

国际金融试验班 2019 年秋 · 时间序列

# 第 9 讲：时间序列回归分析拓展

授课人：刘 岩

武汉大学经管学院金融系

2019 年 11 月 21 日

## 本讲内容

- ① 模型设定与评估
- ② 稳健标准误
- ③ MA 模型的估计

## 本节内容

- 1 模型设定与评估
- 2 稳健标准误
- 3 MA 模型的估计

## 动态回归模型

- 单变量自回归模型:

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad t = p + 1, \dots, T$$
$$\Rightarrow X_t = \mathbf{Y}_{t-1}^\top \boldsymbol{\phi} + \varepsilon_t$$

- 添加其他解释变量  $\mathbf{Z}_t$ :

$$X_t = \mathbf{Y}_{t-1}^\top \boldsymbol{\phi} + \mathbf{Z}_t^\top \boldsymbol{\beta} + \varepsilon_t$$

称为动态回归模型

## 动态回归模型

- 动态回归模型示例：货币政策的 Taylor 规则

$$i_t = \rho i_{t-1} + (1 - \rho)[\phi_\pi(\pi_t - \bar{\pi}) + \phi_y(y_t - y_t^*)] + \varepsilon_t$$

其中  $\bar{\pi}$  为目标通胀率， $y_t - y_t^*$  为产出缺口

- OLS 估计的注意事项
  - $X_t$  平稳性取决于  $\phi$  和  $Z_t$ ，后者通常需要假设为平稳序列
  - OLS 估计的一致性取决于  $\varepsilon_t$  与  $(Y_{t-1}, Z_t)$  的相关性  $\Rightarrow$  通常假设  $\mathbb{E}[\varepsilon_t | Y_{t-1}, Z_t] = 0$
  - OLS 估计的渐近分布：拓展鞅差序列假设为  $\mathbb{E}[Y_{t-1}\varepsilon_t | \Omega_{t-1}] = 0$ ,  $\mathbb{E}[Z_t\varepsilon_t | \Omega_{t-1}] = 0$ , 信息集为  $\Omega_t = \{Y_{t-1}, Y_{t-2}, \dots, Z_t, Z_{t-1}, \dots\}$

## 回归分析的命门：内生性与识别问题

- 加总消费为  $C_t$ ，加总产出（收入）为  $Y_t$ ，考虑两个回归

$$\log(C_t) = \alpha \log(Y_t) + \varepsilon_t, \quad \log(Y_t) = \phi \log(C_t) + \eta_t$$

其中， $\varepsilon_t, \eta_t$  表示两种不同的冲击

- 完全不同的两种解释：
  - ①  $\alpha$  的估计值表示边际消费倾向 (marginal propensity to consume)
  - ②  $\phi$  的估计值表示消费乘数 (consumption multiplier)
- 两个回归方程联立得

$$\log(C_t) = \frac{\varepsilon_t + \alpha \eta_t}{1 - \alpha \phi}, \quad \log(Y_t) = \frac{\phi \varepsilon_t + \eta_t}{1 - \alpha \phi}$$

## 回归分析的命门：内生性与识别问题

- $\alpha$  和  $\phi$  在二战后的凯恩斯主义 (Keynesian) 经济学中都是反映经济运动规律的关键变量
  - 还有投资（储蓄）乘数、政府支出乘数、净出口乘数等
- 问题：两个回归方程的 OLS 估计，都是不一致的！
  - 令  $c_t = \log(C_t)$ ,  $y_t = \log(Y_t)$ , 则

$$\hat{\alpha} = \alpha + \left(\frac{1}{T} \sum_t y_t^2\right)^{-1} \frac{1}{T} \sum_t y_t \varepsilon_t \xrightarrow{\text{a.s.}} \alpha + [\mathbb{E}y_t^2]^{-1} \mathbb{E}y_t \varepsilon_t$$

$$\text{但 } \mathbb{E}y_t \varepsilon_t = \frac{\phi}{1-\alpha\phi} \sigma_\varepsilon^2$$

- $\hat{\phi}$  有同样问题
- 回归分析中的内生性 (endogeneity): 若回归变量与残差项协方差不为 0, 则称回归存在内生性问题  $\Rightarrow$  OLS 估计结果是不一致的

## 回归分析的命门：内生性与识别问题

- 回归存在内生性，又称回归变量的外生性 (exogeneity) 条件不满足
- 此时，OLS 估计方法无法识别 (identify) 模型参数
  - 又称为存在参数识别问题 (identification problem)
- 更重要的影响：内生性带来的参数识别问题，妨碍有效的因果性推断 (causal inference)
  - 所得结论仅为相关性 (correlation)  $\neq$  因果性
  - 而经济学的核心问题之一：研究  $X$  的变动如何引起  $Y$  的变动



