E}[\varepsilon_{t-1}^2]^2 = \alpha_1 \text{var}(\varepsilon_{t-1}^2) = \frac{\alpha_0 \alpha_1}{1-\alpha_1}\end{aligned}$$

从而  $\text{corr}(\varepsilon_t^2, \varepsilon_{t-1}^2) = \alpha_1$ .

(3) 由 (2) 得到  $\mathbb{E}[\varepsilon_t^4] = \mathbb{E}[\sigma_t^4 \mathbb{E}[z_t^4 | I_{t-1}]]$ , 根据概率论可知  $\mathbb{E}[z_t^4] = (\mathbb{E}[z_t^2])^2 + \text{var}(z_t^2) = 3$ , 因

此

$$\mathbb{E}[\varepsilon_t^4] = 3\alpha_0^2 + 6\alpha_0\alpha_1\mathbb{E}[\varepsilon_t^2] + 3\alpha_1^2\mathbb{E}[\varepsilon_t^4] \Rightarrow \mathbb{E}[\varepsilon_t^4] = \frac{3\alpha_0^2(1 + \alpha_1)}{(1 - 3\alpha_1^2)(1 - \alpha_1)}$$

从而当  $0 < \alpha_1 < \frac{1}{\sqrt{3}}$  时

$$K = \frac{\mathbb{E}[\varepsilon_t^4]}{(\mathbb{E}[\varepsilon_t^2])^2} = \frac{3(1 - \alpha_1^2)}{1 - 3\alpha_1^2} > 3$$

□

**5.11** 假设遍历平稳时间序列的线性回归模型为

$$Y_t = X_t' \beta^o + \varepsilon_t$$

$$\varepsilon_t = z_t \sigma_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2, \quad \alpha_0 > 0, \quad 0 < \alpha_1 < 1$$

$$\{z_t\} \sim \text{IID } N(0, 1)$$

其中时间序列  $\{X_t\}$  和  $\{\varepsilon_t\}$  是相互独立的。

(1) OLS 估计量  $\hat{\beta}$  是  $\beta_o$  的一致估计量吗? 请解释。

(2) 估计量  $s^2 \hat{Q}$  是否是渐近方差  $\text{avar}(\sqrt{n} \hat{\beta})$  的一致估计? ARCH 效应是否会影响 OLS 估计量的渐近方差的结构? 请解释。

**解答** (1) 由于模型设定  $\mathbb{E}[\varepsilon_t | X_t] = \sigma_t \mathbb{E}[z_t | X_t] = 0$  仍成立, 故而  $\hat{\beta}$  是  $\beta_o$  的一致估计量。

(2) 由于  $\{X_t\}$  和  $\{\varepsilon_t\}$  是相互独立的, 因此根据问题 5.10 可知

$$\text{var}(\varepsilon_t | X_t) = \text{var}(\varepsilon_t) = \frac{\alpha_0}{1 - \alpha_1}$$

也即随机扰动项是条件同方差的, 估计量  $s^2 \hat{Q}$  是渐近方差  $\text{avar}(\sqrt{n} \hat{\beta})$  的一致估计。

□

**5.12** 假设线性时间序列回归模型

$$Y_t = X_t' \beta^o + \varepsilon_t$$

的随机扰动项直接可观测. 假定假设 5.1 – 5.5 成立, 且时间序列  $\{X_t\}$  和  $\{\varepsilon_t\}$  相互独立, 且  $\{\varepsilon_t\}$  存在 ARCH( $q$ ) 效应, 即

$$\mathbb{E}[\varepsilon_t^2 | I_{t-1}] = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2$$

其中  $I_{t-1}$  是由  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  生成的  $\sigma$ -域. 给出渐近方差  $\text{avar}(\sqrt{n} \hat{\beta})$  的表达形式, 其中  $\hat{\beta}$  是 OLS 估计量。

**解答** 由于  $\{X_t\}$  和  $\{\varepsilon_t\}$  相互独立, 故而满足条件同方差假定  $\mathbb{E}[\varepsilon_t^2 | \sigma_t] = \sigma^2$ , 在假设假设 5.1 – 5.5 成立的条件下, 渐近方差  $\text{avar}(\sqrt{n} \hat{\beta}) = \sigma^2 Q$ , 这里的  $\sigma^2$  可以根据 LIE 得到

$$\mathbb{E}[\varepsilon_t^2] = \mathbb{E}[\mathbb{E}[\varepsilon_t^2 | I_{t-1}]] = \alpha_0 + \sum_{j=1}^p \alpha_j \mathbb{E}[\varepsilon_{t-j}^2] \Rightarrow \sigma^2 = \frac{\alpha_0}{1 - \sum_{j=1}^q \alpha_j}$$

□

### 5.13 假设线性时间序列回归模型

$$Y_t = X_t' \beta^o + \varepsilon_t$$

的随机扰动项直接可观测. 假定假设 5.1 – 5.5 成立, 这个模型分别包括静态模型和动态模型两种情形, 且  $\{\varepsilon_t\}$  存在 ARCH( $q$ ) 效应, 即

$$\mathbb{E}[\varepsilon_t^2 | I_{t-1}] = \alpha_0 + \sum_{j=1}^q \alpha_j \varepsilon_{t-j}^2$$

其中  $I_{t-1}$  是由  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  生成的  $\sigma$ -域. 是否必须使用  $Q^{-1}VQ^{-1}$  作为  $\text{avar}(\sqrt{n}\hat{\beta})$  的渐近方差公式? 请解释.

**解答** 静态模型的情况已在问题 5.12 中讨论过了. 在动态模型的情况下, 由于  $X_t$  中包括了  $Y_t$  的滞后项,  $\mathbb{E}[\varepsilon_t^2 | X_t] \neq \sigma^2$  可能不再成立, 也即  $\{X_t\}$  和  $\{\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}\}$  不独立, 从而条件同方差假定不成立, 此时必须使用  $Q^{-1}VQ^{-1}$  作为  $\text{avar}(\sqrt{n}\hat{\beta})$  的渐近方差公式.

□

### 5.14 假设动态线性时间序列回归模型

$$\begin{aligned} Y_t &= \beta_0^o + \beta_1^o Y_{t-1} + \varepsilon_t \\ &= X_t' \beta^o + \varepsilon_t \end{aligned}$$

满足假设 5.1 – 5.5, 同时假设  $\{\varepsilon_t\}$  存在 ARCH(1), 即

$$\mathbb{E}[\varepsilon_t^2 | I_{t-1}] = \alpha_0 + \alpha_1 Y_{t-1}^2$$

给出渐近方差  $\text{avar}(\sqrt{n}\hat{\beta})$  的表达形式, 其中  $\hat{\beta}$  是 OLS 估计量.

**解答** 即使  $\{\varepsilon_t\}$  存在 ARCH(1) 效应, 假设 5.5 也保证了  $\{\varepsilon_t\}$  也是 MDS, 从而

$$\sqrt{n}(\hat{\beta} - \beta^o) = \hat{Q}^{-1} n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{d} N(0, Q^{-1}VQ^{-1})$$

因此  $\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$ , 其中  $Q = \mathbb{E}[X_t X_t']$ ,  $V = \alpha_0 \mathbb{E}[X_t X_t'] + \alpha_1 \mathbb{E}[X_t X_t' Y_{t-1}^2]$ . 并且由于

$$\mathbb{E}[\varepsilon_t^2 | X_t] = \mathbb{E}[\mathbb{E}[\varepsilon_t^2 | Y_{t-1}] | I_{t-1}] \neq \sigma^2$$

$\text{avar}(\sqrt{n}\hat{\beta})$  的形式也不能化简.

□

### 5.15 假设线性时间序列回归模型

$$Y_t = X_t' \beta^o + \varepsilon_t$$

满足假设 5.1, 5.2 和 5.4, 时间序列  $\{X_t\}$  和  $\{\varepsilon_t\}$  是相互独立的, 且  $\mathbb{E}[\varepsilon_t] = 0$ . 进一步假设  $\{\varepsilon_t\}$  存在序列相关.

(1)  $\{\varepsilon_t\}$  的序列相关是否影响 OLS 估计量  $\hat{\beta}$  的一致性? 请解释.

(2)  $\{\varepsilon_t\}$  的序列相关是否影响 OLS 估计量  $\hat{\beta}$  的渐近方差表达形式  $\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$ ?

其中  $V = \lim_{n \rightarrow \infty} \text{var} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right)$ . 特别是, 是否仍有  $V = \mathbb{E}[X_t X_t' \varepsilon_t^2]$ ? 请解释.

**解答** (1) 不影响. 注意到  $\hat{\beta} - \beta^o = \hat{Q}^{-1} n^{-1} \sum_{t=1}^n X_t \varepsilon_t$ , 由遍历平稳的 WLLN 可知  $\hat{Q} \xrightarrow{p} Q$ , 又因

为假设 5.4 可保证了  $Q^{-1}$  的存在性, 根据连续映射定理即得  $\hat{Q}^{-1} \xrightarrow{p} Q^{-1}$ .

另一方面, 因为  $\{X_t\}$  和  $\{\varepsilon_t\}$  相互独立, 且  $\mathbb{E}[\varepsilon_t] = 0$ , 于是  $\mathbb{E}[X_t \varepsilon_t] = \mathbb{E}[X_t] \mathbb{E}[\varepsilon_t] = \mathbf{0}$ , 从而由遍历平稳的 WLLN 可知

$$n^{-1} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{p} \mathbb{E}[X_t \varepsilon_t] = \mathbf{0}$$

于是  $\hat{\beta} \xrightarrow{p} \beta^o$ .

(2) 影响  $\hat{\beta}$  的渐近方差表达形式  $\text{avar}(\sqrt{n}\hat{\beta}) = Q^{-1}VQ^{-1}$ , 这是由于  $\{\varepsilon_t\}$  存在序列相关, 则  $\{\varepsilon_t\}$  不是白噪声, 这也表明  $\{X_t \varepsilon_t\}$  不是 MDS, 因此不适用于遍历平稳 MDS 的中心极限定理, 并且此时

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{var} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right) &= \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \sum_{s=1}^n \mathbb{E}[X_t \varepsilon_t X_s' \varepsilon_s] \\ &= \lim_{n \rightarrow \infty} n^{-1} \left[ \sum_{t=1}^n \mathbb{E}[X_t X_t' \varepsilon_t^2] + \sum_{t \neq s} \mathbb{E}[X_t \varepsilon_t X_s' \varepsilon_s] \right] \\ &= V + \lim_{n \rightarrow \infty} n^{-1} \sum_{t \neq s} \mathbb{E}[X_t X_s'] \mathbb{E}[\varepsilon_t \varepsilon_s] \neq V \end{aligned}$$

因此  $V = \mathbb{E}[X_t X_t' \varepsilon_t^2]$  不成立.

□

### 5.16 假设线性时间序列回归模型

$$\begin{aligned} Y_t &= \beta_0^o + \beta_1^o Y_{t-1} + \varepsilon_t \\ &= X_t' \beta^o + \varepsilon_t \end{aligned}$$

满足假设 5.1, 5.2 和 5.4,  $|\beta_1^o| < 1$ , 其中  $X_t = (1, Y_{t-1})'$ . 进一步假设  $\{\varepsilon_t\}$  是一个 MA(1) 过程

$$\varepsilon_t = \rho v_{t-1} + v_t$$

其中,  $\{v_t\} \sim \text{IID}(0, \sigma_v^2)$ , 即  $\{\varepsilon_t\}$  存在一阶自相关. 在这种情形下, OLS 估计量  $\hat{\beta}$  是  $\beta^o$  的一致估计量吗? 请解释.

**解答** 利用滞后算子  $L$  可知  $Y_{t-2} = LY_{t-1}$ , 参考问题 5.3 的做法, 利用幂级数展开得到

$$Y_{t-1} = \frac{\beta_0^o}{1 - \beta_1^o} + \sum_{j=0}^{\infty} (\beta_1^o)^j \varepsilon_{t-j-1} \quad (|\beta_1^o| < 1)$$

根据问题 5.2 可知

$$\mathbb{E}[\varepsilon_{t-j} \varepsilon_t] = \begin{cases} \rho \sigma_v^2, & j = 1 \\ 0, & j > 1 \end{cases}$$

于是  $\mathbb{E}[Y_{t-1} \varepsilon_t] = \mathbb{E}[\varepsilon_{t-1} \varepsilon_t] = \rho \sigma_v^2 \neq 0$ , 从而  $\mathbb{E}[X_t \varepsilon_t] \neq \mathbf{0}$ ,  $\hat{\beta}$  不是  $\beta^o$  的一致估计量.

□

## 具有条件异方差和自相关扰动项的线性回归模型

6.1 给定假设 6.1 – 6.3 和 6.5(1), 证明:

$$\text{avar} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right) = \lim_{n \rightarrow \infty} \text{var} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right) = \sum_{j=-\infty}^{\infty} \Gamma(j)$$

**解答** 第一个等号成立是定义.

主要参考本章第三节的内容, 经过一定数学运算后得到

$$\text{var} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right) = \sum_{|j| < n} \left( 1 - \frac{|j|}{n} \right) \Gamma(j) \leq \sum_{|j| < n} |\Gamma(j)|$$

因为  $\sum_{j=-\infty}^{\infty} \|\Gamma(j)\| < \infty$ , 根据 Lebesgue 控制收敛定理<sup>10</sup>可知

$$\lim_{n \rightarrow \infty} \sum_{|j| < n} \left( 1 - \frac{|j|}{n} \right) \Gamma(j) = \sum_{j=-\infty}^{\infty} \Gamma(j)$$

□

6.2 假设对于所有  $j > p_0$ ,  $\Gamma(j) = 0$ , 这里  $p_0$  是一个固定的正整数. 在这种情况下, 长期方差  $V = \sum_{j=-p_0}^{p_0} \Gamma(j)$ , 且可以用以下方差估计量估计它

$$\hat{V} = \sum_{j=-p_0}^{p_0} \hat{\Gamma}(j)$$

其中  $\hat{\Gamma}(j)$  是样本自协方差函数. 证明对于任意给定的滞后阶数  $j \in \{-p_0, \dots, p_0\}$ , 当  $n \rightarrow \infty$  时,  $\hat{\Gamma}(j) \xrightarrow{P} \Gamma(j)$ .

由于  $p_0$  是一固定的整数, 因此对于给定的  $j \in \{-p_0, \dots, p_0\}$ ,  $\hat{\Gamma}(j) \xrightarrow{P} \Gamma(j)$  的一个重要含义是, 当  $n \rightarrow \infty$  时,  $\hat{V} \xrightarrow{P} V$ .

**解答** 当  $j \in \{0, 1, \dots, p_0\}$  时, 由于  $e_t = X_t'(\beta^o - \hat{\beta}) + \varepsilon_t$ , 因此自协方差函数为

$$\begin{aligned} \hat{\Gamma}(j) &= n^{-1} \sum_{t=j+1}^n X_t e_t e_{t-j} X_{t-j}' \\ &= n^{-1} \sum_{t=j+1}^n X_t \left[ X_t'(\beta^o - \hat{\beta}) + \varepsilon_t \right] \left[ X_{t-j}'(\beta^o - \hat{\beta}) + \varepsilon_{t-j} \right] X_{t-j}' \\ &= A_1 + A_2 + A_3 + A_4 \end{aligned}$$

<sup>10</sup> 设  $\{f_n\}$  是测度空间  $(X, \mathcal{F}, \mu)$  上的复可测函数序列, 并且  $f(x) = \lim_{n \rightarrow \infty} f_n(x)$  对于一切  $x \in X$  都成立. 如果存在函数  $g \in L^1(\mu)$ , 使得对于一切  $n = 1, 2, \dots$  以及  $x \in X$  都有  $|f_n(x)| \leq g(x)$ , 那么  $f \in L^1(\mu)$ , 并且

$$\lim_{n \rightarrow \infty} \int_X f_n d\mu = \int_X f d\mu$$

其中

$$\begin{aligned} A_1 &= n^{-1} \sum_{t=j+1}^n X_t \left[ X'_t(\beta^o - \hat{\beta})(\beta^o - \hat{\beta})' X_{t-j} \right] X'_{t-j} \\ A_2 &= n^{-1} \sum_{t=j+1}^n X_t \left[ X'_t(\beta^o - \hat{\beta}) \varepsilon_{t-j} \right] X'_{t-j} \\ A_3 &= n^{-1} \sum_{t=j+1}^n X_t \left[ \varepsilon_t X'_{t-j} (\beta^o - \hat{\beta}) \right] X'_{t-j} \\ A_4 &= n^{-1} \sum_{t=j+1}^n X_t X'_{t-j} \varepsilon_t \varepsilon_{t-j} \end{aligned}$$

由于  $\hat{\beta} \xrightarrow{p} \beta^o$ , 因而  $A_i \xrightarrow{p} \mathbf{0}$ , 这里  $i = 1, 2, 3$ . 另一方面, 如果对于任意的  $j, l \in \{0, \dots, k\}$ ,  $\mathbb{E}[X_{jt}^4]$ ,  $\mathbb{E}[\varepsilon_t^4]$  以及  $\mathbb{E}[\varepsilon_{t-j}^4]$  均有限, 那么根据 Cauchy-Schwarz 不等式可以得到

$$\mathbb{E}|X_{jt} X_{lt} \varepsilon_t \varepsilon_{t-j}| \leq \mathbb{E}[X_{jt}^4]^{\frac{1}{4}} \mathbb{E}[X_{lt}^4]^{\frac{1}{4}} \mathbb{E}[\varepsilon_t^4]^{\frac{1}{4}} \mathbb{E}[\varepsilon_{t-j}^4]^{\frac{1}{4}} < \infty$$

那么由遍历平稳的 WLLN 可知

$$n^{-1} \sum_{t=j+1}^n X_t X'_{t-j} \varepsilon_t \varepsilon_{t-j} \xrightarrow{p} \mathbb{E}[X_t X'_{t-j} \varepsilon_t \varepsilon_{t-j}]$$

也即  $\hat{\Gamma}(j) \xrightarrow{p} \Gamma(j)$ . 由于  $\Gamma(j) = \Gamma(-j)'$ , 因此当  $j \in \{-p_0, \dots, -1\}$  时, 结论也成立.

□

**6.3** 假设  $\{Y_t\}$  是一个单变量平稳时间序列, 且存在谱密度函数

$$h(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \gamma(j) e^{-ij\omega}$$

证明: 当  $p \rightarrow \infty$  时,  $\text{var} \left( p^{-\frac{1}{2}} \sum_{j=1}^p Y_{t-j} \right) \rightarrow 2\pi h(0)$ .

