时间序列分析与R 语言实战
第六章：模型识别（一）：自相关函数

Outline

1. 自相关函数
2. 样本自相关函数

Outline

1. 自相关函数
2. 样本自相关函数

自相关函数：概念与 MA 模型

自相关函数：

$$\rho_k = \frac{Cov(Y_t, Y_{t-k})}{\sqrt{Var(Y_t)Var(Y_{t-k})}}, \quad k = 1, 2, \cdots$$

- MA(1) 模型：

$$\rho_k = \begin{cases} 1 & k = 0 \\ -\dfrac{\theta}{1+\theta^2} & k = 1 \\ 0 & k \geq 2 \end{cases}$$

- MA(q) 模型：

$$\rho_k = \begin{cases} 1 & k = 0 \\ \dfrac{-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{i+k}}{1 + \sum_{i=1}^{q} \theta_i^2} & 1 \leq k \leq q \\ 0 & k > q \end{cases}$$

自相关函数：AR 模型

- AR(1) 模型：

$$\rho_k = \begin{cases} 1 & k = 0 \\ \phi^k & k \geq 1 \end{cases}$$

- AR(2) 模型：

$$\rho_k = \begin{cases} 1 & k = 0 \\ \dfrac{\phi_1}{1 - \phi_2} & k = 1 \\ \phi_1 \rho_{k-1} + \phi_2 \rho_{k-2} & k \geq 2 \end{cases}$$

- AR(p) 模型：ρ_k 呈指数衰减

自相关函数：ARMA模型及小结

- ARMA(1,1) 模型：

$$\rho_k = \begin{cases} 1 & k = 0 \\ \dfrac{(\phi-\theta)(1-\phi\theta)}{1-2\phi\theta+\theta^2}\phi^{k-1} & k \geq 1 \end{cases}$$

- ARMA(p,q) 模型：ρ_k 呈指数衰减

小结：

	AR(p)	MA(q)	ARMA(p, q)
自相关函数	拖尾	q 阶截尾	拖尾

Outline

样本自相关函数：概念

样本自相关函数：

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}, \quad k = 1, 2, \cdots$$

- **序列** $\{Y_t\}$ **：**

$$Y_t = \mu + \sum_{j=0}^{\infty} \psi_j e_{t-j}$$

  其中，e_t 独立同分布并具有零均值和有限的非零方差。

- **假设条件：**

$$\sum_{j=0}^{\infty} |\psi_j| < \infty, \qquad \sum_{j=0}^{\infty} j\psi_j^2 < \infty$$

样本自相关函数

- **结论：** 对任意固定的 m，当 $n \to \infty$ 时，$\sqrt{n}(r_1-\rho_1, r_2-\rho_2, \cdots, r_m-\rho_m)$ 收敛至 $N(\mathbf{0}, \Sigma)$。

$$\Sigma = \begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ c_{m1} & c_{m2} & \cdots & c_{mm} \end{bmatrix}$$

其中，

$$c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + \rho_i\rho_j\rho_k^2)$$

- **推论：** n 充分大时，r_k 近似服从 $N(\rho_k, c_{kk}/n)$，且 $Corr(r_k, r_j) \approx c_{kj}/\sqrt{c_{kk}c_{jj}}$

样本自相关函数：白噪声

$$c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + \rho_i\rho_j\rho_k^2)$$

假设 $\{Y_t\}$ 是白噪声，

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k \geq 1 \end{cases}$$

则，

$$c_{kk} = 1 \quad 和 \quad c_{ij} = 0, i \neq j$$

样本自相关函数 r_k：

- 方差：

$$Var(r_k) \approx \frac{1}{n}$$

- 自相关函数：

$$Corr(r_k, r_j) \approx 0, \qquad k \neq j$$

样本自相关函数：AR(1)

假设 $\{Y_t\}$ 是 AR(1) 序列，$\boxed{c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + \rho_i\rho_j\rho_k^2)}$

$$\rho_k = \begin{cases} 1 & k = 0 \\ \phi^k & k > 0 \\ \phi^{-k} & k < 0 \end{cases}$$

则，

- $i = j$ 时

$$c_{kk} = \frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}$$

- 对于一般的 $0 < i < j$，

$$c_{ij} = \frac{(\phi^{j-i} - \phi^{j+i})(1+\phi^2)}{1-\phi^2} + (j-i)\phi^{j-i} - (j+i)\phi^{j+i}$$

样本自相关函数：AR(1)

AR(1) 模型的样本自相关函数：

- 方差：

$$Var(r_k) \approx \frac{1}{n}[\frac{(1+\phi^2)(1-\phi^{2k})}{1-\phi^2} - 2k\phi^{2k}]$$

  - $k = 1$ 时，

$$Var(r_1) \approx \frac{1-\phi^2}{n}$$

  - k 较大时，

$$Var(r_k) \approx \frac{1}{n}[\frac{1+\phi^2}{1-\phi^2}]$$

- 自相关函数：

$$Corr(r_k, r_j) \approx \frac{c_{kj}}{\sqrt{c_{kk}c_{jj}}}$$

  - 特别地，

$$Corr(r_1, r_2) \approx 2\phi\sqrt{\frac{1-\phi^2}{1+2\phi^2-3\phi^4}}$$

样本自相关函数：AR(1)

