中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

第十章 经典线性回归分析

洪永淼
中国科学院数学与系统科学研究院
中国科学院大学经济与管理学院

目　录

- 假设 $\{Y_i, X_i'\}_{i=1}^n$ 是一个容量为 n 的可观测随机样本，其中
  - ✓ Y_i 是一个**标量**，
  - ✓ $X_i = (1, X_{1i}, \cdots, X_{ki})'$ 是一个 $(k+1)\times 1$ 维的**列向量**，$X_{1i}, \cdots, X_{ki}$ 是**回归变量**，
  - ✓ i 在**截面数据**中代表**个体单元**（如：一个企业、家庭或国家），在**时间序列数据**中代表一个**时期**（如：天、周、月、年）。
- 本章的目的是用随机样本 $\{Y_i, X_i'\}_{i=1}^n$ 生成的数据对条件均值 $E(Y_i \mid X_i)$ 进行建模，估计未知参数值并进行统计推断。
- 为书写方便，本章记 $p = k+1$。其中，k 是不含截距项的回归变量的个数，p 是未知参数的个数。

线性模型假设 (Linearity)

- 本章考察以下**线性回归模型**

$$
\begin{aligned}
Y_i &= \alpha + \textstyle\sum_{j=1}^{k} \beta_j X_{ji} + \varepsilon_i \\
&= X_i'\theta + \varepsilon_i, \qquad i = 1, \cdots, n
\end{aligned}
$$

- 其中
  - ✓ $\theta = (\alpha, \beta_1, \cdots, \beta_k)'$ 是一个 $p \times 1$ 维的**未知参数向量**
  - ✓ ε_i 是一个不可观测的**随机扰动项**
  - ✓ Y_i 称为**因变量**或**被解释变量** (dependent variable or regressand)
  - ✓ $X_i = (1, X_{1i}, \cdots, X_{ki})'$ 称为 $p \times 1$ 维**自变量**或**回归元** (independent variables or regressors)

模型的正确设定 & 模型误设

- 若存在某一参数值 θ_0，有

$$E(Y_i \mid X_i) = X_i'\theta_0$$

则称线性回归模型是关于条件均值 $E(Y_i \mid X_i)$ 的**正确设定。**

- 若对所有参数值 θ，

$$E(Y_i \mid X_i) \neq X_i'\theta$$

则称线性回归模型是关于条件均值 $E(Y_i \mid X_i)$ 的**误设。**

- **本章假设线性回归模型为正确设定。**

- 当且仅当条件均值 $E(Y_i \mid X_i)$ 的线性模型设定正确时，有

$$E(\varepsilon_i \mid X_i) = 0$$

- 此时，参数值

$$\theta_0 = \frac{\mathbf{d}E(Y_i \mid X_i)}{\mathbf{d}X_i}$$

可解释为 X_i 对 Y_i 的**期望边际效应** (expected marginal effect)，并称为**真实参数值。**

- 例如，若 X_i 是收入，Y_i 是消费，则 θ_0 是**期望边际消费倾向** (marginal propensity to consume)。

线性回归模型的正确设定 ≠ 存在因果关系

- 线性回归模型，即使是正确设定，也并不意味着 X_i 和 Y_i 之间存在着因果关系。
- 正如 Kendall & Stuart (1961) 指出的，“统计关系，无论多么强和多么富有启示性，都不能确立因果关系。因果关系的思想必须来自统计学之外，来源于一些理论或其他方面”。
- 线性回归模型描述一种线性预测或推测（predictive）关系，即给定 X_i，是否可用线性模型预测或推测 Y_i?

- 现令

$$\boldsymbol{Y} = (Y_1, \cdots, Y_n)', \qquad n \times 1$$
$$\boldsymbol{X} = (X_1, \cdots, X_n)', \qquad n \times p$$
$$\boldsymbol{\varepsilon} = (\varepsilon_1, \cdots, \varepsilon_n)', \qquad n \times 1$$

- 这里 $\boldsymbol{X}$ 的第 i 行是 $1 \times p$ 维行向量 $X_i' = (1, X_{1i}, \cdots, X_{ki})$。
- 用矩阵符号，线性回归模型可简洁地表示为

$$\boldsymbol{Y} = \boldsymbol{X}\theta_0 + \boldsymbol{\varepsilon}$$

$$n \times 1 = (n \times p)(p \times 1) + n \times 1$$

严格外生性假设 (strict exogeneity)

- 本章假设以下条件成立，即

$$E(\varepsilon_i \mid \boldsymbol{X}) = E(\varepsilon_i \mid X_1, \cdots, X_i, \cdots, X_n) \\ = 0, i = 1, \cdots, n$$

- 这一条件称为**严格外生性** (strict exogeneity)。
- 它隐含线性回归模型为 $E(Y_i \mid X_i)$ 的正确设定，因为通过重复期望法则，可推出 $E(\varepsilon_i \mid X_i) = 0$ 和 $E(\varepsilon_i) = 0$。
- 对于任意 $i, j \in \{1, \cdots, n\}$，有 $E(X_i \varepsilon_j) = 0$，因为

$$\begin{aligned} E(X_i \varepsilon_j) &= E[E(X_i \varepsilon_j \mid \boldsymbol{X})] \\ &= E[X_i E(\varepsilon_j \mid \boldsymbol{X})] \\ &= E(X_i 0) \\ &= \mathbf{0} \end{aligned}$$

严格外生性假设 (Cont.)

- 由于$E(\varepsilon_i)=0$，$E\left(X_i\varepsilon_j\right)=\mathbf{0}$ 意味着对任意的 $i,j\in\{1,\cdots,n\}$，有 $\operatorname{cov}(X_i,\varepsilon_j)=\mathbf{0}$。
- 矩阵 $\boldsymbol{X}$ 包含所有的自变量向量 $X_1,X_2,\cdots,X_n$。如果下标 i 表示时间，则严格外生性假设要求随机扰动项 ε_i 的条件均值不依赖于所有自变量过去和未来的数值。这排除了扰动项 ε_i 与自变量未来值之间的相关性，因此排除了所谓动态时间序列回归模型，即 X_i 包含因变量 Y_i 的滞后项 (如 Y_{i-1},Y_{i-2}) 的回归模型。

例 [first-order autoregression, AR (1)]

- 例如，考虑**一阶自回归 AR(1) 模型**

$$\begin{aligned} Y_i &= \alpha_0 + \beta_0 Y_{i-1} + \varepsilon_i \\ &= X_i'\theta_0 + \varepsilon_i, \qquad i = 1, \cdots, n \end{aligned}$$

$$\{\varepsilon_i\} \sim \mathrm{IID}(0, \sigma^2)$$

其中 $X_i = (1, Y_{i-1})'$ 包含 Y_i 的一阶滞后项 Y_{i-1}。

- 这是一个**动态回归模型**，因为 $\beta_0 Y_{i-1}$ 代表过去的“记忆”或“反馈”对现值的影响。
- 随机扰动项 ε_i 可视为第 i 时期的**“新信息”效应**。由于新信息不能够被预测，因此今天的新信息应该与昨天的信息无关，这就意味着 $E(\varepsilon_i \mid X_i) = 0$。

例 (Cont.) [first-order autoregression, AR (1)]

- 假设新信息 ε_i 为 IID $(0, \sigma^2)$ 序列，则 ε_i 与过去的信息集合互相独立，故有 $E(X_i\varepsilon_i) = E(X_i)E(\varepsilon_i) = \mathbf{0}$
- 但是，$E(X_{i+1}\varepsilon_i) \neq \mathbf{0}$，从而 $E(\varepsilon_i \mid \boldsymbol{X}) \neq 0$，即严格外生性假设不成立。
- 换言之，严格外生性排除了一阶自回归模型。
- 这里，X_i 所包含的 Y_i 的一阶滞后项 Y_{i-1} 称为**先决变量** (predetermined variable)，因为 Y_{i-1} 与 ε_i 正交但依赖于 $\{\varepsilon_i\}$ 过去的信息。

其他外生性的定义

- 在计量经济学中，还有其他外生性的定义。
  - ✓ 比如，可将其定义为 $\{\varepsilon_i\}$ 和 $\boldsymbol{X}$ 相互独立，或 $\boldsymbol{X}$ 是非随机的。这排除了条件异方差 (即 $\mathrm{var}(\varepsilon_i \mid \boldsymbol{X})$ 随 $\boldsymbol{X}$ 的变化而变化) 的可能性。
  - ✓ 在严格外生性假设下，仍允许条件异方差的存在，因为没有假设 ε_i 和 $\boldsymbol{X}$ 是相互独立的，仅仅假设条件均值 $E\,(\varepsilon_i \mid \boldsymbol{X})$ 不依赖于 $\boldsymbol{X}$。

情形(1)：X 是非随机的

◆ **问题**

如果 **X** 是非随机的，这对严格外生性假设将有何影响?

- 如果 **X** 是非随机的，严格外生性假设将变成以下简单的条件

$$E(\varepsilon_i \mid \boldsymbol{X}) = E(\varepsilon_i) = 0$$

- 非随机 **X** 的一个例子是 $X_i = (1, i, \cdots, i^k)'$，其中 i 代表时间。这是一个具有时间趋势的回归模型

$$\begin{aligned} Y_i &= X_i'\theta_0 + \varepsilon_i \\ &= \alpha_0 + \textstyle\sum_{j=1}^{k} \beta_{j0} i^j + \varepsilon_i \end{aligned}$$

情形(2)：$\{Y_i, X_i'\}_{i=1}^n$ 是一个独立同分布随机样本

◆ **问题**

如果 $\{Y_i, X_i'\}_{i=1}^n$ 是一个独立同分布随机样本，即随机向量 $(Y_1, X_1'), \cdots, (Y_n, X_n')$ 是联合独立的，这对严格外生性假设有何影响?

- 当 $\{Y_i, X_i'\}_{i=1}^n$ 为独立同分布随机样本时，严格外生性假设将变为

$$
\begin{aligned}
E(\varepsilon_i \mid \boldsymbol{X}) &= E(\varepsilon_i \mid X_1, \cdots, X_i, \cdots, X_n) \\
&= E(\varepsilon_i \mid X_i) \\
&= 0
\end{aligned}
$$

- 换言之，当 $\{Y_i, X_i'\}_{i=1}^n$ 为独立同分布时，$E(\varepsilon_i \mid \boldsymbol{X}) = 0$ 等价于 $E(\varepsilon_i \mid X_i) = 0$，即线性回归模型正确设定。

球形误差方差 (spherical error variance) 假设

- (a) [**条件同方差** (conditional homoskedasticity)]

$$E\left(\varepsilon_i^2 \mid \boldsymbol{X}\right) = \sigma^2 > 0, \qquad i = 1, \cdots, n$$

- (b) [**条件不相关** (conditional uncorrelatedness)]

$$E\left(\varepsilon_i \varepsilon_j \mid \boldsymbol{X}\right) = 0, \qquad i \neq j, \qquad i, j \in \{1, \cdots, n\}$$

球形误差方差假设的含义

- 其中，条件 (a) 意味着 ε_i **存在条件同方差**，即

$$\begin{aligned}\operatorname{var}(\varepsilon_i \mid \boldsymbol{X}) &= E\left(\varepsilon_i^2 \mid \boldsymbol{X}\right) - [E(\varepsilon_i \mid \boldsymbol{X})]^2 \\ &= E\left(\varepsilon_i^2 \mid \boldsymbol{X}\right) \\ &= \sigma^2\end{aligned}$$

- 条件 (b) 意味着 $\{\varepsilon_i\}$ **不存在条件自相关**，即对所有的 $i \neq j$，有

$$\begin{aligned}\operatorname{cov}\left(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}\right) &= E\left(\varepsilon_i \varepsilon_j \mid \boldsymbol{X}\right) - E(\varepsilon_i \mid \boldsymbol{X}) E\left(\varepsilon_j \mid \boldsymbol{X}\right) \\ &= 0\end{aligned}$$

✓ 如果 i 表示**个体单元**，这意味着**横截面不相关**（cross-sectional uncorrelatedness)

✓ 如果 i 代表**时间**，这意味着**序列不相关**（serial uncorrelatedness)

- 根据重复期望法则，上述**球型方差 (spherical error variance) 假设**意味着对于所有 $i \in \{1,\cdots,n\}$，$\mathrm{var}(\varepsilon_i) = \sigma^2$，这称为**无条件同方差**。同样地，对于所有 $i \neq j$，有 $\mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0$。
- **严格外生性假设**和**球形误差方差假设**结合起来，可简洁表述为

$$E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \mathbf{0}$$

$$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}$$

其中 I 是一个 $n \times n$ 的单位阵。

- 由 $E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \mathbf{0}$ 和 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}$ 并不能推出 ε_i 和 $\boldsymbol{X}$ 是相互独立的，它们只是假设 ε_i 的条件均值和条件方差不依赖于 $\boldsymbol{X}$，但允许 ε_i 的条件高阶矩 (比如条件偏度和峰度) 依赖于 $\boldsymbol{X}$。

非奇异性假设 (non-singularity)

- 为了识别未知参数值 θ_0，本章还将假设 $p\times p$ 维方阵 $\boldsymbol{X}'\boldsymbol{X}=\sum_{i=1}^{n}X_iX_i'$ 是非奇异的。
- 这个假设**排除**了对任意 $i\in\{1,\cdots,n\}$，向量 X_i 中包括截距项在内的 p 个自变量之间存在**多重共线性** (multicollinearity) 的可能性。
- 自变量 X_i 存在**多重共线性**，是指至少存在一个 $j\in\{0,1,\cdots,k\}$ 以及对所有的 $i\in\{1,\cdots,n\}$，变量 X_{ji} 可表示为其他 k 个变量 $\{X_{li},l\neq j\}$ 的线性组合。在这种情况下，$\boldsymbol{X}'\boldsymbol{X}$ 不是非奇异矩阵，其结果将导致线性回归模型的真实参数值 θ_0 不可识别。

非奇异性假设 (Cont.)

- $\boldsymbol{X}'\boldsymbol{X}$ 的非奇异性意味着 $\boldsymbol{X}$ 必须是**满秩**的，即秩为 p。因此，需要 $p \leq n$，即解释变量的个数不能超过样本容量 n，这是真实参数值 θ_0 可识别的必要条件。
- 直观上说，如果 $\{X_i\}_{i=1}^n$ 的值不存在变动或变动很小，将很难确定 Y_i 和 X_i 之间的关系。经典线性回归的目的就是考察 X_i 的变化如何导致 Y_i 的变化，因此，需要不同 i 的 X_i 取值有显著不同，才能较为精确地确定 X_i 和 Y_i 的关系。
- 某种意义上， $\boldsymbol{X}'\boldsymbol{X}$ 可称为样本 $\boldsymbol{X}$ 的“信息矩阵”，因为它测度了 $\boldsymbol{X}$ 中的信息含量。$\boldsymbol{X}'\boldsymbol{X}$ 包含的信息含量影响参数值 θ_0 的估计精度。

近似多重共线性导致的问题

- 如果存在**近似多重共线性**（near-multicollinearity），即自变量 X_i 的样本值之间存在近似的线性关系，这时，虽然 $\boldsymbol{X}'\boldsymbol{X}$ 是非奇异的，但 $\boldsymbol{X}'\boldsymbol{X}$ 接近一个奇异矩阵，其最小特征值不为零但很接近零，且并不随样本容量 n 的增大而增大。
- 其结果是，虽然以下将要介绍的普通最小二乘（ordinary least squares，OLS）估计量总有定义，其有限样本分布也可获得，但当 $n \to \infty$ 时，OLS 估计量的均方误并不趋于零，从而 OLS 估计量不收敛于真实参数值 θ_0。

目　录

◆ 问题

- 如何利用随机样本 $\{Y_i, X_i'\}_{i=1}^n$ 产生的数据估计未知参数值 θ_0?

