国际金融试验班 2018 年秋 • 时间序列

# 第 7 讲：AR 模型的统计推断

授课人：刘　岩

武汉大学经管学院金融系

2018 年 11 月 16 日–12 月 7 日

## 本讲内容

1. 统计推断基础
2. 线性回归模型的推断
3. AR 系数的推断

## 本节内容

1. 统计推断基础
2. 线性回归模型的推断
3. AR 系数的推断

## 统计推断的基本概念

- 假设样本的 $\{\boldsymbol{X}_t\}_{t=1}^T$ 所满足的统计模型为给定的总体分布

$$M_{\boldsymbol{\theta}} = \mathbb{P}(\{\boldsymbol{X}_t\}_{t=1}^T|\boldsymbol{\theta})$$

- 基本任务：从样本 $\{\boldsymbol{X}_t\}_{t=1}^T$ 反推 $\boldsymbol{\theta}$ 未知的真实值 $\boldsymbol{\theta} = \boldsymbol{\theta}_0$
- 估计：从样本 $\{\boldsymbol{X}_t\}_{t=1}^T$ 获得 $\boldsymbol{\theta}_0$ 的估计值 $\hat{\boldsymbol{\theta}}_T$
- 推断：由样本计算得到的 $\hat{\boldsymbol{\theta}}_T$ 判断 $\boldsymbol{\theta}$ 是否为特定取值 $\boldsymbol{\theta}_0$
- 误差：由于 $\hat{\boldsymbol{\theta}}_T$ 继承了样本的随机性，上述推断可能存在误差 $\Longrightarrow$ 统计推断需要控制误差

# 推断的逻辑

- 统计推断和悬疑破案有同样的问题：从已知事实（样本）推断未知真相（参数真实值）
- 真相的可能取值：可能有无穷多中可能，但至少是两种可能

$\Longrightarrow$ 即便是无穷多中可能，总可以使用二分法：把可能真相分为互不相交两组 $G, B$

- 推断的两种结论：支持 $G$/反对 $B$ vs. 支持 $B$/反对 $G$
- 推断有误差 $\Longrightarrow$ 4 种结果：

| | | 真相 $G$ | 真相 $B$ |
|---|---|---|---|
| 推断 | 支持 $G$ | 正确 | 错误 |
| | 支持 $B$ | 错误 | 正确 |

## 统计假设与两类错误

- 将参数真实值（真相）分为互不相交的两类，等价于做两个相互排斥的假设 (hypothesis)
- 统计学的传统：将最希望被排除的假设，称为原假设 (null hypo.)；另一个假设，称为备择假设 (alternative hypo.)
  - 原假设又称为零假设，一般记做 $H_0$
  - 备择假设一般记做 $H_1$ 或 $H_a$
- 假设检验 (hypo. testing)：通过检验决定是否支持假设
  - 第一类错误 (Type-I error)：在 $H_0$ 成立时，支持 $H_1$
  - 第二类错误 (Type-II error)：在 $H_1$ 成立时，支持 $H_0$

| | $H_0$ 成立 | $H_1$ 成立 |
|---|---|---|
| 支持 $H_0$ | | 第二类错误 |
| 支持 $H_1$ | 第一类错误 | |

## 统计推断的原则：优先控制第一类错误

- 原假设是研究者最希望排除的假设

$\Rightarrow$ 优先控制住拒绝 $H_0$、接受 $H_1$ 的可能错误

  - 如果嫌疑人真是个好人，我们希望尽可能避免将其判做坏人的情况 $\Rightarrow$ 减少冤假错案
  - 现代统计学诞生于英国，优先控制第一类错误受到英国普通法 (common law) “疑罪从无”原则的影响

- 统计推断就是一个决策过程

| | 好人 | 坏人 |
|---|---|---|
| 无罪 | | 逍遥法外 第二类错误 |
| 有罪 | 冤假错案 第一类错误 | |

## 统计量、显著性水平与 $p$-值

- 如何支持/接受或反对/拒绝一个统计假设？
- 检验统计量 (test statistics)：从统计模型出发，构造一个统计量 $\xi_T$（r.v.）并确定其在特定假设 $H_x$ 下的分布 $\mathbb{P}(\xi_T|H_x)$
- 利用样本观测值，计算统计量的具体取值 $\bar{\xi}_T$；若在 $H_x$ 下观测到 $\bar{\xi}_T$ 的概率 $\mathbb{P}(\bar{\xi}_T|H_x)$ 小于某个临界值，则拒绝该假设
- 该临界值称为该检验的显著性水平 (significance level)
  - 常用取值包括 1%、5%、10%
- 观测到 $\bar{\xi}_T$ 的概率 $\mathbb{P}(\bar{\xi}_T|H_x)$ 本身，称为检验假设 $H_x$ 的 $p$-值 ($p$-value)
  - 按惯例，$p$-值越低，称“拒绝该假设的显著性越高”