方差与自相关函数：

$$Var(r_1) \approx \frac{1-\phi^2}{n}$$

$$Var(r_2) \approx \frac{1+2\phi^2-3\phi^4}{n}$$

$$Corr(r_1, r_2) \approx 2\phi\sqrt{\frac{1-\phi^2}{1+2\phi^2-3\phi^4}}$$

$$Var(r_{10}) \approx \frac{1}{n}[\frac{(1+\phi^2)(1-\phi^{20})}{1-\phi^2} - 20\phi^{20}]$$

AR(1) 模型选定的 r_k 的大样本结果：

ϕ	$\sqrt{Var(r_1)}$	$\sqrt{Var(r_2)}$	$Corr(r_1, r_2)$	$\sqrt{Var(r_{10})}$
± 0.9	$0.436/\sqrt{n}$	$0.807/\sqrt{n}$	± 0.972	$2.436/\sqrt{n}$
± 0.7	$0.714/\sqrt{n}$	$1.122/\sqrt{n}$	± 0.891	$1.704/\sqrt{n}$
± 0.4	$0.917/\sqrt{n}$	$1.115/\sqrt{n}$	± 0.658	$1.175/\sqrt{n}$
± 0.2	$0.980/\sqrt{n}$	$1.037/\sqrt{n}$	± 0.378	$1.041/\sqrt{n}$

样本自相关函数：MA(1)

方差与自相关函数：

$$c_{ij} = \sum_{k=-\infty}^{\infty} (\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + \rho_i\rho_j\rho_k^2)$$

$$c_{11} = 1 - 3\rho_1^2 + 4\rho_1^4 \Rightarrow Var(r_1) \approx \frac{1 - 3\rho_1^2 + 4\rho_1^4}{n}$$

$$c_{kk} = 1 + 2\rho_1^2, \quad \Rightarrow Var(r_k) \approx \frac{1 + 2\rho_1^2}{n}, \quad k > 1$$

$$c_{12} = 2\rho_1(1 - \rho_1^2) \Rightarrow Corr(r_1, r_2) \approx \frac{2\rho_1(1 - \rho_1^2)}{\sqrt{(1 - 3\rho_1^2 + 4\rho_1^4)(1 + 2\rho_1^2)}}$$

MA(1) 模型选定的 r_k 的大样本结果:

θ	$\sqrt{Var(r_1)}$	$\sqrt{Var(r_k)}, \ k > 1$	$Corr(r_1, r_2)$
± 0.9	$0.709/\sqrt{n}$	$1.222/\sqrt{n}$	∓ 0.864
± 0.7	$0.730/\sqrt{n}$	$1.201/\sqrt{n}$	∓ 0.836
± 0.5	$0.789/\sqrt{n}$	$1.149/\sqrt{n}$	∓ 0.741
± 0.3	$0.892/\sqrt{n}$	$1.073/\sqrt{n}$	∓ 0.532

样本自相关函数：MA(q)

$$c_{ij}=\sum_{k=-\infty}^{\infty}(\rho_{k+i}\rho_{k+j}+\rho_{k-i}\rho_{k+j}-2\rho_i\rho_k\rho_{k+j}-2\rho_j\rho_k\rho_{k+i}+\rho_i\rho_j\rho_k^2)$$

k 阶自相关系数：

$$\rho_k=\begin{cases}1 & k=0\\ \dfrac{-\theta_k+\sum_{i=1}^{q-k}\theta_i\theta_{i+k}}{1+\sum_{i=1}^{q}\theta_i^2} & 1\le k\le q\\ 0 & k>q\end{cases}$$

$i=j$ 时，

$$c_{kk}=1+2\sum_{j=1}^{q}\rho_j^2,\quad k>q$$

则方差，

$$Var(r_k)\approx\frac{1}{n}[1+2\sum_{j=1}^{q}\rho_j^2],\quad k>q$$

时间序列分析与R 语言实战
第六章：模型识别（二）：偏自相关函数和扩展的自相关函数

Outline

1. 偏自相关函数
2. 扩展的自相关函数

Outline

1. 偏自相关函数
    - 偏自相关函数
    - 样本偏自相关函数
2. 扩展的自相关函数

AR(1)：

$$Y_t = \phi Y_{t-1} + e_t$$

那么，

$$\begin{aligned}
\gamma_2 &= Cov(Y_t, Y_{t-2}) \\
&= Cov(\phi Y_{t-1} + e_t, Y_{t-2}) \\
&= Cov(\phi(\phi Y_{t-2} + e_{t-1}) + e_t, Y_{t-2}) \\
&= Cov(\phi^2 Y_{t-2} + \phi e_{t-1} + e_t, Y_{t-2}) \\
&= \phi^2 \gamma_0
\end{aligned}$$

受 Y_{t-1} 的影响，Y_t 与 Y_{t-2} 相关。

偏自相关函数：引例（续）及定义

AR(1)：

$$Y_t = \phi Y_{t-1} + e_t$$

消除 Y_{t-1} 的影响，

$$Cov(Y_t - \phi Y_{t-1}, Y_{t-2} - \phi Y_{t-1}) = Cov(e_t, Y_{t-2} - \phi Y_{t-1}) = 0$$

定义

消除中间介入变量 Y_{t-1}，Y_{t-2}，Y_{t-3}，$\cdots$，Y_{t-k+1} 的影响后，Y_t 和 Y_{t-k} 的相关系数函数，称为 k 阶滞后偏自相关系数。

偏自相关函数：数学表达

对平稳序列 $\{Y_t\}$，

- Y_t 的条件期望：

$$\hat{E}Y_t = E(Y_t|Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1})$$

- Y_{t-k} 的条件期望：

$$\hat{E}Y_{t-k} = E(Y_{t-k}|Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1})$$

