国际金融试验班 2018 年秋 • 时间序列

第 6 讲：AR 模型的估计

授课人：刘　岩

武汉大学经管学院金融系

2018 年 11 月 8 日

本讲内容

1. AR 过程的自协方差结构
2. AR 过程的回归估计

本节内容

AR(1) 过程的例子

- 给定平稳 AR(1) 过程 $X_t = \phi X_{t-1} + \varepsilon_t$，$|\phi| < 1$
- 计算可知

$$\sigma_X^2(k) = \phi^k \frac{\sigma_\varepsilon^2}{1-\phi^2}$$

- 由此，可将高阶自协方差写为

$$\sigma_X^2(k) = \phi^k \sigma_X^2(0)$$

- 另一种写法：首先注意到 $\mathbb{E}[\varepsilon_t X_{t-1}]$，且

$$X_t X_{t-1} = \phi X_{t-1}^2 + \varepsilon_t X_{t-1}$$

$$\Rightarrow \mathbb{E}[X_t X_{t-1}] = \phi \mathbb{E}[X_{t-1}^2] + \mathbb{E}[\varepsilon_t X_{t-1}] \Rightarrow \sigma_X^2(1) = \phi \sigma_X^2(0)$$

$$X_t X_{t-(k+1)} = \phi X_{t-1} X_{t-(k+1)} + \varepsilon_t X_{t-(k+1)}$$

$$\Rightarrow \sigma_X^2(k+1) = \phi \sigma_X^2(k)$$

AR(2) 过程的例子

- 给定平稳 AR(2) 过程 $X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$，特征多项式 $A(z) = 1 - \phi_1 z - \phi_2 z^2$ 的两个零点均位于单位圆之外
- 两端分别乘以 X_{t-1}，X_{t-2} 可得

$$\begin{aligned}
\sigma_X^2(1) &= \phi_1 \sigma_X^2(0) + \phi_2 \sigma_X^2(1) \\
\sigma_X^2(2) &= \phi_1 \sigma_X^2(1) + \phi_2 \sigma_X^2(0)
\end{aligned}$$

- 两端乘以 X_t，注意到 $\mathrm{cov}(X_t, \varepsilon_t) = \sigma_\varepsilon^2$，可得

$$\sigma_X^2(0) = \phi_1 \sigma_X^2(1) + \phi_2 \sigma_X^2(2) + \sigma_\varepsilon^2$$

Yule-Walker 方程

- 给定平稳 AR(p) 过程 $X_t = \sum_{i=1}^{p} \phi_i X_{t-i} + \varepsilon_t$，特征多项式 $A(z) = 1 - \sum_{i=1}^{p} \phi_i z^i$ 的 p 个零点均位于单位圆之外
- 两端依次乘以 $X_{t-1}, \ldots, X_{t-p}$，可得

$$\begin{bmatrix} \sigma_X^2(1) \\ \vdots \\ \sigma_X^2(p) \end{bmatrix} = \begin{bmatrix} \sigma_X^2(0) & \cdots & \sigma_X^2(p-1) \\ \vdots & \ddots & \vdots \\ \sigma_X^2(p-1) & \cdots & \sigma_X^2(0) \end{bmatrix} \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}$$

- 两端乘以 X_t 可得：

$$\sigma_X^2(0) = \sum_{i=1}^{p} \phi_i \sigma_X^2(i) + \sigma_\varepsilon^2$$

- 上述两组方程，称为 Yule-Walker 方程

本节内容

1. AR 过程的自协方差结构

2. AR 过程的回归估计
   - 基础概念
   - AR(1) 过程的估计
   - AR(p) 过程的估计

统计模型：一般概念

- 统计模型：描述数据（样本）的一个联合分布，即总体分布 (population distribution)

$$\mathbb{P}(\{\boldsymbol{X}_t\}|\boldsymbol{\theta})$$

- $\{\boldsymbol{X}_t\}$ 对应一组观测值，即样本；上述联合分布又可理解为数据产生过程 (data generating process, DGP)
- 其中 $\boldsymbol{\theta}$ 表示这个统计模型的一组参数；即决定数据统计性质的一组参数
- 例如 AR(1) 模型 $X_t = \phi X_{t-1} + \varepsilon_t$，对应的参数向量为 $\boldsymbol{\theta} = (\phi, \sigma_\varepsilon^2)$，决定了数据（样本）的一个特定联合分布