## 均值 $t$-检验：简单情形

- 最常见的假设检验为均值 $t$-检验，简称 $t$-检验 ($t$-test)
- 考虑两个 iid 序列 $\{X_t\}_{t=1}^T$ 与 $\{Y_t\}_{t=1}^T$ 分别服从总体分布 $N(\mu_X, \sigma_X^2)$, $N(\mu_Y, \sigma_Y^2)$
- 简化假设：$\mu_X, \mu_Y$ 未知，但 $\sigma_X, \sigma_Y$ 已知
- 需要检验的原假设为 $H_0 : \mu_X = \mu_Y \Leftrightarrow H_0 : \mu_X - \mu_Y = 0$
- 计算可知，两个样本均值 $\hat{\mu}_X, \hat{\mu}_Y$ 分别服从如下分布

$$\hat{\mu}_X \sim N\left(\mu_X, \tfrac{1}{T}\sigma_X^2\right), \quad \hat{\mu}_Y \sim N\left(\mu_Y, \tfrac{1}{T}\sigma_Y^2\right)$$

- 选取统计量为 $\xi_T = \hat{\mu}_X - \hat{\mu}_Y$，在 $H_0$ 下 $\xi_T$ 的分布为

$$\xi_T = \hat{\mu}_X - \hat{\mu}_Y = N\left(\mu_X - \mu_Y, \tfrac{1}{T}(\sigma_X^2 + \sigma_Y^2)\right) = N\left(0, \tfrac{1}{T}(\sigma_X^2 + \sigma_Y^2)\right)$$

## 均值 $t$-检验：简单情形

- 带入样本 $\{X_t, Y_t\}$ 具体取值，可计算统计量 $\xi_T$ 的取值 $\bar{\xi}_T$
- 考虑如下概率，$\Phi$ 为标准正态 cdf：

$$\begin{aligned}\mathbb{P}(|\xi_T| \geq |\bar{\xi}_T|) &= \mathbb{P}\left(\frac{|\xi_T|}{\sigma(\xi_T)} \geq \frac{|\bar{\xi}_T|}{\sigma(\xi_T)}\right) \\ &= 2\Phi\left(-\frac{|\bar{\xi}_T|}{\sigma(\xi_T)}\right)\end{aligned}$$

- 若 $|\bar{\xi}_T|$ 很大，说明观察到 $\xi_T \leq -|\bar{\xi}_T|$ 或 $\xi_T \geq |\bar{\xi}_T|$ 概率很小
- 此时拒绝 $H_0 : \mu_X = \mu_Y$ 出现（第一类）错误的概率很小！
- $\frac{\bar{\xi}_T}{\sigma(\xi_T)}$ 称为 <u>$t$-统计量</u> ($t$-statistic)

## 大样本 $t$-检验与标准误

- 考虑 iid 样本 $\{X_t\}_{t=1}^T$，服从总体分布 $N(\mu_X, \sigma_X^2)$
- 假设此时 $\mu_X, \sigma_X^2$ 均为未知，并考虑原假设 $H_0 : \mu_X = \mu_0 \equiv 0$
- 自然的想法：检验样本均值 $\hat{\mu}_T$ 是否显著不同于 $\mu_0 = 0$
- $H_0$ 之下，$\hat{\mu}_T = \frac{1}{T}\sum_t X_t$ 的分布为 $N\left(\mu_0, \frac{1}{T}\sigma_X^2\right) = N\left(0, \frac{1}{T}\sigma_X^2\right)$

$$\Rightarrow \frac{\hat{\mu}_T}{\frac{1}{\sqrt{T}}\sigma_X} \sim N(0,1) \Rightarrow \text{统计量}\,\xi_T = \frac{\hat{\mu}_T}{\frac{1}{\sqrt{T}}\hat{\sigma}_{X,T}} \xrightarrow{\text{d}} N(0,1)$$