- k 阶滞后偏自相关系数：

$$\phi_{kk} = \frac{E(Y_t - \hat{E}Y_t)(Y_{t-k} - \hat{E}Y_{t-k})}{E(Y_{t-k} - \hat{E}Y_{t-k})^2}$$

偏自相关函数：数学表达

假设序列 $\{Y_t\}$零均值，使用 Y_{t-1}，Y_{t-2}，Y_{t-3}，$\cdots$，Y_{t-k+1} 的线性函数对 Y_t 和Y_{t-k} 进行预测

- Y_t 最优预测表达式为：

$$\beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-k+1}$$

- Y_{t-k} 最优预测表达式为：

$$\beta_1 Y_{t-k+1} + \beta_2 Y_{t-k+2} + \cdots + \beta_{k-1} Y_{t-1}$$

注：

$$\min_{\beta_1, \beta_2, \cdots, \beta_{k-1}} E(Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \cdots - \beta_{k-1} Y_{t-k+1})^2$$

偏自相关函数：数学表达

k 阶滞后偏自相关函数：

- $k = 1$,

$$\phi_{11} = Corr(Y_t, Y_{t-1}) = \rho_1$$

- $k \geq 2$,

$$\begin{aligned}\phi_{kk} = Corr(&Y_t - \beta_1 Y_{t-1} - \beta_2 Y_{t-2} - \cdots - \beta_{k-1} Y_{t-k+1},\\ &Y_{t-k} - \beta_1 Y_{t-k+1} - \beta_2 Y_{t-k+2} - \cdots - \beta_{k-1} Y_{t-1})\end{aligned}$$

特别地：基于正态分布假设

$$\phi_{kk} = Corr(Y_t, Y_{t-k} | Y_{t-1}，Y_{t-2}，Y_{t-3}，\cdots, Y_{t-k+1})$$

偏自相关函数：$k=2$

对平稳序列 $\{Y_t\}$，假设只基于 Y_{t-1} 对 Y_t 和 Y_{t-2} 进行预测，

- 均方误差：

$$E(Y_t-\beta Y_{t-1})^2=\gamma_0-2\beta\gamma_1+\beta^2\gamma_0$$

- 最小化均方误差：

$$\min_{\beta} E(Y_t-\beta Y_{t-1})^2$$

- β 求解：

$$\begin{aligned}\frac{\partial g(\beta)}{\partial \beta} &= 2\beta\gamma_0-2\gamma_1=0\\ \Longrightarrow \qquad \beta &= \frac{\gamma_1}{\gamma_0}=\rho_1\end{aligned}$$

- Y_t 最优预测表达式为： $\rho_1 Y_{t-1}$
- 同样地，Y_{t-2} 最优预测表达式为： $\rho_1 Y_{t-1}$

偏自相关函数：$k=2$

消除 Y_{t-1} 影响，

- 协方差：

$$Cov(Y_t-\rho_1 Y_{t-1}, Y_{t-2}-\rho_1 Y_{t-1}) = \gamma_2-\rho_1\gamma_1-\rho_1\gamma_1+\rho_1^2\gamma_0 = \gamma_0(\rho_2-\rho_1^2)$$

- 方差：

$$Var(Y_t-\rho_1 Y_{t-1}) = Var(Y_{t-2}-\rho_1 Y_{t-1}) = \gamma_0-2\rho_1\gamma_1+\rho_1^2\gamma_0 = \gamma_0(1-\rho_1^2)$$

二阶滞后偏自相关函数：

$$\boxed{\phi_{22} = \frac{\rho_2-\rho_1^2}{1-\rho_1^2}}$$

偏自相关函数：AR 模型

AR(1)：$k=2$ 时，

$$\rho_k = \phi^k, k = 1, 2, \cdots$$

$$\phi_{11} = \rho_1$$

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

$$\phi_{22} = \frac{\phi^2 - \phi^2}{1 - \phi^2} = 0$$

- 对所有的 $k > 1$，$\phi_{kk} = 0$。

AR(p)：$k > p$ 时，基于 Y_{t-1}，Y_{t-2}，Y_{t-3}，$\cdots$，Y_{t-k+1} 对 Y_t 和Y_{t-k} 进行预测，

- Y_t 最优预测表达式为：

$$\phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p}$$

- Y_{t-k} 最优预测表达式为：

$$h(Y_{t-1}, Y_{t-2}, \cdots, Y_{t-p}, \cdots, Y_{t-k+1})$$

偏自相关函数：AR 模型

AR(p)：消除 Y_{t-1}，Y_{t-2}，Y_{t-3}，$\cdots$，Y_{t-k+1} 影响，

- 协方差：

$$\begin{aligned}
&Cov(Y_t - \phi_1 Y_{t-1} - \phi_2 Y_{t-2} - \cdots - \phi_p Y_{t-p}, \\
&\qquad Y_{t-k} - h(Y_{t-k+1}, Y_{t-k+2}, \cdots, Y_{t-1}) \\
=&Cov(e_t, Y_{t-k} - h(Y_{t-k+1}, Y_{t-k+2}, \cdots, Y_{t-1})) \\
=&0
\end{aligned}$$