**解答** 显然  $2\pi h(0) = \sum_{j=-\infty}^{\infty} \gamma(j)$ . 因为  $\{Y_t\}$  是平稳的, 故而  $\gamma(j)$  与起始时间点  $t$  无关, 因此

$$\begin{aligned} \text{var} \left( p^{-\frac{1}{2}} \sum_{j=1}^p Y_{t-j} \right) &= p^{-1} \sum_{k=1}^p \sum_{l=1}^p \mathbb{E}[Y_{t-k} Y'_{t-l}] \\ &= p^{-1} \sum_{k=1}^p \sum_{m=1}^p \gamma(m-k) = \sum_{|j| < p} \left( 1 - \frac{|j|}{p} \right) \gamma(j) \end{aligned}$$

由 Lebesgue 控制收敛定理可知  $\lim_{p \rightarrow \infty} \text{var} \left( p^{-\frac{1}{2}} \sum_{j=1}^p Y_{t-j} \right) = \sum_{j=-\infty}^{\infty} \gamma(j)$ .

□

**6.4** (1) 令  $\gamma(j) = \text{cov}(Y_t, Y_{t-j})$ , 其中  $\{Y_t\}$  是一个弱平稳时间序列. 给出一个  $\{Y_t\}$  的例子,



使得至少存在一个非零滞后阶  $j \neq 0$ , 有  $\gamma(j) \neq 0$ , 但  $\sum_{j=1}^{\infty} \gamma(j) = 0$ .

(2) 方差比检验能否检测出 (1) 中的时间序列自相关特征? 请解释.

**解答** (1) 考虑如下 MA(2) 过程

$$Y_t = \varepsilon_{t-1} - \frac{1}{2}\varepsilon_{t-2} + \varepsilon_t$$

$$\{\varepsilon_t\} \sim \text{WN}(0, \sigma^2)$$

显然它是弱平稳的. 又知  $\gamma(1) = \frac{1}{2}\sigma^2$ ,  $\gamma(2) = -\frac{1}{2}\sigma^2$ , 而当  $j > 2$  时,  $\gamma(j) = 0$ , 因此

$$\sum_{j=1}^{\infty} \gamma(j) = \gamma(1) + \gamma(2) = 0$$

(2) 不能. 方差比检验仅检测长期方差  $\sum_{j=-\infty}^{\infty} \gamma(j)$  是否等于个体方差  $\gamma(0)$ . 考虑 Lo & MacKinlay (1988) 提出的  $\text{VR}_0$  统计量

$$\text{VR}_0 = \frac{1}{2}\pi\sqrt{n/p} \left[ \hat{f}(0) - \frac{1}{2\pi} \right] \xrightarrow{d} N[0, 2(2p-1)(p-1)/3p]$$

其中

$$\hat{f}(0) = \frac{1}{2\pi} \sum_{j=-p}^p \left( 1 - \frac{|j|}{p} \right) \hat{\rho}(j)$$

这里的  $\hat{\rho}(j) = \hat{\gamma}(j)/\hat{\gamma}(0)$  为样本自相关系数函数.

若原假设  $\mathbb{H}_0: \sum_{j=1}^{\infty} \Gamma(j) = 0$  成立, 那么  $\hat{f}(0)$  应该与  $\frac{1}{2\pi}$  相差不大,  $\text{VR}_0$  应该很接近于 0.

换言之, 方差比检验考虑的是鞅差分序列假设成立时的零频率处的谱密度特征, 而没有考虑其它非零频率处的谱密度, 从而无法检验时间序列的自相关特征.

□

# 工具变量回归分析

7.1 考虑以下简单的 Keynes 国民收入决定模型

$$C_t = \beta_0^o + \beta_1^o(Y_t - T_t) + \varepsilon_t \quad (7.1.1)$$

$$T_t = \gamma_0^o + \gamma_1^o Y_t + v_t \quad (7.1.2)$$

$$Y_t = C_t + G_t \quad (7.1.3)$$

其中  $C_t, Y_t, T_t, G_t$  分别是消费、收入、税收和政府支出,  $\{\varepsilon_t\} \sim \text{IID}(0, \sigma_\varepsilon^2)$ ,  $v_t \sim \text{IID}(0, \sigma_v^2)$ , 且  $\{\varepsilon_t\}$  和  $\{v_t\}$  两个序列相互独立. 方程 (7.1.1) 是消费函数, (7.1.2) 是税收函数, (7.1.3) 是收入恒等式.

(1) 模型 (7.1.1) 的 OLS 估计量  $\hat{\beta}_1$  是边际消费倾向  $\beta_1^o$  的一致估计量吗? 请解释.

(2) 假定  $G_t$  是外生变量 (即  $G_t$  不依赖于  $C_t$  和  $Y_t$ ).  $G_t$  是否可作为有效的工具变量? 如果是, 描述二阶段最小二乘估计法; 如果不是, 请解释.

(3) 假设政府要保持预算平衡, 使得

$$G_t = T_t + w_t \quad (7.1.4)$$

其中  $\{w_t\} \sim \text{IID}(0, \sigma_w^2)$ .  $G_t$  是否为有效的工具变量? 如果是, 描述二阶段最小二乘估计法; 如果不是, 请解释.

**解答** (1) 不是. 经过简单运算可知

$$\begin{aligned} Y_t - T_t &= (1 - \gamma_1^o)Y_t - (\gamma_0^o + v_t) \\ &= (1 - \gamma_1^o)[\beta_0^o + \beta_1^o(Y_t - T_t) + \varepsilon_t + G_t] - (\gamma_0^o + v_t) \\ &= (1 - \gamma_1^o)\beta_1^o(Y_t - T_t) + (1 - \gamma_1^o)(\beta_0^o + \varepsilon_t + G_t) - (\gamma_0^o + v_t) \end{aligned}$$

于是

$$Y_t - T_t = \frac{(1 - \gamma_1^o)\beta_0^o - \gamma_0^o}{1 - (1 - \gamma_1^o)\beta_1^o} + \frac{(1 - \gamma_1^o)(\varepsilon_t + G_t) - v_t}{1 - (1 - \gamma_1^o)\beta_1^o}$$

由于  $\mathbb{E}[\varepsilon_t] = 0$  且  $\{\varepsilon_t\}$  和  $\{v_t\}$  相互独立, 因此

$$\mathbb{E}[(Y_t - T_t)\varepsilon_t] = \frac{(1 - \gamma_1^o)(\mathbb{E}[G_t\varepsilon_t] + \sigma_\varepsilon^2)}{1 - (1 - \gamma_1^o)\beta_1^o} \neq 0$$

故而  $\hat{\beta}_1$  不是  $\beta_1^o$  的一致估计量.

(2) 可以. 首先考虑  $G_t$  的外生性:  $\mathbb{E}[G_t\varepsilon_t] = \mathbb{E}[G_t]\mathbb{E}[\varepsilon_t] = 0$ . 再来考虑  $G_t$  与内生解释变量  $Y_t - T_t$  的相关性

$$\mathbb{E}[(Y_t - T_t)G_t] = \frac{[(1 - \gamma_1^o)\beta_0^o - \gamma_0^o]\mathbb{E}[G_t] + (1 - \gamma_1^o)\mathbb{E}[G_t^2]}{1 - (1 - \gamma_1^o)\beta_1^o} \neq 0$$

因此  $G_t$  可以作为工具变量. 2SLS 估计法步骤如下:

(i) 设  $X_t = (1, Y_t - T_t)'$ ,  $Z_t = (1, G_t)'$ , 用  $X_t$  对  $Z_t$  进行 OLS 回归, 并得到拟合值  $\hat{X}_t$ ;

(ii) 用  $C_t$  对  $\hat{X}_t$  进行 OLS 回归, 即可得到 2SLS 估计量.

(3) 不是. 简单运算可知

$$G_t = \frac{\gamma_0^o}{1 - \gamma_1^o} + \frac{\gamma_1^o}{1 - \gamma_1^o} C_t + \frac{v_t + w_t}{1 - \gamma_1^o}$$

因此

$$\mathbb{E}[G_t \varepsilon_t] = \frac{\gamma_1^o}{1 - \gamma_1^o} \mathbb{E}[C_t \varepsilon_t] + \frac{1}{1 - \gamma_1^o} \mathbb{E}[w_t \varepsilon_t] \neq 0$$

故而  $G_t$  不是有效的工具变量.

□

## 7.2 考虑数据生成过程

$$Y_t = X_t' \beta^o + \varepsilon_t \quad (7.2.1)$$

其中  $X_t = (1, X_{1t})'$ , 有

$$X_{1t} = v_t + u_t \quad (7.2.2)$$

$$\varepsilon_t = w_t + u_t \quad (7.2.3)$$

其中  $\{v_t\}$ ,  $\{u_t\}$  和  $\{w_t\}$  均是 IID  $N(0, 1)$ , 并且三个序列相互独立.

(1) OLS 估计量  $\hat{\beta}$  是  $\beta^o$  的一致估计吗? 请解释.

(2) 假定  $Z_{1t} = w_t - \varepsilon_t$  且可以观测到  $Z_{1t}$ , 则  $Z_t = (1, Z_{1t})'$  是有效的工具变量吗? 请解释.

(3) 找一个工具变量, 用该工具变量构造  $\hat{\beta}_{2SLS}$  并推导其渐近分布.

(4) 考虑检验原假设

$$\mathbb{H}_0: R\beta^o = r$$

其中  $R$  是  $J \times 2$  常矩阵,  $r$  是  $J \times 1$  常向量, 且  $J \leq 2$ . 假设  $F$  是 2SLS 第二阶段的经典  $F$  统计量,  $J \cdot F$  在原假设  $\mathbb{H}_0$  下是否服从渐近  $\chi_J^2$  分布? 请解释.

**解答** (1) 不是. 简单运算可知

$$\mathbb{E}[X_{1t} \varepsilon_t] = \mathbb{E}[(v_t + u_t)(w_t + u_t)] = \mathbb{E}[u_t^2] = 1$$

故而 OLS 估计量  $\hat{\beta}$  不是  $\beta^o$  的一致估计.

(2) 不是. 简单运算可知

$$\mathbb{E}[Z_{1t} \varepsilon_t] = \mathbb{E}[(w_t - \varepsilon_t)(w_t + u_t)] = -\mathbb{E}[u_t^2] = -1$$

故而  $Z_t$  不是有效的工具变量.

(3) 假定  $v_t$  可以观测到, 考虑  $Z_{1t} = v_t$  作为工具变量, 这是因为

$$\mathbb{E}[Z_{1t} \varepsilon_t] = \mathbb{E}[v_t(w_t + u_t)] = \mathbb{E}[v_t] \mathbb{E}[w_t + u_t] = 0$$

$$\mathbb{E}[Z_{1t} X_{1t}] = \mathbb{E}[v_t(v_t + u_t)] = \mathbb{E}[v_t^2] = 1$$

容易构建  $\hat{\beta}_{2SLS} = (\hat{\mathbf{X}}' \hat{\mathbf{X}})^{-1} \hat{\mathbf{X}}' \mathbf{Y}$ , 其中  $\hat{\mathbf{X}} = \mathbf{Z} \hat{\gamma}$ ,  $\hat{\gamma} = (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{X}$ , 于是

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) \xrightarrow{d} N\left(\mathbf{0}, Q_{\tilde{\mathbf{X}} \tilde{\mathbf{X}}}^{-1} V_{\tilde{\mathbf{X}} \tilde{\mathbf{X}}} Q_{\tilde{\mathbf{X}} \tilde{\mathbf{X}}}^{-1}\right)$$

其中  $Q_{\tilde{\mathbf{X}} \tilde{\mathbf{X}}} = \mathbb{E}[\tilde{\mathbf{X}}_t \tilde{\mathbf{X}}_t']$ ,  $\tilde{\mathbf{X}}_t = \gamma' Z_t$ ,  $\gamma = \mathbb{E}[Z_t Z_t']^{-1} \mathbb{E}[Z_t X_t']$ , 并且  $V_{\tilde{\mathbf{X}} \tilde{\mathbf{X}}} = \gamma' \mathbb{E}[Z_t Z_t' \varepsilon_t^2] \gamma$ . 现在来考虑具体的渐近分布.

由于  $Z_t = (1, Z_{1t})'$ ,  $X_t = (1, X_{1t})'$ , 可以得到

$$\begin{aligned}\mathbb{E}[Z_t Z_t'] &= \mathbb{E} \begin{bmatrix} 1 & v_t \\ v_t & v_t^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \\ \mathbb{E}[Z_t X_t'] &= \mathbb{E} \begin{bmatrix} 1 & v_t + u_t \\ v_t & v_t^2 + v_t u_t \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}\end{aligned}$$

于是最优线性最小二乘近似系数为

$$\gamma = \mathbb{E}[Z_t Z_t']^{-1} \mathbb{E}[Z_t X_t'] = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

从而

$$Q_{\tilde{X}\tilde{X}} = \gamma' \mathbb{E}[Z_t Z_t'] \gamma = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

以及

$$V_{\tilde{X}\tilde{X}} = \gamma' \mathbb{E}[Z_t Z_t'] \mathbb{E}[\varepsilon_t^2] \gamma = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

这里  $\mathbb{E}[Z_t Z_t' \varepsilon_t^2] = \mathbb{E}[Z_t Z_t'] \mathbb{E}[\varepsilon_t^2]$  是由  $\{v_t\}$ ,  $\{u_t\}$  和  $\{w_t\}$  的独立性得到的. 因此

$$\text{avar}(\sqrt{n} \hat{\beta}_{2\text{SLS}}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

(4) 根据问题 7.6 的结论可知,  $s^2 = e'e/(n-K)$  不会依概率收敛于  $\sigma^2$ , 其中  $e = Y - \hat{\mathbf{X}}' \hat{\beta}_{2\text{SLS}}$ . 故而以下  $J \cdot F$  统计量

$$J \cdot F = n(R \hat{\beta}_{2\text{SLS}} - r)' [s^2 R \hat{Q}_{\tilde{X}\tilde{X}}^{-1} R']^{-1} (R \hat{\beta}_{2\text{SLS}} - r)$$

也不服从渐近  $\chi_J^2$  分布.

□

### 7.3 考虑下列供求模型

$$Y_t = \alpha_0^o + \alpha_1^o P_t + \alpha_2^o S_t + \varepsilon_t$$

$$Y_t = \beta_0^o + \beta_1^o P_t + \beta_2^o C_t + v_t$$

其中第一个方程是某一商品需求模型,  $Y_t$  是商品需求数量,  $P_t$  是商品价格,  $S_t$  是替代品的价格,  $\varepsilon_t$  是需求冲击. 第二个是商品供给模型, 其中  $Y_t$  是商品供应量,  $C_t$  是生产成本,  $v_t$  是供给冲击. 假定  $S_t$  和  $C_t$  是外生变量,  $\{\varepsilon_t\} \sim \text{IID}(0, \sigma_\varepsilon^2)$ ,  $\{v_t\} \sim \text{IID}(0, \sigma_v^2)$ , 并且  $\{\varepsilon_t\}$  和  $\{v_t\}$  两个序列相互独立. 假设市场总是出清, 即需求数量总是等于供给数量.

(1) 假定用工具  $Z_t = (S_t, C_t)'$  并用 2SLS 估计需求模型. 描述 2SLS 估计步骤. 得到的 2SLS 估计量  $\hat{\alpha}_{2\text{SLS}}$  是  $\alpha^o = (\alpha_0^o, \alpha_1^o, \alpha_2^o)'$  的一致估计量吗? 请解释.

(2) 假定用工具  $Z_t = (S_t, C_t)'$  并用 2SLS 估计供给模型. 描述 2SLS 估计步骤. 得到的 2SLS 估计量  $\hat{\beta}_{2\text{SLS}}$  是  $\beta^o = (\beta_0^o, \beta_1^o, \beta_2^o)'$  的一致估计量吗? 请解释.

(3) 假定  $\{\varepsilon_t\}$  和  $\{v_t\}$  是同期相关的, 即  $\mathbb{E}[\varepsilon_t v_t] \neq 0$ . 当商品供求受同一冲击影响时, 就会出现这种情形. 这会影响 (1) 和 (2) 的结论吗? 请解释.

**解答** (1) 设  $\tilde{Z}_t = (1, S_t, C_t)'$ ,  $X_t = (1, P_t, S_t)'$ . 先用  $X_t$  对  $\tilde{Z}_t$  进行 OLS 回归, 并得到拟合值  $\hat{X}_t$ , 再用  $Y_t$  对  $\hat{X}_t$  进行 OLS 回归, 由此得到 2SLS 估计量  $\hat{\alpha}_{2SLS}$ . 由于  $S_t$  和  $C_t$  是外生的, 故而

$$\mathbb{E}[\tilde{Z}_t \varepsilon_t] = \mathbf{0}$$

显然  $\mathbb{E}[\tilde{Z}_t X_t'] \neq \mathbf{0}$ , 因此  $\hat{\alpha}_{2SLS}$  是  $\alpha^o$  的一致估计量.

(2) 设  $\eta_t = (1, P_t, C_t)'$ . 先用  $\eta_t$  对  $\tilde{Z}_t$  进行 OLS 回归, 并得到拟合值  $\hat{\eta}_t$ , 再用  $Y_t$  对  $\hat{\eta}_t$  进行 OLS 回归, 由此得到 2SLS 估计量  $\hat{\beta}_{2SLS}$ . 同理可得  $\mathbb{E}[\tilde{Z}_t v_t'] = \mathbf{0}$ ,  $\mathbb{E}[\tilde{Z}_t X_t'] \neq \mathbf{0}$ , 因此  $\hat{\beta}_{2SLS}$  是  $\beta^o$  的一致估计量.

(3) 不影响. 因为 2SLS 估计量的一致性仅依赖于  $S_t$  和  $C_t$  的外生性, 而与  $\{\varepsilon_t\}$  和  $\{v_t\}$  的相关性无关, 后者只会影响 2SLS 估计量的渐近方差.

□

**7.4** 证明当假设 7.1 – 7.4 成立, 那么当  $n \rightarrow \infty$  时,  $\hat{\beta}_{2SLS} \xrightarrow{p} \beta^o$ .

**解答** 定义记号

$$\begin{aligned}\hat{Q}_{XZ} &= n^{-1} \sum_{t=1}^n X_t Z_t' \\ \hat{Q}_{ZZ} &= n^{-1} \sum_{t=1}^n Z_t Z_t' \\ \hat{Q}_{ZX} &= n^{-1} \sum_{t=1}^n Z_t X_t'\end{aligned}$$

于是

$$\hat{\beta}_{2SLS} - \beta^o = \left( \hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX} \right)^{-1} \hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \frac{\mathbf{Z}' \varepsilon}{n}$$

根据遍历平稳的 WLLN 可知

$$\begin{aligned}\hat{Q}_{XZ} &\xrightarrow{p} Q_{XZ} = \mathbb{E}[X_t Z_t'] \\ \hat{Q}_{ZZ} &\xrightarrow{p} Q_{ZZ} = \mathbb{E}[Z_t Z_t'] \\ \hat{Q}_{ZX} &\xrightarrow{p} Q_{ZX} = \mathbb{E}[Z_t X_t']\end{aligned}$$

并且  $\frac{\mathbf{Z}' \varepsilon}{n} \xrightarrow{p} \mathbb{E}[Z_t \varepsilon_t] = \mathbf{0}$ , 因此

$$\hat{\beta}_{2SLS} - \beta^o \xrightarrow{p} (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1} \cdot \mathbf{0} = \mathbf{0}$$

□

**7.5** 假定假设 7.1 – 7.5 成立, 证明:

(1) 当  $n \rightarrow \infty$  时,  $\sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) \xrightarrow{d} N(\mathbf{0}, \Omega)$ , 其中

$$\Omega = (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1} V Q_{ZZ}^{-1} Q_{ZX} (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$

这里的  $V$  由假设 7.5 给出.