  其中，$\hat{\sigma}_{X,T}^2 = \frac{1}{T}\sum_t (X_t - \hat{\mu}_T)^2$ 为 $\sigma_X^2$ 的一致估计
- $\xi_T$ 为 $\hat{\mu}_T$ 的标准化 (standardization)，有极限分布 $N(0,1)$
- $\hat{\mu}_T$ 的样本标准差 $\frac{1}{\sqrt{T}}\hat{\sigma}_{X,T}$ 称为标准误 (standard error)

## 本节内容

## 线性回归 OLS 估计的大样本理论

- 考虑多元回归模型

$$Y_t = \sum_{k=1}^{K} \beta_k X_{kt} + \varepsilon_t = \boldsymbol{X}_t^{\top}\boldsymbol{\beta} + \varepsilon_t, \quad t = 1, \ldots, T$$

- 矩阵形式为

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

- $\boldsymbol{\beta}$ 的 OLS 估计为

$$\hat{\boldsymbol{\beta}}_T = (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{Y}$$

- 对 $H_0 : \boldsymbol{\beta} = \boldsymbol{\beta}_0$ 等类假设进行推断，需要知道 $\hat{\boldsymbol{\beta}}_T$ 的渐近分布

## 大样本的意义

- 大样本 (large sample) 时，概率极限理论可以直接确定估计量 $\hat{\beta}_T$ 的渐近分布 (asymptotic distribution)
    - 大数定律和中心极限定理
    - 后者的实质：标准化样本均值的分布收敛到标准正态分布
- 小样本时，只有在很强的假设之下，才有可能得到 $\hat{\beta}_T$ 的分布，从而实现相关统计推断
    - 如残差项 $\varepsilon_t$ 是正态分布

## 线性回归 OLS 估计的一致性

- 估计量 $\hat{\boldsymbol{\beta}}_T$ 统计推断的基础是一致性：$\hat{\boldsymbol{\beta}}_T \xrightarrow{\text{a.s.}} \boldsymbol{\beta}_0$
- 假设模型系数真实值为 $\boldsymbol{\beta} = \boldsymbol{\beta}_0$，则 $\boldsymbol{\beta}$ 的 OLS 估计可以写为

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_T &= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y} \\
&= (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top (\boldsymbol{X}\boldsymbol{\beta}_0 + \boldsymbol{\varepsilon}) \\
&= \boldsymbol{\beta}_0 + (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta}_0 + \left( \frac{1}{T} \boldsymbol{X}^\top \boldsymbol{X} \right)^{-1} \frac{1}{T} \boldsymbol{X}^\top \boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta}_0 + \left( \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t \boldsymbol{X}_t^\top \right)^{-1} \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{X}_t \varepsilon_t
\end{aligned}
$$

## 线性回归 OLS 估计的一致性

- OLS 回归的一般默认假设：
  1. $\{\boldsymbol{X}_t\}$ 平稳，故交叉 2-阶矩矩阵 $\boldsymbol{M}_{K\times K} = \mathbb{E}\boldsymbol{X}_t\boldsymbol{X}_t^{\intercal}$ 存在
  2. 进一步，$\boldsymbol{M}$ 可逆，即 $\boldsymbol{M}^{-1}$ 存在
  3. $\{\varepsilon_t\}$ 为 iid 序列，且给定 $\boldsymbol{X}_t$ 时的条件均值为 0，即 $\mathbb{E}[\varepsilon_t|\boldsymbol{X}_t] = 0$；由此知 $\mathbb{E}\boldsymbol{X}_t\varepsilon_t = \boldsymbol{0}_{K\times 1}$
- 在此组假设之下，应用大数定律

$$
\begin{aligned}
\left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t^{\intercal}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\varepsilon_t &\xrightarrow{\text{a.s.}} \boldsymbol{M}_{K\times K}^{-1}\boldsymbol{0}_{K\times 1} \\
&= \boldsymbol{0}_{K\times 1}
\end{aligned}
$$

故 $\hat{\boldsymbol{\beta}}_T \xrightarrow{\text{a.s.}} \boldsymbol{\beta}_0$，即 OLS 估计量具有一致性