- k 阶滞后偏自相关函数：

$$\boxed{\phi_{kk} = 0, \qquad k > p}$$

偏自相关函数：Y-W 方程

Y-W 方程

对具有 ρ_k 的平稳过程，可通过下面的Yule-Walker方程求解 ϕ_{kk}：

$$\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \phi_{k3}\rho_{j-3} + \cdots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \cdots, k$$

k 个线性方程组：

$$\begin{cases} \rho_1 = & \phi_{k1} + & \rho_1\phi_{k2} + & \rho_2\phi_{k3} + & \cdots + & \rho_{k-1}\phi_{kk} \\ \rho_2 = & \rho_1\phi_{k1} + & \phi_{k2} + & \rho_1\phi_{k3} + & \cdots + & \rho_{k-2}\phi_{kk} \\ & \vdots & & & & \\ \rho_k = & \rho_{k-1}\phi_{k1} + & \rho_{k-2}\phi_{k2} + & \rho_{k-3}\phi_{k3} + & \cdots + & \phi_{kk} \end{cases}$$

偏自相关函数：Y-W 方程证明

- 假设平稳序列 $\{Y_t\}$零均值，使用 Y_{t-1}，Y_{t-2}，Y_{t-3}，$\cdots$，Y_{t-k} 的线性函数对 Y_t 进行拟合:

$$Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_k Y_{t-k} + e_t$$

- 取条件期望：

$$\begin{aligned}\hat{E} Y_t =& \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-k+1} + \beta_k \hat{E} Y_{t-k} \\ &+ E(e_t | Y_{t-1}, Y_{t-2}, \cdots, Y_{t-k+1})\end{aligned}$$

等价于，

$$\hat{E} Y_t = \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \cdots + \beta_{k-1} Y_{t-k+1} + \beta_k \hat{E} Y_{t-k}$$

- 于是，

$$Y_t - \hat{E} Y_t = \beta_k (Y_{t-k} - \hat{E} Y_{t-k}) + e_t$$

偏自相关函数：Y-W 方程证明

- 同时乘以 $Y_{t-k} - \hat{E}Y_{t-k}$，求期望：

$$
\begin{aligned}
E(Y_t - \hat{E}Y_t)(Y_{t-k} - \hat{E}Y_{t-k}) = & \beta_k E(Y_{t-k} - \hat{E}Y_{t-k})^2 \\
& + E[e_t(Y_{t-k} - \hat{E}Y_{t-k})]
\end{aligned}
$$

- 系数 β_k，

$$
\beta_k = \frac{E(Y_t - \hat{E}Y_t)E(Y_{t-k} - \hat{E}Y_{t-k})}{E(Y_{t-k} - \hat{E}Y_{t-k})^2}
$$

- Y_t 最佳拟合：

$$
\begin{aligned}
x_t - \mu &= \phi_{11}x_{t-1} + w_{1t}, \\
x_t - \mu &= \phi_{21}x_{t-1} + \phi_{22}x_{t-2} + w_{2t}, \\
&\vdots \\
x_t - \mu &= \phi_{p1}x_{t-1} + \phi_{p2}x_{t-2} + \cdots + \phi_{pp}x_{t-p} + w_{pt},
\end{aligned}
$$

$$
Y_t = \phi_{k1}Y_{t-1} + \phi_{k2}Y_{t-2} + \cdots + \phi_{kk}Y_{t-k} + e_t
$$

- 同乘 Y_{t-j}，求期望，然后除以 γ_0：

$$
\rho_j = \phi_{k1}\rho_{j-1} + \phi_{k2}\rho_{j-2} + \phi_{k3}\rho_{j-3} + \cdots + \phi_{kk}\rho_{j-k}, \quad j = 1, 2, \cdots, k
$$

偏自相关函数：Y-W 方程

线性方程组的矩阵形式：

$$\begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{pmatrix}$$

根据 Cramer 法则，

$$\phi_{kk} = \frac{|D_k|}{|D|}$$

其中，

$$D = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & 1 \end{pmatrix}, \quad D_k = \begin{pmatrix} 1 & \rho_1 & \cdots & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_2 \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_k \end{pmatrix}$$

偏自相关函数

递推公式：

$$\phi_{kk} = \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j}$$

其中，

$$\phi_{k,j} = \phi_{k-1,j} - \phi_{kk}\phi_{k-1,k-j}, \quad j = 1, 2, \cdots, k-1$$

- 一阶滞后偏自相关函数：

$$\phi_{11} = \rho_1$$

- 二阶滞后偏自相关函数：

$$\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1 - \phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

- 三阶滞后偏自相关函数：

$$\phi_{33} = \frac{\rho_3 - \phi_{21}\rho_2 - \phi_{22}\rho_1}{1 - \phi_{21}\rho_1 - \phi_{22}\rho_2}$$

偏自相关函数：MA(1)

MA(1):

$$Y_t = e_t - \theta e_{t-1}$$

$$\boxed{\rho_k = \begin{cases} 1 & k = 0 \\ -\dfrac{\theta}{1+\theta^2} & k = 1 \\ 0 & k \geq 2 \end{cases}}$$

- 一阶滞后偏自相关函数：

$$\phi_{11} = \rho_1 = \frac{-\theta}{1+\theta^2}$$

- 二阶滞后偏自相关函数：

$$\phi_{22} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2} = \frac{-\theta^2}{1+\theta^2+\theta^4}$$

偏自相关函数：MA(1) 模型与小结

MA(1)：

$$Y_t = e_t - \theta e_{t-1}$$

- k 阶滞后偏自相关函数：

$$\phi_{kk} = \frac{\theta^k(1-\theta^2)}{1-\theta^{2(k+1)}}, \quad k \geq 1$$

小结：

	AR(p)	MA(q)	ARMA(p, q)
偏自相关函数	p 阶截尾	拖尾	拖尾

Outline

1. 偏自相关函数
   - 偏自相关函数
   - 样本偏自相关函数
2. 扩展的自相关函数

样本偏自相关函数

样本偏自相关函数：

$$\hat{\phi}_{kk} = \frac{r_k - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_{k-j}}{1 - \sum_{j=1}^{k-1} \hat{\phi}_{k-1,j} r_j}$$

