2020 秋季本科时间序列

第 6 次作业

提交日期：12 月 3 日

1. 考虑一般的线性回归模型
$$\boldsymbol{Y}_{T\times 1} = \boldsymbol{X}_{T\times K}\boldsymbol{\beta}_{K\times 1} + \boldsymbol{e}_{T\times 1},$$
$T > K$，解释变量交叉矩矩阵 $\boldsymbol{X}^\intercal\boldsymbol{X}$ 满秩，且残差项满足独立、同方差假设，即
$$\mathbb{E}\boldsymbol{e}\boldsymbol{e}^\intercal = \sigma_e^2\boldsymbol{I}_{T\times T},$$
$\boldsymbol{I}$ 表示单位阵。

   (a) 请证明 OLS 估计下的残差向量有如下表达式
   $$\hat{\boldsymbol{e}} = \left[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\right]\boldsymbol{e}.$$

   (b) 对于 $n \times n$ 方阵 $\boldsymbol{M} = [m_{ij}]$，定义 $\boldsymbol{M}$ 的迹为 $\mathrm{tr}(\boldsymbol{M}) = m_{11} + \cdots m_{nn}$。假设 $\boldsymbol{A}$, $\boldsymbol{B}$ 分别为 $n \times k$ 与 $k \times n$ 的矩阵，n 与 k 未必相等。请证明
   $$\mathrm{tr}(\boldsymbol{AB}) = \mathrm{tr}(\boldsymbol{BA}).$$
   再假设 $\boldsymbol{A}$, $\boldsymbol{B}$ 均为 $n \times n$ 的矩阵，请证明
   $$\mathrm{tr}(\boldsymbol{A} - \boldsymbol{B}) = \mathrm{tr}(\boldsymbol{A}) - \mathrm{tr}(\boldsymbol{B}).$$

   (c) 请证明样本残差平方和可以写为：
   $$\sum_t \hat{e}_t^2 = \mathrm{tr}(\hat{\boldsymbol{e}}^\intercal\hat{\boldsymbol{e}}) = \mathrm{tr}\big(\boldsymbol{e}^\intercal\left[\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^\intercal\boldsymbol{X})^{-1}\boldsymbol{X}^\intercal\right]\boldsymbol{e}\big).$$

   (d) 请证明，给定 $\boldsymbol{X}$ 时样本残差平方和的条件期望为
   $$\mathbb{E}\left[\textstyle\sum_t \hat{e}_t^2 \middle| \boldsymbol{X}\right] = (T-K)\sigma_e^2,$$
   并进而证明
   $$\frac{1}{T-K}\sum_t \hat{e}_t^2$$
   是 σ_e^2 的无偏估计量。

2. 考虑 MA(2) 过程 $X_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2}$，$t = 1, \ldots, T$，$\{\varepsilon_t\}$ 为 iid $N(0, \sigma_\varepsilon^2)$ 的冲击序列。

(a) 请写出 $\boldsymbol{X} = [X_1, \ldots, X_T]^\intercal$ 的协方差矩阵，并计算其对数似然函数 $\log L(\theta_1, \theta_2, \sigma_\varepsilon^2 | \boldsymbol{X})$ 的表达式。

(b) 请计算极大似然估计对应的 1-阶最优条件。你能写出参数 $\theta_1, \theta_2, \sigma_\varepsilon^2$ 的极大似然估计表达式吗？如果 $\theta_2 = 0$，你能写出此时 $\theta_1, \sigma_\varepsilon^2$ 的极大似然估计表达式吗？

(c) 假设 $\sigma_X^2(k) = a_k$，$k = 0, 1, 2$，请计算 $\theta_1, \theta_2, \sigma_\varepsilon^2$ 的矩估计表达式。

3. 考虑带均值的 AR(1) 过程 $X_t = \mu + \phi X_{t-1} + \varepsilon_t$，$t = 1, \ldots, T$，$\sigma_\varepsilon^2 = 1$。

(a) 固定 $\mu = 1$，考虑 $\phi = 0.2, 0.8, 0.95$ 三个取值，分别记为 $\phi^{(i)}$，$i = 1, 2, 3$。给定 $\phi^{(i)}$，请在 R 或 Python 中编程，生成 $X_t^{(i)}$ 的随机模拟序列，$T = 1000$，起始值为 $X_0^{(i)} = \mathbb{E}^{(i)} X_t = \mu/(1 - \phi^{(i)})$。绘图展示这 3 个模拟序列，并简要讨论其时序特征。

(b) 固定 (a) 中 3 个模拟序列不变，对每个序列 $\left\{X_t^{(i)}\right\}$，按照每新增 100 个样本，进行 OLS 估计；即 $k = 1, \ldots, 10$，对样本 $\left\{X_t^{(i)} : t = 1, \ldots, 100k\right\}$ 进行 OLS 估计，估计结果记为 $\hat{\boldsymbol{\beta}}^{(i),k}$。对每一个 i，绘制 $\hat{\boldsymbol{\beta}}^{(i),k}$ 关于 k 的序列图。注意该向量有两个元素，应当分开绘图。请讨论随着 ϕ 的增大，OLS 估计随样本量 $100k$ 增加时的收敛性有何变化。

(c) 给定 (a) 中 $\phi^{(i)}$ 取值，$i = 1, 2, 3$。按照每次生成 $T = 100$ 个 $\left\{X_t^{(i)}\right\}$ 模拟样本的方式，重复生成 500 组样本 $\left\{X_t^{(i),j}\right\}$，$j = 1, \ldots, 500$。对每组样本 $\left\{X_t^{(i),j}\right\}$，进行 OLS 估计得到系数向量 $\hat{\boldsymbol{\beta}}^{(i),j}$。得到 500 个系数估计向量后，绘制其关于 j 的样本直方图。注意每个估计向量有两个系数，应当分开绘图。计算这两个系数的样本标准差。

(d) 请根据课件内容，对每组参数取值 i，计算 $T = 100$ 期样本 AR(1) 序列系数 OLS 估计值 $\hat{\boldsymbol{\beta}}^{(i)}$ 两个元素各自的渐近标准误。注意，这个渐近标准误的计算可以基于冲击的真实方差 $\sigma_\varepsilon^2 = 1$ 以及解释变量的真实协方差矩阵 $\boldsymbol{M}$。请对比这两个渐近标准误的取值与 (c) 中所得系数估计值样本标准差的关系。

(e) 重复 (c)–(d) 的内容，但取 $T = 900$。比较此时系数估计值渐近标准误与 500 组样本所得样本标准差的关系。此时估计系数的样本标准差是否近似为 (c) 中的 $1/3$？渐近标准误与 (d) 相比呢？