(2) 如果  $\{Z_t \varepsilon_t\}$  是遍历平稳的鞅差分序列, 且  $\mathbb{E}[\varepsilon_t^2 | Z_t] = \sigma^2$ , 则

$$\Omega = \sigma^2 (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$

**解答** (1) 定义符号  $\hat{A} = (\hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX})^{-1} \hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1}$ , 于是

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) = \hat{A} \cdot \frac{\mathbf{Z}'\varepsilon}{\sqrt{n}}$$

根据中心极限定理假设可知

$$\frac{\mathbf{Z}'\varepsilon}{\sqrt{n}} = n^{-\frac{1}{2}} \sum_{t=1}^n Z_t \varepsilon_t \xrightarrow{d} N(\mathbf{0}, V)$$

根据 Slutsky 定理可知

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) \xrightarrow{d} N(\mathbf{0}, A V A')$$

其中  $A = (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1}$  是  $\hat{A}$  的概率极限, 因此

$$\Omega = A V A' = (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1} V Q_{ZZ}^{-1} Q_{ZX} (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}$$

(2) 因为  $\{Z_t \varepsilon_t\}$  是遍历平稳的鞅差分序列, 故而  $V = \Gamma(0) = \mathbb{E}[Z_t Z_t' \varepsilon_t^2]$ , 又因为  $\mathbb{E}[\varepsilon_t^2 | Z_t] = \sigma^2$ , 由此得到

$$V = \mathbb{E}[Z_t Z_t' \varepsilon_t^2] = \mathbb{E}[Z_t Z_t' \mathbb{E}[\varepsilon_t^2 | Z_t]] = \sigma^2 Q_{ZZ}$$

因此  $\Omega = \sigma^2 (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1}$ .

□

**7.6 假定** 假设 7.1 – 7.4, 7.6 和 7.7 成立.

(1) 定义

$$\hat{s}^2 = \frac{\hat{e}'\hat{e}}{n}$$

其中  $\hat{e} = Y - \hat{\mathbf{X}}\hat{\beta}_{2SLS}$ . 证明: 当  $n \rightarrow \infty$  时,  $\hat{s}^2 \xrightarrow{p} \sigma^2 = \text{var}(\varepsilon_t^2)$ .

(2) 定义

$$s^2 = \frac{e'e}{n}$$

其中  $e = Y - \hat{\mathbf{X}}\hat{\beta}_{2SLS}$  是第二阶段  $Y_t$  对  $\hat{X}_t = \hat{\gamma}' Z_t$  回归的估计残差. 证明:  $s^2$  不是  $\sigma^2$  的一致估计量.

**解答** (1) 定义符号

$$Q_{XX} = n^{-1} \sum_{t=1}^n X_t X_t'$$

根据  $\hat{e} = \varepsilon - \mathbf{X}(\hat{\beta}_{2SLS} - \beta^o)$  可知

$$\begin{aligned} \hat{e}'\hat{e} &= [\varepsilon - \mathbf{X}(\hat{\beta}_{2SLS} - \beta^o)]' [\varepsilon - \mathbf{X}(\hat{\beta}_{2SLS} - \beta^o)] \\ &= \varepsilon'\varepsilon - 2(\hat{\beta}_{2SLS} - \beta^o)' X' \varepsilon + (\hat{\beta}_{2SLS} - \beta^o)' \mathbf{X}' \mathbf{X} (\hat{\beta}_{2SLS} - \beta^o) \end{aligned}$$

从而

$$\hat{s}^2 = n^{-1} \sum_{t=1}^n \varepsilon_t^2 - 2(\hat{\beta}_{2SLS} - \beta^o)' \left( n^{-1} \sum_{t=1}^n X_t \varepsilon_t \right) + (\hat{\beta}_{2SLS} - \beta^o)' Q_{XX} (\hat{\beta}_{2SLS} - \beta^o)$$

又因为当  $n \rightarrow \infty$  时

$$\begin{aligned} n^{-1} \sum_{t=1}^n X_t X_t' &\xrightarrow{p} \mathbb{E}[X_t X_t'] \\ n^{-1} \sum_{t=1}^n X_t \varepsilon_t &\xrightarrow{p} \mathbb{E}[X_t \varepsilon_t] \\ n^{-1} \sum_{t=1}^n \varepsilon_t^2 &\xrightarrow{p} \mathbb{E}[\varepsilon_t^2] = \mathbb{E}[\mathbb{E}[\varepsilon_t^2 | Z_t]] = \sigma^2 \end{aligned}$$

注意到  $\hat{\beta}_{2SLS} \xrightarrow{p} \beta^o$ , 因此  $\hat{s}^2 \xrightarrow{p} \sigma^2$ .

(2) 根据  $\hat{e}$  和  $e$  的定义, 可以做等价变换

$$e = Y - \hat{\mathbf{X}} \hat{\beta}_{2SLS} = \hat{e} + (\mathbf{X} - \hat{\mathbf{X}}) \hat{\beta}_{2SLS}$$

从而

$$s^2 - \hat{s}^2 = 2n^{-1} \hat{\beta}_{2SLS}' (\mathbf{X} - \hat{\mathbf{X}})' \hat{e} + n^{-1} \hat{\beta}_{2SLS}' (\mathbf{X} - \hat{\mathbf{X}})' (\mathbf{X} - \hat{\mathbf{X}}) \hat{\beta}_{2SLS}$$

再定义符号

$$\begin{aligned} A_1 &= n^{-1} (\mathbf{X} - \hat{\mathbf{X}})' \hat{e} = n^{-1} \sum_{t=1}^n (X_t - \hat{X}_t) \hat{e}_t \\ A_2 &= n^{-1} (\mathbf{X} - \hat{\mathbf{X}})' (\mathbf{X} - \hat{\mathbf{X}}) = n^{-1} \sum_{t=1}^n (X_t - \hat{X}_t) (X_t - \hat{X}_t)' \end{aligned}$$

先来考虑  $A_1$  项

$$\begin{aligned} A_1 &= n^{-1} \sum_{t=1}^n (X_t - \hat{\gamma}' Z_t) [\varepsilon_t - X_t' (\hat{\beta}_{2SLS} - \beta^o)] \\ &= n^{-1} \sum_{t=1}^n X_t \varepsilon_t - \hat{\gamma}' \left( n^{-1} \sum_{t=1}^n Z_t \varepsilon_t \right) - (\hat{Q}_{XX} - \hat{\gamma}' \hat{Q}_{ZX}) (\hat{\beta}_{2SLS} - \beta^o) \end{aligned}$$

注意到当  $n \rightarrow \infty$  时,  $\hat{\beta}_{2SLS} \xrightarrow{p} \beta^o$ ,  $\mathbb{E}[Z_t \varepsilon_t] = \mathbf{0}$ , 根据遍历平稳的 WLLN 可知  $A_1 \xrightarrow{p} \mathbb{E}[X_t \varepsilon_t] \neq \mathbf{0}$ .

再来考虑  $A_2$  项

$$\begin{aligned} A_2 &= n^{-1} \sum_{t=1}^n (X_t - \hat{\gamma}' Z_t) (X_t - \hat{\gamma}' Z_t)' \\ &= \hat{Q}_{XX} - \hat{\gamma}' \hat{Q}_{ZX} - \hat{Q}_{XZ} \hat{\gamma} + \hat{\gamma}' \hat{Q}_{ZZ} \hat{\gamma} \end{aligned}$$

注意到  $\gamma = Q_{ZZ}^{-1} Q_{ZX}$ , 因此当  $n \rightarrow \infty$  时

$$A_2 \xrightarrow{p} Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{ZX}$$

再定义  $M_Z = I - Z(Z'Z)^{-1}Z'$ , 容易验证

$$Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{ZX} = \lim_{n \rightarrow \infty} \frac{\mathbf{X}' M_Z \mathbf{X}}{n} \sim \text{p.s.d.}$$

由此可知  $A_2$  并不依概率收敛于  $\mathbf{0}$ . 综上所述

$$s^2 - \hat{s}^2 \xrightarrow{p} 2(\beta^o)' \mathbb{E}[X_t \varepsilon_t] + (\beta^o)' (Q_{XX} - Q_{XZ} Q_{ZZ}^{-1} Q_{ZX}) \beta^o$$

也即  $s^2$  不是  $\sigma^2$  的一致估计量.

□

**7.7 [2SLS 假设检验]** 假定假设 7.1 – 7.5 成立. 定义第二阶段回归的  $F$  统计量

$$F = \frac{n(R\hat{\beta}_{2SLS} - r)'[R\hat{Q}_{\hat{X}\hat{X}}^{-1}R']^{-1}(R\hat{\beta}_{2SLS} - r)/J}{e'e/(n-K)}$$

其中  $e_t = Y_t - \hat{X}_t'\hat{\beta}_{2SLS}$  是第二阶段  $Y_t$  对  $\hat{X}_t$  回归的估计残差. 在原假设  $\mathbb{H}_0: R\beta^o = r$  下,  $J \cdot F \xrightarrow{d} \chi_J^2$  吗? 如果是, 给出理由. 如果不是, 对其进行修正, 从而使其在原假设  $\mathbb{H}_0: R\beta^o = r$  成立时, 收敛于  $\chi_J^2$  分布.

**解答** 在原假设  $\mathbb{H}_0: R\beta^o = r$  成立的情况下

$$\sqrt{n}(R\hat{\beta}_{2SLS} - r) \xrightarrow{d} N(0, R\Omega_1 R')$$

其中

$$\Omega_1 = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}Q_{XZ}Q_{ZZ}^{-1}VQ_{ZZ}^{-1}Q_{ZX}(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$$

渐近方差  $V = \text{avar}\left(n^{-1}\sum_{t=1}^n Z_t \varepsilon_t\right)$ . 现在需要分情况讨论渐近方差的形式.

(i) 如果  $\{Z_t \varepsilon_t\}$  是 MDS, 并且条件同方差假设  $\mathbb{E}[\varepsilon_t^2|Z_t] = \sigma^2$  成立, 那么  $V = \sigma^2 Q_{ZZ}$ , 此时

$$\sqrt{n}(R\hat{\beta}_{2SLS} - r) \xrightarrow{d} N[0, \sigma^2 R(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}R']$$

从而可以构建 Wald 统计量

$$W_1 = \frac{n(R\hat{\beta}_{2SLS} - r)'[R\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}R']^{-1}(R\hat{\beta}_{2SLS} - r)}{\hat{e}'\hat{e}/(n-K)} \xrightarrow{d} \chi_J^2$$

其中残差  $\hat{e} = Y - X\hat{\beta}_{2SLS}$ .

(ii) 如果  $\{Z_t \varepsilon_t\}$  是 MDS, 并且条件异方差假设  $\mathbb{E}[\varepsilon_t^2|Z_t] \neq \sigma^2$  成立, 此时

$$\sqrt{n}(R\hat{\beta}_{2SLS} - r) \xrightarrow{d} N(0, R\Omega_2 R')$$

其中

$$\Omega_2 = (Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}Q_{XZ}Q_{ZZ}^{-1}\mathbb{E}[Z_t Z_t' \varepsilon_t^2]Q_{ZZ}^{-1}Q_{ZX}(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$$

此时可以构建稳健 Wald 统计量

$$W_2 = n(R\hat{\beta}_{2SLS} - r)'[R\hat{\Omega}_2 R']^{-1}(R\hat{\beta}_{2SLS} - r) \xrightarrow{d} \chi_J^2$$

其中

$$\hat{\Omega}_2 = (\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX})^{-1}\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{V}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX}(\hat{Q}_{XZ}\hat{Q}_{ZZ}^{-1}\hat{Q}_{ZX})^{-1}$$

这里的  $\hat{V} = n^{-1}\sum_{t=1}^n Z_t Z_t' \hat{e}_t^2$ .

(iii) 如果  $\{Z_t \varepsilon_t\}$  不是 MDS, 那么需要估计长期协方差矩阵, 此时构建的 Wald 统计量和上文中的  $W_2$  具有相同的形式, 只不过需要把  $\hat{V}$  替换为长期协方差矩阵  $V$  的一致估计量,  $V$  的一个可行的一致估计量可以选为

$$\hat{V} = \sum_{j=-p_n}^{p_n} k(j/p_n)\hat{\Gamma}(j)$$



其中  $k(\cdot)$  为某个核函数, 并且当  $n \rightarrow \infty$  时,  $p_n \rightarrow \infty$ ,  $p_n/n \rightarrow 0$ .

□

**7.8 令**

$$\hat{V} = n^{-1} \sum_{t=1}^n Z_t Z_t' \hat{\varepsilon}_t^2$$

其中  $\hat{\varepsilon}_t = Y_t - X_t' \hat{\beta}_{2SLS}$ , 证明: 在假设 7.1 – 7.8 下,  $\hat{V} \xrightarrow{P} V \equiv \mathbb{E}[Z_t Z_t' \varepsilon_t^2]$ .

**解答** 类似之前的做法, 可以将  $\hat{\varepsilon}_t$  记作  $\hat{\varepsilon}_t = \varepsilon_t - X_t'(\hat{\beta}_{2SLS} - \beta^o)$ , 于是

$$\begin{aligned} V &= n^{-1} \sum_{t=1}^n Z_t Z_t' \varepsilon_t^2 - 2n^{-1} \sum_{t=1}^n Z_t Z_t' [\varepsilon_t X_t' (\hat{\beta}_{2SLS} - \beta^o)] \\ &\quad + n^{-1} \sum_{t=1}^n Z_t Z_t' [(\hat{\beta}_{2SLS} - \beta^o)' X_t X_t' (\hat{\beta}_{2SLS} - \beta^o)] \end{aligned}$$

如果对于任意  $i, j, l, m \in \{0, \dots, k\}$ ,  $\mathbb{E}[Z_{it} Z_{jt} \varepsilon_t^2]$ ,  $\mathbb{E}[Z_{it} Z_{jt} X_{lt} \varepsilon_t]$ , 以及  $\mathbb{E}[Z_{it} Z_{jt} X_{lt} X_{mt}]$  均有限, 那么由遍历平稳的 WLLN 可知

$$n^{-1} \sum_{t=1}^n Z_t Z_t' \varepsilon_t^2 \xrightarrow{P} \mathbb{E}[Z_t Z_t' \varepsilon_t^2]$$

加之  $\hat{\beta}_{2SLS} \xrightarrow{P} \beta^o$  便可推之  $\hat{V} \xrightarrow{P} V$  成立.

□

**7.9 假定下列假设成立.**

**假设 7.9.1 [线性]**  $\{Y_t, X_t', Z_t'\}$  是一个可观测的遍历平稳过程

$$Y_t = X_t' \beta^o + \varepsilon_t \quad t = 1, \dots, n$$

$\beta^o$  是  $K \times 1$  未知参数向量,  $\varepsilon_t$  是不可观测扰动项.

**假设 7.9.2 [非奇异性]**  $K \times K$  矩阵

$$Q_{XX} = \mathbb{E}[X_t X_t']$$

是有限、对称和非奇异的矩阵.

**假设 7.9.3 [正交性]**

- (i)  $\mathbb{E}[X_t \varepsilon_t] = \mathbf{0}$ ;
- (ii)  $\mathbb{E}[Z_t \varepsilon_t] = \mathbf{0}$ , 其中  $Z_t$  是  $l \times 1$  随机向量,  $l \geq K$ ;
- (iii)  $l \times l$  矩阵

$$Q_{ZZ} = \mathbb{E}[Z_t Z_t']$$

是有限和非奇异的,  $l \times K$  矩阵

$$Q_{ZX} = \mathbb{E}[Z_t X_t']$$

是有限与满秩的.

**假设 7.9.4 [鞅差分]**  $\{(X_t', Z_t') \varepsilon_t\}$  是 MDS.

**假设 7.9.5 [条件同方差]**  $\mathbb{E}[\varepsilon_t^2 | X_t, Z_t] = \sigma^2$ .

在这些假设下, OLS 估计量

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

和 2SLS 估计量

$$\hat{\beta}_{2SLS} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y}$$

都是  $\beta^o$  的一致估计量.

(1) 证明: 通过选择合适的工具变量,  $\hat{\beta}$  是 2SLS 估计量  $\hat{\beta}_{2SLS}$  的一个特例.

(2) 估计量  $\hat{\beta}$  和  $\hat{\beta}_{2SLS}$  哪一个更渐近有效?

**解答** (1) 这是显然的, 选取  $\mathbf{X}$  自身作为工具变量  $\mathbf{Z}$ , 立即可得  $\hat{\beta} = \hat{\beta}_{2SLS}$ .

(2) 根据题目假设可知

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, \sigma^2 Q_{XX}^{-1})$$

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) \xrightarrow{d} N[\mathbf{0}, \sigma^2(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}]$$

记  $\Omega_1 = \sigma^2 Q_{XX}^{-1}$ ,  $\Omega_2 = \sigma^2(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$ . 于是根据问题 7.6 的结论可知

$$\Omega_1^{-1} - \Omega_2^{-1} = \sigma^{-2}(Q_{XX} - Q_{XZ}Q_{ZZ}^{-1}Q_{ZX}) \sim \text{p.s.d.}$$

从而  $\hat{\beta}$  比  $\hat{\beta}_{2SLS}$  更加渐近有效.

□

### 7.10 考虑线性回归模型

$$Y_t = X_t'\beta^o + \varepsilon_t$$

其中  $\mathbb{E}[X_t\varepsilon_t] \neq \mathbf{0}$ . 我们的目的是对  $\beta^o$  的一致估计.

首先, 考虑辅助回归

$$X_t = \gamma'Z_t + v_t$$

其中  $X_t$  是原始模型的解释变量,  $Z_t$  是工具变量,  $\gamma = \mathbb{E}[Z_tZ_t']^{-1}\mathbb{E}[Z_tX_t']$  是最优线性最小二乘近似系数,  $v_t$  是  $K \times 1$  回归误差项.

假设现在不是分解  $X_t$ , 而是对原始回归模型中的随机扰动项  $\varepsilon_t$  进行分解

$$\varepsilon_t = v_t'\rho^o + u_t$$

其中  $\rho^o = \mathbb{E}[v_tv_t']^{-1}\mathbb{E}[v_t\varepsilon_t]$  是最优线性最小二乘近似系数.