# 线性回归 OLS 估计系数的渐近分布

- 已知系数估计量 $\hat{\boldsymbol{\beta}}_T \xrightarrow{\text{a.s.}} \boldsymbol{\beta}_0$，且

$$
\begin{aligned}
\mathbb{E}\hat{\boldsymbol{\beta}}_T &= \boldsymbol{\beta}_0 + \mathbb{E}(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{X}^\intercal \boldsymbol{\varepsilon} \\
&= \boldsymbol{\beta}_0 + \mathbb{E}\{\mathbb{E}[(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{X}^\intercal \boldsymbol{\varepsilon}|\boldsymbol{X}]\} \\
&= \boldsymbol{\beta}_0 + \mathbb{E}\{(\boldsymbol{X}^\intercal \boldsymbol{X})^{-1}\boldsymbol{X}^\intercal \mathbb{E}[\boldsymbol{\varepsilon}|\boldsymbol{X}]\} = \boldsymbol{\beta}_0
\end{aligned}
$$

- 通常而言 $\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0)$ 收敛到一个多元正态分布

$\Rightarrow$ 中心极限定理 (central limit theorem)

## 中心极限定理 (central limit theorem, CLT)

- 考虑 iid 样本 $\{X_t\}$，期望为 $\mu_X$，方差为 $\sigma_X^2$
- 由大数定律，$\hat{\mu}_{X,T} - \mu_X = T^{-1}\sum_t (X_t - \mu_X) \xrightarrow{\text{a.s.}} 0$
- 计算可得，$\text{var}(\hat{\mu}_{X,T}) = \sigma_X^2/T$，故 $\hat{\mu}_{X,T}$ 的标准化

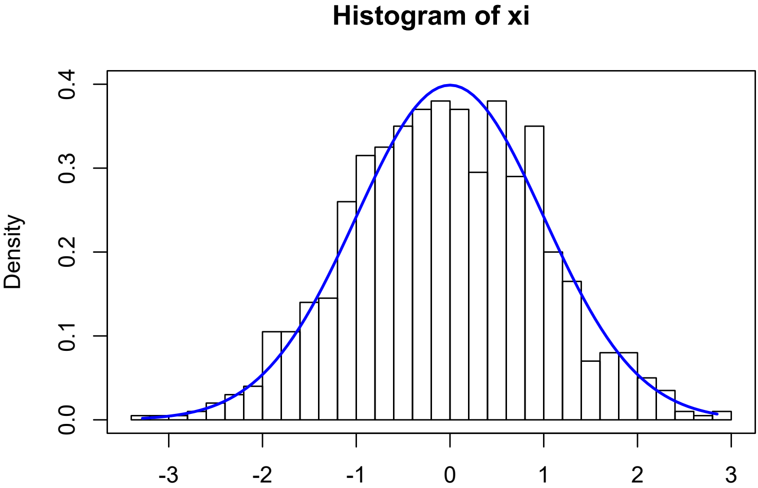
$$\xi_T = \frac{\hat{\mu}_{X,T} - \mu_X}{\sqrt{\text{var}(\hat{\mu}_{X,T})}} = \frac{T^{-1}\sum_t (X_t - \mu_X)}{\sqrt{T}^{-1}\sigma_X} = \frac{1}{\sqrt{T}}\sum_t \frac{X_t - \mu_X}{\sigma_X}$$

方差为 1

### 定理 1 (iid 序列 CLT：1-元情形)

假设有 iid 序列 $\{X_t\}$，则其标准化样本均值 $\xi_T \xrightarrow{\text{d}} N(0,1)$，或等价的，$\sqrt{T}(\hat{\mu}_{X,T} - \mu_X) \xrightarrow{\text{d}} N(0, \sigma_X^2)$

## CLT：Monte Carlo 模拟示例



# 多元情形下的中心极限定理

**定理 2 (iid 序列 CLT：多元情形)**

假设有 iid 向量序列 $\{\boldsymbol{X}_t\}$，期望向量为 $\boldsymbol{\mu}$，协方差矩阵为 $\boldsymbol{\Sigma}$，则其样本均值向量有如下极限分布

$$\sqrt{T}(\hat{\boldsymbol{\mu}}_{X,T} - \boldsymbol{\mu}_X) \xrightarrow{\mathrm{d}} N(0, \boldsymbol{\Sigma})$$