定义 10.1

[OLS 估计量]：定义线性回归模型 $Y_i = X_i'\theta + \varepsilon$ 的残差平方和 (sum of squared residuals, SSR) 为

$$SSR(\theta) \equiv (\boldsymbol{Y} - \boldsymbol{X}\theta)'(\boldsymbol{Y} - \boldsymbol{X}\theta)$$
$$= \sum_{i=1}^n (Y_i - X_i'\theta)^2$$

则普通最小二乘 (OLS) 估计量 $\hat{\theta}$ 是以下最优化问题的解:

$$\hat{\theta} = \arg \min_{\theta \in \mathbb{R}^p} \mathrm{SSR}(\theta)$$

定理 10.1

[OLS 的存在性 (Existence of OLS)]：假设 $p \times p$ 维矩阵 $\boldsymbol{X}'\boldsymbol{X}$ 非奇异，则 OLS 估计量 $\hat{\theta}$ 存在，且

$$\hat{\theta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

证明：

- 假设 a 和 θ 均是 $p \times 1$ 维向量，利用下列等式

$$\frac{\partial(a'\theta)}{\partial\theta} = a$$

定理 10.1 (Cont.)

- 可得

$$
\begin{aligned}
\frac{d\mathrm{SSR}(\theta)}{d\theta} &= \frac{d}{d\theta}\sum_{i=1}^{n}(Y_i - X_i'\theta)^2 \\
&= \sum_{i=1}^{n}\frac{\partial(Y_i - X_i'\theta)^2}{\partial\theta} \\
&= \sum_{i=1}^{n}2(Y_i - X_i'\theta)\frac{\partial(Y_i - X_i'\theta)}{\partial\theta} \\
&= -2\sum_{i=1}^{n}X_i(Y_i - X_i'\theta) \\
&= -2\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\theta)
\end{aligned}
$$

定理 10.1 (Cont.)

- OLS 估计量必须满足**一阶条件**：

$$-2\boldsymbol{X}'(\boldsymbol{Y}-\boldsymbol{X}\hat{\theta})=\boldsymbol{0}$$

- 从而有

$$(\boldsymbol{X}'\boldsymbol{X})\hat{\theta}=\boldsymbol{X}'\boldsymbol{Y}$$

- 由于 $\boldsymbol{X}'\boldsymbol{X}$ 是非奇异的，因此有

$$\hat{\theta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$$

定理 10.1 (Cont.)

- 检查**二阶条件**，有 $p\times p$ 维**样本黑塞矩阵**

$$\frac{\partial^2 \mathrm{SSR}(\theta)}{\partial\theta\partial\theta'} = -2\sum_{i=1}^{n}\frac{\partial}{\partial\theta'}[X_i(Y_i - X_i'\theta)]$$
$$= 2\boldsymbol{X}'\boldsymbol{X}$$

- 由于样本黑塞矩阵为正定，故 $\hat{\theta}$ 是**全局最优解。**

证毕。

※ **注意:** $\hat{\theta}$ 的存在只需要 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵这一条件。

- $\hat{Y}_i \equiv X_i'\hat{\theta}$ 称为观测值 Y_i 的**拟合值** (fitted value) 或**预测值** (predicted value)。
- $e_i \equiv Y_i - \hat{Y}_i$ 是观测值 Y_i 的**估计残差** (estimated residual) 或**预测误差** (predicted error)。
- 注**估计残差**

$$\begin{aligned} e_i &= Y_i - \hat{Y}_i \\ &= (X_i'\theta_0 + \varepsilon_i) - X_i'\hat{\theta} \\ &= \varepsilon_i - X_i'\left(\hat{\theta} - \theta_0\right) \end{aligned}$$

  其中真实扰动项 ε_i 是不可避免的**噪声**，
- $X_i'\left(\hat{\theta} - \theta_0\right)$ 是**估计误差**；当样本容量越大 (从而 $\hat{\theta}$ 越趋近 θ_0) 时，这一项将变得越小，乃至可忽略不计。

- 普通最小二乘法 (OLS) 的一阶条件意味着 $n \times 1$ 维估计残差向量 $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\hat{\theta}$ 与自变量矩阵 $\boldsymbol{X}$ 是正交的，即

$$\boldsymbol{X}'\boldsymbol{e} = \sum_{i=1}^{n} X_i e_i = \boldsymbol{0}$$

- 这是由 OLS 的性质决定的，可由最小化问题 $\min_{\theta \in \mathbb{R}^p} SSR(\theta)$ 的一阶条件直接推出。无论线性回归模型设定是否正确，即不论 $E(\varepsilon_i \mid \boldsymbol{X}) = 0$ 是否成立，该正交条件总成立。
- 另外，如果 X_i 包含截距项，则 $\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$ 意味着 $\boldsymbol{l}'\boldsymbol{e} = 0 = \sum_{i=1}^{n} e_i = 0$，即估计残差 e_i 的样本均值总为 0，其中 $\boldsymbol{l} = (1, 1, \cdots, 1)'$ 为一个 $n \times 1$ 维向量，其每个元素均取值为 1。

目　录

◆ **问题**

线性回归模型对数据的拟合程度如何？换言之，线性回归模型对观测数据 $\{Y_t\}$ 的变动的预测能力如何?

- 为了刻画拟合优度，需要定义一些准则或指标。

定义 10.2

[非中心化 R^2 (Uncentered R^2)]：非中心化多元相关系数平方 R^2 定义为

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

其中，$\widehat{\boldsymbol{Y}}$ 是 $n \times 1$ 维的拟合值向量，第二个等式是由 OLS 估计的一阶条件得到的。

- R_{uc}^2 的含义是：因变量 $\{Y_i\}$ 的非中心化的样本二次型变动可以被预测值 $\{\widehat{Y}_i\}$ 的非中心化样本二次型变动所解释的比例。
- 根据定义，总有 $0 \leq R_{uc}^2 \leq 1$。

定义 10.3

[中心化 R^2 或决定系数 (Coefficient of Determination)]：决定系数定义为

$$R^2 \equiv 1 - \frac{\sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

其中 $\bar{Y} = n^{-1} \sum_{i=1}^{n} Y_i$ 是样本均值。

- 当 X_i 包含截距项即 $X_{0i} = 1$ 时，可进行如下正交分解：

$$\begin{aligned}\sum_{i=1}^{n} (Y_i - \bar{Y})^2 &= \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y} + Y_i - \hat{Y}_i\right)^2 \\ &= \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n} e_i^2 + 2\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)e_i \\ &= \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2 + \sum_{i=1}^{n} e_i^2\end{aligned}$$

定义 10.3 (Cont.)

- 其中**交叉项**

$$\begin{aligned}\sum_{i=1}^{n}(\hat{Y}_i-\bar{Y})e_i &= \sum_{i=1}^{n}\hat{Y}_ie_i-\bar{Y}\sum_{i=1}^{n}e_i \\ &= \hat{\theta}'\sum_{i=1}^{n}X_ie_i-\bar{Y}\sum_{i=1}^{n}e_i \\ &= \hat{\theta}'\boldsymbol{X}'\boldsymbol{e}-\bar{Y}\boldsymbol{l}'\boldsymbol{e} \\ &= \hat{\theta}'\mathbf{0}-\bar{Y}0 \\ &= 0\end{aligned}$$

- 这里使用了 OLS 估计的一阶条件，即 $\boldsymbol{X}'\boldsymbol{e}=\mathbf{0}$ 和 $\boldsymbol{l}'\boldsymbol{e}=\sum_{i=1}^{n} e_i=0$ (因为 X_i 包含截距项)。

定义 10.3 (Cont.)

- 从而

$$R^2 \equiv 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$
$$= \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2 - \sum_{i=1}^{n} e_i^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$
$$= \frac{\sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2}{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}$$

- 并且有

$$0 \le R^2 \le 1$$

定义 10.3 (Cont.)

- 如果 X_i 不包含截距项，则

$$\sum_{i=1}^{n}(Y_i-\bar{Y})^2=\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2+\sum_{i=1}^{n}e_i^2+2\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)e_i$$
$$\neq\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)^2+\sum_{i=1}^{n}e_i^2$$

- 在这种情况下，R^2 可能为负值！因为交叉项 $2\sum_{i=1}^{n}\left(\hat{Y}_i-\bar{Y}\right)e_i$ 可能为负值。
- 当 X_i 包含截距项时，中心化 R^2 和非中心化 R_{uc}^2 有相似的解释，即 R^2 测度 $\boldsymbol{Y}$ 的样本方差中可被线性模型拟合值 $X_i'\hat{\theta}$ 所解释的那部分的比例。

例 10.1：[资本资产定价模型 (Capital Asset Pricing Model, CAPM) 与 R^2 的经济含义]

- 经典资本资产定价模型可由以下方程刻画：

$$r_i - r_{fi} = \alpha + \beta\left(r_{mi} - r_{fi}\right) + \varepsilon_i, \qquad i = 1, \cdots, n$$

- 其中
  - ✓ r_i 是**某资产投资组合**在时期 i 的收益率，r_{fi} 是**无风险资产**在时期 i 的收益率，r_{mi} 是**市场投资组合**在时期 i 的收益率。
  - ✓ $r_i - r_{fi}$ 是资产投资组合的风险溢价。
  - ✓ $r_{mi} - r_{fi}$ 是市场投资组合的风险溢价，是唯一的**系统风险因子。**
  - ✓ ε_i 是资产投资组合的**特质风险**，可通过分散化投资而消除。

例 10.1 (Cont.)：

- 在该模型中，R^2 有很好的经济解释：它是其资产投资组合的风险 (以其风险溢价 $r_i - r_{fi}$ 的样本方差测度) 能够被市场风险因子 $r_{mi} - r_{fi}$ 解释的那部分的比例，而 $1 - R^2$ 是该资产投资组合的风险中特质风险因子 ε_i 所占的比例。
- 实际上，中心化 R^2 是 $\{Y_i\}$ 和 $\{\hat{Y}_i\}$ 之间相关系数的平方。

定理 10.2

$R^2=\hat{\rho}_{Y\hat{Y}}^2$，这里 $\hat{\rho}_{Y\hat{Y}}$ 是 $\{Y_i\}$ 和 $\{\hat{Y}_i\}$ 之间的样本相关系数。

- R^2 被称为多元相关系数平方的原因：由于拟合值 $\hat{Y}_i=X_i'\hat{\theta}=\hat{\alpha}+\sum_{j=1}^{k}\hat{\beta}_j X_{ji}$ 是 $\{X_{ji}\}_{j=1}^{k}$ 的线性组合，其中$\hat{\theta}=(\hat{\alpha},\hat{\beta}_1,\cdots,\hat{\beta}_k)'$，$R^2$ 可视为 Y_i 和自变量 $\{X_{ji}\}_{j=1}^{k}$ 之间的多元样本相关系数的加权平均的平方。

定理 10.3 [R^2 的非减性]

- 假设 $\{Y_i, X_{1i}, \cdots, X_{(k+q)i}\}_{i=1}^{n}$ 是样本容量为 n 的可观测随机样本，R_1^2 是以下线性回归模型的中心化拟合优度

$$Y_i = X_i'\theta + \varepsilon_i$$

其中 $X_i = (1, X_{1i}, \cdots, X_{ki})'$，$\theta$ 是 $p \times 1$ 维的未知参数向量。

- 另假设 R_2^2 是以下扩展的线性回归模型的中心化拟合优度

$$Y_i = \tilde{X}_i'\gamma + u_i$$

其中 $\tilde{X}_i = \left(1, X_{1i}, \cdots, X_{ki}, X_{(k+1)i}, \cdots, X_{(k+q)i}\right)'$，$\gamma$ 是 $(p+q) \times 1$ 维的未知参数向量，q 是正整数。则

$$R_2^2 \geq R_1^2$$

证明：

- 根据拟合优度 R^2 的定义 (即定义 10.3)，有

$$R_1^2 = 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

$$R_2^2 = 1 - \frac{\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}}{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- 其中 $\boldsymbol{e}$ 是 $\boldsymbol{Y}$ 对 $\boldsymbol{X}$ 回归的估计残差向量， $\tilde{\boldsymbol{e}}$ 是 $\boldsymbol{Y}$ 对 $\tilde{\boldsymbol{X}}$ 回归的估计残差向量。
- 因此，只需要证明 $\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}} \le \boldsymbol{e}'\boldsymbol{e}$。
- 因为 OLS 估计量 $\hat{\gamma} = \left(\tilde{\boldsymbol{X}}'\tilde{\boldsymbol{X}}\right)^{-1}\tilde{\boldsymbol{X}}'\boldsymbol{Y}$ 是使扩展回归模型 $Y_i = \tilde{X}_i'\gamma + u_i$ 的 $SSR(\gamma)$ 最小化的最优解，故对于任意的 $\gamma \in \mathbb{R}^{p+q}$，有 $\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}} = \sum_{i=1}^{n}\left(Y_i - \tilde{X}_i'\hat{\gamma}\right)^2 \le \sum_{i=1}^{n}\left(Y_i - \tilde{X}_i'\gamma\right)^2$

证明 (Cont.)：

- 现选择 $\gamma = \left(\widehat{\theta}', \mathbf{0}'\right)'$，其中 $\hat{\theta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ 是第一个回归模型 $Y_i = X_i'\theta + \varepsilon_i$ 的 OLS 估计量，则有

$$\begin{aligned}\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}} &\le \sum_{i=1}^{n} \left(Y_i - X_i'\widehat{\theta} - \sum_{j=k+1}^{k+q} 0X_{ji}\right)^2 \\ &= \sum_{i=1}^{n} \left(Y_i - X_i'\widehat{\theta}\right)^2 \\ &= \boldsymbol{e}'\boldsymbol{e}\end{aligned}$$

- 因此，$R_1^2 \le R_2^2$。

证毕。

定理 10.3 的重要含义

- 首先，R^2 可用于自变量数目相等的线性回归模型的比较，但不适用于比较不同自变量数目的线性回归模型，因为模型的自变量越多，R^2 会越大，即使新增加的自变量对因变量没有真正的解释力，R^2 也会增加。
- 其次，R^2 不是正确模型设定的判断标准。它测度的是抽样变化而非总体。R^2 高并不意味着模型设定正确，同样，正确的模型设定也并不意味着 R^2 高。事实上，给定自变量 X_i，R^2 值的大小与线性回归模型的信噪比 (signal to noise ratio) 有关。

定理 10.3 的重要含义 (Cont.)

- 严格来讲，R^2 只是测度了一种关联性 (association)，与因果关系无关。在经济时间序列实证分析中，高的 R^2 通常容易获得。有时即使两变量间的因果关系很弱或几乎不存在，也能获得高的 R^2。
  - ✓ 例如，在伪回归 (spurious regression) 中，因变量 Y_i 和自变量 X_i 之间不存在因果关系，但由于它们在时间上常常表现出相同的趋势，结果 R^2 接近于 1。

定理 10.3 的重要含义 (Cont.)

- 最后，R^2 测度了因变量 Y_i 和自变量 X_i 之间的线性关联程度 (见定理 10.3)。当 $E(Y_i \mid X_i)$ 是非线性函数时，用 R^2 来测度非线性回归模型的拟合优度并不合适。
  - ✓ 例如，考察线性回归模型
  $$\ln Y_i = \alpha_0 + \beta_{10} \ln L_i + \beta_{20} \ln K_i + \varepsilon_i$$
  - ✓ 其中，Y_i 是产出，L_i 是劳动，K_i 是资本。产出 Y_i 不是投入 L_i 和 K_i 的线性函数。
  - ✓ 在这种情况下，R^2 是 $\ln Y_i$ 的总样本变化能够被 $\ln K_i$ 和 $\ln L_i$ 的样本变化所解释的比例。它并不是 Y_i 的样本二次型变动能够被 L_i 和 K_i 的样本变化所解释的比例。

◆ 问题 10.1

那么，对于线性回归模型，什么是合适的模型选择准则?