其中，

$$\hat{\phi}_{k,j} = \hat{\phi}_{k-1,j} - \hat{\phi}_{kk}\hat{\phi}_{k-1,k-j}, \quad j = 1, 2, \cdots, k-1$$

- AR(p) 模型下，$k > p$ 时，$\hat{\phi}_{kk}$ 近似服从 $N(0, 1/n)$。

Outline

扩展的自相关函数：原理

原理：

- 假设 ARMA 模型的 AR 部分已知
- 在观测时间序列中“滤出”自回归部分，可得到一个纯 MA 过程
- 自回归部分的系数，可通过有限次的回归来估计

扩展的自相关函数：ARMA(1,1) 模型

ARMA(1,1)：

$$Y_t = \phi Y_{t-1} + e_t - \theta e_{t-1}$$

- 假设模型为纯AR模型，进行回归：

  $$Y_t = \phi^{(0)} Y_{t-1} + e_t^{(0)}$$

  估计值 $\hat{\phi}^{(0)} = \rho_1 = (\phi - \theta)(1 - \phi\theta)/(1 - 2\phi\theta + \theta^2)$

- 第一次回归：

  $$Y_t = \phi^{(1)} Y_{t-1} + \beta_1^{(1)} \hat{e}_{t-1}^{(0)} + e_t^{(1)}$$

  其中，$\hat{e}_t^{(0)} = Y_t - \hat{\phi}^{(0)} Y_{t-1}$，估计值 $\hat{\phi}^{(1)} = \widetilde{\phi}$ 为 ϕ 的一致估计值，$W_t = Y_t - \widetilde{\phi} Y_{t-1}$ 近似为 MA(1) 过程。

扩展的自相关函数：ARMA(p, q) 模型

ARMA(p, q)：

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \cdots - \theta_q e_{t-q}$$

- 假设模型为纯AR模型，进行回归：

$$Y_t = \sum_{l=1}^{p} \phi_l^{(0)} Y_{t-l} + e_t^{(0)}$$

- 第一次回归：

$$Y_t = \sum_{l=1}^{p} \phi_l^{(1)} Y_{t-l} + \beta_1^{(1)} \hat{e}_{t-1}^{(0)} + e_t^{(1)}$$

其中 $\hat{e}_t^{(0)} = Y_t - \sum_{l=1}^{p} \hat{\phi}_l^{(0)} Y_{t-l}$

- 第二次回归：

$$Y_t = \sum_{l=1}^{p} \phi_l^{(2)} Y_{t-l} + \beta_1^{(2)} \hat{e}_{t-1}^{(1)} + \beta_2^{(2)} \hat{e}_{t-2}^{(0)} + e_t^{(2)}$$

扩展的自相关函数：ARMA(p, q) 模型

- 第 j 次回归：

$$Y_t = \sum_{l=1}^{k} \phi_l^{(j)} Y_{t-l} + \sum_{i=1}^{j} \beta_i^{(j)} \hat{e}_{t-i}^{(j-i)} + e_t^{(j)}$$

- 通过一系列回归：

$$W_{t,k,j} = Y_t - \widetilde{\phi}_1 Y_{t-1} - \widetilde{\phi}_2 Y_{t-2} - \cdots - \widetilde{\phi}_k Y_{t-k}$$

- 定义 $W_{t,k,j}$ 的样本自相关系数为扩展的样本自相关系数（EACF）。
- 当 $k = p,\ j \geq q$ 时，$\{W_{t,k,j}\}$ 近似为 MA(q) 模型。
- 当 $k > p$ 时，出现过拟合。
- $j + 1$ 阶滞后之后，$W_{t,k,j}$ 的样本自相关系数近似服从 $N(0, 1/(n-k-j))$。

时间序列分析与R 语言实战
第六章：模型识别（三）：模拟案例分析

Outline

1. 识别方法
2. 模拟数据

Outline

1. 识别方法
2. 模拟数据

识别方法：ACF 和 PACF 的一般特征

ARMA 模型 ACF 和 PACF 的一般特征

	AR(p)	MA(q)	ARMA(p, q)
ACF	拖尾	q 阶截尾	拖尾
PACF	p 阶截尾	拖尾	拖尾

假设检验

- **基本思想：** 小概率反证法
- **步骤：**
  - 提出原假设 H_0，则 H_0 的对立面为备择假设 H_1
  - 假设 H_0 成立，计算由样本得到的统计量
  - 根据统计量的大小和分布，做出统计推断

MA(q) 模型

- 原假设 H_0：$r_k = \rho_k$
- 拒绝域：$\{r_k : |r_k - \rho_k| > 2\sqrt{Var(r_k)}\}$
- 统计推断：若计算所得 $r_k \in [\rho_k - 2\sqrt{Var(r_k)}, \rho_k + 2\sqrt{Var(r_k)}]$，则认为是 MA(q) 模型，否则拒绝原假设。

AR(p) 模型

- 若计算所得 $\hat{\phi}_{kk} \in [-2/\sqrt{n}, 2/\sqrt{n}]$, $k > p$，则认为是 AR(p) 模型，否则拒绝原假设。