- 多元情形下仍然可以定义标准化样本均值：当 $\boldsymbol{\Sigma}$ 可逆时，

$$\xi_T = \left(\frac{1}{\sqrt{T}}\boldsymbol{\Sigma}^{\frac{1}{2}}\right)^{-1}\left(\hat{\boldsymbol{\mu}}_{X,T} - \boldsymbol{\mu}_X\right)$$

# 协方差矩阵补充性质

- 协方差阵 $\mathbf{\Sigma}_{n\times n}$ 为非负定对称阵，故特征值 $\lambda_1,\ldots,\lambda_n\geq 0$，且存在矩阵 $C$ 满足 $C^{-1}=C^{\mathsf{T}}$，使得 $\mathbf{\Sigma}$ 可对角化为

$$\mathbf{\Sigma}=C\mathrm{diag}(\lambda_1,\cdots,\lambda_n)C^{-1}$$

  其中 diag 表示由给定数组构造对角矩阵

- 此时有

$$\begin{bmatrix}\lambda_1 & & \\ & \ddots & \\ & & \lambda_n\end{bmatrix}=\begin{bmatrix}\sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n}\end{bmatrix}\begin{bmatrix}\sqrt{\lambda_1} & & \\ & \ddots & \\ & & \sqrt{\lambda_n}\end{bmatrix}$$

- 定义 $\mathbf{\Sigma}^{\frac{1}{2}}=C\mathrm{diag}\left(\sqrt{\lambda_1},\cdots,\sqrt{\lambda_n}\right)C^{-1}$，仍为非负定对称阵，满足 $\mathbf{\Sigma}=\mathbf{\Sigma}^{\frac{1}{2}}\mathbf{\Sigma}^{\frac{1}{2}}$

## 一般情况下的中心极限定理

- 上述定理 1/2 要求 iid 假设；应用中，iid 假设不一定满足
- 仔细观察可见，iid 假设“形式”上主要用于确定样本均值的协方差矩阵；而这一步骤，主要依赖于不同期变量相关性为 0 这一事实
- 如果只假设序列不相关，无法保证极限分布存在

$\Rightarrow$ 弱于 iid 而强于不相关的假设为：鞅差序列 (martingale difference series)

- $\Omega_t$ 表示 $t$ 时信息集；$\mathbb{E}[X_{t+1}|\Omega_t]$ 表示对当期信息的条件期望 (conditional expectation)
- 给定 $\{\Omega_t\}$，$\{X_t\}$ 称为鞅差序列，若 $\mathbb{E}[X_{t+1}|\Omega_t] = 0$
- 前述中心极限定理在鞅差序列假设下均成立！

## 线性回归 OLS 估计系数的统计推断

- 为分析 $\hat{\boldsymbol{\beta}}_T$ 的渐近分布，考虑

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) = \left(\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\boldsymbol{X}_t^\intercal\right)^{-1}\underbrace{\sqrt{T}\,\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\varepsilon_t}_{\text{视作样本均值}}$$

- 注意 $\mathbb{E}[\boldsymbol{X}_t\varepsilon_t\cdot\varepsilon_t\boldsymbol{X}_t^\intercal] = \sigma_\varepsilon^2\mathbb{E}[\boldsymbol{X}_t\boldsymbol{X}_t^\intercal] = \sigma_\varepsilon^2\boldsymbol{M}$
- 在 OLS 常用假设下，及 $\boldsymbol{X}_t\varepsilon_t$ 鞅差序列假设下，由 CLT 可得

$$\sqrt{T}\frac{1}{T}\sum_{t=1}^{T}\boldsymbol{X}_t\varepsilon_t \xrightarrow{\text{d}} N(\boldsymbol{0},\sigma_\varepsilon^2\boldsymbol{M})$$

- 再由 $T^{-1}\sum_t\boldsymbol{X}_t\boldsymbol{X}_t^\intercal \xrightarrow{\text{a.s.}} \boldsymbol{M}$ 可得

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{\text{d}} N\left(\boldsymbol{0},\sigma_\varepsilon^2\boldsymbol{M}^{-1}\right)$$