- 通常存在很多备选的自变量可用于预测因变量，但没有必要在回归模型中包含所有的自变量。
- **权衡：**
  - ✓ 自变量越多，模型的系统偏差越小，如果所有的参数估计都不存在误差，那么该模型的预测能力是最优的。
  - ✓ 在样本容量 n 给定的情形下，参数越多，参数估计的准确性越差。
- “KISS (Keep It Sophistically Simple)”原则：尽量用简单的模型刻画数据所包含的重要信息。

模型选择准则

(1) Akaike 信息准则 (Akaike's information criterion, AIC)

- 线性回归模型可通过选择合适的未知参数数目 p，以最小化 AIC 准则来选择模型，由 Akaike (1973) 提出：

$$AIC = \ln\left(s^2\right) + \frac{2p}{n}$$

- 其中
  - ✓ $s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-p}$，是 $E\left(\varepsilon_i^2\right) = \sigma^2$的残差方差估计量（residual variance estimator)
  - ✓ $p = k + 1$ 是包括截距项在内的**自变量** X_i 的**数目**
  - ✓ $\ln(s^2)$ 测度**模型的拟合优度**
  - ✓ $2p/n$ 测度**模型的复杂程度**

(2) Bayesian 信息准则 (Bayesian information criterion, BIC)

- 线性回归模型也可以通过选择合适的维数 p，以最小化以下所谓的 Bayesian 信息准则即 BIC 准则来选择模型：

$$BIC = \ln\left(s^2\right) + \frac{p\ln(n)}{n}$$

- BIC 准则由 Schwarz (1978) 提出。

- AIC 和 BIC 两个信息准则都试图在模型的拟合优度 $\ln(s^2)$ 和尽量少用参数之间进行权衡。
- 当 $n \geq 7$ 时，$\ln n \geq 2$，与 AIC 相比，BIC 对模型复杂度给予更大的惩罚，这从各式第二项的对比中可以看出。因此，BIC 倾向于选择更加简单的线性回归模型。
- AIC 和 BIC 的区别主要在于它们的构造方式不同，AIC 倾向于选择具有最优预测能力的模型。
  - ✓ 相对于 BIC 而言，AIC 选择的参数更多；
  - ✓ BIC 倾向于选择正确的维数 p。在一定的正则条件下，当样本容量 $n \to \infty$ 时，BIC 可一致地选择正确的维数 p。

(3) 调整的 R^2

- $\bar{R}^2$，称为调整的 R^2，也可用于选择线性回归模型。
- 回顾 R^2 的定义，其可写为

$$R^2 = 1 - \frac{n^{-1}\boldsymbol{e}'\boldsymbol{e}}{n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2}$$

- 可以证明， $n^{-1}\boldsymbol{e}'\boldsymbol{e}$ 和 $n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$ 分别是方差 $\sigma^2 = \text{var}(\varepsilon_i)$ 和 $\sigma_Y^2 = \text{var}(Y_i)$ 的有偏估计，即 $E(n^{-1}\boldsymbol{e}'\boldsymbol{e}) \neq \sigma^2$, $E[n^{-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2] \neq \sigma_Y^2$。

- 这些偏差有时导致 R^2 偏向选择较复杂的模型。可以通过使用无偏估计量 $\boldsymbol{e}'\boldsymbol{e}/(n-p)$ 和 $(n-1)^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$ 而加以消除这些偏差。
- 这样可得偏差调整后的 R^2，即

$$\bar{R}^2=1-\frac{\boldsymbol{e}'\boldsymbol{e}/(n-p)}{(n-1)^{-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2}$$

- 这称为调整后的 R^2。在 $\bar{R}^2$ 中，调整的是自由度或自变量 X_i 的个数。可以证明

$$\bar{R}^2=1-\frac{n-1}{n-p}\left(1-R^2\right)$$

- 应该注意，虽然 X_i 包含截距项，但 $\bar{R}^2$ 也可能取负值。

◆ 问题 10.2

为什么复杂的模型在实际应用中并不一定预测最好?

- 在许多应用中，特别是自变量数目较多时，若自变量存在近似多重共线性，则矩阵 $\boldsymbol{X}'\boldsymbol{X}$ 可能接近于奇异矩阵。因此，OLS 估计量 $\hat{\beta}$ 将不稳定，导致均方误 (MSE) 的方差较大。
- **岭回归 (ridge regression)**：一个解决方法就是考察以下限制 β 的大小的估计量

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}}\,(Y - \boldsymbol{X}\beta)'(Y - \boldsymbol{X}\beta) + \lambda\beta'\beta$$
$$= (\boldsymbol{X}'\boldsymbol{X} + \lambda I)^{-1}\boldsymbol{X}'Y$$

其中 λ 是调整参数，通过平方系数之和来控制参数 β 的大小。

- **LASSO** (least absolute shrinkage and selection operator)：当存在高维自变量集时，特别是当自变量个数 p 可能超过样本容量 n 时，β 的许多参数可能为 0 或足够小可忽略不计，这称为高维线性回归的**稀疏性 (sparsity) 假设**。
- 当参数个数 p 比样本容量 n 大时，无法进行 OLS 估计，因为 $\boldsymbol{X}'\boldsymbol{X}$ 是奇异的。在这种情况下，可考察以下估计量

$$\hat{\beta} = \underset{\beta}{\operatorname{argmin}} (Y - \boldsymbol{X}\beta)'(Y - \boldsymbol{X}\beta) + \lambda|\beta|_1$$

其中 $|\beta|_1 = \sum_{j=0}^{k} |\beta_j|$ 是 β 的 L_1 范数。

目　录

◆ 问题

现考察 OLS 估计量 $\hat{\theta}$ 的概率统计性质:

✓ $\hat{\theta}$ 是 θ_0 的一个很好的估计量吗 (无偏性)?

✓ $\hat{\theta}$ 是 θ_0 的最优估计量吗 (有效性)?

✓ $\hat{\theta}$ 的抽样分布是什么 (正态性)?

- 由于 $\hat{\theta}$ 是随机样本 $\{Y_i, X_i'\}_{i=1}^n$ 的函数，估计量 $\hat{\theta}$ 的分布常称为 $\hat{\theta}$ 的抽样分布。
- $\hat{\theta}$ 的抽样分布对了解 $\hat{\theta}$ 的统计性质以及对未知真实参数值 θ_0 进行统计推断是非常有用的，比如可用于对未知参数值 θ_0 的**置信区间估计**和**假设检验**。

定理 10.4

假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，$E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0}$，以及 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2\boldsymbol{I}$，其中 $\boldsymbol{I}$ 为 $n \times n$ 维单位矩阵。则对所有 $n > p$，

(a) [无偏性 (unbiasedness)] $E(\hat{\theta} \mid \boldsymbol{X}) = \theta_0$

(b) [方差结构 (variance structure)]

$$\begin{aligned}\mathrm{var}(\hat{\theta} \mid \boldsymbol{X}) &= E\{[\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})][\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})]' \mid \boldsymbol{X}\} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\end{aligned}$$

(c) [正交性 (orthogonality between $\hat{\theta}$ and e)]

$$\mathrm{cov}(\hat{\theta}, \boldsymbol{e} \mid \boldsymbol{X}) = \boldsymbol{0}$$

定理 10.4 (Cont.)

(d) [高斯-马尔可夫（Gauss-Markov）定理] 对任意的线性无偏估计量 $\tilde{b}$，$\mathrm{var}(\tilde{b} \mid \boldsymbol{X}) - \mathrm{var}(\hat{\theta} \mid \boldsymbol{X})$ 为半正定（positive semi-definite, PSD)

(e) [残差方差估计量 (residual variance estimator)]

$$s^2 = \frac{\boldsymbol{e}'\boldsymbol{e}}{n-p} = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2$$

是 $\sigma^2 = E\left(\varepsilon_i^2\right)$ 的无偏估计量，即 $E(s^2 \mid \boldsymbol{X}) = \sigma^2$。

证明:

- (a) 由 $\hat{\theta} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ 和 $\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\theta}_0 + \varepsilon$，可推得

$$\hat{\theta} - \theta_0 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\varepsilon$$

- 故有

$$\begin{aligned} E\left[\left(\hat{\theta} - \theta_0\right) \mid \boldsymbol{X}\right] &= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \mid \boldsymbol{X}] \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{0} \\ &= \boldsymbol{0} \end{aligned}$$

- 这里使用了**严格外生性假设。**

证明 (Cont.):

- **(b)** 给定 $\hat{\theta} - \theta_0 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$ 和 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2\boldsymbol{I}$，有

$$
\begin{aligned}
\mathrm{var}(\hat{\theta} \mid \boldsymbol{X}) &= E\{[\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})][\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})]' \mid \boldsymbol{X}\} \\
&= E\left[(\hat{\theta} - \theta_0)(\hat{\theta} - \theta_0)' \mid \boldsymbol{X}\right] \\
&= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

- 这里，假设 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2\boldsymbol{I}$ 是保证 $\mathrm{var}(\hat{\theta} \mid \boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的关键条件。

证明 (Cont.):

- (c) 定义 $n \times n$ 维投影矩阵 $\boldsymbol{P} = \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$ 以及 $\boldsymbol{M} = \boldsymbol{I} - \boldsymbol{P}$ 则 $\boldsymbol{P}$ 和 $\boldsymbol{M}$ 均为对称矩阵，且 $\boldsymbol{PX} = \boldsymbol{X}$，$\boldsymbol{MX} = \boldsymbol{0}$，$\boldsymbol{P}^2 = \boldsymbol{P}$，以及 $\boldsymbol{M}^2 = \boldsymbol{M}$。给定 $\hat{\theta} - \theta_0 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$，有 $\boldsymbol{e} = \boldsymbol{Y} - \boldsymbol{X}\hat{\theta} = \boldsymbol{MY} = \boldsymbol{M\varepsilon}$，$E(\boldsymbol{e} \mid \boldsymbol{X}) = \boldsymbol{M}E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0}$。因此，

$$
\begin{aligned}
\operatorname{cov}(\hat{\theta}, \boldsymbol{e} \mid \boldsymbol{X}) &= E\{[\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})][\boldsymbol{e} - E(\boldsymbol{e} \mid \boldsymbol{X})]' \mid \boldsymbol{X}\} \\
&= E\left[\left(\hat{\theta} - \theta_0\right)\boldsymbol{e}' \mid \boldsymbol{X}\right] \\
&= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{M} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{M} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}\boldsymbol{M} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M} \\
&= \boldsymbol{0}
\end{aligned}
$$

- 假设 $E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2\boldsymbol{I}$ 是保证 $\hat{\theta}$ 和 $\boldsymbol{e}$ 的相关系数为零的关键条件。

证明 (Cont.)：

- (d) 考虑 θ_0 的一个线性估计量

$$\hat{b} = \boldsymbol{C}'\boldsymbol{Y}$$

其中 $\boldsymbol{C} = C(\boldsymbol{X})$ 是可能依赖于 $\boldsymbol{X}$ 的 $n \times p$ 维矩阵。

- 若有

$$\begin{aligned} E(\hat{b} \mid \boldsymbol{X}) &= \boldsymbol{C}'\boldsymbol{X}\theta_0 + \boldsymbol{C}'E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) \\ &= \boldsymbol{C}'\boldsymbol{X}\theta_0 \\ &= \theta_0 \end{aligned}$$

则无论 θ_0 的值是多少， $\hat{b}$ 都是 θ_0 的无偏估计。

证明 (Cont.)：

- 而 $E(\hat{b} \mid \boldsymbol{X}) = \theta_0$ 成立，当且仅当以下 $\boldsymbol{C}'\boldsymbol{X} = \boldsymbol{I}$ 成立：
- 现在求 $\hat{b}$ 的条件方差-协方差。因为

$$\begin{aligned}\hat{b} &= \boldsymbol{C}'\boldsymbol{Y} \\ &= \boldsymbol{C}'(\boldsymbol{X}\theta_0 + \boldsymbol{\varepsilon}) \\ &= \boldsymbol{C}'\boldsymbol{X}\theta_0 + \boldsymbol{C}'\boldsymbol{\varepsilon} \\ &= \theta_0 + \boldsymbol{C}'\boldsymbol{\varepsilon}\end{aligned}$$

- b 的条件方差-协方差

$$\begin{aligned}\operatorname{var}(\hat{b} \mid \boldsymbol{X}) &= E\{[\hat{b} - E(\hat{b} \mid \boldsymbol{X})][\hat{b} - E(\hat{b} \mid \boldsymbol{X})]' \mid \boldsymbol{X}\} \\ &= E\left[\left(\hat{b} - \theta_0\right)\left(\hat{b} - \theta_0\right)' \mid \boldsymbol{X}\right] \\ &= E(\boldsymbol{C}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{C} \mid \boldsymbol{X}) \\ &= \boldsymbol{C}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{C} \\ &= \boldsymbol{C}'\sigma^2\boldsymbol{I}\boldsymbol{C} \\ &= \sigma^2\boldsymbol{C}'\boldsymbol{C}\end{aligned}$$

证明 (Cont.):

- 利用 $\boldsymbol{C}'\boldsymbol{X}=\boldsymbol{I}$ 和 $\boldsymbol{M}^2=\boldsymbol{M}$，有

$$\begin{aligned}
\operatorname{var}(\hat{b}\mid \boldsymbol{X})-\operatorname{var}(\hat{\theta}\mid \boldsymbol{X}) &= \sigma^2\boldsymbol{C}'\boldsymbol{C}-\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\\
&= \sigma^2[\boldsymbol{C}'\boldsymbol{C}-\boldsymbol{C}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{C}]\\
&= \sigma^2\boldsymbol{C}'[\boldsymbol{I}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\boldsymbol{C}\\
&= \sigma^2\boldsymbol{C}'\boldsymbol{M}\boldsymbol{C}\\
&= \sigma^2\boldsymbol{C}'\boldsymbol{M}\boldsymbol{M}\boldsymbol{C}\\
&= \sigma^2\boldsymbol{C}'\boldsymbol{M}'\boldsymbol{M}\boldsymbol{C}\\
&= \sigma^2(\boldsymbol{M}\boldsymbol{C})'(\boldsymbol{M}\boldsymbol{C})\\
&= \sigma^2\boldsymbol{D}'\boldsymbol{D}\\
&= \sigma^2\textstyle\sum_{i=1}^{n} D_iD_i'\\
&\sim \text{半正定 (PSD)}
\end{aligned}$$

- 这里，利用了一个基本事实：对于任意实值向量 $D_i\in\mathbb{R}^p$，方阵 $\boldsymbol{D}'\boldsymbol{D}\equiv\sum_{i=1}^{n} D_iD_i'$ 总是半正定的，这里 $\boldsymbol{D}=(D_1,D_2,\cdots,D_n)'$ 为一个 $n\times p$ 维矩阵。(问题：如何证明？)

证明 (Cont.):

- **(e)** 现在证明 $E[\boldsymbol{e}'\boldsymbol{e}/(n-p)]=\sigma^2$。因为 $\boldsymbol{e}'\boldsymbol{e}=\boldsymbol{\varepsilon}'\boldsymbol{M}^2\boldsymbol{\varepsilon}=\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}$，以及 $\mathrm{tr}(\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{BA})$，其中 $\mathrm{tr}(\cdot)$ 是迹运算 (trace) 的算符，可得

$$
\begin{aligned}
E(\boldsymbol{e}'\boldsymbol{e}\mid\boldsymbol{X}) &= E(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}\mid\boldsymbol{X})\\
&= E[\mathrm{tr}(\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon})\mid\boldsymbol{X}]\\
&= E[\mathrm{tr}(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{M})\mid\boldsymbol{X}]\\
&= \mathrm{tr}[E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\mid\boldsymbol{X})\boldsymbol{M}]\\
&= \mathrm{tr}\left(\sigma^2\boldsymbol{I}\boldsymbol{M}\right)\\
&= \sigma^2\mathrm{tr}(\boldsymbol{M})\\
&= \sigma^2(n-p)
\end{aligned}
$$

证明 (Cont.)：

- 其中，利用 $\mathrm{tr}(\boldsymbol{AB})=\mathrm{tr}(\boldsymbol{BA})$，有

$$\begin{aligned}\mathrm{tr}(\boldsymbol{M})&=\mathrm{tr}(\boldsymbol{I})-\mathrm{tr}[\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}']\\&=\mathrm{tr}(\boldsymbol{I})-\mathrm{tr}[\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]\\&=n-p\end{aligned}$$

- 因此，

$$\begin{aligned}E\left(s^2\mid\boldsymbol{X}\right)&=\frac{E\left(\boldsymbol{e}'\boldsymbol{e}|\boldsymbol{X}\right)}{n-p}\\&=\frac{\sigma^2(n-p)}{n-p}\\&=\sigma^2\end{aligned}$$

