国际金融试验班 2020 年秋 · 时间序列

第 8 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

2020 年 11 月 5 日

本讲内容

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

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

- 矩阵形式为

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

- $\boldsymbol{\beta}$ 的 OLS 估计为

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

- 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

OLS 估计系数的注记

- OLS 估计系数 $\hat{\boldsymbol{\beta}}_T$ 是样本 $\mathcal{X} = \{Y_t, \boldsymbol{X}_t\}$ 的一个函数，因此当 $\mathcal{X}$ 作为随机变量时，$\hat{\boldsymbol{\beta}}_T$ 也是一个随机变量
- 但当样本 $\mathcal{X}$ 给定，即样本为其实现值 $\{y_t, \boldsymbol{x}_t\}$ 时，相应的估计系数有确定取值，非随机，此时通常将 OLS 估计值写为 $\boldsymbol{b}_T$
- $\hat{\boldsymbol{\beta}}_T$ 是一个随机变量，需要进一步确定其分布，而 $\boldsymbol{b}_T$ 是这个随机变量的实现值

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 ε_t 是正态分布

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 μ_X，方差为 σ_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

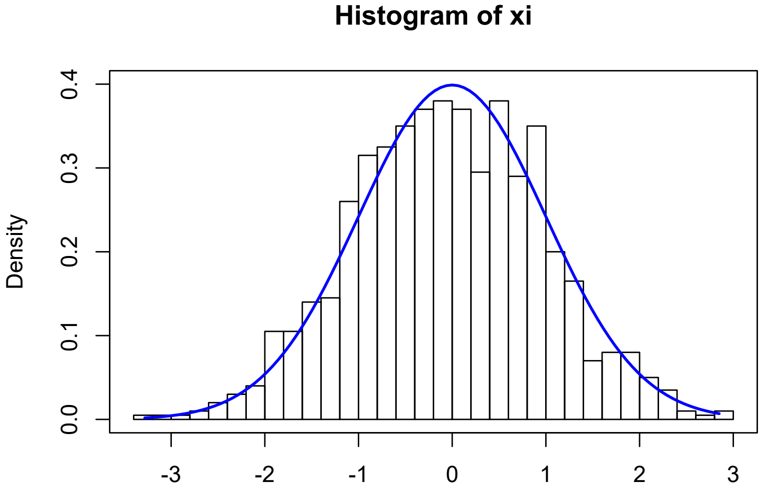
$$\xi_T = \frac{\hat{\mu}_{X,T} - \mu_X}{\sqrt{\text{var}(\hat{\mu}_{X,T})}} = \frac{T^{-1}\sum_t (X_t - \mu_X)}{\sqrt{T}^{-1}\sigma_X} = \frac{1}{\sqrt{T}}\sum_t \frac{X_t - \mu_X}{\sigma_X}$$

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 $\{\mathbf{X}_t\}$，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 $\mathbf{\Sigma}_{n\times n}$ 为半正定对称阵，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 $\mathbf{\Sigma}^{\frac{1}{2}} = C\mathrm{diag}\left(\sqrt{\lambda_1}, \cdots, \sqrt{\lambda_n}\right)C^{-1}$, 仍为半正定对称阵，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)

- Ω_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

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

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

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

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

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

回归系数渐近分布的样本估计

- 总体分布未知，故 σ_ε^2 与 $\boldsymbol{M}$ 均需从样本 $\{\boldsymbol{Y}, \boldsymbol{X}\}$ 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 ε_t 的一致估计，LLN $\Rightarrow \hat{\sigma}_{\varepsilon,T}^2 \equiv T^{-1}\sum_t \hat{\varepsilon}_t^2 \xrightarrow{\text{a.s.}} \sigma_\varepsilon^2$
  - 有限样本下，$\hat{\sigma}_{\varepsilon,T}^2$ 有偏差更小的无偏估计量
- 结合这两方面，可知

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

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 (standard error)，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

系数检验的进一步说明

- ξ_T 是做为随机变量的样本 $\{Y_t, \boldsymbol{X}_t\}$ 的函数，因此也是随机变量
- 确定了 OLS 估计量 $\hat{\beta}_{k,T}$ 及相应的 ξ_T 的分布后，即可以根据作为实现值的样本取值 $\{y_t, \boldsymbol{x}_t\}$ 来计算 $\hat{\beta}_{k,T}$ 的实际取值 $\hat{b}_{k,T}$，进而计算 t-统计量 ξ_T 的实际取值

$$\bar{\xi}_T = \frac{\hat{b}_{k,T} - \beta_{k,0}}{\sigma(\hat{\beta}_{k,T})}$$

- 回归系数 $\hat{\beta}_{k,T}$ 关于原假设 $\beta_{k,0}$ 检验的 p-值为如下双尾概率

$$\mathbb{P}(|\xi_T| \geq |\bar{\xi}_T|)$$

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）

$$\boldsymbol{R} = [\ldots, 0, \underbrace{1}_{i^{\text{th}}}, 0, \ldots, 0, \underbrace{-1}_{j^{\text{th}}}, 0, \ldots],$$

