国际金融试验班 2021 年秋・时间序列

第 4 讲：时间序列数据基本特征

授课人：刘　岩

武汉大学经管学院金融系

2021 年 9 月 28 日

本讲内容

1. 基本统计特征
2. 平稳性的基本概念
3. 季节性与趋势性

本节内容

时间序列数据主要统计特征

与普通数据序列一样，时间序列数据 $\{X_t\}$ 主要统计特征包括：

- 期望：$\mathbb{E}X_t$
- 方差、标准差：$\mathrm{var}(X_t), \sigma_X$
- 以及两个变量间的协方差：$\mathrm{cov}(X_t, Y_t)$ 及相关系数 ρ_{XY}

但与普通数据不同，时间序列数据的主要统计特征还包括

- 自协方差 (autocovariance)：$\mathrm{cov}(X_t, X_{t-k}),\ k \geq 1$
- 对应的，自相关系数 (autocorrelation)：

$$\rho_k = \frac{\mathrm{cov}(X_t, X_{t-k})}{\sigma_X^2}$$

- 注意，上述各阶矩和 t 无关 $\Rightarrow$ 平稳性假设

矩的估计

- 给定时间序列数据 $\{X_t\}_{t=1}^T$，各阶矩的估计 (estimate) 如下
- 期望（均值）：$\hat{\mu}_X = \frac{1}{T}\sum_t X_t$；$\sum_t$ 表示对 $t = 1, \ldots, T$ 求和
- 方差：

$$\hat{\sigma}_X^2 = \begin{cases} \frac{1}{T-1}\sum_t (X_t - \hat{\mu}_X)^2, & \text{unbiased if iid} \\ \frac{1}{T}\sum_t (X_t - \hat{\mu}_X)^2, & \text{biased} \end{cases}$$

  大样本之下，无偏 (unbiased) 估计与有偏 (biased) 估计等价
- 协方差：

$$\hat{\sigma}_{XY} = \begin{cases} \frac{1}{T-1}\sum_t (X_t - \hat{\mu}_X)(Y_t - \hat{\mu}_Y), & \text{unbiased if iid} \\ \frac{1}{T}\sum_t (X_t - \hat{\mu}_X)(Y_t - \hat{\mu}_Y), & \text{biased} \end{cases}$$

- 相关系数：$\hat{\rho}_{XY} = \hat{\sigma}_{XY}/(\hat{\sigma}_X\hat{\sigma}_Y)$

矩的估计

- 自协方差：

$$\hat{\sigma}_k^2 = \frac{1}{T-k} \sum_{t=k+1}^{T} (X_t - \hat{\mu}_X)(X_{t-k} - \hat{\mu}_X)$$

  注意求和中 t 的取值范围

- 自相关系数：

$$\hat{\rho}_k = \frac{\hat{\sigma}_k^2}{\hat{\sigma}_X^2}$$

练习　估计值 $\hat{\rho}_{XY}, \hat{\rho}_k$ 是否仍然介于 ± 1 之间？

- 在上下文区分明确时，也可以略去“ˆ”符号
- 但有时为明确样本量，也会增加 T 下标，如 $\hat{\mu}_{X,T}$

平稳时间序列的大数定律

给定时间序列数据 $\{X_t\}_{t=1}^{T}$，上述矩估计的理论基础在于大数定律 (law of large number)

定理 1

当 X_t 满足特定条件时，样本均值以概率 1 收敛到总体均值

$$\lim_{T\to\infty} \hat{\mu}_{X,T} = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} X_t = \mathbb{E}X_t$$

大数定律：进一步讨论

- 特定条件：弱于 $\{X_t\}$ iid；本课程中，可以理解为“平稳性条件”——严格说法是“严平稳遍历性条件”
- 收敛：随机变量的收敛概念有 4 种，依分布收敛 $\xrightarrow{d}$，依概率收敛 $\xrightarrow{p}$，矩收敛 $\xrightarrow{m}$，概率-1 收敛（几乎处处收敛）$\xrightarrow{a.s.}$
  - 假设有 r.v. 序列 $\{Z_t\}_{t=1}^{\infty}$ 和 r.v. Z，则称
    1. $Z_t \xrightarrow{d} Z$，若 $\lim_{t\to\infty} F_{Z_t}(z) = F_Z(z)$ 对几乎所有 z 成立
    2. $Z_t \xrightarrow{p} Z$，若 $\forall \varepsilon > 0$，$\lim_{t\to\infty} \mathbb{P}(|Z_t - Z| \geq \varepsilon) = 0$
    3. $Z_t \xrightarrow{m} Z$，若 $\lim_{t\to\infty} \mathbb{E}[(Z_t - Z)^2] = 0$
    4. $Z_t \xrightarrow{a.s.} Z$，若 $\mathbb{P}(\lim_{t\to\infty} Z_t = Z) = 1$
  - 4 种收敛的关系

$$\left\{ \begin{array}{c} \xrightarrow{a.s.} \\ \xrightarrow{m} \end{array} \right\} \Rightarrow \xrightarrow{p} \Rightarrow \xrightarrow{d}$$