- 样本残差方差 $s^2=\boldsymbol{e}'\boldsymbol{e}/(n-p)$ 可视为第六章所研究的随机样本 $\{Y_i\}_{i=1}^n$ 的样本方差 $S_n^2=(n-1)^{-1}\sum_{i=1}^n(Y_i-\bar{Y})^2$ 的推广。

证毕。

定理 10.4 的含义

- 定理 10.4 (a) 和 (b) 意味着 $\hat{\theta}$ 的条件均方误

$$
\begin{aligned}
\mathrm{MSE}(\hat{\theta} \mid \boldsymbol{X}) &= E\left[\left(\hat{\theta} - \theta_0\right)\left(\hat{\theta} - \theta_0\right)' \mid \boldsymbol{X}\right] \\
&= \mathrm{var}(\hat{\theta} \mid \boldsymbol{X}) + \mathrm{Bias}(\hat{\theta} \mid \boldsymbol{X})\mathrm{Bias}(\hat{\theta} \mid \boldsymbol{X})' \\
&= \mathrm{var}(\hat{\theta} \mid \boldsymbol{X})
\end{aligned}
$$

- 其中偏差 $\mathrm{Bias}(\hat{\theta} \mid \boldsymbol{X}) \equiv E(\hat{\theta} \mid \boldsymbol{X}) - \theta_0 = \mathbf{0}$。均方误的大小衡量了估计量 $\hat{\theta}$ 和 θ_0 的接近程度。

定义 10.4

[有效性]：若 $\text{var}(\hat{b} \mid \boldsymbol{X}) - \text{var}(\hat{\theta} \mid \boldsymbol{X})$ 为半正定，则参数 θ_0 的无偏估计量 $\hat{\theta}$ 比其无偏估计量 $\hat{b}$ 更有效。

- 当 $\hat{\theta}$ 比 $\hat{b}$ 更有效时，对任意 $\tau \in \boldsymbol{R}^p$，有 $\tau'\tau = 1$，

$$\tau'[\text{var}(\hat{b} \mid \boldsymbol{X}) - \text{var}(\hat{\theta} \mid \boldsymbol{X})]\tau \geq 0$$

- 例如，令 $\tau = (0,\cdots,1,0,\cdots,0)'$，当第 j 个元素为 1 且其他元素为 0 时，有

$$\text{var}(\hat{b}_j) - \text{var}(\hat{\theta}_j) \geq 0, 1 \leq j \leq p$$

- 值得注意，即使存在**近似多重共线性**，在定理 10.4 假设下，OLS 估计量 $\hat{\theta}$ 仍是最优线性无偏估计量，其中，$\boldsymbol{X}'\boldsymbol{X}$ 是非奇异矩阵但其最小特征值不会随样本容量 n 的增大而增大。
- 近似多重共线性本质上是一个数据问题，若目的是估计未知参数 θ_0，则无法纠正或改善该问题。

目　录

- 为了推导 $\hat{\theta}$ 的有限样本条件下的抽样分布，现在假设 $\boldsymbol{\varepsilon}$ 服从条件正态分布，即

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N\left(\boldsymbol{0}, \sigma^2 \boldsymbol{I}\right)$$

- 由此条件正态分布假设可推出严格外生性假设（$E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \boldsymbol{0}$）和球型误差方差假设 ($E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) = \sigma^2 \boldsymbol{I}$) 。
- 事实上，在条件正态分布下，$\boldsymbol{\varepsilon}$ 的条件概率密度函数

$$f(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \frac{1}{\left(\sqrt{2\pi\sigma^2}\right)^n} \exp\left(-\frac{\boldsymbol{\varepsilon}'\boldsymbol{\varepsilon}}{2\sigma^2}\right) = f(\boldsymbol{\varepsilon})$$

不依赖于 $\boldsymbol{X}$，从而随机扰动项 $\boldsymbol{\varepsilon}$ 独立于 $\boldsymbol{X}$。
- 因此，$\boldsymbol{\varepsilon}$ 的任何条件矩均不依赖于 $\boldsymbol{X}$。

◆ 问题 10.3

$\hat{\theta}$ 的抽样分布是什么?

- 定义**权重** $C_i = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}_i$，有

$$\begin{aligned}\hat{\theta} - \theta_0 &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\sum_{i=1}^{n} X_i\varepsilon_i \\ &= \sum_{i=1}^{n} C_i\varepsilon_i\end{aligned}$$

- 因此，给定 $\boldsymbol{X}$，$\hat{\theta} - \theta_0$ 是 $\boldsymbol{\varepsilon}$ 的线性组合，当 $\boldsymbol{\varepsilon}$ 服从联合正态分布时，$\hat{\theta} - \theta_0$ 也服从正态分布。

定理 10.5

[$\hat{\theta}$ 的条件正态分布 (Conditional Normality of $\hat{\theta}$)]：假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，以及 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$。则对所有 $n > p$，

$$(\hat{\theta} - \theta_0) \mid \boldsymbol{X} \sim N[\mathbf{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

证明：

- 给定 $\boldsymbol{X}$，$\hat{\theta} - \theta_0$ 是相互独立正态分布的随机变量 $\{\varepsilon_i\}_{i=1}^{n}$ 的线性加权和，从而 $\hat{\theta} - \theta_0$ 也服从正态分布。**证毕。**

注意：

- 当存在近似多重共线性时，只要 $\boldsymbol{X}'\boldsymbol{X}$ 还是非奇异矩阵，OLS 估计量 $\hat{\theta}$ 在有限样本下仍服从条件正态分布 $N[\theta_0, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]$。

推论 10.1

[$\boldsymbol{R(\hat{\theta}-\theta_0)}$ 的条件正态分布]：假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，以及 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(0, \sigma^2 \boldsymbol{I})$。则对于任何非随机 $J \times p$ 维矩阵 R 以及所有 $n > p$，有

$$R(\hat{\theta} - \theta_0) \mid \boldsymbol{X} \sim N[\boldsymbol{0}, \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1} R']$$

证明：

- 因为给定 $\boldsymbol{X}$， $\hat{\theta}-\theta_0$ 服从正态分布。因此，线性组合 $R(\hat{\theta}-\theta_0)$ 的条件分布也服从正态分布，并且

$$E\left[R\left(\hat{\theta}-\theta_0\right)\mid \boldsymbol{X}\right]=RE\left[\left(\hat{\theta}-\theta_0\right)\mid \boldsymbol{X}\right]=R\mathbf{0}=\mathbf{0}$$

和

$$\begin{aligned}\operatorname{var}\left[R\left(\hat{\theta}-\theta_0\right)\mid \boldsymbol{X}\right]&=E\left\{R\left(\hat{\theta}-\theta_0\right)\left[R\left(\hat{\theta}-\theta_0\right)\right]'\mid \boldsymbol{X}\right\}\\&=E\left[R\left(\hat{\theta}-\theta_0\right)\left(\hat{\theta}-\theta_0\right)'R'\mid \boldsymbol{X}\right]\\&=RE\left[\left(\hat{\theta}-\theta_0\right)\left(\hat{\theta}-\theta_0\right)'\mid \boldsymbol{X}\right]R'\\&=R\operatorname{var}(\hat{\theta}\mid \boldsymbol{X})R'\\&=\sigma^2R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'\end{aligned}$$

- 从而，$R(\hat{\theta}-\theta_0)\mid \boldsymbol{X}\sim N[\mathbf{0},\sigma^2R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']$。**证毕。**

◆ 问题 10.4

在推论 10.1 中，非随机矩阵 R 起什么作用呢？为什么需要知道 $R(\hat{\theta}-\theta_0)$ 的条件分布？

- 非随机 $J \times p$ 维矩阵 R 可视为一个选择矩阵（selection matrix)。
  - ✓ 比如，如果 $R=(1,0,\cdots,0)$，则 $R(\hat{\theta}-\theta_0)=\hat{\alpha}-\alpha_0$。
- 置信区间估计和假设检验需要用到 $R(\hat{\theta}-\theta_0)$ 的抽样分布。但是，因为 $\operatorname{var}(\varepsilon_i)=\sigma^2$ 是未知的，$\operatorname{var}\left[R\left(\hat{\theta}-\theta_0\right)\middle|\boldsymbol{X}\right]=\sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'$ 也是未知的，故需要估计 σ^2。

目　录

◆ 问题 10.5

如何估计 $\mathrm{var}(\hat{\theta} \mid \boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$?

- 为了估计 σ^2，可用残差方差的估计量 $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-p)$。

引理 10.1

[正态随机变量的二次型 (Quadratic Form of Normal Random Variables)]：

假设 $m \times 1$ 维随机向量 $v \sim N(\mathbf{0}, \boldsymbol{I})$，并且 Q 是 $m \times m$ 维的非随机对称幂等矩阵，其秩为 $q \leq m$，则二次型

$$v'Qv \sim \chi_q^2$$

- 在以下应用中，$v = \boldsymbol{\varepsilon}/\sigma \sim N(\mathbf{0}, \boldsymbol{I})$，$Q = \boldsymbol{M}$。因为 $\operatorname{rank}(\boldsymbol{M}) = n - p$，所以

$$\left.\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}\right| \boldsymbol{X} \sim \chi_{n-p}^2$$

定理 10.6

[残差方差估计量 (Residual Variance Estimator)]：假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，以及 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})$。则对任意的 $n > p$，

(a) $\left.\dfrac{(\mathrm{n}-p)s^2}{\sigma^2}\right|\boldsymbol{X} = \left.\dfrac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}\right|\boldsymbol{X} \sim \chi^2_{n-p}$

(b) 给定 $\boldsymbol{X}$ 的条件下，s^2 和 $\hat{\theta}$ 是相互独立的。

证明:

- (a) 因为 $\boldsymbol{e}=\boldsymbol{M\varepsilon}$，所以

$$\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}=\frac{\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}}{\sigma^2}=\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)'\boldsymbol{M}\left(\frac{\boldsymbol{\varepsilon}}{\sigma}\right)$$

- 另外，由于 $\boldsymbol{\varepsilon}\mid\boldsymbol{X}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{I})$，且 $\boldsymbol{M}$ 为一个秩为 $n-p$ 的幂等矩阵，根据引理 10.1，有

$$\left.\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}\right|\boldsymbol{X}=\left.\frac{\boldsymbol{\varepsilon}'\boldsymbol{M}\boldsymbol{\varepsilon}}{\sigma^2}\right|\boldsymbol{X}\sim\chi^2_{n-p}$$

证明 (Cont.):

- (b) 以下证明在给定 $\boldsymbol{X}$ 的条件下，s^2 和 $\hat{\theta}$ 是相互独立的。因为 $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-p)$ 是 $\boldsymbol{e}$ 的函数，只需要证明 $\boldsymbol{e}$ 和 $\hat{\theta}$ 相互独立。
- 首先证明 $\boldsymbol{e}$ 和 $\hat{\theta}$ 服从条件联合正态分布。

$$\begin{bmatrix} \boldsymbol{e} \\ \hat{\theta} - \theta_0 \end{bmatrix} = \begin{bmatrix} \boldsymbol{M\varepsilon} \\ (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon} \end{bmatrix} = \begin{bmatrix} \boldsymbol{M} \\ (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \end{bmatrix} \boldsymbol{\varepsilon} = \boldsymbol{A\varepsilon}$$

- 其中，$(n+p)\times 1$ 维向量 $\boldsymbol{A}$ 依赖于 $\boldsymbol{X}$。
- 因为 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(0, \sigma^2\boldsymbol{I})$，线性组合

$$\boldsymbol{A}\varepsilon = \begin{bmatrix} \boldsymbol{M} \\ (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}' \end{bmatrix} \boldsymbol{\varepsilon}$$

- 在给定 $\boldsymbol{X}$ 条件下也服从正态分布。另外，从定理 10.4 (c)，有 $\mathrm{cov}(\hat{\theta}, \boldsymbol{e} \mid \boldsymbol{X}) = \boldsymbol{0}$。对联合正态分布而言，零相关等价于相互独立，因此有 $\hat{\theta}$ 和 $\boldsymbol{e}$ 相互独立。**证毕。**

定理 10.6 的含义

- 定理 10.6 是第六章定理 6.6 和定理 6.7 的推广。为了讨论定理 10.6 的含义，首先回顾 χ_q^2 分布的性质。假设 $\{Z_i\}_{i=1}^q$ 是 IID $N(0,1)$ 的随机变量，则随机变量

$$\chi^2 = \sum_{i=1}^{q} Z_i^2 \sim \chi_q^2$$

- χ_q^2 分布有两个重要性质，即： $E\left(\chi_q^2\right) = q$， $\mathrm{var}\left(\chi_q^2\right) = 2q$。
- 因此，定理 10.6 (a) 意味着

$$E\left[\left.\frac{(n-p)s^2}{\sigma^2}\right| \boldsymbol{X}\right] = n - p$$

- 从而有 $E(s^2 \mid \boldsymbol{X}) = \sigma^2$ 。

定理 10.6 的含义 (Cont.)

- 从定理 10.6 (a) 还可推出，

$$\operatorname{var}\left[\left.\frac{(n-p)s^2}{\sigma^2}\right|\boldsymbol{X}\right]=2(n-p)$$

- 因此有

$$\operatorname{var}(s^2 \mid \boldsymbol{X})=\frac{2\sigma^4}{n-p}$$

- 这样，当 $n\to\infty$ 时， s^2 的条件均方误

$$\begin{aligned}\operatorname{MSE}(s^2 \mid \boldsymbol{X}) &= E\left[\left(s^2-\sigma^2\right)^2 \mid \boldsymbol{X}\right]\\ &= \operatorname{var}(s^2 \mid \boldsymbol{X})+\left[E(s^2 \mid \boldsymbol{X})-\sigma^2\right]^2\\ &= \frac{2\sigma^4}{n-p}\\ &\to 0\end{aligned}$$

定理 10.6 (Cont.)

- ✓ 样本残差方差 $s^2 = \boldsymbol{e}'\boldsymbol{e}/(n-p)$ 是随机样本 $\{Y_i\}_{i=1}^n$ 的方差 $S_n^2 = (n-1)^{-1}\sum_{i=1}^n (Y_i - \bar{Y})^2$ 的推广。这里，估计残差样本 $\{e_i\}_{i=1}^n$ 的自由度为 $n-p$，这是因为随机样本 $\{Y_i, X_i'\}_{i=1}^n$ 有 n 个观测值，可视为有 n 个自由度。
- ✓ 当估计 σ^2 时，需要用到估计残差样本 $\{e_i\}_{i=1}^n$。这 n 个估计残差并不是线性独立的，因为它们必须满足 OLS 估计的一阶条件，即

$$\boldsymbol{X}'\boldsymbol{e} = \boldsymbol{0}$$
$$(p\times n)\times(n\times 1) = (p\times 1)$$

定理 10.6 (Cont.)

- 为了估计 p 个未知参数 θ_0，一阶条件对估计残差样本 $\{e_i\}_{i=1}^n$ 施加了 p 个约束。从而使得 $\{e_i\}_{i=1}^n$ 失去 p 个自由度，因此，$\{e_i\}_{i=1}^n$ 剩下 $n-p$ 个自由度。
- 值得注意，样本方差 S_n^2 可视为只有一个截距项的简单线性回归模型的残差方差估计量：$Y_i=\theta_0+\varepsilon_i$。

◆ 问题 10.6

为什么 $\hat{\theta}$ 和 s^2 的抽样分布比较实用?