ARMA(p, q) 模型

- 假设 p 给定，计算扩展的样本自相关关系数，进而得到 EACF 数值表格。
- 若扩展的样本自相关关系数不在区间 $[-2/\sqrt{n-k-j}, 2/\sqrt{n-k-j}]$，认为显著不为零，用 x 表示，否则记为 0，可得到 EACF 符号表格。
- 根据 EACF 表格，做出统计推断。

识别方法：ARMA(1,1) 模型的理论 EACF

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	x	x	x	x	x	x	x	x	x	x	x	x	x	x
1	x	0^*	0	0	0	0	0	0	0	0	0	0	0	0
2	x	x	0	0	0	0	0	0	0	0	0	0	0	0
3	x	x	x	0	0	0	0	0	0	0	0	0	0	0
4	x	x	x	x	0	0	0	0	0	0	0	0	0	0
5	x	x	x	x	x	0	0	0	0	0	0	0	0	0
6	x	x	x	x	x	x	0	0	0	0	0	0	0	0
7	x	x	x	x	x	x	x	0	0	0	0	0	0	0

Outline

1. 识别方法
2. 模拟数据

Outline

1. 识别方法
2. 模拟数据
   - MA 模型
   - AR 模型
   - ARMA 模型

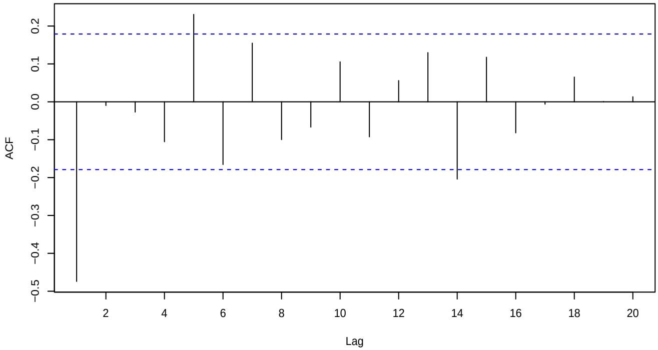
模拟数据：MA(1) 模型

例 6.1：$\theta = 0.9$ 的 MA(1) 模型：$\rho_1 = -0.497$，$r_1 = -0.474$，$r_1 \in [-0.497 - 2 \times 0.71/\sqrt{120}, -0.497 + 2 \times 0.71/\sqrt{120}] = [-0.627, -0.367]$，$2/\sqrt{n} = 0.1826$

$$\rho_1 = -\frac{\theta}{1+\theta^2}$$

样本 ACF

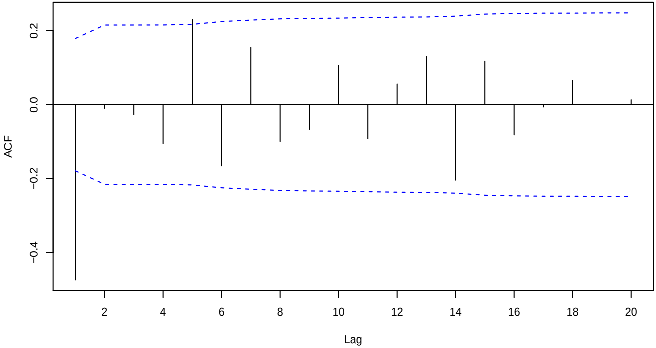
$$Var(r_1) \approx \frac{1 - 3\rho_1^2 + 4\rho_1^4}{n}$$



例 6.1（续 1）：

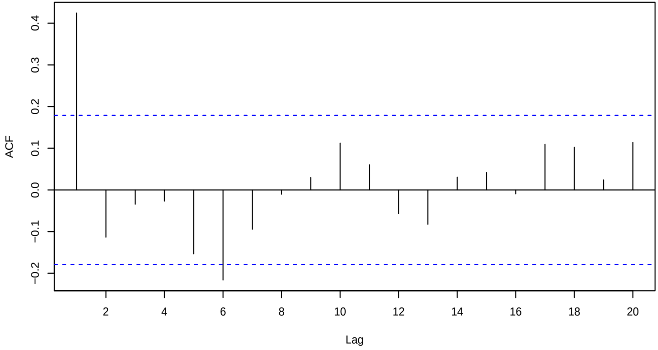
$$Var(r_k) \approx \frac{1}{n}[1 + 2\sum_{j=1}^{q} \rho_j^2], \quad k > q$$

具有另一种边界的样本 ACF



例 6.2：$\theta = -0.9$ 的 MA(1) 模型：$\rho_1 = 0.497$，$r_1 = 0.424$，$r_1 \in [0.367, 0.627]$，$2/\sqrt{n} = 0.1826$