## 内生性问题的 3 大来源

- ① 遗漏变量 (omitted variable): 真实模型是  $Y_t = \alpha X_t + \gamma Z_t + \varepsilon_t$ , 但回归模型仅考虑  $X_t$ ,

$$Y_t = \alpha X_t + \underbrace{u_t}_{=\gamma Z_t + \varepsilon_t}$$

- ② 倒向因果 (inverse causality): 回归模型  $Y_t = \alpha X_t + \gamma Z_t + \varepsilon_t$ , 但实际上  $Z_t$  是由  $Y_t$  所决定

$$Z_t = \beta Y_t + v_t \Rightarrow \mathbb{E}Z_t \varepsilon_t \neq 0$$

- ③ 共同因素 (common/confounding factor): 前述消费-产出回归就是一个典型例子——宏观经济变量几乎都受到共同冲击(因素)的作用

## 内生性问题的解决

- 近 20 年：经济学实证研究领域的因果推断 (causal inference) 革命
  - 今年的 3 位诺奖 Banerjee, Duflo, and Kreme 就在发展经济学微观实证研究领域推广随机控制实验 (randomized controlled trials, RCT) 方法取得突出成就
- 常见方法：非随机实验方法
  - 工具变量 (instrumental variables)
  - 双重差分 (difference in difference, DID)
  - 断点回归 (regression discontinuity)
- 参考书
  - Angrist and Pischke, *Mostly Harmless Econometrics*, 2009
  - 赵西亮, 《基本有用的计量经济学》, 2017

回归模型的评估： $R^2$  与调整  $R^2$ 

- 给定回归方程  $Y_t = X_t^T \beta + \varepsilon_t$ , 称  $\hat{Y}_t = X_t^T \hat{\beta}$  为回归估计的拟合值 (fitted value)
- 用矩阵表示  $\hat{Y} = X\hat{\beta} = X(X^T X)^{-1} X^T Y$
- 回归残差的估计值为  $\hat{\varepsilon} = Y - \hat{Y}$ , 可验证  $\varepsilon \perp \hat{Y}$ , 即  $\varepsilon^T \hat{Y} = 0 \Rightarrow \{Y_t\}$  样本变动 (方差) 可分解为

$$\sum_t Y_t^2 = Y^T Y = \hat{Y}^T \hat{Y} + \hat{\varepsilon}^T \hat{\varepsilon} = \sum_t \hat{Y}_t^2 + \sum_t \hat{\varepsilon}_t^2$$

- 定义

$$R^2 = \frac{\sum_t \hat{Y}_t^2}{\sum_t Y_t^2} = \frac{\hat{Y}^T \hat{Y}}{Y^T Y} = 1 - \frac{\hat{\varepsilon}^T \hat{\varepsilon}}{Y^T Y} \in [0, 1]$$

- $R^2$  可理解为回归模型对  $Y_t$  的解释力 (explanatory power)

回归模型的评估： $R^2$  与调整  $R^2$ 

- 只有一个常数回归变量即  $\mathbf{X}_t^\top = 1$  时， $\hat{Y}_t = \mu_Y$ ，此时有  $R^2 = T\mu_Y^2 / \mathbf{Y}^\top \mathbf{Y}$
- 增加回归变量可以提高  $R^2$ ：定义  $\tilde{\mathbf{X}} = [\mathbf{X}, \mathbf{X}_{K+1}]$ ， $\mathbf{b} = [\boldsymbol{\beta}^\top, \beta_{K+1}]^\top$ ，考虑回归  $\mathbf{Y} = \tilde{\mathbf{X}}\mathbf{b} + \mathbf{e}$ ，则

$$\tilde{\mathbf{Y}} = \tilde{\mathbf{X}}\hat{\mathbf{b}} \Rightarrow \tilde{\mathbf{Y}}^\top \tilde{\mathbf{Y}} \geq \hat{\mathbf{Y}}^\top \hat{\mathbf{Y}} \Rightarrow R^2 \uparrow$$

- 极端情况：如果  $K = T$ ，且  $\mathbf{X}$  满秩，则由  $\mathbf{X}\boldsymbol{\beta} = \mathbf{Y}$  可直接解出  $\boldsymbol{\beta} \Rightarrow \boldsymbol{\varepsilon} = \mathbf{0}$ ， $R^2 = 1$
- 克服解释变量个数对解释力的影响：定义调整  $R^2$

$$\bar{R}^2 = 1 - \frac{\hat{\boldsymbol{\varepsilon}}^\top \hat{\boldsymbol{\varepsilon}} / (T - K - 1)}{\mathbf{Y}^\top \mathbf{Y} / (T - 1)} = 1 - (1 - R^2) \frac{T - 1}{T - K - 1}$$