- $\hat{\theta}$ 和 s^2 的抽样分布在关于真实模型参数 θ_0 的置信区间估计和假设检验中十分有用。

目　录

- 考虑以下线性参数假设：

$$\mathbb{H}_0: R\theta_0 = r$$
$$(J \times p)(p \times 1) = (J \times 1)$$

其中

✓ R 是一个非随机 $J \times p$ 的选择矩阵，

✓ r 是一个非随机 $J \times 1$ 维向量，

✓ J 是 p 个参数 θ_0 的限制条件个数。假设 $J \leq p$，即参数的限制条件个数不超过未知参数 θ_0 的维数。

- 需要强调，$\mathbb{H}_0$ 假设检验的前提是关于条件均值 $E\,(Y_i\,|X_i)$ 的模型正确设定。

例 10.2：[转型经济改革效果评估]

- 考虑以下扩展的生产函数

$$\ln(Y_i) = \alpha_0 + \beta_{10}\ln(L_i) + \beta_{20}\ln(K_i) + \beta_{30}AU_i + \beta_{40}PS_i + \varepsilon_i,$$
$$i = 1, \cdots, n$$

- AU_i 是一个反映企业 i 是否有自主权 (autonomy) 的虚拟变量，
- PS_i 是企业 i 与国家的利润分成 (profit sharing) 比例。

情况1：

- 如果有兴趣检验自主权 AU_i 对企业生产率是否有影响，由于 $\theta_0 = (\alpha_0, \beta_{10}, \beta_{20}, \beta_{30}, \beta_{40})'$，则原假设可写成

$$\mathbb{H}_0^a: \beta_{30} = 0$$

- 这等价于选择 $R = (0, 0, 0, 1, 0)$ 和 $r = 0$。

情况2：

- 如果有兴趣检验利润分成比例是否影响生产率，可考虑原假设

$$\mathbb{H}_0^b: \beta_{40} = 0$$

情况3：

- 如果有兴趣检验生产技术是否为规模报酬不变（constant return to scale，CRS)，则可考虑原假设

$$\mathbb{H}_0^c: \beta_{10} + \beta_{20} = 1$$

- 这等价于选择 $R = (0,1,1,0,0)$ 和 $r = 1$。

情况4：

- 如果有兴趣检验自主权和利润分成对生产率的联合影响，可设定原假设：

$$\mathbb{H}_0^d\colon \beta_{30} = \beta_{40} = 0$$

- 这等价于选择

$$R = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$r = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

例 10.3: [未来即期汇率的无偏预测]

- 考虑线性回归模型

$$S_{i+\tau} = \alpha_0 + \beta_0 F_i(\tau) + \varepsilon_{i+\tau}, \qquad i = 1, \cdots, n$$

其中

✓ $S_{i+\tau}$ 是第 $i+\tau$ 期的即期汇率,

✓ $F_i(\tau)$ 是远期汇率，即第 $i+\tau$ 期到期的汇率在第 i 期的外汇价格。

- 原假设是远期汇率 $F_i(\tau)$ 是即期汇率 $S_{i+\tau}$ 的无偏预测，即

$$E(S_{i+\tau} \mid I_i) = F_i(\tau)$$

其中 I_i 是第 i 期的信息集。

- 这在经济学和金融学称为**预期假说**（expectations hypothesis）。

例 10.3 (Cont.):

- 给定上述回归模型，这一经济假说可写成以下原假设:

$$\mathbb{H}_0^e: \alpha_0 = 0, \qquad \beta_0 = 1$$

并且 $E(\varepsilon_{i+\tau} \mid I_i) = 0$。

- 这等价于选择

$$R = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$r = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

假设检验的基本思想

- 为了检验原假设

$$\mathbb{H}_0: R\theta_0 = r，其中\ \theta_0 = (\alpha_0, \beta_0)'$$

可考虑统计量

$$R\hat{\theta} - r$$

并检验它是否显著不等于零。

假设检验的基本思想 (Cont.)

- 若原假设 $\mathbb{H}_0: R\theta_0 = r$ 成立，
  - ✓ 则当 $n \to \infty$ 时，

$$R\hat{\theta} - r = R\left(\hat{\theta} - \theta_0\right) \to \mathbf{0}$$

  因为，当 $n \to \infty$ 时， $\hat{\theta}$ 在 MSE 意义上趋近于 θ_0。

- 若备选假设 $R\theta_0 \neq r$ 成立，
  - ✓ 则当 $n \to \infty$ 时，仍然在 MSE 意义上有 $\hat{\theta} - \theta_0 \to \mathbf{0}$ (因为线性回归模型设定正确)，从而有

$$R\hat{\theta} - r = R\left(\hat{\theta} - \theta_0\right) + R\theta_0 - r$$
$$\xrightarrow{p} R\theta_0 - r \neq \mathbf{0}$$

  即 $R\hat{\theta} - r$ 将依概率收敛于非零的极限 $R\theta_0 - r$。

假设检验的基本思想 (Cont.)

- $R\hat{\theta}-r$ 在原假设 $\mathbb{H}_0$ 和备选假设 $\mathbb{H}_A$ 下的行为完全不同。因此，可通过检验 $R\hat{\theta}-r$ 和 $\mathbf{0}$ 之间是否存在显著差异来检验 $\mathbb{H}_0$ 是否成立。
  - ✓ 如果 $R\hat{\theta}-r$ 接近 $\mathbf{0}$，则没有证据拒绝原假设 $\mathbb{H}_0$。
  - ✓ 如果 $R\hat{\theta}-r$ 显著地偏离 $\mathbf{0}$，则可拒绝原假设 $\mathbb{H}_0$。
- 怎么判定 $R\hat{\theta}-r$ 显著接近/偏离 $\mathbf{0}$?
  - ✓ 用 $R\hat{\theta}-r$ 抽样分布的临界值来判断 $R\hat{\theta}-r$ 的实现值是否与 $\mathbf{0}$ 相接近，而该临界值依赖于样本容量 n 以及事先选定的显著性水平 $\alpha \in (0,1)$。

◆ 问题

在原假设 $\mathbb{H}_0$ 下 $R\hat{\theta}-r$ 的抽样分布是什么?

- 根据推论 10.1,

$$R(\hat{\theta}-\theta_0) \mid \boldsymbol{X} \sim N[\mathbf{0}, \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']$$

- 则给定 $\boldsymbol{X}$ 的条件下，有

$$\begin{aligned} R\hat{\theta}-r &= R(\hat{\theta}-\theta_0)+R\theta_0-r \\ &\sim N[R\theta_0-r, \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'] \end{aligned}$$

推论 10.2

假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，以及 $\varepsilon \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$。则当原假设 $\mathbb{H}_0$: $R\theta_0 = r$ 成立时，对于所有 $n > p$，有

$$\left(R\hat{\theta} - r\right) \mid \boldsymbol{X} \sim N[\boldsymbol{0}, \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1} R']$$

- $R\hat{\theta} - r$ 并不能作为检验 $\mathbb{H}_0$ 的一个统计量，因为其条件方差 $\mathrm{var}(R\hat{\theta} \mid \boldsymbol{X}) = \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1} R'$ 涉及未知参数 σ^2，因此无法计算 $R\hat{\theta} - r$ 抽样分布的临界值。

◆ 问题 10.7

如何构建一个可计算的检验统计量呢?

- 检验统计量的构造依赖于参数限制数目 J 的取值，现在分 $J=1$ 和 $J>1$ 两种情形进行讨论。

t- 检验 ($J = 1$)

- 在原假设 $\mathbb{H}_0$ 下，

$$\left(R\hat{\theta} - r\right) \mid \boldsymbol{X} \sim N[\boldsymbol{0}, \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']$$

- 当 $J = 1$ 时，$R\hat{\theta} - r$ 的条件方差是一个标量：

$$\operatorname{var}[(R\hat{\theta} - r) \mid \boldsymbol{X}] = \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'$$

- 因此，给定 $\boldsymbol{X}$ 的条件下，

$$\frac{R\hat{\theta} - r}{\sqrt{\operatorname{var}[(R\hat{\theta} - r) \mid \boldsymbol{X}]}} = \frac{R\hat{\theta} - r}{\sqrt{\sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$$

$$\sim N(0,1)$$

◆ 问题

$\frac{R\hat{\theta}-r}{\sqrt{\sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$ 的无条件分布是什么?

- $\frac{R\hat{\theta}-r}{\sqrt{\sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$ 的无条件分布是 $N(0,1)$。
- 但是，由于 σ^2 是未知的，不能使用比率

$$\frac{R\hat{\theta}-r}{\sqrt{\sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$$

作为检验统计量。

- 需要将 σ^2 替换为 s^2，后者是 σ^2 的一个无偏估计量。这样，获得以下可计算的检验统计量

$$T=\frac{R\hat{\theta}-r}{\sqrt{s^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$$

- 但是，这一替换，导致检验统计量 T 不再服从标准正态分布，而是服从学生 t-分布 (Student's t-distribution)，即

$$T=\frac{R\hat{\theta}-r}{\sqrt{s^2R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$$

$$=\frac{\frac{R\hat{\theta}-r}{\sqrt{\sigma^2R\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1}R'}}}{\sqrt{\frac{(n-p)s^2}{\sigma^2}/(n-p)}}$$

$$\sim\frac{N(0,1)}{\sqrt{\chi^2_{n-p}/(n-p)}}$$

$$\sim t_{n-p}$$

其中 t_{n-p} 表示自由度为 $n-p$ 的学生 t-分布。

- 这里，在给定 $\boldsymbol{X}$ 的条件下，检验统计量 T 中的分子

$$\frac{R\hat{\theta}-r}{\sqrt{\sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}} \sim N(0,1)$$

- 而统计量 T 中的分母

$$\sqrt{\frac{(n-p)s^2}{\sigma^2}/(n-p)} \sim \sqrt{\chi^2_{n-p}/(n-p)}$$

- 由于给定 $\boldsymbol{X}$，$\hat{\theta}$ 和 s^2 相互独立，统计量 T 中分子与分母也是相互独立的。
- 根据 t-分布定义，统计量 T 因此服从学生 t_{n-p}-分布。

- t_q-分布有一个重要性质，即当自由度 $q\to\infty$，$t_q\to N(0,1)$。这表明，在 $\mathbb{H}_0$ 下，当 $n\to\infty$ 时，

$$T=\frac{R\hat{\theta}-r}{\sqrt{s^2R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}\xrightarrow{d}N(0,1)$$

- 这一结论在实际应用中具有重要意义：即当样本容量 n 足够大时，t_{n-p} 和 $N(0,1)$ 的临界值几乎没有差异。因此，在大样本条件下，可使用 $N(0,1)$ 的临界值。

基于抽样分布临界值的 t- 检验判断法则

- 推导出检验统计量 T 的抽样分布之后，现在可描述当 $J=1$ 时，一个基于抽样分布临界值的检验原假设 $\mathbb{H}_0$ 的判断法则：
  - ✓ (a) 拒绝 $\mathbb{H}_0$: $R\theta_0 = r$ 的条件：在事先给定的显著性水平 $\alpha \in (0,1)$ 下，如果 $|T| > C_{t_{n-p},\frac{\alpha}{2}}$。$C_{t_{n-p},\frac{\alpha}{2}}$ 是 t_{n-p}-分布在 $\frac{\alpha}{2}$ 水平上的右侧临界值，满足

$$P\left(t_{n-p} > C_{t_{n-p},\frac{\alpha}{2}}\right) = \frac{\alpha}{2} \text{ 或等价地 } P\left(\left|t_{n-p}\right| > C_{t_{n-p},\frac{\alpha}{2}}\right) = \alpha$$

  - ✓ (b) 无法拒绝原假设 $\mathbb{H}_0$: $R\theta_0 = r$ 的条件：如果 $|T| \leq C_{t_{n-p},\frac{\alpha}{2}}$。

假设检验的两类错误

- 随机样本 $\{Y_i, X_i'\}_{i=1}^n$ 的总体信息有限，$\mathbb{H}_0$ 检验存在**两类错误。**
  - ✓ 一类是 $\mathbb{H}_0$ 为真但是被拒绝，这称为**第一类错误。**显著性水平 α 是发生第一类错误的概率。若

$$\mathrm{P}\left(|\mathrm{T}| > C_{t_{n-p,\frac{\alpha}{2}}} \middle| \mathbb{H}_0\right) = \alpha$$

则称该决策规则是**尺度**为 α 的检验。

  - ✓ 另外，概率 $P(|T| > C_{t_{n-p,\frac{\alpha}{2}}} \mid \mathbb{H}_0$为假) 称为尺度为 α 的检验的**功效函数。**若

$$\mathrm{P}\left(|\mathrm{T}| > C_{t_{n-p,\frac{\alpha}{2}}} \middle| \mathbb{H}_0\text{为假}\right) < 1$$

则存在 $\mathbb{H}_0$ 为假但未拒绝的可能性，这称为**第二类错误。**

假设检验的两类错误 (Cont.)

- 在理想情况下，应将第一类与第二类错误最小化，但对于任意给定有限样本，这是无法实现的。
- 在实际应用中，通常先预设第一类错误的水平，即所谓的显著性水平，然后最小化第二类错误。
- 常见的显著性水平 α 有 10%，5%，1%。

t- 检验统计量 T 的 P- 值

- 假设观测数据 $\boldsymbol{z}^n=\{y_i,x_i'\}_{i=1}^n$ 是随机样本 $\boldsymbol{Z}^n=\{Y_i,X_i'\}_{i=1}^n$ 的一个实现，则可计算 t-检验统计量 T 相应的实现值，即

$$T(\boldsymbol{z}^n)=\frac{R(\boldsymbol{x}'\boldsymbol{x})^{-1}\boldsymbol{x}'\mathbf{y}-r}{\sqrt{s^2R(\boldsymbol{x}'\boldsymbol{x})^{-1}R'}}$$

- 概率

$$\begin{aligned}P(\boldsymbol{z}^n)&=P(|T|>|T(\boldsymbol{z}^n)|\mid\mathbb{H}_0)\\&=P\left(\left|t_{n-p}\right|>|T(\boldsymbol{z}^n)|\right)\end{aligned}$$

称为给定数据 $\boldsymbol{Z}^n=\boldsymbol{z}^n$ 时，检验统计量 T 的 P-值，其中 t_{n-p} 代表一个自由度为 $n-p$ 的学生 t-分布随机变量。

基于 P- 值的判定法则

- (a) 如果 $P(\boldsymbol{z}^n) < \alpha$，在 α 显著性水平下**拒绝** $\mathbb{H}_0$: $R\theta_0 = r$;
- (b) 如果 $P(\boldsymbol{z}^n) \geq \alpha$，在 α 显著性水平下**无法拒绝** $\mathbb{H}_0$: $R\theta_0 = r$。

例 10.4：[转型经济改革效果评估]

- 考虑例 10.2 中扩展的生产函数模型。

$$\ln(Y_i) = \alpha_0 + \beta_{10}\ln(L_i) + \beta_{20}\ln(K_i) + \beta_{30}AU_i + \beta_{40}PS_i + \varepsilon_i,$$
$$i = 1, \cdots, n$$

- AU_i 是一个反映企业 i 是否有自主权（autonomy）的虚拟变量，
- PS_i 是企业 i 与国家的利润分成 (profit sharing) 比例。

- 首先考虑检验以下原假设

$$\mathbb{H}_0^a: \beta_{30} = 0$$

- 其中 β_{30} 是扩展生产函数回归模型中的自主权变量 AU_i 的系数，这等价于选择 $R = (0, 0, 0, 1, 0)$ 和 $r = 0$。

例 10.4 (Cont.):

- 这时，有

$$s^2R(\boldsymbol{X}'\boldsymbol{X})^{-1}R' = [s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{(4,4)}$$
$$= S_{\hat{\beta}_3}^2$$

- 它是 $\mathrm{var}\left(\hat{\beta}_3 \mid \boldsymbol{X}\right)$ 的估计量，因此 ***t*-检验统计量**为

$$T = \frac{R\hat{\theta} - r}{\sqrt{s^2R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}}$$
$$= \frac{\hat{\beta}_3}{S_{\hat{\beta}_3}}$$
$$\sim t_{n-5}$$

例 10.4 (Cont.):

- 接着，可**检验不变规模报酬假设**

$$\mathbb{H}_0^c: \beta_{10} + \beta_{20} = 1$$

- 这对应于$R = (0,1,1,0,0)$ 和 $r = 1$。此时，有

$$\begin{aligned} s^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R' &= S^2_{\hat{\beta}_1} + S^2_{\hat{\beta}_2} + 2\widehat{\mathrm{cov}}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \boldsymbol{X}\right) \\ &= [s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{(2,2)} + [s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{(3,3)} + 2[s^2(\boldsymbol{X}'\boldsymbol{X})^{-1}]_{(2,3)} \\ &= S^2_{\hat{\beta}_1+\hat{\beta}_2} \end{aligned}$$

- 它是 $\mathrm{var}(\hat{\beta}_1 + \hat{\beta}_2 \mid \boldsymbol{X})$ 的估计量。这里，$\widehat{\mathrm{cov}}\left(\hat{\beta}_1, \hat{\beta}_2 \mid \boldsymbol{X}\right)$ 是 $\hat{\beta}_1$ 和 $\hat{\beta}_2$ 之间条件协方差 $\mathrm{cov}(\hat{\beta}_1, \hat{\beta}_2 \mid \boldsymbol{X})$ 的估计量。

例 10.4 (Cont.):

- 这里的 t-检验统计量为

$$T = \frac{R\hat{\theta} - r}{\sqrt{s^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'}} = \frac{\hat{\beta}_1 + \hat{\beta}_2 - 1}{S_{\hat{\beta}_1 + \hat{\beta}_2}} \sim t_{n-5}$$

F- 检验 ($J > 1$)

◆ 问题 10.8

如果参数限制个数 $J > 1$，$R\hat{\theta} - r$ 是一个随机向量。在这种情形下，如何构造检验统计量?