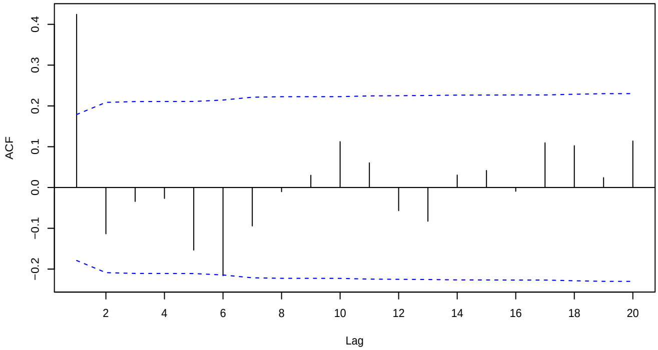
样本 ACF



模拟数据：MA(1) 模型

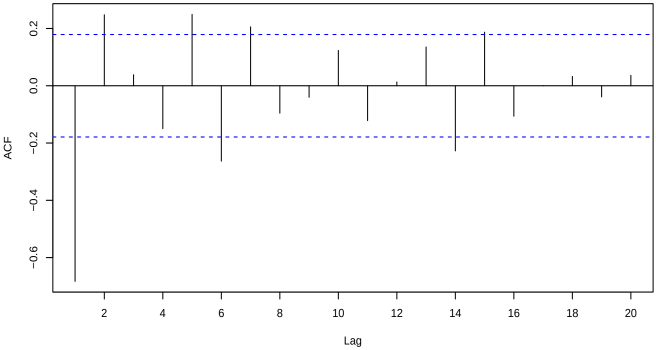
例 6.2（续 1）：

具有另一种边界的样本 ACF



例 6.3：$\theta_1 = 1$，$\theta_2 = -0.6$ 的 MA(2) 模型

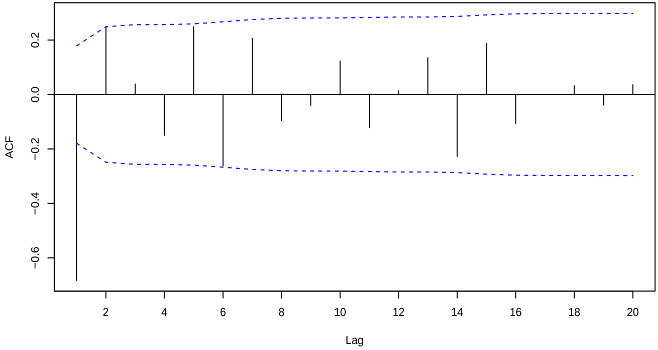
样本 ACF



模拟数据：MA(2) 模型

例 6.3（续 1）：

具有另一种边界的样本 ACF

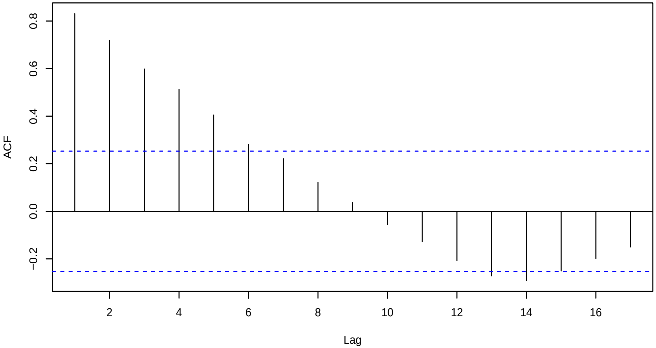


Outline

1. 识别方法
2. 模拟数据
   - MA 模型
   - AR 模型
   - ARMA 模型

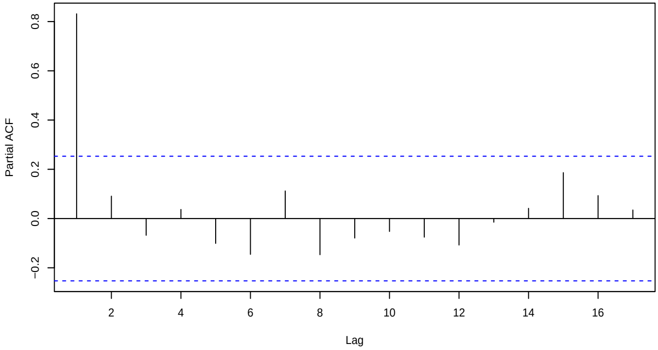
例 6.4： $\phi = 0.9$ 的 AR(1) 模型

样本 ACF



例 6.4（续 1）：$\phi = 0.9$ 的 AR(1) 模型

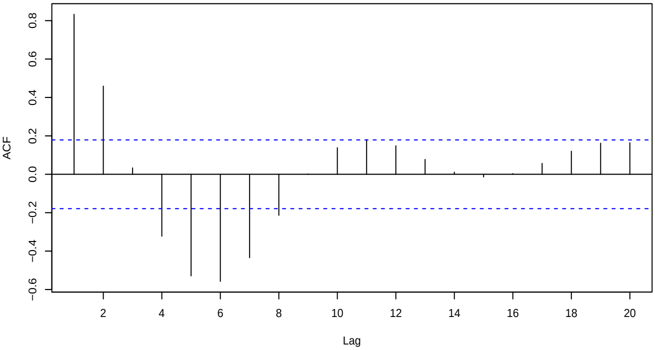
样本 PACF



模拟数据：AR(2) 模型

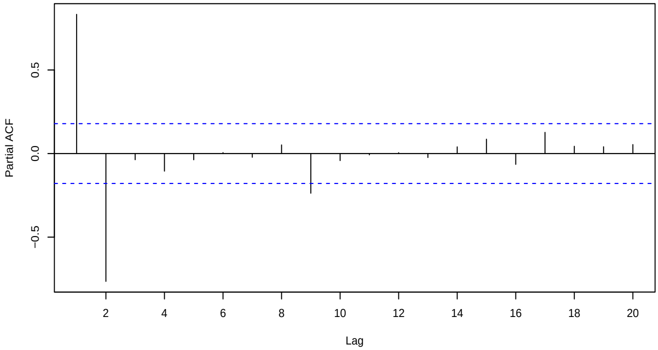
例 6.5：$\phi_1 = 1.5$，$\phi_2 = -0.75$ 的 AR(2) 模型