$$\mathbb{P}(\{X_t\}|\boldsymbol{\theta})$$

统计模型的估计：一般概念

- 统计学家（或一般研究者）关心的是从数据反推参数

$$\{\boldsymbol{X}_t\}_{t=1}^T \Rightarrow \boldsymbol{\theta}$$

  换言之，总体分布的参数未知，需要用数据反推
- 这个推测过程称为估计 (estimation)
- 一般来说，问题是需要找到一个函数 $f : \{\boldsymbol{X}_t\}_{t=1}^T \mapsto \hat{\boldsymbol{\theta}}$，从数据出发得到参数 $\boldsymbol{\theta}$ 的估计值 $\hat{\boldsymbol{\theta}}$
- 这个估计值 $\hat{\boldsymbol{\theta}} = f(\{\boldsymbol{X}_t\}_{t=1}^T)$ 需要满足一些条件，如一致性：

$$\lim_{T\to\infty} f(\{\boldsymbol{X}_t\}_{t=1}^T) = \boldsymbol{\theta}$$

  在某种概率收敛意义下成立
  - 例如，$\{X_t\}$ iid 序列，均值为 μ（未知），则大数定律保证，简单的算术平均 $\hat{\mu}_T = \frac{1}{T}\sum_{t=1}^T X_t \xrightarrow{\text{a.s.}} \mu$

AR(1) 的估计：自回归系数

- 假设数据（观测值）序列 $\{X_t\}_{t=1}^{T}$ 满足一个 AR(1) 过程 $X_t = \phi X_{t-1} + \varepsilon_t$，但参数 ϕ, σ_ε 未知
  - 这个 AR(1) 过程就是 $\{X_t\}$ 的 DGP
- 问题：如何估计 ϕ 以及 σ_ε？
- 联想 AR(1) 的 Yule-Walker 方程 $\phi = \sigma_X^2(1)/\sigma_X^2(0)$，可以得到一个 ϕ 的一个估计值

$$\hat{\phi} = \frac{\hat{\sigma}_X^2(1)}{\hat{\sigma}_X^2(0)}$$

其中，$\hat{\sigma}_X^2(1), \hat{\sigma}_X^2(0)$ 分别为 $\{X_t\}$ 的 1-阶样本自协方差与样本方差；后两者的一致性保证了 $\hat{\phi}$ 的一致性

AR(1) 自回归系数的 OLS 估计

- 上面的估计方法利用了 AR(1) 过程的特殊性质
- 更一般的方法，是直接进行线性回归 (linear regression)：

$$X_t = \beta X_{t-1} + e_t$$

  回归系数 β 可以使用一般最小二乘 (ordinary least square, OLS) 进行估计
- OLS 法：最小化残差平方和

$$\min_{\beta} \sum_{t=2}^{T} (X_t - \beta X_{t-1})^2$$

  该最小化问题的解称为 OLS 估计量，记为 $\hat{\beta}$

AR(1) 自回归系数的 OLS 估计

- OLS 估计量的直接推导：最小化问题 1-阶条件

$$\sum_{t=2}^{T} X_{t-1}(X_t - \beta X_{t-1}) = 0 \Rightarrow \hat{\beta}_T = \frac{\sum_{t=2}^{T} X_{t-1}X_t}{\sum_{t=2}^{T} X_{t-1}^2}$$

- 上式中分子分母同时除以 T 可得

$$\hat{\beta}_T = \frac{\frac{1}{T}\sum_{t=2}^{T} X_{t-1}X_t}{\frac{1}{T}\sum_{t=2}^{T} X_{t-1}^2} \xrightarrow{\text{a.s.}} \frac{\sigma_X^2(1)}{\sigma_X^2(0)} = \phi$$

故 X_t 自回归方程系数的 OLS 估计值具有一致性

- 这里的 $\hat{\beta}$ 就是之前介绍的（1-阶）偏自相关系数

OLS 估计的矩阵形式：单一解释变量

- 单一解释变量线性回归的一般写法

$$Y_t = \beta X_t + e_t$$

其向量形式为

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

其中 $\boldsymbol{Y} = [Y_1, \ldots, Y_T]^\intercal$ 为被解释变量，$\boldsymbol{X} = [X_1, \ldots, X_T]^\intercal$ 解释变量，$\boldsymbol{e} = [e_1, \ldots, e_T]^\intercal$ 残差项

- 最小化残差和可表示为

$$\min_{\beta} \boldsymbol{e}^\intercal \boldsymbol{e} = \min_{\beta} (\boldsymbol{Y} - \boldsymbol{X}\beta)^\intercal (\boldsymbol{Y} - \boldsymbol{X}\beta)$$