  - 大数定律中“收敛”可以达到 $\xrightarrow{a.s.}$

本节内容

平稳性的两个定义

给定双边无穷时间序列 $\{X_t\}_{t=-\infty}^{\infty}$

1. 协方差平稳性 (covariance stationarity)：$\{X_t\}$ 称为协方差平稳序列，如果 $\mathbb{E}X_t = \mu$，$\text{cov}(X_t, X_{t-k}) = \sigma_k^2$，$k = 0, \ldots, \infty$，均不依赖于 t
   - 协方差平稳性又称为 2-阶矩平稳或矩平稳
2. 严平稳性 (strict stationarity)：$\{X_t\}$ 称为严平稳序列，如果 $\forall k = 0, \ldots, \infty$,
$$(X_t, \ldots, X_{t-k})$$
的分布与 t 无关
3. 本门课中，“平稳”序列同时满足上述两个定义
   - 这两个定义互相不蕴含对方

平稳性与趋势、季节性

- 平稳性与趋势不兼容
  - X_t 平稳，则 $\mathbb{E}X_t = \mu$ 为常数，且大数定律意味着 $\frac{1}{T}\sum_{t=1}^{T} X_t \to \mu$；均值回归倾向
  - 趋势意味着 X_t 的样本均值不会收敛
- 平稳性与季节性兼容
  - 季节性变量：$\{S_t\}$ 满足 $S_{t+p} = S_t, \forall t$，p 表示季节周期
  - 对任意 t，$\sum_{0\le k\le p-1} S_{t+k}$ 为常数
  - 如果把 $X_t = Y_t + S_t$，Y_t 平稳，且将 X_t 的期望理解为一个季节周期上的均值，则 X_t 也平稳
  - 例如，令 $U \sim \mathcal{U}([-\pi,\pi])$，$X_t = \cos(\pi t + U)$，$t \in \mathbb{Z}$，可验证 $\mathbb{E}X_t$ 及 $\sigma_k^2 = \mathrm{cov}(X_{t+k}, X_t)$ 与 t 无关；但 σ_k^2 呈现周期性

平稳性和平稳分布

- 平稳性定义中的期望和协方差，都是在平稳分布 (stationary distribution) 下计算的
    - 这个分布就是严平稳概念中的 1-元变量 X_t 服从的分布
- 一般而言，随机过程是否存在平稳分布是个问题
    - $X_0 = 0$，$X_t = X_{t-1} + \varepsilon_t$，$\varepsilon_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\varepsilon^2)$，就不存在平稳分布
    - 给定一个有限状态马氏链 $X_t \in \{1, 2\}$，转移矩阵为

$$\mathbf{P} = \begin{bmatrix} \rho & 1-\rho \\ 1-\rho & \rho \end{bmatrix}$$

$\rho \in (0, 1)$，则 X_t 有平稳分布 $\mathbf{u} = \left[\frac{1}{2}, \frac{1}{2}\right]$

- 即便随机过程存在平稳分布，从任何一个起点 $X_0 = \bar{x}$ 开始，X_t 的分布会不会逐渐收敛到平稳分布，也是一个问题

平稳性的关键性质

定理 2

若 $\{X_t\}$ 是平稳序列，对任意的 $k+1$-元可测函数 f，定义

$$Y_t = f(X_t, \ldots, X_{t-k})$$

则 $\{Y_t\}$ 也是平稳序列。

随机收敛的性质

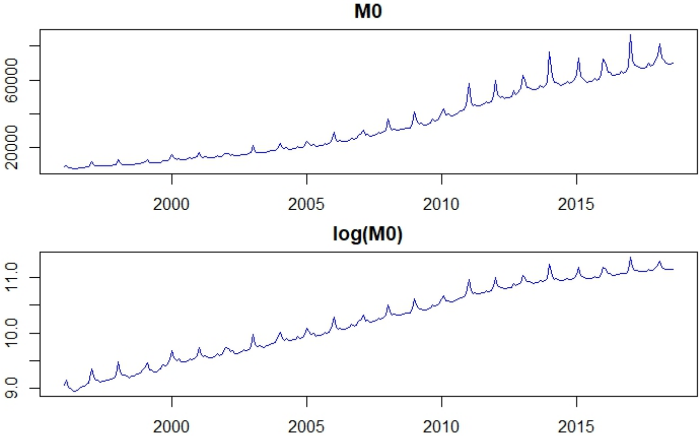
定理 3

假设 $X_t \xrightarrow{\text{a.s.}} X$, $Y_t \xrightarrow{\text{a.s.}} Y$，则对任意的可测函数 f，有