假设  $v_t$  是可观测的, 考虑下面扩展的线性回归模型

$$Y_t = X_t'\beta^o + v_t'\rho^o + u_t$$

证明:  $\mathbb{E}[(X_t', v_t)u_t] = \mathbf{0}$ . 这一正交条件的一个重要含义是, 如果  $v_t$  是可观测的, 则  $Y_t$  对  $X_t$  和  $v_t$  的 OLS 估计量是  $(\beta^o, \rho^o)'$  的一致估计.

**解答** 根据定理 2.6 可知  $\mathbb{E}[Z_tv_t'] = \mathbf{0}$ ,  $\mathbb{E}[v_tu_t] = \mathbf{0}$ , 故而有

$$\begin{aligned}\mathbb{E}[X_tu_t] &= \mathbb{E}[(\gamma'Z_t + v_t)u_t] = \mathbb{E}[\gamma'Z_tu_t] \\ &= \mathbb{E}[\gamma'Z_t(\varepsilon_t - v_t'\rho^o)] = \mathbf{0}\end{aligned}$$

因此  $\mathbb{E}[(X_t', v_t)'u_t] = \mathbf{0}$ .

□

**7.11** 考虑问题 7.10 中对  $\beta^o$  的估计. 在实际应用中,  $v_t$  是不可观测的. 但是, 可用其估计量

$$\hat{v}_t = X_t - \hat{\gamma}'Z_t = X_t - \hat{X}_t$$

现在考虑以下拓展的可行线性回归模型

$$Y_t = X_t'\beta^o + \hat{v}_t'\rho^o + \tilde{u}_t$$

并将对应的 OLS 估计量记为  $\hat{\alpha} = (\hat{\beta}', \hat{\rho}')'$ , 其中  $\hat{\beta}$  是  $\beta^o$  的 OLS 估计量,  $\hat{\rho}$  是  $\rho^o$  的 OLS 估计量. 证明:  $\hat{\beta} = \hat{\beta}_{2SLS}$ .

**解答** 方法一: 首先有

$$\mathbf{X} = \mathbf{Z}\hat{\gamma} + \hat{\mathbf{V}} = \hat{\mathbf{X}} + \hat{\mathbf{V}}$$

易知  $\hat{\mathbf{X}}'\hat{\mathbf{V}} = \hat{\mathbf{V}}'\hat{\mathbf{X}} = \mathbf{0}$ . 现在将拓展的可行线性回归模型改写为矩阵形式

$$Y = \mathbf{X}\beta^o + \hat{\mathbf{V}}\rho^o + U$$

利用分块回归<sup>11</sup>可知

$$\hat{\beta} = (\mathbf{X}'M_2\mathbf{X})^{-1}\mathbf{X}'M_2Y$$

其中,  $M_2 = I_n - \hat{\mathbf{V}}(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}'$ . 根据线性代数可知

$$\begin{aligned}\mathbf{X}'M_2\mathbf{X} &= \mathbf{X}'[I - \hat{\mathbf{V}}(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}']\mathbf{X} \\ &= \mathbf{X}'\mathbf{X} - \mathbf{X}'\hat{\mathbf{V}}(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}'(\hat{\mathbf{X}} + \hat{\mathbf{V}}) \\ &= \mathbf{X}'\mathbf{X} - \mathbf{X}'\hat{\mathbf{V}} = \mathbf{X}'\hat{\mathbf{X}} = \hat{\mathbf{X}}'\hat{\mathbf{X}}\end{aligned}$$

另一方面

$$\begin{aligned}\mathbf{X}'M_2Y &= (\hat{\mathbf{X}} + \hat{\mathbf{V}})'[I - \hat{\mathbf{V}}(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}']Y \\ &= \hat{\mathbf{X}}'Y + \hat{\mathbf{V}}'[I - \hat{\mathbf{V}}(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}']Y = \hat{\mathbf{X}}'Y\end{aligned}$$

因此  $\hat{\beta} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'Y = \hat{\beta}_{2SLS}$ .

方法二: 为了便于利用矩阵分解公式, 这里将回归模型改写为

$$Y = \hat{\mathbf{V}}\rho^o + \mathbf{X}\beta^o + U$$

于是可以将 OLS 估计量  $(\hat{\rho}, \hat{\beta})'$  记作

$$\begin{bmatrix} \hat{\rho} \\ \hat{\beta} \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{V}}'\hat{\mathbf{V}} & \hat{\mathbf{V}}'\mathbf{X} \\ \mathbf{X}'\hat{\mathbf{V}} & \mathbf{X}'\mathbf{X} \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{V}}' \\ \mathbf{X}' \end{bmatrix} Y = \begin{bmatrix} B & C' \\ C & D \end{bmatrix}^{-1} \begin{bmatrix} \hat{\mathbf{V}}' \\ \mathbf{X}' \end{bmatrix} Y$$

<sup>11</sup> 设总体线性回归模型为

$$Y = \mathbf{X}_1\beta_1^o + \mathbf{X}_2\beta_2^o + \varepsilon$$

则  $(\beta_1^o, \beta_2^o)$  的 OLS 估计量为

$$\begin{aligned}\hat{\beta}_1 &= (\mathbf{X}_1'M_2\mathbf{X}_1)^{-1}\mathbf{X}_1'M_2Y \\ \hat{\beta}_2 &= (\mathbf{X}_2'M_1\mathbf{X}_2)^{-1}\mathbf{X}_2'M_1Y\end{aligned}$$

其中  $M_1 = I_n - \mathbf{X}_1(\mathbf{X}_1'\mathbf{X}_1)^{-1}\mathbf{X}_1'$ ,  $M_2 = I_n - \mathbf{X}_2(\mathbf{X}_2'\mathbf{X}_2)^{-1}\mathbf{X}_2'$ .

设  $E = D - CB^{-1}C'$ , 于是

$$E = \mathbf{X}'\mathbf{X} - \mathbf{X}'\hat{\mathbf{V}}(\hat{\mathbf{V}}'\hat{\mathbf{V}})^{-1}\hat{\mathbf{V}}'\mathbf{X} = \hat{\mathbf{X}}'\hat{\mathbf{X}}$$

根据矩阵分解公式<sup>12</sup>可知

$$\hat{\beta} = -E^{-1}CB^{-1}\hat{\mathbf{V}}'Y + E^{-1}\mathbf{X}'Y = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'Y$$

□

**7.12** 假设  $n \times 1$  向量  $\hat{Y}$  是  $Y$  对  $\mathbf{Z}$  回归的拟合值,  $n \times K$  矩阵  $\hat{\mathbf{X}}$  是  $\mathbf{X}$  对  $\mathbf{Z}$  回归的拟合值. 证明:  $\hat{\beta}_{2SLS}$  等价于  $\hat{Y}$  对  $\hat{\mathbf{X}}$  回归的 OLS 估计量, 即  $\hat{\beta}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\hat{Y}$ .

**解答** 将  $Y$  对  $\mathbf{Z}$  进行 OLS 回归得到

$$Y = \mathbf{Z}\hat{\delta} + \hat{U}$$

其中  $\hat{\delta}$  和  $\hat{U}$  分别为 OLS 估计量和残差. 易知  $\mathbf{Z}'\hat{U} = \mathbf{0}$ , 从而  $\hat{\mathbf{X}}'\hat{U} = \mathbf{0}$ , 因此

$$\hat{\beta}_{2SLS} = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'(\hat{Y} + \hat{U}) = (\hat{\mathbf{X}}'\hat{\mathbf{X}})^{-1}\hat{\mathbf{X}}'\hat{Y}$$

□

**7.13 [Hausman 检验]** 假定问题7.9中的假设 7.1, 7.2, 7.3(2, 3), 7.4 和 7.5 成立. 一个检验原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t | X_t] = \mathbf{0}$  的检验统计量可通过比较 OLS 估计量  $\hat{\beta}$  和 2SLS 估计量  $\hat{\beta}_{2SLS}$  的大小来构造, 因为在原假设  $\mathbb{H}_0$  下, 它们将收敛于相同的极限, 而在备择假设下它们一般将收敛于不同的极限.

(1) 证明

$$\sqrt{n}(\hat{\beta} - \beta^o) = Q_{XX}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n X_t \varepsilon_t + o_p(1)$$

其中  $Q_{XX} = \mathbb{E}[X_t X_t']$ .

(2) 证明

$$\sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) = Q_{\tilde{X}\tilde{X}}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{X}_t \varepsilon_t + o_p(1)$$

其中  $Q_{\tilde{X}\tilde{X}} = \mathbb{E}[\tilde{X}_t \tilde{X}_t']$ ,  $\tilde{X}_t = \gamma' Z_t$ ,  $\gamma = \mathbb{E}[Z_t Z_t']^{-1} \mathbb{E}[Z_t X_t']$ .

(3) 证明

$$\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}) = \frac{1}{\sqrt{n}} \sum_{t=1}^n (Q_{XX}^{-1} X_t - Q_{\tilde{X}\tilde{X}}^{-1} \tilde{X}_t) \varepsilon_t + o_p(1)$$

(4)  $\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta})$  的渐近分布是由 (3) 中的主导项决定的. 推导其渐近分布.

<sup>12</sup> 设矩阵  $B \in \mathbb{R}^{m \times m}$ ,  $C \in \mathbb{R}^{n \times m}$ ,  $D \in \mathbb{R}^{n \times n}$ , 若矩阵

$$A = \begin{bmatrix} B & C' \\ C & D \end{bmatrix}$$

可逆, 那么它的逆矩阵为

$$A^{-1} = \begin{bmatrix} B^{-1}(I + C'E^{-1}CB^{-1}) & -B'^{-1}C'E^{-1} \\ -E^{-1}CB^{-1} & E^{-1} \end{bmatrix}$$

其中  $E = D - CB^{-1}C'$ .

(5) 构造一个检验  $H_0$  的渐近  $\chi^2$  检验统计量. 这里  $\chi^2$  分布的自由度是多少? 假设  $Q_{XX} - Q_{\tilde{X}\tilde{X}}$  是正定的.

**解答** (1) 显然有

$$\sqrt{n}(\hat{\beta} - \beta^o) = \hat{Q}_{XX}^{-1} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right)$$

由于  $\hat{Q}_{XX}^{-1} \xrightarrow{p} Q_{XX}^{-1}$ , 故而  $\hat{Q}_{XX}^{-1} = Q_{XX}^{-1} + o_p(1)$ , 又因为  $n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \xrightarrow{d} N(0, \sigma^2 Q_{XX}) = O_p(1)$ , 且  $O_p(1) \cdot o_p(1) = o_p(1)$ , 因此

$$\sqrt{n}(\hat{\beta} - \beta^o) = Q_{XX}^{-1} \left( n^{-\frac{1}{2}} \sum_{t=1}^n X_t \varepsilon_t \right) + o_p(1)$$

(2) 类似可以得到

$$\begin{aligned} \sqrt{n}(\hat{\beta}_{2SLS} - \beta^o) &= \left( n^{-1} \sum_{t=1}^n \hat{X}_t \hat{X}_t' \right)^{-1} \left( n^{-\frac{1}{2}} \sum_{t=1}^n \hat{X}_t \varepsilon_t \right) \\ &= \left( n^{-1} \sum_{t=1}^n \hat{\gamma}' Z_t Z_t' \hat{\gamma} \right)^{-1} \left( n^{-\frac{1}{2}} \sum_{t=1}^n \hat{\gamma}' Z_t \varepsilon_t \right) \\ &= Q_{\tilde{X}\tilde{X}}^{-1} \frac{1}{\sqrt{n}} \sum_{t=1}^n \tilde{X}_t \varepsilon_t + o_p(1) \end{aligned}$$

(3) 根据 (1) 和 (2) 的结果可知

$$\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( Q_{\tilde{X}\tilde{X}}^{-1} \tilde{X}_t - Q_{XX}^{-1} X_t \right) \varepsilon_t + o_p(1)$$

(4) 由于  $\left\{ \left( Q_{\tilde{X}\tilde{X}}^{-1} \tilde{X}_t - Q_{XX}^{-1} X_t \right) \varepsilon_t \right\}$  是遍历平稳的 MDS, 于是根据中心极限定理可知

$$\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}) \xrightarrow{d} N(0, \Omega)$$

其中  $\Omega = \text{var} \left[ \left( Q_{\tilde{X}\tilde{X}}^{-1} \tilde{X}_t - Q_{XX}^{-1} X_t \right) \varepsilon_t \right]$ . 由于条件同方差假定成立, 将  $\Omega$  展开可得

$$\Omega = \sigma^2 \left( Q_{\tilde{X}\tilde{X}}^{-1} - Q_{\tilde{X}\tilde{X}}^{-1} Q_{\tilde{X}X} Q_{XX}^{-1} - Q_{XX}^{-1} Q_{X\tilde{X}} Q_{\tilde{X}\tilde{X}}^{-1} + Q_{XX}^{-1} \right)$$

其中  $Q_{\tilde{X}X} = \mathbb{E}[\tilde{X}_t X_t']$ ,  $Q_{X\tilde{X}} = \mathbb{E}[X_t \tilde{X}_t']$ . 考虑  $X_t$  对  $Z_t$  的线性投影

$$X_t = \gamma' Z_t + u_t$$

易知  $\mathbb{E}[Z_t u_t'] = 0$ , 于是

$$Q_{\tilde{X}X} = \gamma' \mathbb{E}[Z_t Z_t'] \gamma + \gamma' \mathbb{E}[Z_t u_t'] = Q_{\tilde{X}\tilde{X}}$$

同理可证  $Q_{X\tilde{X}} = Q_{\tilde{X}\tilde{X}}$ , 因此

$$\Omega = \sigma^2 \left( Q_{\tilde{X}\tilde{X}}^{-1} - Q_{XX}^{-1} \right)$$

也即

$$\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}) \xrightarrow{d} N \left[ 0, \sigma^2 \left( Q_{\tilde{X}\tilde{X}}^{-1} - Q_{XX}^{-1} \right) \right]$$

(5) 显然可以构建一个 Hausman 检验统计量

$$H = n(\hat{\beta}_{2SLS} - \hat{\beta})' \left[ s^2 \left( \hat{Q}_{\tilde{X}\tilde{X}}^{-1} - \hat{Q}_{XX}^{-1} \right) \right]^{-1} (\hat{\beta}_{2SLS} - \hat{\beta}) \xrightarrow{d} \chi_K^2$$

其中

$$\begin{aligned} \hat{Q}_{\tilde{X}\tilde{X}} &= n^{-1} \sum_{t=1}^n \tilde{X}_t \tilde{X}_t', & \hat{Q}_{XX} &= n^{-1} \sum_{t=1}^n X_t X_t' \\ s^2 &= e'e/(n-K), & e &= Y - X\hat{\beta} \end{aligned}$$

注意到  $\hat{Q}_{\tilde{X}\tilde{X}} = \hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX}$ , 这其实就是定理 7.12 中的 Hausman 检验统计量.

□

**7.14** 假定问题 7.9 中的假设 7.1, 7.2, 7.3(2, 3) 和 7.4 成立, 且  $\mathbb{E}[X_{jt}^2] < \infty, 1 \leq j \leq K$ ,  $\mathbb{E}[Z_{jt}^4] < \infty, 1 \leq j \leq l$ . 另外假设  $\mathbb{E}[\varepsilon_t^2 | X_t, Z_t] \neq \sigma^2$ . 构造一个稳健的 Hausman 检验统计量以检验原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t | X_t] = 0$ , 并在原假设  $\mathbb{H}_0$  下推导出它的渐近分布.

**解答** 根据问题 7.13 的结论可知

$$\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}) = \frac{1}{\sqrt{n}} \sum_{t=1}^n \left( Q_{\tilde{X}\tilde{X}}^{-1} \tilde{X}_t - Q_{XX}^{-1} X_t \right) \varepsilon_t + o_p(1)$$

根据中心极限定理得

$$\sqrt{n}(\hat{\beta}_{2SLS} - \hat{\beta}) \xrightarrow{d} N(0, \Omega)$$

其中  $\Omega = \text{var} \left[ \left( Q_{\tilde{X}\tilde{X}}^{-1} \tilde{X}_t - Q_{XX}^{-1} X_t \right) \varepsilon_t \right]$ .

利用  $\tilde{X}_t = \gamma' Z_t, \gamma = Q_{ZZ}^{-1} Q_{ZX}$  以及  $A = (Q_{XZ} Q_{ZZ}^{-1} Q_{ZX})^{-1} Q_{XZ} Q_{ZZ}^{-1}$  可以将  $\Omega$  展开为

$$\Omega = AV_1A' - AV_2Q_{XX}^{-1} - Q_{XX}^{-1}V_3A' + Q_{XX}^{-1}V_4Q_{XX}^{-1}$$

其中  $V_1 = \mathbb{E}[Z_t Z_t' \varepsilon_t^2], V_2 = \mathbb{E}[Z_t X_t' \varepsilon_t^2], V_3 = \mathbb{E}[X_t Z_t' \varepsilon_t^2], V_4 = \mathbb{E}[X_t X_t' \varepsilon_t^2]$ . 由于当  $n \rightarrow \infty$  时

$$\begin{aligned} \hat{Q}_{XX} &\xrightarrow{p} Q_{XX}^{-1} \\ \hat{A} &= (\hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX})^{-1} \hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \xrightarrow{p} A \end{aligned}$$

现在只需要得到  $V_i$  的一致估计量. 按照之前的做法, 用残差  $e_t = Y_t - X_t' \hat{\beta}$  替代  $\varepsilon_t$ , 由于相关的矩条件成立, 根据遍历平稳的 WLLN 可知

$$\begin{aligned} \hat{V}_1 &= n^{-1} \sum_{t=1}^n Z_t Z_t' e_t^2 \xrightarrow{p} V_1 \\ \hat{V}_2 &= n^{-1} \sum_{t=1}^n Z_t X_t' e_t^2 \xrightarrow{p} V_2 \\ \hat{V}_3 &= n^{-1} \sum_{t=1}^n X_t Z_t' e_t^2 \xrightarrow{p} V_3 \\ \hat{V}_4 &= n^{-1} \sum_{t=1}^n X_t X_t' e_t^2 \xrightarrow{p} V_4 \end{aligned}$$

于是  $\Omega$  的一致估计量为

$$\hat{\Omega} = \hat{A} \hat{V}_1 \hat{A}' - \hat{A} \hat{V}_2 \hat{Q}_{XX}^{-1} - \hat{Q}_{XX}^{-1} \hat{V}_3 \hat{A}' + \hat{Q}_{XX}^{-1} \hat{V}_4 \hat{Q}_{XX}^{-1}$$

在原假设  $\mathbb{H}_0$  成立的条件下, 可以构建稳健 Hausman 检验统计量

$$H_{\text{robust}} = n(\hat{\beta}_{2\text{SLS}} - \hat{\beta})' \hat{\Omega}^{-1} (\hat{\beta}_{2\text{SLS}} - \hat{\beta}) \xrightarrow{d} \chi_K^2$$

□

**7.15** 在定理 7.12 中, Hausman 检验统计量定义为

$$H = \frac{n(\hat{\beta}_{2\text{SLS}} - \hat{\beta})' \left[ (\hat{Q}_{XZ} \hat{Q}_{ZZ}^{-1} \hat{Q}_{ZX})^{-1} - \hat{Q}_{XX}^{-1} \right]^{-1} (\hat{\beta}_{2\text{SLS}} - \hat{\beta})}{s^2}$$

这里  $s^2 = e'e/(n-K)$ ,  $e = Y - \mathbf{X}\hat{\beta}$  为 OLS 估计残差. 再定义另一个 Hausman 检验统计量, 记为  $\hat{H}$ , 其定义与  $H$  一样, 只是将  $s^2$  换成  $\hat{s}^2 = \hat{e}'\hat{e}/(n-K)$ ,  $\hat{e} = Y - \mathbf{X}\hat{\beta}_{2\text{SLS}}$ .