引理 10.2

如果 $q \times 1$ 维随机向量 $Z \sim N(\mathbf{0}, V)$，其中 $V = \text{var}(Z)$ 是 $q \times q$ 维的对称、有限、非奇异的方差-协方差矩阵，则

$$Z'V^{-1}Z \sim \chi_q^2$$

证明：

- 因为 V 是对称和正定的，可找到一个对称和可逆的矩阵 $V^{1/2}$，使得

$$V^{1/2}V^{1/2}=V$$
$$V^{-1/2}V^{-1/2}=V^{-1}$$

- 现在，定义一个新的随机变量 $Y=V^{-1/2}Z$，则 $E(Y)=\mathbf{0}$，且

$$\begin{aligned}\operatorname{var}(Y)&=E\{[Y-E(Y)][Y-E(Y)]'\}\\&=E(YY')\\&=E\left(V^{-1/2}ZZ'V^{-1/2}\right)\\&=V^{-1/2}E(ZZ')V^{-1/2}\\&=V^{-1/2}VV^{-1/2}\\&=V^{-1/2}V^{1/2}V^{1/2}V^{-1/2}\\&=\boldsymbol{I}\end{aligned}$$

证明 (Cont.)：

- 从而，$Y=(Y_1,\cdots,Y_q)'\sim N(\mathbf{0},\boldsymbol{I})$。因此，$Y_1,\cdots,Y_q$ 是相互独立的正态分布随机变量，故有 $Y'Y=\sum_{i=1}^{q}Y_i^2\sim\chi_q^2$。**证毕。**

- 根据推论 10.2，当原假设 $\mathbb{H}_0$: $R\theta_0 = r$ 成立时，

$$(R\hat{\theta} - r) \mid \boldsymbol{X} \sim N[\mathbf{0}, \sigma^2 R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']$$

- 因此，由引理 10.2，二次型随机变量

$$\left.\frac{(R\hat{\theta} - r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta} - r)}{\sigma^2}\right| \boldsymbol{X} \sim \chi_J^2$$

- 因为 χ_J^2 分布不依赖于 $\boldsymbol{X}$，二次型随机变量的无条件分布也是 χ_J^2 分布：

$$\frac{(R\hat{\theta} - r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta} - r)}{\sigma^2} \sim \chi_J^2$$

- 由于 σ^2 是未知的，与构造 t-检验统计量 T 一样，需要将 σ^2 替换为 s^2，从而得**二次型统计量**

$$\frac{(R\hat{\theta}-r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta}-r)}{s^2}$$

- 这一替换导致二次型的分布不再是 χ^2 分布，它在除以参数限制个数 J 后，将服从自由度为 $(J,n-p)$ 的 $\mathcal{F}$- 分布。

$\mathcal{F}$ -分布

- 假设 $U \sim \chi_p^2$，$V \sim \chi_q^2$，并且 U 和 V 是相互独立的，则比率
$$\frac{U/p}{V/q} \sim \mathcal{F}_{p,q}$$
称为服从自由度 (p,q) 的 $\mathcal{F}_{p,q}$-分布。

- 二次型在除以 J 后，会服从 $\mathcal{F}_{J,n-p}$-分布。将二次型表示为

$$\frac{(R\hat{\theta}-r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta}-r)/J}{s^2} = \frac{\dfrac{(R\hat{\theta}-r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta}-r)}{\sigma^2}/J}{\dfrac{(n-p)s^2}{\sigma^2}/(n-p)}$$
$$\sim \frac{\chi_J^2/J}{\chi_{n-p}^2/(n-p)}$$
$$\sim \mathcal{F}_{J,n-p}$$

F- 检验统计量

- 其中，$\mathcal{F}_{J,n-p}$ 表示自由度为 J 和 $n-p$ 的 $\mathcal{F}$-分布。
  - ✓ 分子

$$\frac{(R\hat{\theta}-r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta}-r)}{\sigma^2} \sim \chi_J^2$$

  - ✓ 分母

$$\frac{(n-p)s^2}{\sigma^2} \sim \chi_{n-p}^2$$

  - ✓ 分子分母相互独立。
- 根据 $\mathcal{F}$-分布的定义，即可得 $\mathcal{F}_{J,n-p}$-分布。
- 根据以上讨论，可定义以下 F-检验统计量

$$F \equiv \frac{(R\hat{\theta}-r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta}-r)/J}{s^2}$$

定理 10.7

假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，以及 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(0, \sigma^2 \boldsymbol{I})$。则当原假设 $\mathbb{H}_0$: $R\theta_0 = r$ 成立时，对任意的 $n > p$，有

$$F \sim \mathcal{F}_{J,n-p}$$

- 事实上，F-检验也适用 $J = 1$。$\mathcal{F}$-分布有一个重要的性质，

$$\mathcal{F}_{1,q} \sim t_q^2$$

- 这表明，当 $J = 1$ 时，使用 t-统计检验或 F-统计检验将得到完全相同的结论。

定理 10.8

假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，令 $SSR_u = \boldsymbol{e}'\boldsymbol{e}$ 是以下无约束回归模型的残差平方和

$$\boldsymbol{Y} = \boldsymbol{X}\theta + \boldsymbol{\varepsilon}$$

另外，令 $SSR_r = \tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}$ 是以下有约束回归模型的残差平方和

$$\boldsymbol{Y} = \boldsymbol{X}\theta + \boldsymbol{\varepsilon}$$

其约束条件为

$$R\theta = r$$

其中 $\tilde{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\widetilde{\boldsymbol{\theta}}$， $\tilde{\theta}$ 是有约束回归模型的 OLS 估计量。则 F-检验统计量可写为

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

证明：

- $\tilde{\theta}$ 是在原假设 $\mathbb{H}_0$ 成立时有约束线性回归模型的 OLS 估计量，即

$$\tilde{\theta} = \arg\min_{\theta \in \mathbb{R}^p} (\boldsymbol{Y} - \boldsymbol{X}\theta)'(\boldsymbol{Y} - \boldsymbol{X}\theta)$$

- 其中约束条件为 $R\theta = r$。首先，构建拉格朗日函数，

$$L(\theta, \lambda) = (\boldsymbol{Y} - \boldsymbol{X}\theta)'(\boldsymbol{Y} - \boldsymbol{X}\theta) + 2\lambda'(r - R\theta)$$

- 其中 λ 是一个 $J \times 1$ 的向量，称为拉格朗日乘子向量。
- 拉格朗日函数的一阶条件为：

$$\frac{\partial L(\tilde{\theta}, \tilde{\lambda})}{\partial \theta} = -2\boldsymbol{X}'(\boldsymbol{Y} - \boldsymbol{X}\tilde{\theta}) - 2R'\tilde{\lambda} = \boldsymbol{0}$$

$$\frac{\partial L(\tilde{\theta}, \tilde{\lambda})}{\partial \lambda} = 2(r - R\tilde{\theta}) = \boldsymbol{0}$$

证明 (Cont.)：

- 另一方面，无约束回归模型的 OLS 估计量是 $\hat{\theta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$，结合上述一阶条件的第一个方程，可得

$$-(\hat{\theta}-\tilde{\theta})=(\boldsymbol{X}'\boldsymbol{X})^{-1}R'\tilde{\lambda}$$

$$R(\boldsymbol{X}'\boldsymbol{X})^{-1}R'\tilde{\lambda}=-R(\hat{\theta}-\tilde{\theta})$$

- 因此，拉格朗日乘子为

$$\begin{aligned}\tilde{\lambda}&=-[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}R(\hat{\theta}-\tilde{\theta})\\&=-[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta}-r)\end{aligned}$$

这里使用了约束条件 $R\tilde{\theta}=r$。

- 从 $\tilde{\lambda}$ 的表达式可看出，$\tilde{\lambda}$ 的大小揭示了 $R\hat{\theta}-r$ 与 $\mathbf{0}$ 之间差距的大小。

证明 (Cont.):

- 现在，将 $\tilde{\lambda}$ 表达式代入 $\hat{\theta}-\tilde{\theta}$ 的表达式，可得

$$\hat{\theta}-\tilde{\theta}=(\boldsymbol{X}'\boldsymbol{X})^{-1}R'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}\left(R\hat{\theta}-r\right)$$

- 根据定义，有约束的回归模型的估计残差

$$\begin{aligned}\tilde{\boldsymbol{e}}&=\boldsymbol{Y}-\boldsymbol{X}\tilde{\theta}\\&=\boldsymbol{Y}-\boldsymbol{X}\hat{\theta}+\boldsymbol{X}\left(\hat{\theta}-\tilde{\theta}\right)\\&=\boldsymbol{e}+\boldsymbol{X}\left(\hat{\theta}-\tilde{\theta}\right)\end{aligned}$$

- 故有

$$\begin{aligned}\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}&=\boldsymbol{e}'\boldsymbol{e}+\left(\hat{\theta}-\tilde{\theta}\right)'\boldsymbol{X}'\boldsymbol{X}\left(\hat{\theta}-\tilde{\theta}\right)\\&=\boldsymbol{e}'\boldsymbol{e}+\left(R\hat{\theta}-r\right)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}\left(R\hat{\theta}-r\right)\end{aligned}$$

- 因此

$$\left(R\hat{\theta}-r\right)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}\left(R\hat{\theta}-r\right)=\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}-\boldsymbol{e}'\boldsymbol{e}$$

证明 (Cont.)：

- 由 F – 检验统计量的定义，得

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

$$=\frac{(\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}-\boldsymbol{e}'\boldsymbol{e})/J}{\boldsymbol{e}'\boldsymbol{e}/(n-p)}$$

证毕。

- 若原假设 $\mathbb{H}_0$ 为真，则有约束模型的残差平方和 SSR_r 几乎等于无约束模型的残差平方和，只受抽样变化导致的差异所影响。
- 若 SSR_r 显著大于 SSR_u，则存在证据拒绝 $\mathbb{H}_0$，而 SSR_r 与 SSR_u 相差多大才算足够大则取决于 $\mathcal{F}_{J,n-p}$-分布的临界值。

定理 10.9

[沃尔德检验]： 假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，以及 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$。则当原假设 $\mathbb{H}_0$: $R\theta_0 = r$ 成立且 $n \to \infty$ 时，沃尔德检验统计量

$$W = \frac{(R\hat{\theta} - r)'[R(\boldsymbol{X}'\boldsymbol{X})^{-1}R']^{-1}(R\hat{\theta} - r)}{s^2} = JF \xrightarrow{d} \chi_J^2$$

- 这里定义的沃尔德检验统计量 W 与 F-检验统计量 F 只相差一个比例常数 J，这是因为假设条件同方差成立。
- 如果存在条件异方差，仍然可以定义稳健沃尔德检验统计量，但是 $W = JF$ 这一关系将不再成立，更多讨论参见洪永淼(2011，第四章)。

目　录

10.8.1 检验所有解释变量的联合显著性

- 考虑以下线性回归模型

$$\begin{aligned} Y_i &= X_i'\theta_0 + \varepsilon_i \\ &= \alpha_0 + \sum_{j=1}^{k} \beta_{j0} X_{ji} + \varepsilon_i, \qquad i = 1, \cdots, n \end{aligned}$$

- 这里的目的是检验除截距 α_0 外，所有自变量 $X_{1i}, \cdots, X_{ki}$ 的联合影响是否为零。
  - ✓ $\mathbb{H}_0: \beta_{10} = \cdots = \beta_{k0} = 0$
  - ✓ $\mathbb{H}_A$:至少存在一个 $j \in \{1, \cdots, k\}, \beta_{j0} \neq 0$
- 可使用 F-检验 $F \sim \mathcal{F}_{k,n-(k+1)}$。

- $\mathbb{H}_0: \beta_{10} = \cdots = \beta_{k0} = 0$ 成立时的约束回归模型可简化为

$$Y_i = \alpha_0 + \varepsilon_i,\ i = 1, \cdots, n$$

- 这个约束模型的 OLS 估计量为 $\tilde{\theta} = (\bar{Y}, 0, \cdots, 0)'$，这里 $\bar{Y}$ 为 $\{Y_t\}_{i=1}^{n}$ 的样本均值。从而，

$$\tilde{\boldsymbol{e}} = \boldsymbol{Y} - \boldsymbol{X}\tilde{\theta} = \boldsymbol{Y} - \bar{Y}\boldsymbol{l}$$

- 这里 $\boldsymbol{l} = (1, \cdots, 1)'$ 为 $n \times 1$ 维向量。因此，有

$$\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}} = (\boldsymbol{Y} - \bar{Y}\boldsymbol{l})'(\boldsymbol{Y} - \bar{Y}\boldsymbol{l})$$

- 根据 R^2 的定义：

$$R^2 = 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{(\boldsymbol{Y} - \bar{Y}\boldsymbol{l})'(\boldsymbol{Y} - \bar{Y}\boldsymbol{l})}$$

$$= 1 - \frac{\boldsymbol{e}'\boldsymbol{e}}{\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}}$$

- 可得

$$F=\frac{(\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}-\boldsymbol{e}'\boldsymbol{e})/k}{\boldsymbol{e}'\,\boldsymbol{e}/(n-k-1)}$$

$$=\frac{\left(1-\frac{\boldsymbol{e}'\boldsymbol{e}}{\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}}\right)/k}{\frac{\boldsymbol{e}'\boldsymbol{e}}{\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}}}/(n-k-1)}$$

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

- 因此，为了检验上述联合假设 $\mathbb{H}_0{:}\beta_{10}=\cdots=\beta_{k0}=0$，只需估计无约束模型并获得其 R^2。需要强调，这一公式仅在检验除截距项外所有自变量系数都为零的原假设时才能适用。

例 10.5：[有效市场假说 (Efficient Market Hypothesis, EMH)]

- 假设
  - ✓ Y_i 是第 i 期的某一资产收益率，
  - ✓ I_{i-1} 是第 $i-1$ 期的历史信息集。
- 有效市场假说的经典版本可表述如下：

$$E(Y_i \mid I_{i-1}) = E(Y_i)$$

- 为了检验资产收益率是否可利用其历史信息进行预测，设定以下线性回归模型：

$$Y_i = X_i'\theta_0 + \varepsilon_i$$

其中 $X_i = (1, Y_{i-1}, \cdots, Y_{i-k})'$，这是一个 k 阶自回归模型。

例 10.5 (Cont.):

- 在有效市场假设下，有 $\mathbb{H}_0: \beta_{10} = \cdots = \beta_{k0} = 0$
- 如果备选假设 $\mathbb{H}_A$: 至少存在某个 $j \in \{1, \cdots, k\}$, $\beta_{j0} \neq 0$ 成立，则可利用资产收益率的历史信息预测未来的资产收益率。
- **注意：**有这样一种可能性，即 Y_i 与 $\{Y_{i-j},\ j = 1, \cdots, k\}$ 不存在线性关系，但却存在非线性关系。因此，使用 F-检验，如果没有拒绝 $\mathbb{H}_0$，只能称没有发现拒绝有效市场假说的证据，不能得到有效市场假说成立的结论。

例 10.6：[消费函数和财富效应]

- 令
  - ✓ Y_i 为消费，
  - ✓ X_{1i} 为劳动力收入，
  - ✓ X_{2i} 为流动性资产财富。
- 假设使用某一数据，得到以下线性回归模型 OLS 估计结果：

$$\underset{[1.77]}{Y_i = 33.88} - \underset{[-0.74]}{26.00}X_{1i} + \underset{[0.77]}{6.71}X_{2i} + e_i, \qquad R^2 = 0.742, \qquad n = 25$$

其中方括号中的数字为 t-检验统计量。

例 10.6 (Cont.):

- 假设想考察劳动收入或流动性资产财富是否对消费有影响。劳动收入与财富的单独 t-检验统计量在 5% 显著性水平上均不显著。由于可能存在近似多重共线性，也必须检验劳动收入与财富是否不具有联合显著性。为此，计算 F-检验统计量，

$$F=\frac{R^2/2}{(1-R^2)/(n-3)}$$
$$=(0.742/2)/[(1-0.742)/(25-3)]$$
$$\approx 31.636$$

- 与分布 $\mathcal{F}_{k,n-(k+1)}=\mathcal{F}_{2,22}$ 在 5% 显著性水平的临界值 4.38 相比较，可在 5% 显著性水平上拒绝收入和流动性资产对消费均没有影响的联合原假设。

10.8.2 检验遗漏变量

- 假设 $X_i = \left(X_i^{(1)\prime}, X_i^{(2)\prime}\right)'$，其中，$X_i^{(1)}$ 是 $(k+1)\times 1$ 维向量，$X_i^{(2)}$ 是 $q \times 1$ 维向量。
  - ✓ 如果 $E(Y_i \mid X_i) = E\left(Y_i \mid X_i^{(1)}\right)$，则 $X_i^{(2)}$ 对 Y_i 的条件均值没有解释力，
  - ✓ 如果 $E(Y_i \mid X_i) \neq E\left(Y_i \mid X_i^{(1)}\right)$，则说明 $X_i^{(2)}$ 对 Y_i 的条件均值有解释力。
- 当 $X_i^{(2)}$ 对 Y_i 有解释力，但却没有包含在回归方程中时，则称 $X_i^{(2)}$ 为遗漏变量。

◆ 问题 10.11

如何在线性回归模型的框架下检验 $X_i^{(2)}$ 是否为遗漏变量?