## 模型选择

- 建立回归模型：首先决定选取哪些变量作为解释变量
  - 理想状态：经济、金融理论“完全决定”选取哪些解释变量
  - 实际情况：理论只能做部分决定  $\Rightarrow$  部分变量仍需人工决策
- 模型选择 (model selection)：根据给定的信息准则 (information criterion) 进行变量选取
  - AIC (Akaike's IC):  $AIC = \log \hat{\sigma}_{\varepsilon, K}^2 + 2K/T$ ; 小样本效果较好
  - BIC (Bayesian IC):  $BIC = \log \hat{\sigma}_{\varepsilon, K}^2 + (K \log T)/T$ ; 大样本效果较好
- 从备选解释变量中，选择对应 IC 最小的一组变量

## 本节内容

- 1 模型设定与评估
- 2 稳健标准误**
- 3 MA 模型的估计

## 残差项：同方差 v.s. 异方差

- 到目前为止，均使用同方差 (homoskedasticity) 假设：

$$\mathbb{E}[\varepsilon_t^2 | \mathbf{X}_t] = \sigma_\varepsilon^2$$

- 该假设排除了残差项波动率随时间或解释变量的改变而改变的情况
- 更一般假设：条件异方差 (conditional heteroskedasticity)

$$\mathbb{E}[\varepsilon_t^2 | \mathbf{X}_t] = \sigma_{\varepsilon,t}^2$$

## 中心极限定理：异方差情形

## 定理 1 (鞅差序列 CLT：多元、异方差情形)

假设  $\{\mathbf{X}_t\}$  为 (向量) 鞅差序列,  $\mathbb{E}[\mathbf{X}_t | \Omega_{t-1}] = \mathbf{0}$ ,  $\Omega_t$  为  $t$ -时的信息集:  $\mathbb{E}[\mathbf{X}_t \mathbf{X}_t^\top] = \Sigma_t$  为  $t$ -时协方差矩阵。若序列平均协方差矩阵收敛

$$\frac{1}{T} \sum_t \Sigma_t \rightarrow \Sigma,$$

且样本交叉矩收敛到同样矩阵

$$\frac{1}{T} \sum_t \mathbf{X}_t \mathbf{X}_t^\top \xrightarrow{\text{a.s.}} \Sigma,$$

则对应的样本均值向量有如下极限分布

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



## OLS 估计的渐近分布：异方差情形

- 考察  $\hat{\beta}_T$  的渐近分布

$$\sqrt{T}(\hat{\beta}_T - \beta_0) = \left( \frac{1}{T} \sum_{t=1}^T \mathbf{X}_t \mathbf{X}_t^\top \right)^{-1} \sqrt{T} \frac{1}{T} \sum_{t=1}^T \underbrace{\mathbf{X}_t \varepsilon_t}_{\text{鞅差序列}}$$

- 样本均值为  $\mu_{\mathbf{X}\varepsilon, T} = \frac{1}{T} \sum_t \mathbf{X}_t \varepsilon_t$ , 需要确定  $\sqrt{T} \mu_{\mathbf{X}\varepsilon, T}$  的分布  
 $\Rightarrow$  确定其极限协方差矩阵
- 将 OLS 估计同方差假设替换为如下 **异方差假设**:

$$\frac{1}{T} \sum_t \varepsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top \xrightarrow{\text{a.s.}} \Sigma \equiv \lim_T \frac{1}{T} \sum_t \mathbb{E}[\varepsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top]$$

- 由前页 CLT 知

$$\sqrt{T} \frac{1}{T} \sum_t \mathbf{X}_t \varepsilon_t \xrightarrow{d} N(\mathbf{0}, \Sigma)$$

## OLS 估计的渐近分布：异方差情形

- 继续假设  $\frac{1}{T} \sum_t \mathbf{X}_t \mathbf{X}_t^\top = \mathbf{M}$ ，则此时

$$\sqrt{T}(\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0) \xrightarrow{d} N(\mathbf{0}, \mathbf{M}^{-1} \boldsymbol{\Sigma} \mathbf{M}^{-1})$$