(1) 假设原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t|X_t] = 0$  成立, 证明: 当  $n \rightarrow \infty$  时,  $\hat{H}$  也服从渐近  $\chi_K^2$  分布.

(2) 假设样本容量  $n$  是有限的,  $H$  和  $\hat{H}$  中的哪一个检验统计量在备择假设成立时的第 II 类错误会小些? 请解释.

**解答** (1) 这结论是显然的, 为了让  $n \rightarrow \infty$  时,  $\hat{H} \xrightarrow{d} \chi_K^2$ , 只需证明此时  $\hat{s}^2 \xrightarrow{p} \sigma^2$  即可, 这在问题 7.6 中已经证明过了.

(2) 在备择假设  $\mathbb{H}_1: \mathbb{E}[\varepsilon_t|X_t] \neq 0$  成立的情况下,  $\hat{H}$  检验统计量利用了  $\mathbb{H}_1$  这一信息, 而  $H$  检验统计量没有, 因此  $\hat{H}$  犯第 II 类错误的可能性更小.

□

# 广义矩方法

## 8.1 GMM 估计量定义为

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

其中  $\beta$  是  $K \times 1$  参数向量,  $\hat{W}$  是  $l \times l$  随机对称矩阵, 样本矩

$$\hat{m}(\beta) = n^{-1} \sum_{t=1}^n m_t(\beta)$$

其中  $m_t(\beta)$  是  $l \times 1$  矩函数,  $l \geq K$ . 有下列假设:

假设 8.1.1  $\beta^o$  是  $\mathbb{E}[m_t(\beta^o)] = \mathbf{0}$  的唯一解, 且  $\beta^o$  是  $\Theta$  的内点.

假设 8.1.2  $\{m_t(\beta^o)\}$  是遍历平稳的 MDS, 即

$$\mathbb{E}[m_t(\beta^o) | I_{t-1}] = \mathbf{0}$$

其中  $I_{t-1}$  是第  $t-1$  期的信息集.

假设 8.1.3  $m_t(\beta)$  相对于  $\beta \in \Theta$  连续可微的概率为 1, 且

$$\sup_{\beta \in \Theta} \left\| \frac{d\hat{m}(\beta)}{d\beta} - D(\beta) \right\| \xrightarrow{p} \mathbf{0}$$

其中  $D(\beta) \equiv \frac{d}{d\beta} \mathbb{E}[m_t(\beta)] = \mathbb{E} \left[ \frac{\partial}{\partial \beta} m_t(\beta) \right]$ .

假设 8.1.4 存在某一  $l \times l$  有限、对称与正定的矩阵  $V_o$ , 使得  $\sqrt{n}\hat{m}(\beta^o) \xrightarrow{d} N(\mathbf{0}, V_o)$ .

假设 8.1.5  $\hat{W} \xrightarrow{p} W$ , 其中  $W$  是  $l \times l$  有限、对称与正定的矩阵.

由这些假设可以证明  $\hat{\beta} \xrightarrow{p} \beta^o$ , 这一结论可以回答下列问题. 如有必要, 也可增加新假设.

- (1) 给出  $V_o$  的表达式.
- (2) 给出上述 GMM 最小化问题的一阶条件.
- (3) 推导  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近分布.
- (4) 给出渐近最优解的权重矩阵  $\hat{W}$ . 解释为什么所选择的权重矩阵  $\hat{W}$  是最优的.

**解答** (1) 由于  $\{m_t(\beta^o)\}$  是遍历平稳的, 因此

$$V_o = \text{avar} \left[ n^{-\frac{1}{2}} \sum_{t=1}^n \hat{m}_t(\beta^o) \right] = \lim_{n \rightarrow \infty} \text{var} \left[ n^{-\frac{1}{2}} \sum_{t=1}^n \hat{m}_t(\beta^o) \right] = \mathbb{E}[m_t(\beta^o)m_t(\beta^o)']$$

(2) 设  $\hat{Q}(\beta) = \hat{m}(\beta)' \hat{W}^{-1} \hat{m}(\beta)$ , 于是一阶条件为

$$\left. \frac{d\hat{Q}(\beta)}{d\beta} \right|_{\beta=\hat{\beta}} = 2 \frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1} \hat{m}(\hat{\beta}) = \mathbf{0}$$

(3) 将  $\sqrt{n}\hat{m}(\hat{\beta})$  在  $\beta = \beta^o$  处一阶 Taylor 展开

$$\sqrt{n}\hat{m}(\hat{\beta}) = \sqrt{n}\hat{m}(\beta^o) + \frac{d\hat{m}(\bar{\beta})}{d\beta} \sqrt{n}(\hat{\beta} - \beta^o)$$

其中  $\bar{\beta} = \lambda \hat{\beta} + (1 - \lambda)\beta^o$ ,  $\lambda \in [0, 1]$ . 将 GMM 最小化的一阶条件代入后得



$$\frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1} \sqrt{n} \hat{m}(\beta^o) + \frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1} \frac{d\hat{m}(\bar{\beta})}{d\beta} \sqrt{n} (\hat{\beta} - \beta^o) = \mathbf{0}$$

也即

$$\sqrt{n}(\hat{\beta} - \beta^o) = - \left[ \frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1} \frac{d\hat{m}(\bar{\beta})}{d\beta} \right]^{-1} \frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1} \sqrt{n} \hat{m}(\beta^o)$$

先来考虑  $\frac{d\hat{m}(\hat{\beta})}{d\beta}$ , 假定函数  $D(\cdot)$  对于  $\beta \in \Theta$  是连续的, 根据假设 8.1.3 和  $\hat{\beta} \xrightarrow{p} \beta^o$  可知

$$\left\| \frac{d\hat{m}(\hat{\beta})}{d\beta} - D(\beta^o) \right\| \leq \sup_{\beta \in \Theta} \left\| \frac{d\hat{m}(\hat{\beta})}{d\beta} - D(\hat{\beta}) \right\| + \left\| D(\hat{\beta}) - D(\beta^o) \right\| \xrightarrow{p} \mathbf{0}$$

其中  $D(\beta^o) = D_o = \mathbb{E} \left[ \frac{dm_t(\beta^o)}{d\beta} \right]$ . 由于

$$\|\bar{\beta} - \beta^o\| = \|\lambda(\hat{\beta} - \beta^o)\| \leq \|\hat{\beta} - \beta^o\| \xrightarrow{p} \mathbf{0}$$

故而  $\frac{d\hat{m}(\bar{\beta})}{d\beta} \xrightarrow{p} D_o$ . 记  $\hat{A} = - \left[ \frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1} \frac{d\hat{m}(\bar{\beta})}{d\beta} \right]^{-1} \frac{d\hat{m}(\hat{\beta})}{d\beta'} \hat{W}^{-1}$ , 于是当  $n \rightarrow \infty$  时

$$\hat{A} \xrightarrow{p} -(D_o' W^{-1} D_o)^{-1} D_o' W^{-1}$$

根据假设 8.1.4 和 (1) 的结论

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, \Omega)$$

其中  $\Omega = (D_o' W^{-1} D_o)^{-1} D_o' W^{-1} V_o W^{-1} D_o (D_o' W^{-1} D_o)^{-1}$ ,  $V_o = \mathbb{E}[m_t(\beta^o) m_t(\beta^o)']$ .

(4) 最优权重矩阵  $W = V_o = \text{avar}[\sqrt{n} \hat{m}(\beta^o)]$ , 此时  $\Omega_o = (D_o' V_o^{-1} D_o)^{-1}$ . 设  $\Omega$  为 (3) 中的那个, 其中  $W$  是任意  $l \times l$  有限、对称与正定的矩阵, 现在需要证明  $\Omega_o^{-1} - \Omega^{-1}$  是半正定的, 可以证明

$$\Omega_o^{-1} - \Omega^{-1} = \left( G V_o^{-\frac{1}{2}} D_o \right)' \left( G V_o^{-\frac{1}{2}} D_o \right)$$

其中  $G = I - V_o^{-\frac{1}{2}} W^{-1} D_o (D_o' W^{-1} V_o W^{-1} D_o)^{-1} D_o' W^{-1} V_o^{-\frac{1}{2}}$ .

因此通过选择  $W = \text{avar}[\sqrt{n} \hat{m}(\beta^o)]$  得到的  $\Omega_o$  是渐近有效的. 注意,  $W$  并不是唯一的, 对于任意  $c \neq 0$ ,  $\tilde{W} = c V_o$  也是最优权重矩阵.

□

**8.2** (1) 证明: 线性回归模型  $Y_t = X_t' \beta^o + \varepsilon_t$  中参数  $\beta^o$  的 2SLS 估计量  $\hat{\beta}_{2SLS}$  是通过选择合适的矩函数  $m_t(\beta)$  和权重矩阵  $\hat{W}$  而得到的 GMM 估计量.

(2) 假设  $\{Z_t \varepsilon_t\}$  是一个遍历平稳的 MDS. 请分别比较在条件同方差  $\mathbb{E}[\varepsilon_t^2 | Z_t] = \sigma^2$  和条件异方差  $\mathbb{E}[\varepsilon_t^2 | Z_t] \neq \sigma^2$  情形下, 通过选择最优权重矩阵  $\hat{W} = \tilde{V}$  得到的 GMM 估计量与  $\hat{\beta}_{2SLS}$  的有效性.

**解答** (1) 选取矩函数  $m_t(\beta) = Z_t(Y_t - X_t' \beta)$  与权重矩阵  $W = \mathbb{E}[Z_t Z_t']$ , 此时样本矩为

$$\hat{m}_t(\beta) = \frac{\mathbf{Z}'(Y - \mathbf{X}\beta)}{n}, \quad \hat{W} = \frac{\mathbf{Z}'\mathbf{Z}}{n}$$

于是  $\beta^o$  的 GMM 估计量为

$$\hat{\beta}_{\text{GMM}} = \arg \min_{\beta \in \Theta} \hat{m}(\beta)' \hat{W}^{-1} \hat{m}(\beta)$$

一阶条件为

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

从而

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

将  $\hat{W} = \mathbf{Z}'\mathbf{Z}/n$  代入上式

$$\hat{\beta}_{\text{GMM}} = [\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{X}]^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'Y = \hat{\beta}_{2\text{SLS}}$$

(2) 问题 8.1 中已经证明, 选取最优权重矩阵  $W = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$  得到的 GMM 估计量具有渐近有效性. 因为  $\{Z_t\varepsilon_t\}$  是遍历平稳的 MDS, 故而  $\text{avar}[\sqrt{n}\hat{m}(\beta^o)] = \mathbb{E}[m_t(\beta^o)m_t(\beta^o)']$ .

(i) 当条件同方差  $\mathbb{E}[\varepsilon_t^2|Z_t] = \sigma^2$  成立时, 选取最优权重矩阵  $W = V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)] = \sigma^2 Q_{ZZ}$ .

由于矩函数为  $m_t(\beta) = Z_t(Y_t - X_t'\beta)$ , 由矩阵微分可知

$$D_o = \frac{d}{d\beta} \mathbb{E}[m_t(\beta)] = -Q_{ZX}$$

此时  $\Omega_o = (D_o'V_o^{-1}D_o)^{-1} = \sigma^2(Q_{XZ}Q_{ZZ}^{-1}Q_{ZX})^{-1}$ , 与条件同方差假定下  $\sqrt{n}\hat{\beta}_{2\text{SLS}}$  的渐近方差一致, 因此 GMM 估计量和 2SLS 估计量具有相同的渐近有效性.

(ii) 当条件异方差  $\mathbb{E}[\varepsilon_t^2|Z_t] \neq \sigma^2$  成立时, 通过选择权重矩阵  $W = Q_{ZZ}$  得到的 2SLS 估计量与最优权重矩阵  $\tilde{W} = \mathbb{E}[Z_t Z_t' \varepsilon_t^2]$  不一致, 根据 GMM 估计量的渐近有效性定理可知, 此时 GMM 估计量更加渐近有效.

□

**8.3** 假设  $\{m_t(\beta)\}$  是一个遍历平稳的 MDS, 其中  $m(\cdot)$  在紧集  $\Theta$  上连续,  $\{m_t(\beta)m_t(\beta)'\}$  在  $\Theta$  上满足 UWLLN,  $V(\beta) = \mathbb{E}[m_t(\beta)m_t(\beta)']$  在  $\Theta$  连续, 且  $V_o = \mathbb{E}[m_t(\beta^o)m_t(\beta^o)']$  是有限、对称与非奇异的矩阵. 定义  $\hat{V} = n^{-1} \sum_{t=1}^n m_t(\hat{\beta})m_t(\hat{\beta})'$ , 其中  $\hat{\beta}$  是  $\beta^o$  的一致估计量. 证明: 当  $n \rightarrow \infty$

时,  $\hat{V} \xrightarrow{p} V_o$ .

**解答** 设  $V = \mathbb{E}[m_t(\hat{\beta})m_t(\hat{\beta})']$ , 于是由三角不等式可知

$$\|\hat{V} - V_o\| \leq \sup_{\beta \in \Theta} \|\hat{V} - V\| + \|V - V_o\| \xrightarrow{p} \mathbf{0}$$

其中  $\|\hat{V} - V\| \xrightarrow{p} \mathbf{0}$  是由 UWLLN 保证的, 而  $\|\hat{V} - V_o\| \xrightarrow{p} \mathbf{0}$  是由  $\hat{\beta} \xrightarrow{p} \beta^o$  和  $m(\cdot)$  在  $\Theta$  上的连续性保证的.

□

**8.4** 给定权重矩阵  $\hat{W} \xrightarrow{p} W$ , 这里  $W$  是一个  $l \times l$  有限、对称与正定的矩阵, 且不一定等于  $V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$ . 用相对于  $\hat{W}$  的渐近次优 GMM 估计量  $\hat{\beta}$  构建一个 Wald 检验统计量, 以检验原假设  $\mathbb{H}_0: R(\beta^o) = r$ , 并推导出其渐近分布. 假设所有必要的正则条件成立.

**解答** 在原假设  $\mathbb{H}_0 : R(\beta^o) = r$  成立的条件下, 将  $\sqrt{n}[R(\hat{\beta}) - r]$  在  $\sqrt{n}[R(\beta^o) - r]$  处一阶 Taylor 展开

$$\begin{aligned}\sqrt{n}[R(\hat{\beta}) - r] &= \sqrt{n}[R(\beta^o) - r] + R'(\bar{\beta})\sqrt{n}(\hat{\beta} - \beta^o) \\ &= R'(\bar{\beta})\sqrt{n}(\hat{\beta} - \beta^o)\end{aligned}$$

其中  $\bar{\beta} = \lambda\hat{\beta} + (1 - \lambda)\beta^o, \lambda \in [0, 1]$ .

假定  $R(\cdot)$  在紧空间  $\Theta$  上连续, 根据问题 8.1 又知  $\bar{\beta} \xrightarrow{p} \beta^o$ , 故而  $R'(\bar{\beta}) \xrightarrow{p} R'(\beta^o)$ . 在必要正则条件下

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, \Omega)$$

其中

$$\Omega = (D_o'W^{-1}D_o)^{-1}D_o'W^{-1}V_oW^{-1}D_o(D_o'W^{-1}D_o)^{-1}$$

再根据 Slutsky 定理得

$$\sqrt{n}[R(\hat{\beta}) - r] \xrightarrow{d} N[\mathbf{0}, R'(\beta^o)\Omega R'(\beta^o)']$$

由此构建 Wald 统计量

$$W = n[R(\hat{\beta}) - r]'[R'(\hat{\beta})\hat{\Omega}R'(\hat{\beta})']^{-1}[R(\hat{\beta}) - r] \xrightarrow{d} \chi_J^2$$

其中  $J$  为约束条件的个数, 并且

$$\hat{\Omega} = (\hat{D}_o'\hat{W}^{-1}\hat{D}_o)^{-1}\hat{D}_o'\hat{W}^{-1}\hat{V}_o\hat{W}^{-1}\hat{D}_o(\hat{D}_o'\hat{W}^{-1}\hat{D}_o)^{-1}$$

这里的  $\hat{D}_o = \frac{d\hat{m}(\hat{\beta})}{d\beta}$ ,  $\hat{V}_o$  是  $V_o$  的一致估计量.

值得注意的是, 由于  $W$  不等于  $\text{avar}[\sqrt{n}\hat{m}(\beta^o)]$ , 那么  $\Omega - \Omega_o$  是半正定矩阵, 由此得到的 Wald 检验统计量在有限样本下会更容易犯第 II 类错误.

□

**8.5** 考虑在 GMM 框架下检验原假设  $\mathbb{H}_0 : R(\beta^o) = r$ , 其中  $R(\beta^o)$  是  $J \times 1$  非随机矩阵,  $r$  是  $J \times 1$  非随机向量,  $R'(\beta^o)$  是  $J \times K$  满秩矩阵,  $J \leq K$ . 现在根据 Lagrange 乘子  $\hat{\lambda}^*$  构造一个 LM 检验, 其中  $\hat{\lambda}^*$  是下列有约束的 GMM 最小化问题的最优解

$$(\hat{\beta}^*, \hat{\lambda}^*) = \arg \min_{\beta \in \Theta, \lambda \in \mathbb{R}^J} \{\hat{m}(\beta)' \tilde{V}^{-1} \hat{m}(\beta) + \lambda'[r - R(\beta)]\}$$

其中  $\tilde{V}$  是  $V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$  的预设一致估计量, 它不依赖于  $\beta$ . 构造一个 LM 检验统计量, 并推导其在原假设  $\mathbb{H}_0$  成立下的渐近分布. 假设所有正则条件成立.