$$f(X_t, Y_t) \xrightarrow{\text{a.s.}} f(X, Y).$$

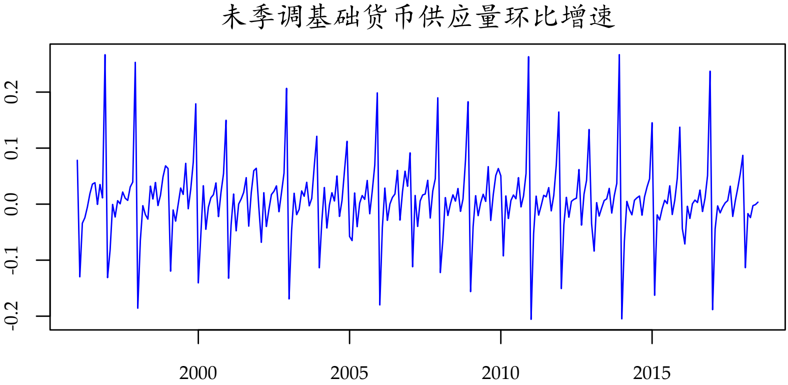
本节内容

季节性示例 1



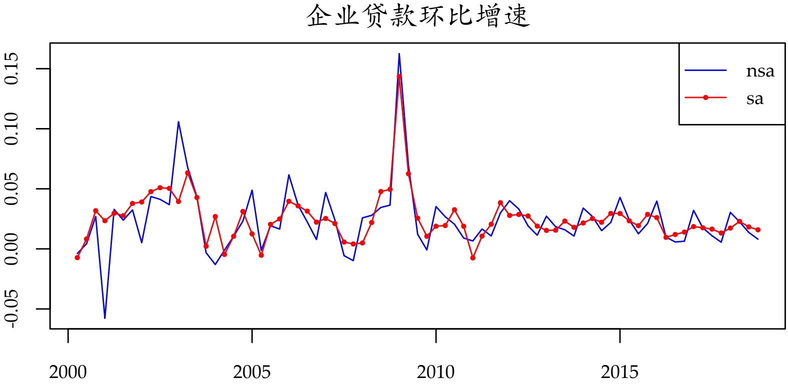
需要对原始数据进行季节调整 (seasonal adjustment)

示例 2：M0 环比增速



- 该序列未经过季调，可以看到明显的季节性周期性波动特征；R 中季节调整的包为 `seas`

示例 3：贷款环比增速

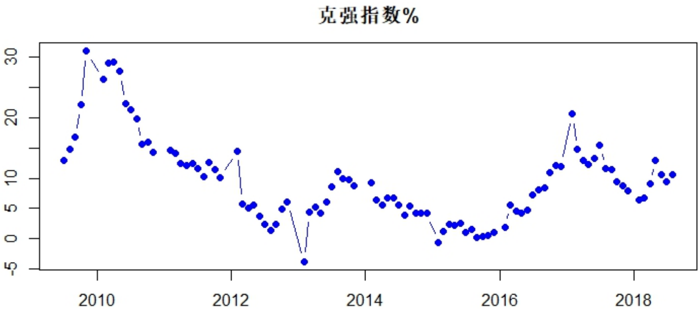


nsa: not seanonally adjusted　　sa: seasonally adjusted

- 未经季节调整的序列，有年度（4 季度）周期波动特征

时间序列的趋势性

- 时间序列关心数据在时间上的变动

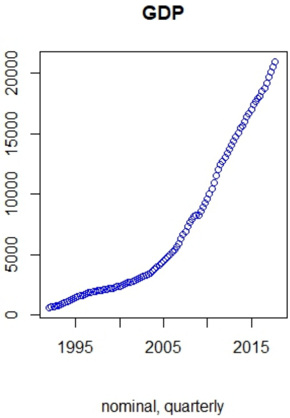


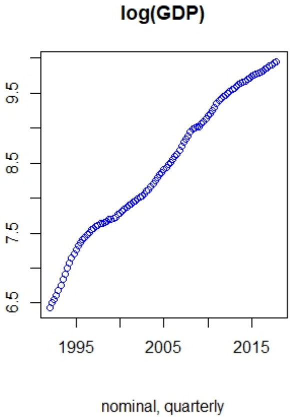
$\hat{\mu} = 9.53,\ \hat{\sigma} = 6.98,\ \hat{\rho} = 0.92$

- 有趋势 vs. 无趋势

趋势的常见类型

- 指数趋势 vs. 线性趋势





- 两者之间差一个对数变换：$X_t \Rightarrow \log(X_t)$

线性趋势调整：基本方法

- 原始数据 $\{X_t\}_{t=1}^T$ 包含线性趋势 $X_t = \alpha t + \tilde{X}_t$，而 $\tilde{X}_t$ 无趋势
- 两端取差分