- 这里有约束的回归模型为

$$Y_i = \alpha_0 + \beta_{10}X_{1i} + \cdots + \beta_{k0}X_{ki} + \varepsilon_i, \qquad i = 1, \cdots, n$$

- 假设 $X_i^{(2)} = (Z_{1i}, \cdots, Z_{qi})'$，则无约束回归模型为

$$Y_i = \alpha_0 + \beta_{10}X_{1i} + \cdots + \beta_{k0}X_{ki} + \gamma_{10}Z_{1i} + \cdots + \gamma_{q0}Z_{qi} + \varepsilon_i, \quad i = 1, \cdots, n$$

- 原假设是新增变量对 Y_i 没有影响。可表述为以下参数假设:

$$\mathbb{H}_0: \gamma_{10} = \gamma_{20} = \cdots = \gamma_{q0} = 0$$

- 备选假设 $\mathbb{H}_A$：至少有一个新增解释变量的系数不为零。
- ***F*-检验统计量**：$F = \dfrac{(\tilde{\boldsymbol{e}}'\tilde{\boldsymbol{e}} - \boldsymbol{e}'\boldsymbol{e})/q}{\boldsymbol{e}'\boldsymbol{e}/(n-k-q-1)} \sim \mathcal{F}_{q,n-(k+q+1)}$

例 10.7：[转型经济改革效果评估]

- 考虑以下扩展生产函数 (参见本章例 10.2)

$$\ln(Y_i) = \alpha_0 + \beta_{10}\ln(L_i) + \beta_{20}\ln(K_i) + \beta_{30}AU_i + \beta_{40}PS_i + \beta_{50}CM_i + \varepsilon_i,$$

其中

  - ✓ AU_i 是自主权 (autonomy) 虚拟变量，
  - ✓ PS_i 是利润分成 (profit sharing) 比例，
  - ✓ CM_i 是经理更换 (change of manager) 虚拟变量。

- 原假设是这三项改革均没有效果。即

$$\mathbb{H}_0: \beta_{30} = \beta_{40} = \beta_{50} = 0$$

- 可用 F-检验。在原假设 $\mathbb{H}_0$ 下，$F \sim \mathcal{F}_{3,n-6}$。

例 10.8：[格兰杰因果关系 (Granger Causality) 检验]

- 考虑二元时间序列 $\{Y_i, X_i\}$，其中
  - ✓ i 表示时间，
  - ✓ $I_{i-1}^{(Y)}$ 是由 $\{Y_{i-1}, Y_{i-2}, \cdots\}$ 生成的 σ-域，
  - ✓ $I_{i-1}^{(X)}$ 是由 $\{X_{i-1}, X_{i-2}, \cdots\}$ 生成的 σ-域。
- 一个例子是 Y_i 表示 GDP 增长率，X_i 表示货币供应增长率。若

$$E\left(Y_i \mid I_{i-1}^{(Y)}, I_{i-1}^{(X)}\right) = E\left(Y_i \mid I_{i-1}^{(Y)}\right)$$

则称 X_i 不是 Y_i 的格兰杰原因。

- 换言之，在给定 $I_{i-1}^{(Y)}$ 的条件下，X_i 的任何滞后变量对 Y_i 的条件均值都没有影响。

例 10.8 (Cont.):

- 格兰杰因果关系是定义在预测能力上的，而不是真正的经济因果关系。从计量经济学的观点看，它是时间序列回归动态模型是否存在遗漏变量的一种检验。

◆ 问题 10.12

如何进行格兰杰因果关系检验呢?

- Granger（1969）最早提出用 F- 检验来检验格兰杰因果关系。考虑以下线性回归模型：

$$Y_i = \alpha_0 + \beta_{10}Y_{i-1} + \cdots + \beta_{p0}Y_{i-p} + \gamma_{10}X_{i-1} + \cdots + \gamma_{q0}X_{i-q} + \varepsilon_i$$

- 如果不存在格兰杰原因，有

$$\mathbb{H}_0: \gamma_{10} = \cdots = \gamma_{q0} = 0$$

- 根据经典线性回归理论，F-检验统计量

$$F \sim \mathcal{F}_{q,n-(p+q+1)}.$$

- 需要指出，经典线性回归理论其实并不适用，因为这里的线性回归模型是一个动态回归模型，不满足严格外生性条件 $(E(\varepsilon_i \mid \boldsymbol{X}) = \mathbf{0})$。
- 但是，渐近理论可证明，如果满足条件同方差假设，当原假设 $\mathbb{H}_0 : \gamma_{10} = \cdots = \gamma_{q0} = 0$ 成立时，即使是线性动态回归模型，当 $n \to \infty$，仍有 $qF \xrightarrow{d} \chi_q^2$。
- 详细讨论参见洪永淼 (2011，第五章)。

例 10.9：[检验结构变化]

- 考虑以下双变量回归模型：$Y_i = \alpha_0 + \beta_0 X_{1i} + \varepsilon_i$，$i = 1, \cdots, n$
  - ✓ 其中 i 代表时间，假设 $\{X_{1i}\}$ 和 $\{\varepsilon_i\}$ 相互独立。
- 假设在时间点 $i = i_0$，可能发生了突变性结构性变化 (structural break)。因此，考虑以下扩展的回归模型：

$$
\begin{aligned}
Y_i &= (\alpha_0 + \alpha_{10} D_i) + (\beta_0 + \beta_{10} D_i) X_{1i} + \varepsilon_i \\
&= \alpha_0 + \beta_0 X_{1i} + \alpha_{10} D_i + \beta_{10} (D_i X_{1i}) + \varepsilon_i, i = 1, \cdots, n
\end{aligned}
$$

- 其中，变量 D_i 称为结构变化虚拟变量，刻画结构变化前后不同的时期。
  - ✓ 如果 $i > i_0$，$D_i = 1$；
  - ✓ 如果 $i \le i_0$，$D_i = 0$。

例 10.9 (Cont.)

- 当没有结构变化时，以下参数原假设成立 $\mathbb{H}_0: \alpha_{10} = \beta_{10} = 0$
- 如果备选假设 $\mathbb{H}_A: \alpha_{10} \neq 0$ 或 $\beta_{10} \neq 0$ 成立，则条件均值 $E(Y_i \mid X_i)$ 存在着结构变化。
- F - 检验统计量 $F \sim \mathcal{F}_{2,n-4}$
  - ✓ 这一检验 (Chow’s test) 最早是由 Chow (1960) 提出的。

例 10.10：[检验固定规模报酬]

- 考虑以下扩展生产函数 (参见例 10.7)：

$$\ln(Y_i) = \alpha_0 + \beta_{10}\ln(L_i) + \beta_{20}\ln(K_i) + \beta_{30}AU_i + \beta_{40}PS_i + \beta_{50}CM_i + \varepsilon_i$$

  - ✓ CRS 等价于原假设 $\mathbb{H}_0: \beta_{10} + \beta_{20} = 1$
  - ✓ 备选假设为 $\mathbb{H}_A: \beta_{10} + \beta_{20} \neq 1$

- 在原假设 $\mathbb{H}_0$: $\beta_{10} + \beta_{20} = 1$ 下的有约束的回归模型为

$$\ln(Y_i) = \alpha_0 + \beta_{10}\ln(L_i) + (1 - \beta_{10})\ln(K_i) + \beta_{30}AU_i + \beta_{40}PS_i + \beta_{50}CM_i + \varepsilon_i$$

  这等价于以下回归模型

$$\ln(Y_i/K_i) = \alpha_0 + \beta_{10}\ln(L_i/K_i) + \beta_{30}AU_i + \beta_{40}PS_i + \beta_{50}CM_i + \varepsilon_i$$

- 相应的 F- 检验统计量：$F \sim \mathcal{F}_{1,n-6}$。这里由于原假设只有一个约束，$t$- 检验和 F- 检验均可用于检验固定规模收益假说。

例 10.11：[工资决定机制]

- 在时间序列下考察工资函数

$$W_i = a_0 + \beta_{10}P_i + \beta_{20}P_{i-1} + \beta_{30}U_i + \beta_{40}V_i + \beta_{50}W_{i-1} + \varepsilon_i,\quad i = 1,\cdots,n$$

- 其中，
  - ✓ i 表示时间，
  - ✓ W_i 表示工资，
  - ✓ P_i 表示价格，
  - ✓ U_i 表示失业，
  - ✓ V_i 表示空缺岗位数。
- 检验原假设 $\mathbb{H}_0: \beta_{10} + \beta_{20} = 0, \beta_{30} + \beta_{40} = 0,$ 且 $\beta_{50} = 1$

例 10.11 (Cont.):

- 原假设 $\mathbb{H}_0$ 提供了一个很好的经济解释，从原假设可得以下有约束的工资等式

$$\Delta W_i = a_0 + \beta_{10}\Delta P_i + \beta_{40} D_i + \varepsilon_i$$

- 其中，
  - ✓ $\Delta W_i = W_i - W_{i-1}$ 是工资增长率，
  - ✓ $\Delta P_i = P_i - P_{i-1}$ 是通货膨胀率，
  - ✓ $D_i = V_i - U_i$ 是就业市场岗位供应过剩指数。
- 因此，原假设 $\mathbb{H}_0$ 表示工资增长取决于通货膨胀率和劳动力供应过剩。
- $\mathbb{H}_0$ 的 F-检验统计量： $F \sim \mathcal{F}_{3,n-6}$

目　录

经典线性回归模型的关键假设

- 线性：$Y_i = X_i'\theta + \varepsilon_i, \quad i = 1, \cdots, n$
- 条件正态分布： $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$
  - ✓ 严格外生：$E(\boldsymbol{\varepsilon} \mid \mathbf{X}) = \mathbf{0}$;
  - ✓ 条件同方差：$E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \mathbf{X}) = \sigma^2 I$;
- $\lambda_{min}(\mathbf{X}'\boldsymbol{X}) > 0$;
- 条件不相关: $\text{cov}(\varepsilon_t, \varepsilon_s \mid \mathbf{X}) = 0 \text{ for } t \neq s$.

问题

经典线性回归模型依赖于关键假设 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{I})$。这里除了条件正态分布外，还包含不存在条件异方差和条件自相关。如果某些假设 (如条件同方差或条件不相关) 不成立，会出现什么问题?

条件正态分布假设的放宽

由条件正态分布假设 ($\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$) 可推导出 OLS 估计量 $\hat{\theta}$ 及其相关统计量的有限样本分布，但这一假设对很多经济金融数据并不合适。

- 现在假设以下更一般的条件：

$$\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V})$$

- 其中 $0 < \sigma^2 < \infty$ 是未知的，但 $\boldsymbol{V} = V(\boldsymbol{X})$ 是一个已知的 $n \times n$ 维对称有限正定矩阵。

条件异方差或条件自相关

- 这里 $\boldsymbol{\varepsilon}$ 的条件分布仍为正态分布，且严格外生性条件成立，但是存在**条件异方差或条件自相关**，因为

$$\begin{aligned}\operatorname{var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) &= E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X}) \\ &= \sigma^2 \boldsymbol{V} \\ &= \sigma^2 V(\boldsymbol{X})\end{aligned}$$

- $\operatorname{var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X})$ 仅包含一个未知常数 σ^2，但它允许存在已知形式的条件异方差 $V(\boldsymbol{X})$。
- $\boldsymbol{V} = V(\boldsymbol{X})$ 有可能不是对角阵，即对 $i \neq j, \operatorname{cov}\left(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}\right)$ 可能不为零，因此，允许存在已知形式的条件异方差与条件自相关。
  - ✓ 若 i 表示**时间**，则存在**序列相关**。
  - ✓ 若 i 表示**截面单位**，则存在**空间相关**（spatial dependence）。

◆ 问题

在上述更一般的条件正态分布假设下，OLS 估计量 $\hat{\theta}$ 的统计性质是怎样的?