**解答** 首先构造 Lagrange 函数

$$L(\beta, \lambda) = \hat{m}(\beta)' \tilde{V}^{-1} \hat{m}(\beta) + 2\lambda'[r - R(\beta)]$$

得到一阶条件为

$$\begin{aligned}\frac{\partial L(\hat{\beta}^*, \hat{\lambda}^*)}{\partial \beta} &= 2 \frac{d\hat{m}(\hat{\beta}^*)}{d\beta'} \tilde{V}^{-1} \hat{m}(\hat{\beta}^*) - 2 \frac{dR(\hat{\beta}^*)}{d\beta'} \hat{\lambda}^* = \mathbf{0} \\ \frac{\partial L(\hat{\beta}^*, \hat{\lambda}^*)}{\partial \lambda} &= R(\hat{\beta}^*) - r = \mathbf{0}\end{aligned}$$

将  $\sqrt{n}\hat{m}(\hat{\beta}^*)$  在  $\beta = \beta^o$  处一阶 Taylor 展开得

$$\sqrt{n}\hat{m}(\hat{\beta}^*) = \sqrt{n}\hat{m}(\beta^o) + \frac{d\hat{m}(\bar{\beta})}{d\beta} \sqrt{n}(\hat{\beta}^* - \beta^o)$$

将其代入到一阶条件得到

$$\sqrt{n}\hat{\lambda}^* = \left( \frac{dR(\bar{\beta})}{d\beta} \hat{A}^{-1} \frac{dR(\hat{\beta})}{d\beta'} \right)^{-1} \frac{dR(\bar{\beta})}{d\beta} \hat{A}^{-1} \frac{d\hat{m}(\hat{\beta}^*)}{d\beta'} \tilde{V}^{-1} \sqrt{n}\hat{m}(\beta^o)$$

其中  $\hat{A} = \frac{d\hat{m}(\hat{\beta}^*)}{d\beta'} \tilde{V}^{-1} \frac{d\hat{m}(\bar{\beta})}{d\beta}$ . 注意到  $\sqrt{n}\hat{m}(\beta^o) \xrightarrow{d} N(\mathbf{0}, V_o)$ , 这里  $V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$ . 由之前的结论

$$\begin{aligned} \frac{d\hat{m}(\hat{\beta}^*)}{d\beta} &\xrightarrow{p} D_o = \frac{dm(\beta^o)}{d\beta} \\ \frac{d\hat{m}(\bar{\beta})}{d\beta} &\xrightarrow{p} D_o = \frac{dm(\beta^o)}{d\beta} \\ \frac{dR(\hat{\beta}^*)}{d\beta} &\xrightarrow{p} R'(\beta^o) = \frac{dR(\beta^o)}{d\beta} \\ \frac{dR(\bar{\beta})}{d\beta} &\xrightarrow{p} R'(\beta^o) = \frac{dR(\beta^o)}{d\beta} \end{aligned}$$

因此

$$\sqrt{n}\hat{\lambda}^* \xrightarrow{d} N(\mathbf{0}, \Omega_o)$$

这里的

$$\Omega_o = [R'(\beta^o)(D_o' V_o^{-1} D_o)^{-1} R'(\beta^o)']^{-1}$$

由此可以构建 LM 统计量

$$LM = n(\hat{\lambda}^*)' \left\{ \frac{dR(\hat{\beta})}{d\beta} \left[ \frac{d\hat{m}(\hat{\beta})}{d\beta'} \tilde{V}^{-1} \frac{d\hat{m}(\hat{\beta})}{d\beta} \right]^{-1} \frac{dR(\hat{\beta})}{d\beta'} \right\} \hat{\lambda}^* \xrightarrow{d} \chi_J^2$$

□

**8.6 [非线性工具变量估计]** 考虑一个非线性回归模型

$$Y_t = g(X_t, \beta^o) + \varepsilon_t$$

其中  $g(X_t, \cdot)$  对于  $\beta$  二阶连续可微,  $\mathbb{E}[\varepsilon_t | X_t] \neq 0$  但  $\mathbb{E}[\varepsilon_t | Z_t] = 0$ . 这里  $Y_t$  是随机变量,  $X_t$  是  $K \times 1$  随机向量,  $Z_t$  是  $l \times 1$  随机向量,  $l \geq K$ . 假定  $\{Y_t, X_t', Z_t'\}$  是一个遍历平稳过程,  $\{Z_t \varepsilon_t\}$  是 MDS.

定义矩函数  $m_t(\beta) = Z_t[Y_t - g(X_t, \beta)]$ . 根据矩条件

$$\mathbb{E}[m_t(\beta^o)] = \mathbf{0}$$

未知参数  $\beta^o$  可得到一致估计. 定义非线性工具变量估计量为下面最小化问题的最优解

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

其中样本矩  $\hat{m}_t(\beta) = n^{-1} \sum_{t=1}^n m_t(\beta)$ ,  $\hat{W} \xrightarrow{p} W$ , 这里  $W$  是一个  $l \times l$  的有限、对称和正定的矩阵.

(1) 证明: 当  $n \rightarrow \infty$  时,  $\hat{\beta} \xrightarrow{p} \beta^o$ .

(2) 推导 GMM 最小化问题的一阶条件.

(3) 推导  $\hat{\beta}$  的渐近分布. 分别就条件同方差  $\mathbb{E}[\varepsilon_t^2|Z_t] = \sigma^2$  和条件异方差  $\mathbb{E}[\varepsilon_t^2|Z_t] \neq \sigma^2$  两种情况进行讨论.

(4) 使  $\hat{\beta}$  渐近有效的权重矩阵  $W$  是什么?

(5) 构造一个统计量检验原假设  $\mathbb{H}_0: R(\beta^o) = r$ , 其中  $R(\beta^o)$  是  $J \times 1$  非随机矩阵,  $R'(\beta^o)$  是  $J \times K$  满秩矩阵,  $r$  是  $J \times 1$  非随机向量,  $J \leq K$ . 推导出检验统计量的渐近分布.

**解答** (1) 定义

$$\hat{Q}(\beta) = -\hat{m}(\beta)' \hat{W}^{-1} \hat{m}(\beta), \quad Q(\beta) = -m(\beta)' W^{-1} m(\beta)$$

假设样本矩  $\hat{m}(\beta)$  在  $\Theta$  依概率一致收敛于  $m(\beta)$ , 那么

$$|\hat{Q}(\beta) - Q(\beta)| \leq |A_1| + 2|A_2| + |A_3| \xrightarrow{p} 0$$

对一切  $\beta \in \Theta$  成立. 其中

$$A_1 = [\hat{m}(\beta) - m(\beta)]' \hat{W}^{-1} [\hat{m}(\beta) - m(\beta)] \xrightarrow{p} 0$$

$$A_2 = m(\beta)' \hat{W}^{-1} [\hat{m}(\beta) - m(\beta)] \xrightarrow{p} 0$$

$$A_3 = m(\beta)' (\hat{W}^{-1} - W^{-1}) m(\beta) \xrightarrow{p} 0$$

故而  $\sup_{\beta \in \Theta} |\hat{Q}(\beta) - Q(\beta)| \xrightarrow{p} 0$ . 由  $g(X_t, \beta)$  的连续性可知  $\hat{Q}(\beta)$  也是  $\Theta$  上的连续函数, 假定  $\beta^o$

是  $\Theta$  中唯一一个使得  $\mathbb{E}[m_t(\beta)] = \mathbf{0}$  的参数, 那么由极值估计量引理可知当  $n \rightarrow \infty$  时,  $\hat{\beta} \xrightarrow{p} \beta^o$ .

(2) 根据  $\frac{d\hat{Q}(\hat{\beta})}{d\beta} = \frac{d\hat{m}(\hat{\beta})}{d\beta} \hat{W}^{-1} \hat{m}(\hat{\beta}) = \mathbf{0}$ , 由矩阵微分可知一阶条件为

$$\left[ n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} Z_t' \right] \hat{W}^{-1} \hat{m}(\hat{\beta}) = \mathbf{0}$$

其中  $\frac{\partial g(X_t, \hat{\beta})}{\partial \beta}$  是  $K \times 1$  向量.

(3) 令  $\hat{D}(\beta) = n^{-1} \sum_{t=1}^n Z_t \frac{\partial g(X_t, \beta)}{\partial \beta'}$ , 由问题 8.1 可知

$$\sqrt{n}(\hat{\beta} - \beta) = [\hat{D}(\hat{\beta})' \hat{W}^{-1} \hat{D}(\bar{\beta})]^{-1} \hat{D}(\hat{\beta})' \hat{W}^{-1} \sqrt{n} \hat{m}(\beta^o)$$

其中  $\bar{\beta} = \lambda \hat{\beta} + (1 - \lambda) \beta^o$ ,  $\lambda \in [0, 1]$ . 假定中心极限定理成立, 即  $\sqrt{n} \hat{m}(\beta^o) \xrightarrow{d} N(\mathbf{0}, V_o)$ , 那么

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, \Omega)$$

由于  $\{Y_t, X_t', Z_t'\}$  是遍历平稳的, 并且  $\{Z_t \varepsilon_t\}$  是 MDS, 故而  $V_o = \mathbb{E}[Z_t Z_t' \varepsilon_t^2]$ .

若  $\mathbb{E}[\varepsilon_t^2|Z_t] = \sigma^2$ , 那么  $V_o = \sigma^2 \mathbb{E}[Z_t Z_t']$ , 即

$$\Omega = \sigma^2 [D(\beta^o)' W^{-1} D(\beta^o)]^{-1} D(\beta^o)' W^{-1} \mathbb{E}[Z_t Z_t'] W^{-1} D(\beta^o) [D(\beta^o)' W^{-1} D(\beta^o)]^{-1}$$

若  $\mathbb{E}[\varepsilon_t^2|Z_t] \neq \sigma^2$ , 那么  $V_o$  不能化简, 即

$$\Omega = [D(\beta^o)'W^{-1}D(\beta^o)]^{-1}D(\beta^o)'W^{-1}\mathbb{E}[Z_tZ_t'\varepsilon_t^2]W^{-1}D(\beta^o)[D(\beta^o)'W^{-1}D(\beta^o)]^{-1}$$

其中  $D(\beta^o) = \mathbb{E}\left[Z_t \frac{\partial g(X_t, \beta^o)}{\partial \beta'}\right]$ .

(4) 由 GMM 估计量的渐近有效性可知最优权重矩阵为  $W = \text{avar}[\sqrt{n}\hat{m}(\beta^o)] = \mathbb{E}[Z_tZ_t'\varepsilon_t^2]$ , 在条件同方差下可以化简为  $W = \sigma^2\mathbb{E}[Z_tZ_t']$ , 在条件异方差下则不能化简.

(5) 类似问题 8.4 的做法, 选取最优权重矩阵  $W = V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$  得到

$$\Omega_o = [D(\beta^o)'V_o^{-1}D(\beta^o)]^{-1}$$

此时可以构建 Wald 检验统计量

$$W = n[R(\hat{\beta}) - r]'[R'(\hat{\beta})\hat{\Omega}_oR'(\hat{\beta})]^{-1}[R(\hat{\beta}) - r] \xrightarrow{d} \chi_r^2$$

其中

$$\hat{\Omega}_o = [\hat{D}(\hat{\beta})'\tilde{V}^{-1}\hat{D}(\hat{\beta})]^{-1}$$

并且这里  $\tilde{V}$  是  $V_o$  的一致估计量, 由于  $\{Z_t\varepsilon_t\}$  是 MDS, 故而可以将其写为

$$\tilde{V} = n^{-1} \sum_{t=1}^n m_t(\hat{\beta})m_t(\hat{\beta})'$$

□

**8.7 [非线性最小二乘估计]** 考虑以下非线性回归模型

$$Y_t = g(X_t, \beta^o) + \varepsilon_t$$

其中  $\beta^o$  是未知的  $K \times 1$  参数向量, 且  $\mathbb{E}[\varepsilon_t|X_t] = 0$ . 假设  $g(X_t, \cdot)$  是关于  $\beta$  的二阶连续可微函数, 且对于所有  $\beta \in \Theta$ ,  $K \times K$  矩阵  $\mathbb{E}\left[\frac{\partial}{\partial \beta}g(X_t, \beta)\frac{\partial}{\partial \beta'}g(X_t, \beta)\right]$  与  $\mathbb{E}\left[\frac{\partial^2}{\partial \beta \partial \beta'}g(X_t, \beta)\right]$  是对称、有限和正定的矩阵.

定义非线性最小二乘 (nonlinear least squares, NLS) 估计量为使残差平方和最小化的解, 即

$$\hat{\beta} = \arg \min_{\beta \in \Theta} \sum_{t=1}^n [Y_t - g(X_t, \beta)]^2$$

一阶条件为

$$\hat{D}(\beta)'e = \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} [Y_t - g(X_t, \hat{\beta})] = 0$$

其中  $\hat{D}(\beta)$  是  $n \times K$  矩阵, 其第  $t$  行为  $\frac{\partial}{\partial \beta}g(X_t, \beta)$ , 而  $e_t = Y_t - g(X_t, \hat{\beta})$ . 这个一阶条件可看成是在恰好识别 ( $l = K$ ) 的情形下, GMM 估计量的一阶条件

$$\hat{m}(\hat{\beta}) = 0$$

其中

$$m_t(\beta) = \frac{\partial g(X_t, \beta)}{\partial \beta} [Y_t - g(X_t, \beta)]$$

一般而言,  $\hat{\beta}$  没有显性解. 假设所有的正则条件成立.

(1) 证明: 当  $n \rightarrow \infty$  时,  $\hat{\beta} \xrightarrow{p} \beta^o$ .

(2) 推导  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近分布.

(3) 如果  $\left\{ \frac{\partial}{\partial \beta} g(X_t, \beta) \varepsilon_t \right\}$  是 MDS, 且  $\mathbb{E}[\varepsilon_t^2 | X_t] = \sigma^2$ ,  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近方差表达式是什么? 给出理由.

(4) 如果  $\left\{ \frac{\partial}{\partial \beta} g(X_t, \beta) \varepsilon_t \right\}$  是 MDS, 但  $\mathbb{E}[\varepsilon_t^2 | X_t] \neq \sigma^2$ ,  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近方差表达式是什么? 给出理由.

(5) 假定  $\left\{ \frac{\partial}{\partial \beta} g(X_t, \beta) \varepsilon_t \right\}$  是 MDS, 且  $\mathbb{E}[\varepsilon_t^2 | X_t] = \sigma^2$ , 构造一个检验统计量检验原假设  $\mathbb{H}_0: R(\beta^o) = r$ , 其中  $R(\beta)$  是一个  $J \times K$  非随机矩阵, 且  $R'(\beta^o) = \frac{d}{d\beta} R(\beta^o)$  是  $J \times K$  满秩矩阵,  $J \leq K$ , 而  $r$  是  $J \times 1$  非随机向量.

(6) 假定  $\left\{ \frac{\partial}{\partial \beta} g(X_t, \beta) \varepsilon_t \right\}$  是 MDS, 但  $\mathbb{E}[\varepsilon_t^2 | X_t] \neq \sigma^2$ , 构造一个检验统计量检验原假设  $\mathbb{H}_0: R(\beta^o) = r$ , 其中  $R(\beta)$  是一个  $J \times K$  非随机矩阵, 且  $R'(\beta^o) = \frac{d}{d\beta} R(\beta^o)$  是  $J \times K$  满秩矩阵,  $J \leq K$ , 而  $r$  是  $J \times 1$  非随机向量.

**解答** (1) 将  $g_t(X_t, \hat{\beta})$  在  $\beta = \beta^o$  处一阶 Taylor 展开

$$g_t(X_t, \hat{\beta}) = g(X_t, \beta^o) + \frac{\partial g(X_t, \bar{\beta})}{\partial \beta} (\hat{\beta} - \beta^o)$$

其中  $\bar{\beta} = \lambda \hat{\beta} + (1 - \lambda) \beta^o$ ,  $\lambda \in [0, 1]$ . 将其代入到一阶条件可知

$$\hat{\beta} - \beta^o = \left[ n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} \frac{\partial g(X_t, \bar{\beta})}{\partial \beta'} \right]^{-1} \left[ n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \bar{\beta})}{\partial \beta} \varepsilon_t \right]$$

设  $\hat{Q} = n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} \frac{\partial g(X_t, \bar{\beta})}{\partial \beta'}$ , 由相应的 WLLN 和给定条件可知  $\hat{Q} \xrightarrow{p} Q = O_p(1)$ . 又有

$$n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \bar{\beta})}{\partial \beta} \varepsilon_t \xrightarrow{p} \mathbb{E} \left[ \frac{\partial g(X_t, \beta^*)}{\partial \beta} \mathbb{E}[\varepsilon_t | X_t] \right] = \mathbf{0}$$

这里的  $\beta^*$  是  $\bar{\beta}$  依概率收敛的值, 因此  $\hat{\beta} - \beta^o = o_p(1)$ . 事实上, 由此可以推断出  $\beta^* = \beta^o$ , 以及

$$Q = \mathbb{E} \left[ \frac{\partial g(X_t, \beta^o)}{\partial \beta} \frac{\partial g(X_t, \beta^o)}{\partial \beta'} \right]$$

(2) 由 (1) 的结论可知

$$\sqrt{n}(\hat{\beta} - \beta^o) = \left[ n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} \frac{\partial g(X_t, \bar{\beta})}{\partial \beta'} \right]^{-1} \left[ n^{-\frac{1}{2}} \sum_{t=1}^n \frac{\partial g(X_t, \bar{\beta})}{\partial \beta} \varepsilon_t \right]$$

假定相应的中心极限定理成立, 即

$$n^{-\frac{1}{2}} \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} \xrightarrow{d} N(\mathbf{0}, V)$$

其中  $V = \text{avar} \left[ n^{-\frac{1}{2}} \sum_{t=1}^n \frac{\partial g(X_t, \beta^o)}{\partial \beta} \varepsilon_t \right]$ . 于是

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, \Omega)$$

这里  $\Omega = Q^{-1}VQ^{-1}$ .

(3) 此时  $V = \sigma^2 \mathbb{E} \left[ \frac{\partial g(X_t, \beta^o)}{\partial \beta} \frac{\partial g(X_t, \beta^o)}{\partial \beta'} \right]$ , 因此

$$\text{avar}(\sqrt{n}\hat{\beta}) = \sigma^2 \mathbb{E} \left[ \frac{\partial g(X_t, \beta^o)}{\partial \beta} \frac{\partial g(X_t, \beta^o)}{\partial \beta'} \right]^{-1}$$

(4) 此时  $V = \mathbb{E} \left[ \frac{\partial g(X_t, \beta^o)}{\partial \beta} \frac{\partial g(X_t, \beta^o)}{\partial \beta'} \varepsilon_t^2 \right]$ ,  $\text{avar}(\sqrt{n}\hat{\beta}) = \Omega$  保持不变.

(5) 可以构建一个 Wald 检验统计量

$$W = n[R(\hat{\beta}) - r]' \left[ s^2 R'(\hat{\beta}) \tilde{Q}^{-1} R'(\hat{\beta})' \right]^{-1} [R(\hat{\beta}) - r] \xrightarrow{d} \chi_J^2$$

其中  $\tilde{Q} = n^{-1} \sum_{t=1}^n \frac{\partial g(X_t, \hat{\beta})}{\partial \beta} \frac{\partial g(X_t, \hat{\beta})}{\partial \beta'}$ ,  $R'(\hat{\beta}) = \frac{dR(\hat{\beta})}{d\beta}$ , 以及  $s^2 = \frac{1}{n-K} \sum_{t=1}^n e_t^2$ .