- 故当  $T$  足够大时， $\hat{\boldsymbol{\beta}}_T - \boldsymbol{\beta}_0 \sim N(\mathbf{0}, \frac{1}{T} \mathbf{M}^{-1} \boldsymbol{\Sigma} \mathbf{M}^{-1})$
- 由条件异方差假设，

$$\begin{aligned} \mathbb{E}[\varepsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top] &= \mathbb{E} \mathbb{E}[\varepsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top | \mathbf{X}_t] \\ &= \mathbb{E}[\mathbf{X}_t \mathbf{X}_t^\top \mathbb{E}(\varepsilon_t^2 | \mathbf{X}_t)] = \mathbb{E}[\sigma_{\varepsilon,t}^2 \mathbf{X}_t \mathbf{X}_t^\top] \end{aligned}$$

- 对比同方差假设： $\mathbb{E}[\varepsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top] = \sigma_\varepsilon^2 \mathbb{E}[\mathbf{X}_t \mathbf{X}_t^\top] = \sigma_\varepsilon^2 \mathbf{M}$

## OLS 估计的渐近分布：异方差情形

- 给定样本  $\{Y_t, \mathbf{X}_t\}_{t=1}^T$ ，需要估计  $M$  与  $\Sigma$
- $M$  的估计与同方差情形一致： $\hat{M} = \frac{1}{T} \sum_t \mathbf{X}_t \mathbf{X}_t^\top = \frac{1}{T} \mathbf{X}^\top \mathbf{X}$
- $\Sigma$  的估计：首先利用  $\hat{\beta}$  得到  $\hat{\varepsilon}_t = Y_t - \mathbf{X}_t^\top \hat{\beta}$ ，进而由

$$\frac{1}{T} \sum_t \varepsilon_t^2 \mathbf{X}_t \mathbf{X}_t^\top \xrightarrow{\text{a.s.}} \Sigma$$

$$\Rightarrow \hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T \hat{\varepsilon}_t^2 \mathbf{X}_t \mathbf{X}_t^\top = \frac{1}{T} \mathbf{X}^\top \text{diag}(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_T^2) \mathbf{X}$$

- 样本渐近分布： $\hat{\beta}_T - \beta_0 \sim N(0, \frac{1}{T} \hat{M}^{-1} \hat{\Sigma} \hat{M}^{-1})$
- 样本渐近协方差矩阵  $\frac{1}{T} \hat{M}^{-1} \hat{\Sigma} \hat{M}^{-1}$  对角线第  $k$  个元素平方根：  
 $\hat{\beta}_{k,T}$  的 稳健标准误 (robust s.e.)

## 本节内容

- 1 模型设定与评估
- 2 稳健标准误
- 3 MA 模型的估计**

## MA 模型的估计

- 考虑 MA( $q$ ) 模型:  $X_t = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} = B(\mathcal{L})\varepsilon_t$ 
  - $X_t$  可逆  $\Leftrightarrow B(z) = 1 + \theta_1 z + \cdots + \theta_q z^q$  的零点  $z_1, \dots, z_q$  均位于单位圆之外
- 若  $X_t$  可逆, 则  $B^{-1}(\mathcal{L})X_t = \varepsilon_t$ , 是一个 AR( $\infty$ ) 过程:

$$B^{-1}(z) = \frac{1}{\left(1 - \frac{z}{z_1}\right) \cdots \left(1 - \frac{z}{z_q}\right)} = 1 + \sum_{i=1}^{\infty} \phi_i z^i$$

$$\Rightarrow X_t = \sum_{i=1}^{\infty} \phi_i X_{t-i} + \varepsilon_t$$

## MA 模型的近回归估计

- 首先将  $X_t$  看做一个 AR( $p$ ) 过程,  $p$  较大, 进行 OLS 估计, 得到残差序列  $\{\hat{\varepsilon}_t\}$
- 再用  $\hat{\varepsilon}_t$  代替  $\varepsilon_t$ , 估计回归方程

$$X_t = \hat{\varepsilon}_t + \vartheta_1 \hat{\varepsilon}_{t-1} + \cdots + \vartheta_q \hat{\varepsilon}_{t-q} + e_t$$

- 可以证明, 当样本量  $T \rightarrow \infty$  时,  $\hat{\vartheta}_j \xrightarrow{\text{a.s.}} \theta_j$
- 也可以直接计算  $X_t$  自协方差关于  $\theta_1, \dots, \theta_q$  的表达式, 然后用  $X_t$  的样本自协方差对  $\{\theta_j\}$  进行估计