- 单一解释变量不易体现矩阵形式的优越性；多变量则不同

AR(1) 的估计：冲击项方差

- 暂时回归 σ_ε^2 的估计问题
- 利用大数律，自然的想法是利用样本方差 $\hat{\sigma}_\varepsilon^2 = \frac{1}{T}\sum_{t=1}^T \varepsilon_t^2$ 来估计 σ_ε^2；但唯一的观测值为 $\{X_t\}$，而非 $\{\varepsilon_t\}$
- 但是，若有一致估计量 $\hat{\phi}_T$，则大样本下

$$e_t = X_t - \hat{\phi}_T X_{t-1} \xrightarrow{\text{a.s.}} X_t - \phi X_{t-1} = \varepsilon_t$$

因此，可利用冲击项 ε_t 的样本估计值 e_t 来估计 σ_ε^2:

$$\hat{\sigma}_e^2 = \frac{1}{T-1}\sum_{t=2}^{T} e_t^2 = \frac{1}{T-1}\sum_{t=2}^{T}(X_t - \hat{\phi}_T X_{t-1})^2 \xrightarrow{\text{a.s.}} \sigma_\varepsilon^2$$

AR(p) 的估计：自回归系数

- 假设数据（观测值）序列 $\{X_t\}_{t=1}^T$ 满足一个 AR(p) 过程 $X_t = \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t$，但参数 $\phi_1, \ldots, \phi_p, \sigma_\varepsilon$ 未知
- 类似于 AR(1)，如果有 ϕ_i 的一致估计 $\hat{\phi}_i$，则可计算样本冲击 e_t，从而用 $\hat{\sigma}_e^2$（一致）估计 σ_ε^2
- 从 Yule-Walker 方程出发，可以得到

$$\begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix} = \begin{bmatrix} \sigma_X^2(0) & \cdots & \sigma_X^2(p-1) \\ \vdots & \ddots & \vdots \\ \sigma_X^2(p-1) & \cdots & \sigma_X^2(0) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_X^2(1) \\ \vdots \\ \sigma_X^2(p) \end{bmatrix}$$

AR(p) 自回归系数的 Yule-Walker 估计

- 给定样本 T 足够大，则样本自协方差是对 X_t 总体自协方差的一致估计
- 在 Yule-Walker 方程中应用样本自协方差，则可以得到 $\phi_1, \ldots, \phi_p$ 的一致估计

$$\begin{bmatrix} \hat{\phi}_1 \\ \vdots \\ \hat{\phi}_p \end{bmatrix} = \begin{bmatrix} \hat{\sigma}_X^2(0) & \cdots & \hat{\sigma}_X^2(p-1) \\ \vdots & \ddots & \vdots \\ \hat{\sigma}_X^2(p-1) & \cdots & \hat{\sigma}_X^2(0) \end{bmatrix}^{-1} \begin{bmatrix} \hat{\sigma}_X^2(1) \\ \vdots \\ \hat{\sigma}_X^2(p) \end{bmatrix}$$

AR(p) 自回归系数的 OLS 估计

- AR(p) 的自回归系数也可在线性回归框架下，用 OLS 方法进行估计
- 将 X_t 写为线性回归形式

$$X_t = \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p} + e_t$$

- 则 OLS 估计为最小化残差平方和问题的解

$$\min_{\{\beta_i\}} \sum_{t=p+1}^{T} (X_t - \beta_1 X_{t-1} + \cdots + \beta_p X_{t-p})^2$$

- 若仍然用微积分求解极值问题的 1-阶最优条件来求解，会比较繁琐

OLS 估计的矩阵形式：多解释变量

- 多解释变量线性回归的一般写法

$$Y_t = \mathbf{X}_t^{\mathsf{T}}\boldsymbol{\beta} + e_t, \qquad t = 1, \ldots, T$$

其中 $\mathbf{X}_t^{\mathsf{T}} = [X_{1t}, \ldots, X_{Kt}]$ 为 K 个回归变量

- 其矩阵形式为

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

其中

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ \vdots \\ Y_T \end{bmatrix}_{T\times 1}, \quad \boldsymbol{X} = \begin{bmatrix} \boldsymbol{X}_1^{\mathsf{T}} \\ \vdots \\ \boldsymbol{X}_T^{\mathsf{T}} \end{bmatrix}_{T\times K}, \quad \boldsymbol{e} = \begin{bmatrix} e_1 \\ \vdots \\ e_T \end{bmatrix}_{T\times 1}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_K \end{bmatrix}_{K\times 1}$$