(6) 可以构建一个稳健 Wald 检验统计量

$$W_{\text{robust}} = n[R(\hat{\beta}) - r]' \left[ R'(\hat{\beta}) \tilde{Q}^{-1} \tilde{V} \tilde{Q}^{-1} R'(\hat{\beta})' \right]^{-1} [R(\hat{\beta}) - r] \xrightarrow{d} \chi_J^2$$

其中  $\tilde{V}$  是  $V$  的一致估计量.

□

**8.8** 假设  $\hat{V}$  是  $V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$  的一致估计量. 证明: 用  $\hat{V}$  替换  $\tilde{V}$  对过度识别检验统计量的渐近分布没有影响, 即当原假设  $\mathbb{H}_0: \mathbb{E}[m_t(\beta^o)] = \mathbf{0}$  成立, 且  $n \rightarrow \infty$  时, 有

$$n\hat{m}(\hat{\beta})' \tilde{V}^{-1} \hat{m}(\hat{\beta}) - n\hat{m}(\hat{\beta})' \hat{V}^{-1} \hat{m}(\hat{\beta}) \xrightarrow{p} 0$$

假设定理 8.11 的条件成立.

**解答** 由于  $\hat{V}$  与  $\tilde{V}$  都是  $V$  的一致估计, 于是

$$n\hat{m}(\hat{\beta})' [(\tilde{V} - V) - (\hat{V} - V)]^{-1} \hat{m}(\hat{\beta}) = o_p(1)$$

因此原式成立.

□

**8.9** 假设  $\tilde{\beta}$  是渐近次优但一致的 GMM 估计量. 如果用  $\tilde{\beta}$  替代两阶段 GMM 估计量  $\hat{\beta}$ , 从而得到过度识别检验统计量  $\tilde{J} = n\hat{m}(\tilde{\beta})' \tilde{V}^{-1} \hat{m}(\tilde{\beta})$ . 这一检验统计量在模型正确设定时服从渐近  $\chi_{l-K}^2$  分布吗? 给出理由. 假设所有必要的正则条件成立.

**解答** 不服从. 在推导  $J$  检验统计量的分布时

$$\hat{\Pi} = I - \tilde{V}^{-\frac{1}{2}} \frac{d\hat{m}(\tilde{\beta})}{d\beta} \left[ \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{V}^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta} \right]^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{V}^{-\frac{1}{2}}$$

此时若将  $\hat{\beta}$  替换为  $\tilde{\beta}$ , 将会得到一个新的  $\tilde{\Pi}$ , 然而它依概率收敛于一个非幂等矩阵, 所以

$$\begin{aligned} \tilde{J} &= [\tilde{V}^{-\frac{1}{2}} \sqrt{n}\hat{m}(\beta^o)]' \tilde{\Pi}' \tilde{\Pi} [\tilde{V}^{-\frac{1}{2}} \sqrt{n}\hat{m}(\beta^o)] \\ &\neq [\tilde{V}^{-\frac{1}{2}} \sqrt{n}\hat{m}(\beta^o)]' \hat{\Pi}' \hat{\Pi} [\tilde{V}^{-\frac{1}{2}} \sqrt{n}\hat{m}(\beta^o)] \xrightarrow{d} G' \Pi G \sim \chi_{l-K} \end{aligned}$$

□

**8.10** 假设  $\tilde{\beta}$  是一个一致但可能渐近次优的 GMM 估计量, 对应于使用权重矩阵  $\tilde{W} \xrightarrow{p} W$ , 这里  $W$  是一个  $l \times l$  有限、对称和正定的矩阵. 现在考虑用  $\tilde{\beta}$  构造一个检验统计量, 以检验原



假设

$$\mathbb{H}_0 : \mathbb{E}[m_t(\beta^o)] = \mathbf{0}$$

基本思想是考虑  $l \times 1$  样本矩  $\hat{m}(\bar{\beta})$ , 检验  $\hat{m}(\bar{\beta})$  是否显著地偏离零向量. 假设所有必要的正则条件成立.

(1) 推导  $\sqrt{n}\hat{m}(\tilde{\beta})$  在  $\mathbb{H}_0$  下的渐近分布.

(2) 构造一个  $\sqrt{n}\hat{m}(\tilde{\beta})$  的二次型, 记为  $\tilde{J}$ , 并推导其在原假设  $\mathbb{H}_0$  下的渐近分布.

(3) 与使用了二阶段最优 GMM 估计量  $\hat{\beta}$  的过度识别检验统计量  $J$  相比,  $\tilde{J}$  检验是比  $J$  检验更有效还是更无效? 即哪个检验在有限样本下有更大的概率拒绝错误模型?

**解答** (1) 这里的渐近次优 GMM 估计量为

$$\tilde{\beta} = \arg \min_{\beta \in \Theta} \hat{m}(\tilde{\beta})' \tilde{W}^{-1} \hat{m}(\tilde{\beta})$$

一阶条件为

$$\frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \hat{m}(\tilde{\beta}) = \mathbf{0}$$

将  $\sqrt{n}\hat{m}(\tilde{\beta})$  在  $\beta = \beta^o$  处一阶 Taylor 展开

$$\sqrt{n}\hat{m}(\tilde{\beta}) = \sqrt{n}\hat{m}(\beta^o) + \frac{d\hat{m}(\tilde{\beta})}{d\beta} \sqrt{n}(\tilde{\beta} - \beta^o)$$

代入到一阶条件得

$$\frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \sqrt{n}\hat{m}(\beta^o) + \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta} \sqrt{n}(\tilde{\beta} - \beta^o) = \mathbf{0}$$

当  $n$  充分大时

$$\sqrt{n}(\tilde{\beta} - \beta^o) = - \left[ \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta} \right]^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \sqrt{n}\hat{m}(\beta^o)$$

再将上式代入到  $\sqrt{n}\hat{m}(\tilde{\beta})$  的 Taylor 展开式中

$$\begin{aligned} \sqrt{n}\hat{m}(\tilde{\beta}) &= \left\{ I - \frac{d\hat{m}(\tilde{\beta})}{d\beta} \left[ \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta} \right]^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \right\} \sqrt{n}\hat{m}(\beta^o) \\ &= \hat{A} \sqrt{n}\hat{m}(\beta^o) \end{aligned}$$

假设中心极限定理成立

$$\sqrt{n}\hat{m}(\beta^o) \xrightarrow{d} N(\mathbf{0}, V_o)$$

其中  $V_o = \text{avar}[\sqrt{n}\hat{m}(\beta^o)]$  是有限、对称和正定的矩阵. 于是

$$\sqrt{n}\hat{m}(\tilde{\beta}) \xrightarrow{d} N(\mathbf{0}, \Omega)$$

其中  $\Omega = AV_oA'$ , 并且

$$A = I - D_o(D_o'W^{-1}D_o)^{-1}D_o'W^{-1}$$

以及  $D_o = \frac{dm(\beta^o)}{d\beta}$ .

(2) 由 (1) 的结论, 可以构建一个 Wald 检验统计量

$$\tilde{J} = n\hat{m}(\tilde{\beta})'(\tilde{A}\tilde{V}^{-1}\tilde{A}')^{-1}\hat{m}(\tilde{\beta}) \xrightarrow{d} \chi_l^2$$

其中

$$\tilde{A} = \left\{ I - \frac{d\hat{m}(\tilde{\beta})}{d\beta} \left[ \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta} \right]^{-1} \frac{d\hat{m}(\tilde{\beta})}{d\beta'} \tilde{W}^{-1} \right\}$$

并且  $\tilde{V}$  是  $V_o$  的一致估计量. 注意, Wald 检验统计量也是一个二次型.

(3) 由于  $\tilde{J}$  服从渐近  $\chi_l^2$  分布,  $l > l - K$ , 故而相同分位数下的临界值比  $\chi_{l-K}^2$  的更大, 更不易拒绝错误的模型假定, 因此相较于  $J$  检验更无效.

□

**8.11** 给定假设 7.1 – 7.3, 7.4(3), 7.6 和 7.7, 为了检验原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t|Z_t] = 0$ , 其中  $Z_t$  是  $l \times 1$  的工具向量, 可考虑以下辅助回归

$$\hat{e}_t = \alpha' Z_t + w_t, \quad t = 1, \dots, n$$

其中  $\hat{e}_t = Y_t - X_t' \hat{\beta}_{2SLS}$ , 假设  $Z_t$  包含截距项,  $R^2$  和  $R_{uc}^2$  分别是上述辅助回归的中心化和非中心化的决定系数. 证明: 在原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t|Z_t] = 0$  成立, 且当  $n \rightarrow \infty$  时,  $nR_{uc}^2 = nR^2 + o_p(1)$ .

**解答** 在原假设  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t|Z_t] = 0$  成立的条件下

$$\mathbb{E}[\hat{e}_t|Z_t] = -\mathbb{E}[X_t'(\hat{\beta}_{2SLS} - \beta^o)|Z_t] = 0$$

因此只需检验辅助回归中  $\alpha$  是否显著不等于 0. 设  $\hat{\alpha}$  为 OLS 估计量,  $\tilde{e}$  是关于  $\hat{e}$  的  $n \times 1$  拟合值向量, 于是  $\tilde{e} = \mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{e}$ , 于是

$$nR_{uc}^2 = \frac{\tilde{e}'\tilde{e}}{\hat{e}'\hat{e}/n} = \frac{\hat{e}'\mathbf{Z}(\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\hat{e}}{\hat{e}'\hat{e}/n}$$

这实际上就是 Sargan 检验统计量, 因而在条件同方差假定  $\mathbb{E}[\varepsilon_t^2|Z_t] = \sigma^2$  下,  $nR_{uc}^2 \xrightarrow{d} \chi_{l-K}^2$ .

比较  $R_{uc}^2$  和  $R^2$  的定义, 即

$$R_{uc}^2 = 1 - \frac{SSR}{\hat{e}'\hat{e}}$$

$$R^2 = 1 - \frac{SSR}{\hat{e}'M_0\hat{e}}$$

其中  $M_0 = I - \iota(\iota'\iota)^{-1}\iota'$ ,  $\iota$  是  $n \times 1$  元素全为 1 的向量, 当  $n \rightarrow \infty$  时,  $M_0 \xrightarrow{p} I$ , 于是  $nR_{uc}^2 = nR^2 + o_p(1)$ .

□

**8.12** 给定假设 7.1 – 7.3, 7.4(3) 和 7.6,  $\mathbb{E}[Z_{jt}^4] < \infty$ ,  $1 \leq j \leq l$ ,  $\mathbb{E}[\varepsilon_t^4] < \infty$  及  $l > K$ . 考虑构建一个检验统计量, 以检验工具变量  $Z_t$  的有效性. 这里, 原假设为  $\mathbb{H}_0: \mathbb{E}[\varepsilon_t|Z_t] = 0$ . 推导检验统计量在原假设成立时的渐近分布. 这事实上是一个稳健化的 Sargan 检验, 在存在条件异方差时也可适用.

**解答** 由于存在条件异方差  $\mathbb{E}[\varepsilon_t^2|Z_t] \neq \sigma^2$ , 因此需要重新选择最优权重矩阵估计量  $\hat{W}$ . 又因为  $\{Z_t\varepsilon_t\}$  是 MDS, 故而  $V_o = \mathbb{E}[Z_t Z_t' \varepsilon_t^2]$ , 选取  $\tilde{V} = n^{-1} \sum_{t=1}^n Z_t Z_t' \hat{e}_t^2$ , 其中  $\hat{e}_t = Y_t - X_t' \hat{\beta}_{2SLS}$ , 在相

---

关矩条件成立的情况下, 易证当  $n \rightarrow \infty$  时,  $\tilde{V} \xrightarrow{p} V_o$ . 选取矩函数  $m_t(\beta) = Z_t(Y_t - X_t'\beta)$ , 由定理 8.11 可知当原假设  $\mathbb{H}_0 : \mathbb{E}[\varepsilon_t|Z_t] = 0$  成立时

$$S_{\text{robust}} = n\hat{m}(\hat{\beta})\tilde{V}^{-1}\hat{m}(\hat{\beta}) = \hat{e}'\mathbf{Z}\left(\sum_{t=1}^n Z_t Z_t' \hat{e}_t^2\right)^{-1} \mathbf{Z}'\hat{e} \xrightarrow{d} \chi_{l-K}^2$$

□

## 最大似然估计和拟最大似然估计

**9.1** 对于 Probit 模型  $\mathbb{P}[Y_t = y|X_t] = \Phi(X_t'\beta^o)^y[1 - \Phi(X_t'\beta^o)]^{1-y}$ , 其中  $\Phi(\cdot)$  是标准正态分布  $N(0, 1)$  的累积分布函数,  $y = 0, 1$ , 证明:

$$(1) \mathbb{E}[Y_t|X_t] = \Phi(X_t'\beta^o).$$

$$(2) \text{var}(Y_t|X_t) = \Phi(X_t'\beta^o)[1 - \Phi(X_t'\beta^o)].$$

**解答** (1)  $\mathbb{E}[Y_t|X_t] = \mathbb{P}[Y_t = 1|X_t] = \Phi(X_t'\beta^o)$ .

$$(2) \text{var}(Y_t|X_t) = \mathbb{E}[Y_t^2|X_t] - \mathbb{E}[Y_t|X_t]^2 = \Phi(X_t'\beta^o)[1 - \Phi(X_t'\beta^o)].$$

□

**9.2** 对于归并 (censored) 回归模型

(1) 证明:  $\mathbb{E}[X_t\varepsilon_t|Y_t > c] \neq 0$ . 因此, 基于归并随机样本的 OLS 估计量不是真实模型参数  $\beta^o$  的一致估计.

(2) 给出  $\beta^o$  的一致估计方法, 并写出估计过程的目标函数.

**解答** (1) 给定 i.i.d. 随机样本  $\mathbf{Z}^n$ ,  $Y_t$  的观测值存在归并问题, 即

$$Y_t = \begin{cases} Y_t^*, & Y_t^* > c \\ c, & Y_t^* \leq c \end{cases}$$

其中  $Y_t^* = X_t'\beta^o + \varepsilon_t$ , 并且  $\varepsilon_t|X_t \sim N(0, \sigma^2)$ . 假设门槛参数  $c$  已知. 根据标准正态的断尾期望<sup>13</sup>可以得到

$$\begin{aligned} \mathbb{E}[\varepsilon_t|X_t, Y_t > c] &= \mathbb{E}[\varepsilon_t|X_t, \varepsilon_t > c - X_t'\beta^o] \\ &= \sigma\lambda\left(\frac{c - X_t'\beta^o}{\sigma}\right) \neq 0 \end{aligned}$$

其中  $\lambda(\cdot) = \frac{\phi(\cdot)}{1 - \Phi(\cdot)}$  为逆 Mills 函数,  $\phi(\cdot)$  和  $\Phi(\cdot)$  分别是标准正态分布的概率密度函数和累积分布函数.

(2)  $Y_t|X_t$  的条件概率密度为

$$f_{Y_t|X_t}(y|X_t, \beta) = \left[1 - \Phi\left(\frac{c - X_t'\beta}{\sigma}\right)\right]^{\mathbb{1}(y=c)} \left[\frac{1}{\sigma}\phi\left(\frac{Y_t - X_t'\beta}{\sigma}\right)\right]^{\mathbb{1}(y>c)}$$

对数似然函数为

$$\begin{aligned} \ln L(\beta, \sigma) &= \frac{n}{2} \ln(2\pi\sigma^2) + \sum_{t=1}^n \left\{ \mathbb{1}(y_t = c) \ln \left[1 - \Phi\left(\frac{c - X_t'\beta}{\sigma}\right)\right] \right. \\ &\quad \left. + \mathbb{1}(y_t > c) \left[1 - \phi\left(\frac{y_t - X_t'\beta}{\sigma}\right)\right] \right\} \end{aligned}$$

□

**9.3** 假设  $f(y|\Psi_t, \beta)$  是  $Y_t$  相对于  $\Psi_t$  的条件概率密度, 其中  $\beta \in \Theta$ . 证明: 对于所有的  $\beta$ ,

<sup>13</sup> 假设  $z \sim N(0, 1)$ ,  $c \in \mathbb{R}$ , 则  $\mathbb{E}[z|z > c] = \int_c^\infty z \frac{\phi(z)}{1 - \Phi(c)} dz = \frac{\phi(c)}{1 - \Phi(c)}$ .

$\tilde{\beta} \in \Theta$ , 有

$$\int_{-\infty}^{\infty} \ln[f(y|\Psi_t, \beta)] f(y|\Psi_t, \tilde{\beta}) dy \leq \int_{-\infty}^{\infty} \ln[f(y|\Psi_t, \tilde{\beta})] f(y|\Psi_t, \tilde{\beta}) dy$$

**解答** 由 Jensen 不等式可知, 由于  $\ln(\cdot)$  是凹函数, 故而不等式变号, 因此

$$\int_{-\infty}^{\infty} \ln \left[ \frac{f(y|\Psi_t, \beta)}{f(y|\Psi_t, \tilde{\beta})} \right] f(y|\Psi_t, \tilde{\beta}) dy \leq \ln \left[ \int_{-\infty}^{\infty} f(y|\Psi_t, \tilde{\beta}) dy \right] = 0$$

整理即可得到原式.

□

**9.4** (1) 假设  $\beta \in \Theta$ , 模型  $f(y|\Psi_t, \beta)$  是  $Y_t$  的条件概率密度的正确设定, 从而存在某一参数值  $\beta^o \in \Theta$  使得  $f(y|\Psi_t, \beta^o)$  等于  $Y_t$  的真实条件概率密度. 假设  $f(y|\Psi_t, \beta)$  相对于  $\beta \in \Theta$  是连续可微的, 且  $\beta^o$  是  $\Theta$  的内点. 证明:

$$\mathbb{E}[S_t(\beta^o)|\Psi_t] = \mathbb{E} \left[ \frac{\partial \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta} \middle| \Psi_t \right] = \mathbf{0}$$

(2) 假设 (1) 成立. 能够判定模型  $f(y|\Psi_t, \beta)$  是  $Y_t$  的条件概率分布的正确设定? 如果是, 给出理由; 如果不是, 举一个反例.

**解答** (1) 对于任意  $\beta \in \Theta$  有

$$\int_{-\infty}^{\infty} f(y|\Psi_t, \beta) dy = 1$$

于是

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta} \int_{-\infty}^{\infty} f(y|\Psi_t, \beta) dy \\ &= \int_{-\infty}^{\infty} \frac{\partial f(y|\Psi_t, \beta)}{\partial \beta} dy \\ &= \int_{-\infty}^{\infty} \frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta} f(y|\Psi_t, \beta) dy \end{aligned}$$

对一切  $\beta \in \Theta$  都成立, 取  $\beta = \beta^o$  即可证得结论.