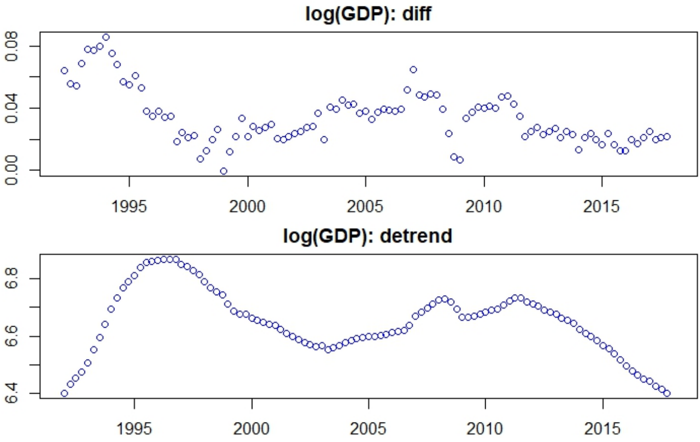
$$\Delta X_t = X_t - X_{t-1} = \alpha + \tilde{X}_t - \tilde{X}_{t-1} = \alpha + \Delta \tilde{X}_t$$

  其中 Δ 表示差分算子，则 ΔX_t 不包含趋势
- 或者，如果要求 $\tilde{X}_1 = \tilde{X}_T$，则有

$$\alpha = \frac{X_T - X_1}{T - 1}$$

  从而可以推算 $\tilde{X}_t$
- 类似的过程称为去趋势 (detrend)

去趋势示例


log(GDP): diff
0.00
0.04
0.08
1995
2000
2005
2010
2015
log(GDP): detrend
6.4
6.6
6.8
1995
2000
2005
2010
2015

时间序列的周期项与趋势项

- 应用时间序列分析的一个重要内容：区分时间序列的周期项 (cycle) 与趋势项 (trend)

$$X_t = X_t^{\text{cycle}} + X_t^{\text{trend}}$$

- 时间序列的谱函数：揭示出时间序列的许多周期信息
  - 高频部分：短周期，$1/\omega$ 小
  - 低频部分：长周期，$1/\omega$ 大
- 时间序列的趋势：可以理解为序列中对应低频周期项的部分

$$X_t^{\text{trend}} = X_t^{\text{low freq}} \qquad X_t^{\text{cycle}} = X_t^{\text{high freq}}$$

- 时间序列低频与高频、趋势与周期的分离，称为时间序列的滤波 (filter)

HP 滤波

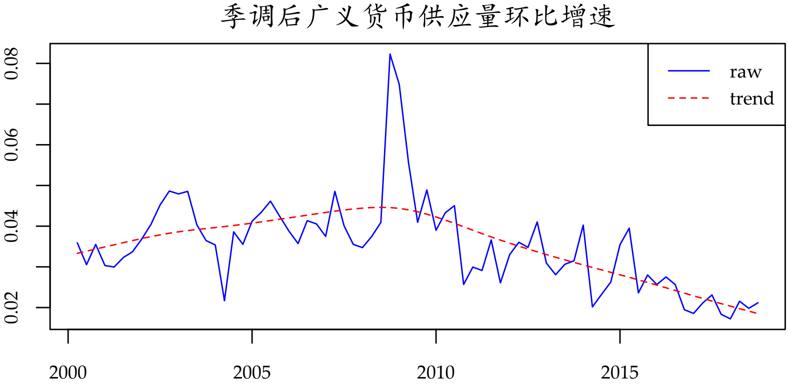
- 给定样本序列 $\{X_t\}_{t=1}^{T}$，HP 滤波 (HP filter) 通过求解最小化问题获得趋势项 $\{\tilde{X}_t\}$：

$$\{\tilde{X}_t\} = \operatorname{argmin}\left\{\sum_{t=1}^{T}(X_t - \tilde{X}_t)^2 + \lambda \sum_{t=2}^{T-1}(\tilde{X}_{t+1} - 2\tilde{X}_t + \tilde{X}_{t-1})\right\}$$

  其中 $X_t - \tilde{X}_t$ 即为滤波所得周期项

- 对年度、季度、月度数据，λ 分别取值 6.25、1,600 与 129,600
- HP 滤波是经济中最流行的滤波算法，为 Hodrick & Prescott 的缩写
  - R 中 `mFilter` 包有 HP 滤波函数 `hpfilter`
  - 更理想的滤波算法为 BP (band pass) 滤波：只保留特定频率区间的数据周期波动

HP 滤波示例：M2 增速



- HP 滤波有效去除了原始增速序列中低频、长周期（趋势）部分；且允许非线性趋势