## 回归系数渐近分布的样本估计

- 总体分布未知，故 $\sigma_\varepsilon^2$ 与 $\boldsymbol{M}$ 均需从样本 $\{Y, \boldsymbol{X}\}$ 进行估计
- $\{\boldsymbol{X}_t\}$ 平稳，LLN $\Rightarrow \hat{\boldsymbol{M}}_T \equiv T^{-1}\sum_t \boldsymbol{X}_t\boldsymbol{X}_t^\intercal \xrightarrow{\text{a.s.}} \boldsymbol{M}$
- 由 $\hat{\boldsymbol{\beta}}_T$ 的一致性，可知样本残差

$$\hat{\varepsilon}_t = Y_t - \boldsymbol{X}_t^\intercal \hat{\boldsymbol{\beta}}_T$$

是 $\varepsilon_t$ 的一致估计，LLN $\Rightarrow \hat{\sigma}_{\varepsilon,T}^2 \equiv T^{-1}\sum_t \hat{\varepsilon}_t^2 \xrightarrow{\text{a.s.}} \sigma_\varepsilon^2$

- 结合这两方面，可知

$$\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 \sim N\left(\boldsymbol{0}, \frac{1}{T}\hat{\sigma}_{\varepsilon,T}^2 \hat{\boldsymbol{M}}_T^{-1}\right)$$

## 回归系数 OLS 估计：单系数假设检验

- $H_0 : \hat{\beta}_{k,T} - \beta_{k,0} = 0 \Rightarrow$ 标准 $t$-检验

$$\xi_T = \frac{\hat{\beta}_{k,T} - \beta_{k,0}}{\sigma(\hat{\beta}_{k,T})}$$

- 其中 $\sigma(\hat{\beta}_{k,T})$ 表示 $\hat{\beta}_{k,T}$ 的标准误，对应矩阵

$$\frac{1}{T}\hat{\sigma}^2_{\varepsilon,T}\hat{\boldsymbol{M}}_T^{-1}$$

  对角线上第 $k$ 个元素的平方根

- 给定 $\hat{\beta}_{k,T}$ 与 $\hat{\sigma}^2_{\varepsilon,T}\hat{\boldsymbol{M}}_T^{-1}$ 收敛，则 $T$ 越大，标准误越小，从而检验的显著性水平越高，$p$-值越小

## 回归系数 OLS 估计：多系数联合检验

- 考虑 $H_0 : \beta_i - \beta_j = 0$，即第 $i$ 和 $j$ 个回归系数相等

  $\Longrightarrow$ 只需从渐近分布中提取出 $\hat{\beta}_{i,T}, \hat{\beta}_{j,T}$ 的联合分布

  - 考虑矩阵（行向量）

$$J = [\dots, 0, \underbrace{\ \ 1\ \ }_{i^{\text{th}}}, 0, \dots, 0, \underbrace{\ -1\ }_{j^{\text{th}}}, 0, \dots],$$

则 $$\boldsymbol{J}\hat{\boldsymbol{\beta}}_T = \hat{\beta}_{i,T} - \hat{\beta}_{j,T} \sim N\left(0, \underbrace{\frac{1}{T}\hat{\sigma}^2_{\varepsilon,T}\boldsymbol{J}\hat{\boldsymbol{M}}_T^{-1}\boldsymbol{J}^{\intercal}}_{1\times 1}\right)$$

## 回归系数 OLS 估计：多系数联合检验

- 一般情况，多系数线性联合假设检验，可引入联合选取矩阵 (joint selection matrix) $\boldsymbol{J}_{L\times K}$：

$$H_0 : \boldsymbol{J}\boldsymbol{\beta} = \boldsymbol{0}_{L\times 1}$$

此时，检验统计（向）量 $\boldsymbol{\xi}_T = \boldsymbol{J}\hat{\boldsymbol{\beta}}_T$ 的渐近分布为

$$N\left(\boldsymbol{0}, \frac{1}{T}\hat{\sigma}^2_{\varepsilon,T}\boldsymbol{J}\hat{\boldsymbol{M}}_T^{-1}\boldsymbol{J}^{\mathsf{T}}\right)$$