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

多系数联合检验

- 一般情况，多系数线性联合假设检验，可引入线性约束矩阵 (linear restriction matrix) $\boldsymbol{R}_{L\times K}$ 及约束目标向量 $\boldsymbol{r}_{L\times 1}$：

$$H_0 : \boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r} \quad \Leftrightarrow \quad H_0 : \boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r} = \boldsymbol{0}_{L\times 1}$$

  此时，检验统计（向）量 $\boldsymbol{\xi}_T = \boldsymbol{R}\hat{\boldsymbol{\beta}}_T - \boldsymbol{r}$ 的渐近分布近似等于

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

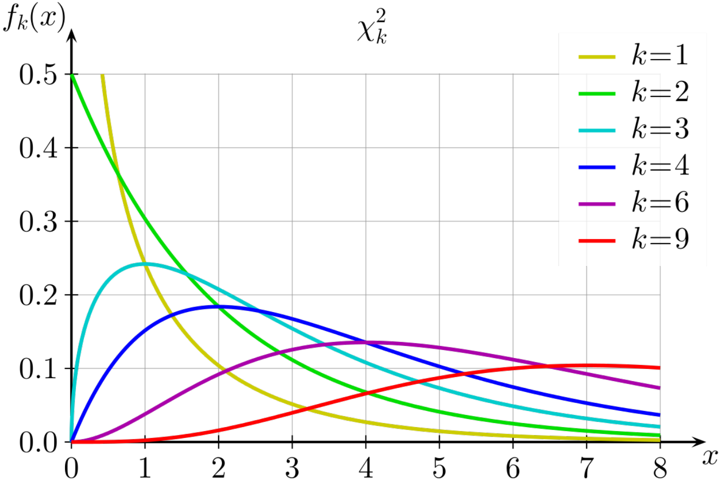
  - $L \leq K$，最多同时检验所有 K 个回归系数的取值

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

χ^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 $Y = \sum_\ell X_\ell^2$ 的分布称做自由度为 L 的 χ^2-分布 (χ^2-dist. with the degree of freedom of L)，记做 $\chi^2(L)$
  - PDF 如下
$$\frac{1}{2^{L/2}\Gamma(L/2)} x^{L/2-1} e^{-x/2}$$
  - χ^2-分布是一种特殊的 Γ-分布
  - 取值范围为 $(0, +\infty)$
- 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
$$(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X})^\intercal(\boldsymbol{\Sigma}^{-1/2}\boldsymbol{X}) = \boldsymbol{X}^\intercal\boldsymbol{\Sigma}^{-1}\boldsymbol{X} \sim \chi^2(L)$$

χ^2-分布的密度函数



多系数检验：Wald 检验与 F-检验

- 上述多系数联合检验，可使用如下 Wald 统计量

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

服从 $\chi^2(L)$-分布，又称为 Wald 检验

  - 当 W_T 的样本取值 $\bar{W}_T = \frac{T}{\hat{\sigma}_{\varepsilon,T}^2}\hat{\boldsymbol{b}}_T^{\intercal}\boldsymbol{R}^{\intercal}\left(\boldsymbol{R}\hat{\boldsymbol{M}}_T^{-1}\boldsymbol{R}^{\intercal}\right)^{-1}\boldsymbol{R}\hat{\boldsymbol{b}}_T$ 大于 α-显著性临界值 $\chi_\alpha^2(L)$，或者该检验 p-值 $\Pr(W_T \geq \bar{W}_T)$ 小于显著性水平 α 时，拒绝 $H_0 : \boldsymbol{R\beta} = \boldsymbol{r}$
  - α 通常为 $1\%, 5\%, 10\%$；$\chi_\alpha^2(L)$ 为分布对应的 α-分位数
- F-检验：统计量为 $F_T = W_T/L$，分布为 F-分布 $F(L, T-K)$
  - 但这是一个小样本检验，需要残差项 ε_t 为 iid 正态分布才能得出 F-分布

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 $\mathbb{E}[\varepsilon_t|\mathbf{X}_t^\intercal] = 0$，故 $T^{-1}\sum_t \mathbf{X}_t\varepsilon_t = \mathbf{0}$
- 渐近分布：关键在于验证 $\mathbf{X}_t^\intercal\varepsilon_t$ 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 ε_{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！