OLS 估计的矩阵形式：多解释变量

- OLS 估计量为最小化残差和问题的解

$$\min_{\boldsymbol{\beta}} \boldsymbol{e}^{\intercal}\boldsymbol{e} = \min_{\boldsymbol{\beta}} (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\intercal}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

- **结论**：若 $\hat{\boldsymbol{\beta}}$ 满足 $\mathbf{X}^{\intercal}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta}) = \mathbf{0}$（零向量），则 $\hat{\boldsymbol{\beta}}$ 为上述最小化问题的解；若 $\mathbf{X}^{\intercal}\mathbf{X}$ 可逆，则 $\hat{\boldsymbol{\beta}} = (\mathbf{X}^{\intercal}\mathbf{X})^{-1}\mathbf{X}^{\intercal}\boldsymbol{Y}$
  - 只需要说明，$\forall \boldsymbol{\beta} \in \mathbb{R}^K$，下列不等式成立即可

$$(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}})^{\intercal}(\boldsymbol{Y} - \boldsymbol{X}\hat{\boldsymbol{\beta}}) \leq (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})^{\intercal}(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{\beta})$$

  - 为此，只需考虑下式的展开

$$(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta})^{\intercal}(\boldsymbol{Y}-\boldsymbol{X}\boldsymbol{\beta}) = (\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta})^{\intercal}(\boldsymbol{Y}-\boldsymbol{X}\hat{\boldsymbol{\beta}}+\boldsymbol{X}\hat{\boldsymbol{\beta}}-\boldsymbol{X}\boldsymbol{\beta})$$

  及 $\hat{\boldsymbol{\beta}}$ 所满足的性质即可

OLS 估计的几何视角

- 将 $\boldsymbol{X}$ 看做 $\mathbb{R}^T$ 中的 K 个列向量，则 $\boldsymbol{X\beta}$ 为这 K 个向量张成的子空间 $\mathcal{X}$ 中的一个向量
- $(\boldsymbol{Y}-\boldsymbol{X\beta})^{\mathsf{T}}(\boldsymbol{Y}-\boldsymbol{X\beta})$ 为 $\boldsymbol{Y}$ 到 $\boldsymbol{X\beta}$ 的距离
- 最小二乘估计就等价于寻找 $\boldsymbol{\beta}$ 使得 $\boldsymbol{Y}$ 到 $\mathcal{X}$ 距离最短
- 这又等价于使得残差向量 $\boldsymbol{e}=\boldsymbol{Y}-\boldsymbol{X\beta}$ 与子空间 $\mathcal{X}$ 垂直，即 $\boldsymbol{X}^{\mathsf{T}}\boldsymbol{e}=\boldsymbol{0}$
- OLS 估计可看做被解释变量对解释变量的投影 (projection)

AR(p) 自回归系数 OLS 估计的一致性

- 将 AR(p) 模型写为 OLS 回归的矩阵形式

$$\boldsymbol{Y} = \boldsymbol{X}^{\top}\boldsymbol{\beta} + \boldsymbol{e}$$

其中

$$\boldsymbol{Y} = \begin{bmatrix} X_{p+1} \\ \vdots \\ X_T \end{bmatrix}_{(T-p)\times 1}, \quad \boldsymbol{X} = \begin{bmatrix} X_p & \cdots & X_1 \\ \vdots & \ddots & \vdots \\ X_{T-1} & \cdots & X_{T-p} \end{bmatrix}_{(T-p)\times K}$$

AR(p) 自回归系数 OLS 估计的一致性

- AR(p) 自回归系数的 OLS 估计为

$$
\begin{aligned}
\hat{\beta}_T &= (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{Y} \\
&= \left(\frac{1}{T}\boldsymbol{X}^\intercal \boldsymbol{X}\right)^{-1} \frac{1}{T}\boldsymbol{X}^\intercal \boldsymbol{Y} \\
&\xrightarrow{\text{a.s.}} \begin{bmatrix} \sigma_X^2(0) & \cdots & \sigma_X^2(p-1) \\ \vdots & \ddots & \vdots \\ \sigma_X^2(p-1) & \cdots & \sigma_X^2(0) \end{bmatrix}^{-1} \begin{bmatrix} \sigma_X^2(1) \\ \vdots \\ \sigma_X^2(p) \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \vdots \\ \phi_p \end{bmatrix}
\end{aligned}
$$

满足一致性

- 一致性的关键：$\mathbb{E}[\varepsilon_t X_{t-i}] = 0,\ \forall i \geq 1$