- 如何检验？$\Longrightarrow$ 将 $\boldsymbol{\xi}_T$ 变形为 $\chi^2$-分布

## $\chi^2$-分布与正态分布的关系

- 给定 $L$ 个独立标准正态 r.v. $X_1, \ldots, X_L$，平方和 $Y = \sum_\ell X_\ell^2$ 的分布称做<u>自由度为 $L$ 的 $\chi^2$-分布</u> ($\chi^2$-dist. with the degree of freedom of $L$)，记做 $\chi^2(L)$
    - $\chi^2$-分布是一种特殊的 $\Gamma$-分布
- 若 $\boldsymbol{X} = [X_1, \ldots, X_L]^\intercal \sim N(\mathbf{0}, \boldsymbol{\Sigma})$，则 $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X} \sim N(\mathbf{0}, \boldsymbol{I}_{L\times L})$，故

$$(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X})^\intercal(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}) = \boldsymbol{X}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \sim \chi^2(L)$$

- 上述多系数联合检验，可使用如下统计量

$$\zeta_T \equiv \boldsymbol{\xi}_T^\intercal\left(\tfrac{1}{T}\hat{\sigma}_{\varepsilon,T}^2 \boldsymbol{J}\hat{\boldsymbol{M}}_T^{-1}\boldsymbol{J}^\intercal\right)^{-1}\boldsymbol{\xi}_T = \frac{T}{\hat{\sigma}_{\varepsilon,T}^2}\hat{\boldsymbol{\beta}}_T^\intercal\boldsymbol{J}^\intercal\left(\boldsymbol{J}\hat{\boldsymbol{M}}_T^{-1}\boldsymbol{J}^\intercal\right)^{-1}\boldsymbol{J}\hat{\boldsymbol{\beta}}_T$$

服从 $\chi^2(L)$-分布；当 $\zeta_T$ 大于临界值时，拒绝 $H_0 : \boldsymbol{J\beta} = \mathbf{0}$

# 本节内容

## AR 模型的 OLS 估计

- 平稳 AR($p$) 模型：$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$，$\{\varepsilon_t\}$ iid，0-均值
- 给定样本 $\{X_t\}_{t=1}^T$，AR($p$) 模型的 OLS 回归矩阵形式：

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

$$\boldsymbol{Y} = \begin{bmatrix} X_{p+1} \\ \vdots \\ X_T \end{bmatrix}, \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_{p+1}^\intercal \\ \vdots \\ \boldsymbol{X}_T^\intercal \end{bmatrix} = \begin{bmatrix} X_p & \cdots & X_1 \\ \vdots & \ddots & \vdots \\ X_{T-1} & \cdots & X_{T-p} \end{bmatrix}, \boldsymbol{\phi} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}$$

- OLS 估计量：$\hat{\boldsymbol{\phi}}_T = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y}$

## AR 模型 OLS 估计系数的渐近分布

- 一致性：平稳性 $\Rightarrow \mathbb{E}[\varepsilon_t|X_{t-1}, X_{t-2}, \ldots, X_{t-p}] = 0$，即 $\mathbb{E}[\varepsilon_t|\boldsymbol{X}_t^\intercal] = 0$，故 $T^{-1}\sum_t \boldsymbol{X}_t\varepsilon_t = \boldsymbol{0}$
- 渐近分布：关键在于验证 $\boldsymbol{X}_t^\intercal\varepsilon_t$ 是鞅差序列，等价于各个分量 $Z_t \equiv X_{t-i}\varepsilon_t$ 是鞅差序列，$\forall i = 1, \ldots, p$
  - 需要验证 $\mathbb{E}[Z_{t+1}|\Omega_t] = 0$；$t$-期信息集 $\Omega_t = \{X_t, X_{t-1}, \ldots\} = \{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$
  - $\mathbb{E}[Z_{t+1}|\Omega_t] = \mathbb{E}[X_t\varepsilon_{t+1}|\Omega_t] = \mathbb{E}[X_t\varepsilon_{t+1}|X_t, \ldots] = X_t\mathbb{E}[\varepsilon_{t+1}|X_t] = 0$
  - $\mathbb{E}[\varepsilon_{t+1}|X_t] = 0$ 是因为 $X_t$ 由 $\{\varepsilon_t, \varepsilon_{t-1}, \ldots\}$ 决定，后者与 $\varepsilon_{t+1}$ 相互独立，故 $\mathbb{E}[\varepsilon_{t+1}|X_t] = \mathbb{E}\varepsilon_{t+1} = 0$
- 前面 OLS 回归统计推断的所有结论均适用！