定理 10.10

假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，且 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{V})$，其中 $\boldsymbol{V} \equiv V(\boldsymbol{X})$ 为一个已知的 $n \times n$ 维有界对称非奇异矩阵。则对所有 $n > p$，

(a) [无偏性 (unbiasedness)] $E(\hat{\theta} \mid \boldsymbol{X}) = \theta_0$

(b) [方差 (variance)]

$$\mathrm{var}(\hat{\theta} \mid \boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \neq \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$$

(c) [正态分布 (normality)]

$$(\hat{\theta} - \theta_0) \mid \boldsymbol{X} \sim N[\boldsymbol{0}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

(d) [相关性 (correlatedness)] $\mathrm{cov}(\hat{\theta}, \boldsymbol{e} \mid \boldsymbol{X}) \neq \boldsymbol{0}$

证明:

- (a) 由 $\hat{\theta} - \theta_0 = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$，得

$$
\begin{aligned}
E\left[\left(\hat{\theta}-\theta_0\right) \mid \boldsymbol{X}\right] &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\mathbf{0} \\
&= \mathbf{0}
\end{aligned}
$$

- (b)

$$
\begin{aligned}
\operatorname{var}(\hat{\theta} \mid \boldsymbol{X}) &= E\left\{[\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})][\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})]' \mid \boldsymbol{X}\right\} \\
&= E\left[\left(\hat{\theta}-\theta_0\right)\left(\hat{\theta}-\theta_0\right)' \mid \boldsymbol{X}\right] \\
&= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}
\end{aligned}
$$

- 由于 $\boldsymbol{V} \neq \boldsymbol{I}$，无法进一步简化上述表达式。

证明 (Cont.):

- (c) 在给定 $\boldsymbol{X}$ 的条件下，$\hat{\theta}-\theta_0$ 仍是 $\boldsymbol{\varepsilon}$ 的线性组合：

$$\hat{\theta}-\theta_0=(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}$$
$$=\sum_{i=1}^{n}C_i\varepsilon_i$$

其中权重向量 $C_i=(\boldsymbol{X}'\boldsymbol{X})^{-1}X_i$。

- 因此，在假设 $\boldsymbol{\varepsilon}\mid\boldsymbol{X}\sim N(\boldsymbol{0},\sigma^2\boldsymbol{V})$ 条件下，$\hat{\theta}-\theta_0$ 仍服从条件正态分布，即给定 $\boldsymbol{X}$，

$$\hat{\theta}-\theta_0\sim N[\boldsymbol{0},\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}]$$

证明 (Cont.):

- (d) 因为 $\boldsymbol{X}'\boldsymbol{V}\boldsymbol{M} \neq 0$，故

$$\begin{aligned}\operatorname{cov}(\hat{\theta}, \boldsymbol{e} \mid \boldsymbol{X}) &= E\{[\hat{\theta} - E(\hat{\theta} \mid \boldsymbol{X})][\boldsymbol{e} - E(\boldsymbol{e} \mid \boldsymbol{X})]' \mid \boldsymbol{X}\} \\ &= E\left[\left(\hat{\theta} - \theta_0\right)\boldsymbol{e}' \mid \boldsymbol{X}\right] \\ &= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}'\boldsymbol{M} \mid \boldsymbol{X}] \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}' \mid \boldsymbol{X})\boldsymbol{M} \\ &= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{M} \\ &\neq \boldsymbol{0}\end{aligned}$$

- 可以看出，由于随机扰动项序列 $\{\varepsilon_i\}$ 存在条件异方差 ($\operatorname{var}(\varepsilon_i \mid \boldsymbol{X}) \neq \sigma^2$) 或者条件自相关 ($\operatorname{cov}(\varepsilon_i, \varepsilon_j \mid \boldsymbol{X}) \neq 0$)，导致了 $\hat{\theta}$ 和 $\boldsymbol{e}$ 之间存在相关性。

证毕。

定理 10.10 的含义

- 在条件正态分布 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{V})$ 假设下，OLS 估计量 $\hat{\theta}$ 仍然是**无偏**的。
- 但是，$\hat{\theta}$ 的**方差**不再具有简单的形式 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$，而是比较复杂的 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$。因此，建立在简单方差形式 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 基础上的经典 t-检验和 F-检验均无效。
- 定理 10.10 结论 (d) 还表明，即使使用正确方差 $\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ 的一致估计量，并用它来构造检验统计量，仍然无法获得有限样本条件下的学生 t-分布和 F-分布。因为，由于给定 $\boldsymbol{X}$，$\hat{\theta}$ 和 $\boldsymbol{e}$ 存在相关性，t-检验统计量和 F-检验统计量定义中的分子和分母不再相互独立。

引理 10.3

[乔里斯基分解（Cholesky factorization)]：对任意 $n \times n$ 维对称有限正定矩阵 $\boldsymbol{V}$，总可分解为

$$\boldsymbol{V}^{-1} = \boldsymbol{C}'\boldsymbol{C}$$
$$\boldsymbol{V} = \boldsymbol{C}^{-1}(\boldsymbol{C}')^{-1}$$

其中，$\boldsymbol{C}$ 是一个 $n \times n$ 维的非奇异矩阵。

- 这里，$\boldsymbol{C}$ 可能是非对称的矩阵。

广义最小二乘 (GLS) 估计量

- 回顾线性回归模型

$$\boldsymbol{Y} = \boldsymbol{X}\theta_0 + \boldsymbol{\varepsilon}$$

- 如果方程两边同时左乘 $\boldsymbol{C}$，可得到以下变换回归模型

$$\boldsymbol{CY} = (\boldsymbol{CX})\theta_0 + \boldsymbol{C\varepsilon}$$

或

$$\boldsymbol{Y}^* = \boldsymbol{X}^*\theta_0 + \boldsymbol{\varepsilon}^*$$

其中 $\boldsymbol{Y}^* = \boldsymbol{CY}$，$\boldsymbol{X}^* = \boldsymbol{CX}$，$\boldsymbol{\varepsilon}^* = \boldsymbol{C\varepsilon}$。

- 变换后的线性回归模型的 OLS 估计量

$$\begin{aligned}\hat{\theta}^* &= (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{Y}^* \\ &= (\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{CX})^{-1}(\boldsymbol{X}'\boldsymbol{C}'\boldsymbol{CY}) \\ &= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y}\end{aligned}$$

- 称为广义最小二乘 (GLS) 估计量。

◆ 问题

GLS 估计量的概率统计性质是怎样的?

- 观察到

$$
\begin{aligned}
E(\boldsymbol{\varepsilon}^* \mid \boldsymbol{X}) &= E(\boldsymbol{C\varepsilon} \mid \boldsymbol{X}) \\
&= \boldsymbol{C}E(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) \\
&= \boldsymbol{C0} \\
&= \boldsymbol{0}
\end{aligned}
$$

- 并且

$$
\begin{aligned}
\operatorname{var}(\boldsymbol{\varepsilon}^* \mid \boldsymbol{X}) &= E(\boldsymbol{\varepsilon}^* \boldsymbol{\varepsilon}^{*\prime} \mid \boldsymbol{X}) \\
&= E(\boldsymbol{C\varepsilon\varepsilon}'\boldsymbol{C}' \mid \boldsymbol{X}) \\
&= \boldsymbol{C}E(\boldsymbol{\varepsilon\varepsilon}' \mid \boldsymbol{X})\boldsymbol{C}' \\
&= \sigma^2 \boldsymbol{CVC}' \\
&= \sigma^2 \boldsymbol{C}[\boldsymbol{C}^{-1}(\boldsymbol{C}')^{-1}]\boldsymbol{C}' \\
&= \sigma^2 \boldsymbol{I}
\end{aligned}
$$

- 由假设 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\mathbf{0}, \sigma^2 \boldsymbol{V})$，可得

$$\boldsymbol{\varepsilon}^* \mid \boldsymbol{X} \sim N\left(\mathbf{0}, \sigma^2 \boldsymbol{I}\right)$$

- 上述变换使新随机扰动项 $\boldsymbol{\varepsilon}^*$ 变成具有**条件同方差**和**不存在条件自相关**的扰动项，且**服从条件正态分布**，因此它满足高斯-马尔可夫定理的条件。

例 10.12：[消除异方差]

- 假设

$$\boldsymbol{V}=\begin{bmatrix}\sigma_1^2 & 0 & \cdots & 0\\ 0 & \sigma_2^2 & \cdots & 0\\ \vdots & \vdots & \vdots & 0\\ 0 & \cdots & \cdots & \sigma_n^2\end{bmatrix}$$

- 则

$$\boldsymbol{C}=\begin{bmatrix}\sigma_1^{-1} & 0 & \cdots & 0\\ 0 & \sigma_2^{-1} & \cdots & 0\\ \vdots & \vdots & \vdots & 0\\ 0 & \cdots & \cdots & \sigma_n^{-1}\end{bmatrix}$$

其中 $\sigma_i^2=\sigma_i^2(\boldsymbol{X})$，$i=1,\cdots,n$，并且

$$\boldsymbol{\varepsilon}^*=\boldsymbol{C}\boldsymbol{\varepsilon}=\left(\frac{\varepsilon_1}{\sigma_1},\frac{\varepsilon_2}{\sigma_2},\cdots,\frac{\varepsilon_n}{\sigma_n}\right)'$$

例 10.12 (Cont.):

- 变换回归模型为

$$Y_i^* = X_i^{*\prime}\theta_0 + \varepsilon_i^*, i = 1, \cdots, n$$

- 其中

$$Y_i^* = \frac{Y_i}{\sigma_i}$$

$$X_i^* = \frac{X_i}{\sigma_i}$$

$$\varepsilon_i^* = \frac{\varepsilon_i}{\sigma_i}$$

- 可以看出，变换 $\boldsymbol{C}$ 通过对 ε_i 除以其条件标准差 σ_i 而消除其条件异方差。

例 10.13：[消除自相关]

- 令 $|\rho| < 1$。假设

$$
\boldsymbol{V} = \begin{bmatrix} 1 & \rho & \rho^2 & \cdots & \rho^{n-2} & \rho^{n-1} \\ \rho & 1 & \rho & \cdots & \rho^{n-3} & \rho^{n-2} \\ \rho^2 & \rho & 1 & \cdots & \rho^{n-4} & \rho^{n-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \rho^{n-2} & \rho^{n-3} & \rho^{n-4} & \cdots & 1 & \rho \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \cdots & \rho & 1 \end{bmatrix}
$$

- 这矩阵实际上是因为回归模型 $Y_i = X_i'\theta_0 + \varepsilon_i$ 中的随机扰动项 $\{\varepsilon_i\}$ 存在一阶自相关 $(AR(1))$ 而造成的，即

$$\varepsilon_i = \rho\varepsilon_{i-1} + v_i$$

其中，$\{v_i\}$ 是独立同分布序列，且 $E(v_i) = 0$，$\mathrm{var}(v_i) = \sigma^2$。

例 10.13 (Cont.):

- 通过运算，可得

$$V^{-1}=\begin{bmatrix} 1 & -\rho & 0 & \cdots & 0 & 0 \\ -\rho & 1+\rho^2 & -\rho & \cdots & 0 & 0 \\ 0 & -\rho & 1+\rho^2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1+\rho^2 & -\rho \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

和

$$C=\begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \cdots & 0 & 0 \\ -\rho & 1 & 0 & \cdots & 0 & 0 \\ 0 & -\rho & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -\rho & 1 \end{bmatrix}$$

例 10.13 (Cont.):

- 则

$$\boldsymbol{\varepsilon}^* = \boldsymbol{C\varepsilon} = \left(\sqrt{1-\rho^2}\varepsilon_1, \varepsilon_2 - \rho\varepsilon_1, \cdots, \varepsilon_n - \rho\varepsilon_{n-1}\right)'$$

- 变换回归模型为

$$Y_i^* = X_i^{*\prime}\theta_0 + \varepsilon_i^*, i = 1, \cdots, n$$

- 其中

$$Y_1^* = \sqrt{1-\rho^2}Y_1, Y_i^* = Y_i - \rho Y_{i-1}, i = 2, \cdots, n$$

$$X_1^* = \sqrt{1-\rho^2}X_1, X_i^* = X_i - \rho X_{i-1}, i = 2, \cdots, n$$

$$\varepsilon_1^* = \sqrt{1-\rho^2}\varepsilon_1, \varepsilon_i^* = \varepsilon_i - \rho\varepsilon_{i-1}, i = 2, \cdots, n$$

- 第一个观察值（$i=1$）的变换 $\sqrt{1-\rho^2}$ 称为**普莱斯-温斯登** (Prais-Winsten) 变换。可通过差分消除 ε_i 的自相关。

定理 10.11

假设 $\boldsymbol{X}'\boldsymbol{X}$ 为非奇异矩阵，且 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{V})$ 为 $n \times n$ 维已知的有限对称非奇异矩阵。则对所有的 $n > p$，

(a) [无偏性 (unbiasedness)]

$$E(\hat{\theta}^* \mid \boldsymbol{X}) = \theta_0$$

(b) [方差结构 (variance structure)]

$$\mathrm{var}(\hat{\theta}^* \mid \boldsymbol{X}) = \sigma^2\left(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*\right)^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}$$

(c) [不相关 (uncorrelatedness)]

$$\mathrm{cov}\left(\hat{\theta}^*, \boldsymbol{e}^* \middle| \boldsymbol{X}\right) = \boldsymbol{0}\text{，其中 } \boldsymbol{e}^* = \boldsymbol{Y}^* - \boldsymbol{X}^*\hat{\theta}^*$$

定理 10.11 (Cont.)

(d) [高斯-马尔可夫 (Gauss-Markov) 定理]

$$\hat{\theta}^* \text{ 是最优线性无偏估计量 (BLUE)}$$

(e) [残差方差估计量 (residual variance estimator)]

$$E(s^{*2} \mid \boldsymbol{X}) = \sigma^2\text{，其中} s^{*2} = \boldsymbol{e}^{*\prime}\boldsymbol{e}^*/(n-p)$$

证明：

- 变换模型 $\boldsymbol{Y}^* = \boldsymbol{X}^*\theta_0 + \boldsymbol{\varepsilon}^*$ 满足定理 10.4 关于经典回归模型所有假设，同时，GLS 估计量 $\hat{\theta}^*$ 是变换模型 $\boldsymbol{Y}^* = \boldsymbol{X}^*\theta_0 + \boldsymbol{\varepsilon}^*$ 的 OLS 估计量。因而，根据定理 10.4，结论 (a)- (e) 成立。

证毕。

GLS 估计量的 t- 检验和 F- 检验

- 由于 $\hat{\theta}^*$ 是变换后的线性回归模型 $\boldsymbol{Y}^* = \boldsymbol{X}^*\theta_0 + \boldsymbol{\varepsilon}^*$ 的 OLS 估计量，而且给定 $\boldsymbol{X}$，新随机扰动项 $\boldsymbol{\varepsilon}^* \sim \text{IID}\, N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$，$t$-检验和 F-检验是可用的。这些检验统计量分别定义如下：

$$T^* = \frac{R\hat{\theta}^* - r}{\sqrt{s^{*2}R(\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}R'}} \sim t_{n-p}$$

$$F^* = \frac{\left(R\hat{\theta}^* - r\right)'[R(\boldsymbol{X}^{*\prime}\boldsymbol{X})^{-1}R']^{-1}\left(R\hat{\theta}^* - r\right)/J}{s^{*2}} \sim \mathcal{F}_{J,n-p}$$

- 注意在假设 $\boldsymbol{\varepsilon} \mid \boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{V})$ 中，尽管 $\boldsymbol{V} = V(\boldsymbol{X})$ 已知，仍需要估计 σ^2。

- 因为 GLS 估计量 $\hat{\theta}^*$ 是 BLUE，而 OLS 估计量 $\hat{\theta}$ 不同于 $\hat{\theta}^*$，即

$$
\begin{aligned}
\hat{\theta}^* &= (\boldsymbol{X}^{*\prime}\boldsymbol{X}^*)^{-1}\boldsymbol{X}^{*\prime}\boldsymbol{Y}^* \\
&= (\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{V}^{-1}\boldsymbol{Y} \\
&\neq (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \hat{\theta}
\end{aligned}
$$

- 因而 OLS 估计量 $\hat{\theta}$ 不是 BLUE。
- GLS 估计最重要的启示在于，在线性回归模型中，通过对条件异方差和条件自相关的适当处理，可以得到有效的估计。
- 实践中，GLS 估计通常是不可行的，因为在实际应用中，$\mathrm{var}(\boldsymbol{\varepsilon} \mid \boldsymbol{X}) = \sigma^2\boldsymbol{V}$ 中的 $n \times n$ 维矩阵 $\boldsymbol{V}$ 往往是未知的。

- 存在若干可行的 GLS 估计使用了矩阵 $\boldsymbol{V}$ 的一致估计量 (参见 Robinson, 1988; White & Stinchcombe, 1991)，但对这些解决方法的讨论已超出本书范围，有兴趣读者可参考洪永淼 (2011)。

目　录

经典线性回归模型理论

- 经典线性回归模型及其**基本假设**
  - ✓ 线性假设、严格外生性、条件同方差、无序列自相关
- **模型选择准则**：
  - ✓ R^2 并不是合适的模型选择准则，因为它总是解释变量维数的非减函数。
  - ✓ 合适的模型选择准则：比如 AIC、BIC 与**调整后的** $\boldsymbol{R^2}$ 等。
- 普通最小二乘 (OLS) 估计量的统计性质：**BLUE**
- 在有限样本假设下的 $\boldsymbol{t}$**-检验**和 $\boldsymbol{F}$**-检验**

GLS

- 当随机扰动项存在条件异方差或者条件自相关时，OLS 估计量仍是无偏的，但不再是 BLUE，并且 $\hat{\theta}$ 和 s^2 不再相互独立。
  - ✓ 在具有已知形式的方差-协方差矩阵 (但存在一个未知尺度参数) 的假设下，可以通过纠正条件异方差和消除自相关的方法，对线性回归模型进行变换，使之转化成满足条件同方差和条件不相关的线性回归模型。
  - ✓ 这种变换后的线性回归模型的 OLS 估计量称为**广义最小二乘 (GLS) 估计量**。
  - ✓ GLS 估计量是 BLUE，相应的 t - 检验和 $F-$ 检验 可以使用。

中国科学院数学与系统科学研究院
Academy of Mathematics and Systems Science
Chinese Academy of Sciences

Thank You！