2021 秋季本科时间序列

第 7 次作业

提交日期：11 月 30 日

1. 请计算 AR(2, 1) 过程 $X_t = 1.3X_{t-1} - 0.4X_{t-2} + \varepsilon_t + 0.6\varepsilon_{t-1}$ 的谱密度函数并绘图，其中 $\sigma_\varepsilon^2 = 1$。比较 AR(2) 过程 $X_t = 1.3X_{t-1} - 0.4X_{t-2} + \varepsilon_t$ 的谱密度，从而说明冲击（残差）项为 MA(1) 时，序列具有更多的高频波动特征。

2. 将 AR(2) 过程 $X_t = 1.3X_{t-1} - 0.4X_{t-2} + \varepsilon_t$ 写为

$$\begin{bmatrix} X_{t+1} \\ X_t \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} X_t \\ X_{t-1} \end{bmatrix} + \begin{bmatrix} \varepsilon_{t+1} \\ 0 \end{bmatrix}$$

   的形式，并说明对任意的 $s \geq 2$，最优线性预测（向量）具有如下递归形式：

$$\begin{bmatrix} \hat{X}_{t+s|t} \\ \hat{X}_{t+s-1|t} \end{bmatrix} = \boldsymbol{A} \begin{bmatrix} \hat{X}_{t+s-1|t} \\ \hat{X}_{t+s-2|t} \end{bmatrix},$$

   从而利用 $\boldsymbol{A}$ 的特征值分解，对任意 $s \geq 1$，计算 $\hat{X}_{t+s|t}$ 的通项表达式，并计算 $\lim_{s\to\infty} \hat{X}_{t+s|t}$。

3. [附加题 10 分，即本次作业满分 110] 计算上问中 s-步均方预测误差方差的极限

$$\lim_{s\to\infty} \mathbb{E}[X_{t+s} - \hat{X}_{t+s|t}]^2.$$

4. 考虑 MA(1) 过程 $X_t = \varepsilon_t + \theta\varepsilon_{t-1}$，其中 $\sigma_\varepsilon^2 = 1$。

   (a) 计算最优线性预测 $\hat{X}_{t+s|t}$ 的表达式，$s \geq 1$。提示：t 期信息集为 $\{X_t, X_{t-1}, X_{t-2}, \ldots\}$，但真正对预测 X_{t+s} 有用的变量个数有限，对不同的 s 取值，首先确定信息集中哪些 X_{t-i} 对预测有用。

   (b) 给定样本 $\{X_t : t = 1, \ldots, T\}$，请设计一个利用 $\{X_t\}$ 样本矩估计 θ 估计量，并说明该估计量是否具有一致性。提示：首先计算 X_t 的自协方差函数 $\gamma(k)$，并挑选合适的自协方差作为目标矩。

   (c) 延续上问，若 σ_ε^2 未知，请设计一个可以同时估计 θ 与 σ_ε^2 的矩估计方法。

   (d) 延续上问并假设 $\varepsilon_t \overset{\text{iid}}{\sim} N(0, \sigma_\varepsilon^2)$，请计算 $\boldsymbol{X} = [X_1, \ldots, X_T]$ 的协方差矩阵，以及 $\boldsymbol{X}$ 的对数似然函数 $\log L(\theta, \sigma_\varepsilon^2 | \boldsymbol{X})$。

5. 请验证第 10 讲课件第 7 页的若干推导。

   (a) 请证明 $\mathrm{cov}(Z, X) = 0$，进而说明 X 与 Z 独立。

   (b) 请证明 $Y - Z = \mathbb{E}Y + \frac{\mathrm{cov}(X,Y)}{\mathrm{var}(X)}(X - \mathbb{E}X)$ 是 X 对 Y 的最小均方预测误差函数，进而 $Y - Z = \mathbb{E}(Y|X)$。

(c) 对于任意的 2-元随机变量 (Y, X)，Y 关于 X 的条件方差定义为

$$\mathrm{var}(Y|X) = \mathbb{E}\left[(Y - \mathbb{E}(Y|X))^2|X\right].$$

在 2-元正态分布假设下，请利用 $Y - \mathbb{E}(Y|X)$ 与 X 独立这一事实，计算 $\mathrm{var}(Y|X)$。提示：说明外层条件期望可以简化为无条件期望，并考虑 $Y - \mathbb{E}Y + \mathbb{E}Y - \mathbb{E}(Y|X)$ 的平方项展开。

6. 从 CMTS 数据库http://cmf.cafr.cn/data/listpage 获取 1992Q1 开始的中国季度 GDP（支出法）、GDP 平减指数与 CPI（居民消费价格）数据，并对名义 GDP 进行价格以获得实际 GDP。

(a) 构造 GDP 季度同比增速序列 $\{y_t\}$，并转换单位为百分比%，使用 `arima` 函数对 $\{y_t\}$ 进行 ARMA 建模估计（默认使用极大似然估计，且不使用差分设定），利用 AIC、BIC 信息准则确定最优的 AR 与 MA 滞后阶数 (p, q)，并汇报模型估计结果（参数取值、标准误、R^2 等）。

(b) 构造 CPI 季度同比增速序列 $\{\pi_t\}$，并转换单位为百分比%，同上使用 `arima` 函数对 $\{\pi_t\}$ 进行 ARMA 建模估计，利用 AIC、BIC 信息准则确定最优的 AR 与 MA 滞后阶数 (p, q)，并汇报模型估计结果（参数取值、标准误、R^2 等）。提示：π_t 在 90 年代初期很高，可能造成样本整体不符合平稳性条件，因此可以在估计时只考虑 95 或者 97 后的数据。

(c) 使用 `TSA` 包的 `ARMAspec` 函数计算上两问中估计出来的 y_t, π_t ARMA 模型对应的理论谱密度，再使用 `spec.pgram` 函数直接对 $\{y_t\}, \{\pi_t\}$ 序列估计样本谱密度。比较理论谱密度与样本谱密度，并说明上两问中估计所得 ARMA 模型是否能够充分刻画两个序列的周期波动特征。

(d) 将 y_t, π_t 限制为 AR 模型，使用 OLS 进行估计，同样利用 AIC、BIC 确定之后阶数。汇报估计结果，并比较同模型（如之后阶数）的 OLS 估计与极大似然估计间的异同。

(e) 对 y_t, π_t，你认为是否有必要在时序建模时，考虑 MA 项？请说明理由。