(2) 不能. 考虑一个平稳时间序列  $\{Y_t\}$ , 服从以下 DGP

$$Y_t = \mu_t(\beta^o) + \sigma_t(\beta^o) z_t$$

其中  $\mu_t(\beta^o) = \mathbb{E}[Y_t|\Psi_t]$ ,  $\sigma_t^2(\beta^o) = \text{var}(Y_t|\Psi_t)$ ,  $\{z_t\}$  服从 i.i.d. 标准正态分布. 考虑用

$$Y_t = \mu_t(\beta) + \varepsilon_t, \quad \{\varepsilon_t\} \sim \text{i.i.d. } N(0, 1)$$

可知存在  $\beta^o \in \Theta$ , 使得  $\mu(\beta^o) = \mathbb{E}[Y_t|\Psi_t]$ , 但是  $\text{var}(Y_t|\Psi_t) = \sigma_t^2(\beta^o) \neq 1$ , 因此条件概率分布  $f(y|\Psi_t, \beta)$  存在误设.

□

**9.5** 假设模型  $f(y|\Psi_t, \beta)$ ,  $\beta \in \Theta \subset \mathbb{R}^K$ , 是  $Y_t$  相对于  $\Psi_t$  的条件概率密度的正确设定, 从而存在某一参数值  $\beta^o \in \Theta$ ,  $f(y|\Psi_t, \beta^o)$  等于  $Y_t$  的真实条件概率密度函数. 假设  $f(y|\Psi_t, \beta)$  相对于  $\beta \in \Theta$  是连续可微的, 且  $\beta^o$  是  $\Theta$  的内点. 证明:

$$\mathbb{E} \left[ \frac{\partial \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta} \frac{\partial \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta'} \middle| \Psi_t \right] + \mathbb{E} \left[ \frac{\partial^2 \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta \partial \beta'} \middle| \Psi_t \right] = \mathbf{0}$$

其中  $\frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta}$  是  $K \times 1$  向量,  $\frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta'}$  是  $\frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta}$  的转置,  $\frac{\partial^2 \ln f(y|\Psi_t, \beta)}{\partial \beta \partial \beta'}$  是  $K \times K$  矩阵,  $\mathbb{E}[\cdot|\Psi_t]$  是对  $Y_t$  的真实条件分布的期望.

**解答** 根据问题 9.4 可知

$$\int_{-\infty}^{\infty} \frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta} f(y|\Psi_t, \beta) dy = 0$$

对上式求导得

$$\begin{aligned} 0 &= \frac{\partial}{\partial \beta'} \int_{-\infty}^{\infty} \frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta} f(y|\Psi_t, \beta) dy \\ &= \int_{-\infty}^{\infty} \frac{\partial}{\partial \beta'} \left[ \frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta} f(y|\Psi_t, \beta) \right] dy \\ &= \int_{-\infty}^{\infty} \frac{\partial^2 \ln f(y|\Psi_t, \beta)}{\partial \beta \partial \beta'} f(y|\Psi_t, \beta) dy \\ &\quad + \int_{-\infty}^{\infty} \frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta} \frac{\partial \ln f(y|\Psi_t, \beta)}{\partial \beta'} f(y|\Psi_t, \beta) dy \end{aligned}$$

由于模型  $f(y|\Psi_t, \beta)$  是  $Y_t$  相对于  $\Psi_t$  的条件概率密度的正确设定, 因此结论成立.

□

**9.6** 令  $V_* = \mathbb{E}[S_t(\beta^*)S_t(\beta^*)']$ ,  $H_* = \mathbb{E}\left[\frac{\partial}{\partial \beta} S_t(\beta^*)\right]$ , 其中  $S_t(\beta) = \frac{\partial}{\partial \beta} \ln f(y|\Psi_t, \beta)$ , 且  $\beta^* = \arg \min_{\beta \in \Theta} l(\beta) = \mathbb{E}[\ln f(y|\Psi_t, \beta)]$ . 解释在什么情形下  $H_*^{-1}V_*H_*^{-1} - (-H_*^{-1})$  是半正定的.

**解答** 在模型  $f(y|\Psi_t, \beta)$  模型设定正确时, MLE 估计量  $\sqrt{n}\hat{\beta}$  的渐近方差为  $-H_*^{-1}$ , 达到了 Cramer-Rao 下界, 而当模型  $f(y|\Psi_t, \beta)$  误设时, QMLE 估计量  $\sqrt{n}\hat{\beta}$  的渐近方差为  $H_*^{-1}V_*H_*^{-1}$ , 没有达到 Cramer-Rao 下界.

因此当模型  $f(y|\Psi_t, \beta)$  模型设定正确时,  $H_*^{-1}V_*H_*^{-1} - (-H_*^{-1})$  是半正定的.

□

**9.7** 假设条件概率密度或条件质量模型  $f(y|\Psi_t, \beta)$  是  $Y_t$  相对于  $\Psi_t$  的条件概率密度的错误设定, 即不存在  $\beta \in \Theta$ , 使得  $f(y|\Psi_t, \beta)$  与  $Y_t$  的真实条件分布一致. 证明:

$$\mathbb{E}\left[\frac{\partial \ln f(Y_t|\Psi_t, \beta^*)}{\partial \beta} \frac{\partial \ln f(Y_t|\Psi_t, \beta^*)}{\partial \beta'} \middle| \Psi_t\right] + \mathbb{E}\left[\frac{\partial^2 \ln f(Y_t|\Psi_t, \beta^*)}{\partial \beta \partial \beta'} \middle| \Psi_t\right] = 0$$

一般不成立, 其中  $\beta^*$  满足假设 9.4 和 9.5. 换言之, 当条件概率模型  $f(y|\Psi_t, \beta)$  不是  $Y_t$  条件概率分布的正确设定时, 条件信息等式一般不成立. 可举例说明.

**解答** 根据问题 9.5 可知, 当模型  $f(y|\Psi_t, \beta)$  是  $Y_t$  相对于  $\Psi_t$  的条件概率密度的正确设定时

$$\mathbb{E}\left[\frac{\partial \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta} \frac{\partial \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta'} \middle| \Psi_t\right] + \mathbb{E}\left[\frac{\partial^2 \ln f(Y_t|\Psi_t, \beta^o)}{\partial \beta \partial \beta'} \middle| \Psi_t\right] = 0$$

对一切  $\beta \in \Theta$  成立. 然而当模型设定错误时, 不存在  $\beta \in \Theta$  使得  $f(y|\Psi_t, \beta)$  与  $Y_t$  的真实条件分布一致, 因此上式一般不成立.

□

**9.8** 考虑以下 MLE 问题.

**假设 9.8.1**  $\{Y_t, X_t'\}$  是遍历平稳过程, 模型  $f(y|\Psi_t, \beta)$  是  $Y_t$  条件概率密度的正确设定, 其中  $\Psi_t = (X_t', \mathbf{Z}_{t-1}')'$ ,  $\mathbf{Z}^{t-1} = (\mathbf{Z}_{t-1}', \mathbf{Z}_{t-2}', \dots, \mathbf{Z}_1')'$ ,  $\mathbf{Z}_t = (Y_t, X_t')'$ . 对于每一个  $\beta$ ,  $\ln f(y|\Psi_t, \beta)$  可

测, 对于每一个  $t$ ,  $\ln f(y|\Psi_t, \cdot)$  相对于  $\beta \in \Theta$  是二阶连续可微的, 其中  $\Theta$  是紧集.

假设 9.8.2  $l(\beta) = \mathbb{E}[\ln f(y|\Psi_t, \beta)]$  是  $\beta \in \Theta$  的连续函数.

假设 9.8.3 (i)  $\beta^o = \arg \max_{\beta \in \Theta} l(\beta)$  是  $l(\beta)$  在  $\Theta$  上的唯一最大化解; (ii)  $\beta^o$  是  $\Theta$  的内点.

假设 9.8.4 (i)  $\left\{ S_t(\beta^o) \equiv \frac{\partial}{\partial \beta} \ln f(Y_t|\Psi_t, \beta^o) \right\}$  服从中心极限定理, 即

$$\sqrt{n} \hat{S}_t(\beta^o) = n^{-\frac{1}{2}} \sum_{t=1}^n S_t(\beta^o) \xrightarrow{d} N(\mathbf{0}, V_o)$$

其中  $V_o$  是一个  $K \times K$  有限、对称和正定的方差 - 协方差矩阵.

(ii)  $\left\{ H_t(\beta) \equiv \frac{\partial^2}{\partial \beta \partial \beta'} \ln f(Y_t|\Psi_t, \beta) \right\}$  在  $\Theta$  上服从 UWLLN, 即

$$\sup_{\beta \in \Theta} \left\| n^{-1} \sum_{t=1}^n H_t(\beta) - H(\beta) \right\| \xrightarrow{p} \mathbf{0}$$

其中  $K \times K$  Hessian 矩阵  $H(\beta) \equiv \mathbb{E}[H_t(\beta)]$  是有限、对称和非奇异的矩阵, 且是  $\beta \in \Theta$  的连续函数.

MLE 估计量定义为  $\hat{\beta} = \arg \max_{\beta \in \Theta} \hat{l}(\beta)$ , 其中  $l(\beta) \equiv n^{-1} \sum_{t=1}^n \ln f(Y_t|\Psi_t, \beta)$ . 假设有  $\hat{\beta} \xrightarrow{p} \beta^o$ ,

这个一致性结果可用于回答以下问题. 给出每一小题的推理过程.

(1) 给出 MLE 估计量  $\hat{\beta}$  的一阶条件.

(2) 推导出  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近分布. 注意  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近分布应该表示为 Hessian 矩阵  $H(\beta^o)$  形式.

(3) 给出  $\sqrt{n}(\hat{\beta} - \beta^o)$  的渐近方差的一致估计量, 并证明它为什么是一致估计量.

(4) 构造一个 Wald 检验统计量以检验原假设  $\mathbb{H}_0: R(\beta^o) = r$ , 其中  $r$  是  $J \times 1$  常向量,  $R(\cdot)$  是  $J \times 1$  向量,  $R'(\beta)$  是  $J \times K$  导数矩阵,  $R'(\beta^o)$  是满秩的,  $J \times K$ . 推导出在原假设  $\mathbb{H}_0: R(\beta^o) = r$  下, Wald 检验统计量的渐近分布.

**解答** (1) 当  $n$  充分大时,  $\hat{\beta}$  是  $\Theta$  的内点, 且最大化对数条件似然函数  $\sum_{t=1}^n \ln f(Y_t|\Psi_t, \beta)$  的一阶条件成立

$$\hat{S}(\hat{\beta}) \equiv n^{-1} \sum_{t=1}^n \frac{\partial \ln f(Y_t|\Psi_t, \hat{\beta})}{\partial \beta} = \mathbf{0}$$

(2) 将  $\hat{S}(\hat{\beta})$  在  $\beta = \beta^o$  处一阶 Taylor 展开并代入一阶条件得

$$\sqrt{n} \hat{S}(\beta^o) + \hat{H}(\bar{\beta}) \sqrt{n}(\hat{\beta} - \beta^o) = \mathbf{0}$$

其中  $\bar{\beta} = \lambda \hat{\beta} + (1 - \lambda) \beta^o$ ,  $\lambda \in [0, 1]$ , 并且 Hessian 矩阵

$$\hat{H}(\beta) = n^{-1} \sum_{t=1}^n \frac{\partial^2 \ln f(Y_t|\Psi_t, \beta)}{\partial \beta \partial \beta'}$$

另一方面

$$\left\| \hat{H}(\bar{\beta}) - H_o \right\| \leq \sup_{\beta \in \Theta} \left\| \hat{H}(\beta) - H(\beta) \right\| + \left\| H(\bar{\beta}) - H(\beta^o) \right\| \xrightarrow{p} 0$$

因为  $H_o \equiv H(\beta^o)$  非奇异, 因而当  $n$  充分大时,  $\hat{H}(\bar{\beta})$  非奇异, 从而

$$\sqrt{n}(\hat{\beta} - \beta^o) = -\hat{H}^{-1}(\bar{\beta})\sqrt{n}S(\beta^o)$$

又因为当  $f(y|\Psi_t, \beta)$  是正确设定的条件概率分布时,  $V_o = \text{avar}[\sqrt{n}\hat{S}(\beta^o)] = -H_o$ , 根据中心极限定理和 Slutsky 定理可知

$$\sqrt{n}(\hat{\beta} - \beta^o) \xrightarrow{d} N(\mathbf{0}, -H_o^{-1})$$

(3) 可以选取  $\hat{\Omega} = -\hat{H}^{-1}(\hat{\beta})$  作为  $-H_o$  的一致估计量, 其中

$$\hat{H}(\beta) = n^{-1} \sum_{t=1}^n \frac{\partial^2 \ln f(Y_t|\Psi_t, \hat{\beta})}{\partial \beta \partial \beta'}$$

由于  $H(\cdot)$  是关于  $\beta \in \Theta$  的连续函数, 因而当  $\hat{\beta} \xrightarrow{p} \beta$  时,  $-\hat{H}^{-1}(\hat{\beta}) \xrightarrow{p} -H^{-1}(\beta^o) = -H_o$ .

(4) 类似第 8 章的做法, 可以构建 Wald 检验统计量

$$W = n[R(\hat{\beta}) - r]' \left[ -\frac{dR(\hat{\beta})}{d\beta'} \hat{H}^{-1}(\hat{\beta}) \frac{dR(\hat{\beta})}{d\beta} \right]^{-1} [R(\hat{\beta}) - r] \xrightarrow{d} \chi_J^2$$

□

**9.9** 考虑独立同分布随机样本线性回归模型  $Y_t = X_t' \alpha^o + \varepsilon_t$ , 其中  $\varepsilon_t|X_t \sim N(0, \sigma_o^2)$ . 令  $\beta = (\alpha', \sigma^2)'$ , 且

$$\begin{aligned} f(Y_t|X_t, \beta) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[ -\frac{(Y_t - X_t' \alpha)^2}{2\sigma^2} \right] \\ \hat{l}(\beta) &= n^{-1} \sum_{t=1}^n \ln f(Y_t|X_t, \beta) \\ &= -\frac{1}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \left[ n^{-1} \sum_{t=1}^n (Y_t - X_t' \beta)^2 \right] \end{aligned}$$

现在考虑检验原假设  $\mathbb{H}_0 : R\beta^o = r$  是否成立, 这里  $R$  是  $J \times K$  满秩矩阵,  $r$  是  $J \times 1$  常向量,  $J \leq K$ .

(1) 证明:

$$\begin{aligned} \hat{l}(\hat{\beta}) &= -\frac{1}{2} \left[ 1 + \ln(2\pi) + \ln \left( \frac{e'e}{n} \right) \right] \\ \hat{l}(\tilde{\beta}) &= -\frac{1}{2} \left[ 1 + \ln(2\pi) + \ln \left( \frac{\tilde{e}'\tilde{e}}{n} \right) \right] \end{aligned}$$

其中  $e$  和  $\tilde{e}$  分别是  $n \times 1$  无约束和有约束模型的残差向量,  $\tilde{\beta}$  是原假设  $\mathbb{H}_0 : R\beta^o = r$  下的有约束 MLE 估计量.

(2) 证明在原假设  $\mathbb{H}_0 : R\beta^o = r$  下, 似然比检验统计量

$$\begin{aligned} LR &= 2n[\hat{l}(\hat{\beta}) - \hat{l}(\tilde{\beta})] \\ &= n \ln(\tilde{e}'\tilde{e}/e'e) \\ &= J \cdot \frac{(\tilde{e}'\tilde{e} - e'e)/J}{e'e/n} + o_p(1) \\ &= J \cdot F + o_p(1) \end{aligned}$$



**解答** (1) 首先有

$$\frac{\partial \hat{l}(\hat{\beta})}{\partial \sigma^2} = -\frac{1}{2\sigma^2} + \frac{1}{2\sigma^4} \left[ n^{-1} \sum_{t=1}^n (Y_t - X_t' \alpha)^2 \right]$$

因而  $\sigma^2 = e'e/n$ , 将其代入到  $\hat{l}(\hat{\beta})$  的表达式中即可得到

$$\hat{l}(\hat{\beta}) = -\frac{1}{2} \left[ 1 + \ln(2\pi) + \ln \left( \frac{e'e}{n} \right) \right]$$

类似可证有约束情形的情况.

(2) 由 (1) 可知

$$LR = 2n[\hat{l}(\hat{\beta}) - \hat{l}(\tilde{\beta})] = n \left( \frac{\tilde{e}'\tilde{e} - e'e}{e'e} + 1 \right)$$

在原假设  $H_0: R\beta^o = r$  成立的条件下, 当  $n \rightarrow \infty$  时,  $\tilde{e}'\tilde{e}$  和  $e'e$  会收敛于同一个值, 由 Maclaurin 展开式

$$\ln(1+x) = x + o(x) \quad (-1 < x \leq 1)$$

于是

$$n \left( \frac{\tilde{e}'\tilde{e} - e'e}{e'e} + 1 \right) = J \cdot \frac{(\tilde{e}'\tilde{e} - e'e)/J}{e'e/n} + o_p(1)$$

整理即可得  $LR = J \cdot F + o_p(1)$ .

□

**9.10 证明:** 如果时间序列条件概率密度模型  $f(y|I_{t-1}, \beta)$  是随机变量  $Y_t$  相对于  $I_{t-1} = \{Y_{t-1}, Y_{t-2}, \dots\}$  的条件概率分布的正确设定, 则存在某一参数值  $\beta^o \in \Theta$ , 动态概率积分变换  $\{U_t(\beta^o)\}$  服从 i.i.d.  $U[0, 1]$ .

**解答** 注意到

$$U_t(\beta^o) = \int_{-\infty}^{Y_t} f(y|I_{t-1}, \beta^o) dy$$

是一个累积分布函数, 因此对于每个  $t = 1, 2, \dots$ ,  $U_t(\beta^o)$  的值域为  $[0, 1]$ , 可以得到

$$\begin{aligned} \mathbb{P}[U_t(\beta^o) \leq u] &= \mathbb{P}[F_Y(y) \leq u] = \mathbb{P}[F_Y^{-1}[F_Y(y)] \leq F_Y^{-1}(u)] \\ &= \mathbb{P}[y \leq F_Y^{-1}(u)] = F_Y[F_Y^{-1}(u)] = u \end{aligned}$$

因此  $U_t \sim U[0, 1]$ ,  $t = 1, 2, \dots$ . 因为条件概率密度模型  $f(y|I_{t-1}, \beta)$  设定正确,  $\{U_t(\beta^o)\}$  不存在任何形式的序列相关, 故而  $\{U_t(\beta^o)\} \sim \text{i.i.d. } U[0, 1]$ .

□