## MA 模型的极大似然估计

- 另一种估计办法是 ML 估计，但需要额外的分布假设
- 若  $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ ，则  $\mathbf{X} = [X_{q+1}, \dots, X_T]^\top$  服从多元正态分布  $N(0, \Sigma)$ 
  - $\Sigma$  的每个元素都是  $\boldsymbol{\theta} = [\theta_1, \dots, \theta_q]^\top$  与  $\sigma_\varepsilon^2$  的函数
  - 可以记做  $\Sigma(\boldsymbol{\theta}, \sigma_\varepsilon^2)$
- 可以写出对数似然函数

$$\log L(\mathbf{X} | \boldsymbol{\theta}, \sigma_\varepsilon^2) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma - \frac{1}{2} \mathbf{X}^\top \Sigma^{-1} \mathbf{X}$$

## 线性回归模型的极大似然估计

- 考虑线性回归模型:  $Y_t = \mathbf{X}_t^\top \boldsymbol{\beta} + \varepsilon_t, t = 1, \dots, T$
- 若残差为正态独立同分布  $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ , 则可将样本数据  $(\mathbf{Y}, \mathbf{X}) = \{Y_t, \mathbf{X}_t\}$  的似然值写为

$$L(\mathbf{Y}, \mathbf{X} | \boldsymbol{\beta}, \sigma_\varepsilon^2) = \frac{1}{\sqrt{(2\pi)^T \sigma_\varepsilon^{2T}}} \exp \left\{ -\frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T (Y_t - \mathbf{X}_t^\top \boldsymbol{\beta})^2 \right\}$$

- ML 估计:  $\max_{\boldsymbol{\beta}, \sigma_\varepsilon^2} L(\mathbf{Y}, \mathbf{X} | \boldsymbol{\beta}, \sigma_\varepsilon^2)$ , 等价于最大化对数似然值

$$\max_{\boldsymbol{\beta}, \sigma_\varepsilon^2} \log L(\mathbf{Y}, \mathbf{X} | \boldsymbol{\beta}, \sigma_\varepsilon^2)$$



## 线性回归模型的极大似然估计

- 对数似然函数：

$$\begin{aligned} \log L(\mathbf{Y}, \mathbf{X} | \boldsymbol{\beta}, \sigma_\varepsilon^2) = & \\ & -\frac{T}{2} \log 2\pi - \frac{T}{2} \log \sigma_\varepsilon^2 - \frac{1}{2\sigma_\varepsilon^2} \sum_{t=1}^T (Y_t - \mathbf{X}_t^\top \boldsymbol{\beta})^2 \end{aligned}$$

- 考虑  $\boldsymbol{\beta}$  的 ML 估计：

$$\max_{\boldsymbol{\beta}} \log L \Leftrightarrow \min_{\boldsymbol{\beta}} \sum_{t=1}^T (Y_t - \mathbf{X}_t^\top \boldsymbol{\beta})^2$$

故， $\hat{\boldsymbol{\beta}}_{ML,T} = \hat{\boldsymbol{\beta}}_{OLS,T} = (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{Y}$

- 考虑  $\hat{\sigma}_{\varepsilon,ML}^2 \Rightarrow$  计算  $\partial_{\sigma_\varepsilon^2} \log L(\mathbf{Y}, \mathbf{X} | \boldsymbol{\beta}, \sigma_\varepsilon^2)$

## ARMA 模型的极大似然估计

- 考虑 ARMA( $p, q$ ) 模型  $A(\mathcal{L})X_t = B(\mathcal{L})\varepsilon_t$

$$X_t = \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}, \quad t = 1, \dots, T$$

- 假设  $\varepsilon_t \stackrel{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$ , 则  $A(\mathcal{L})X_t, t = 1, \dots, T$ , 服从多元正态分布  $N(\mathbf{0}, \Sigma)$ , 其中  $\Sigma$  为  $\theta_1, \dots, \theta_q, \sigma_\varepsilon^2$  的函数
- 进一步的, 定义  $\mathbf{Y}_t = X_t - \phi_1 X_{t-1} - \dots - \phi_p X_{t-p}$ , 以及  $\mathbf{Y} = [\mathbf{Y}_1, \dots, \mathbf{Y}_T]^\top$ , 则可以写出样本  $\mathbf{X} = \{X_t\}_{1-p}^T$  对应的 (对数) 似然函数

$$\log L(\mathbf{X} | \boldsymbol{\theta}, \sigma_\varepsilon^2) = -\frac{T}{2} \log 2\pi - \frac{1}{2} \log \det \Sigma - \frac{1}{2} \mathbf{Y}^\top \Sigma^{-1} \mathbf{Y}$$

- 数值求解系数  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2$  的极大似然估计