样本 ACF



例 6.5（续 1）：$\phi_1 = 1.5$，$\phi_2 = -0.75$ 的 AR(2) 模型

样本 PACF



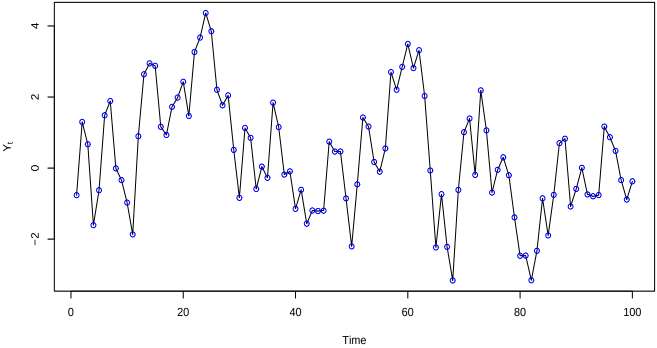
Outline

1. 识别方法
2. 模拟数据
    - MA 模型
    - AR 模型
    - ARMA 模型

模拟数据：ARMA(1,1) 模型

例 6.6：

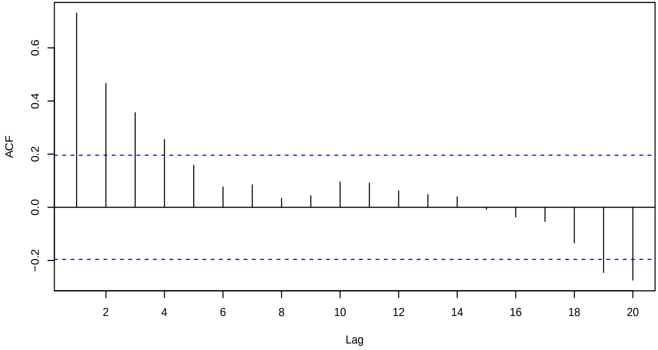
模拟的 $\phi = 0.6$，$\theta = -0.3$ 的 ARMA(1,1) 序列



模拟数据：ARMA(1,1) 模型

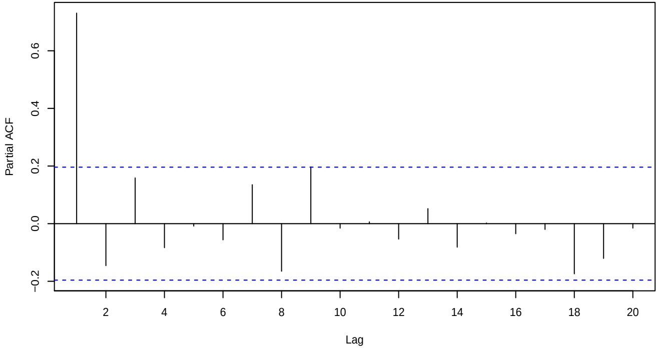
例 6.6（续 1）：

模拟的 ARMA(1,1) 序列的样本 ACF



例 6.6（续 2）：

模拟的 ARMA(1,1) 序列的样本 PACF



例 6.6（续 3）：

模拟的 ARMA(1,1) 序列的样本 EACF

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	X	X	X	X	O	O	O	O	O	O	O	O	O	O
1	X	O	O	O	O	O	O	O	O	O	O	O	O	O
2	X	O	O	O	O	O	O	O	O	O	O	O	O	O
3	X	X	O	O	O	O	O	O	O	O	O	O	O	O
4	X	O	X	O	O	O	O	O	O	O	O	O	O	O
5	X	O	O	O	O	O	O	O	O	O	O	O	O	O
6	X	O	O	O	X	O	O	O	O	O	O	O	O	O
7	X	O	O	O	X	O	O	O	O	O	O	O	O	O

时间序列分析与R 语言实战
第六章：模型识别（四）：非平稳性与其他识别方法

Outline

1. 非平稳性
2. 其他识别方法

Outline

1. 非平稳性
2. 其他识别方法

Outline

1. 非平稳性
   - 平稳序列与非平稳序列
   - 过度差分
   - 单位根检验
2. 其他识别方法

平稳序列与非平稳序列

平稳时间序列：

- 均值为常数，序列围绕均值上下波动；
- 自协方差只依赖于时间间隔；
- 多数平稳序列，随滞后项的增加，样本 ACF 与 PACF 呈指数衰减或截尾。

非平稳时间序列：

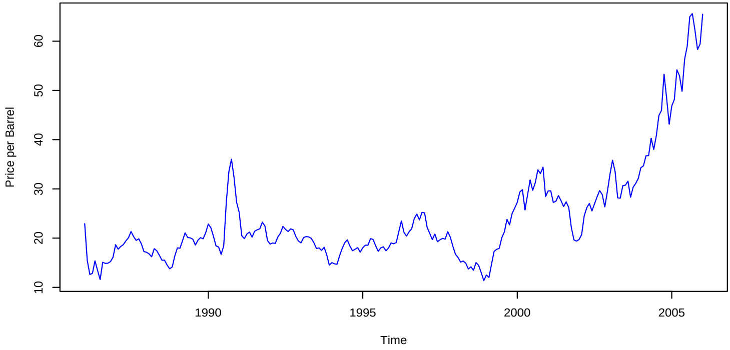
- 或不具有固定的均值；
- 或自协方差不只依赖于时间间隔；
- 随滞后项的增加，样本 ACF 一般缓慢衰减，倾向于缓慢移动，或者向上或者向下，有着明显的“趋势”。

非平稳序列的识别

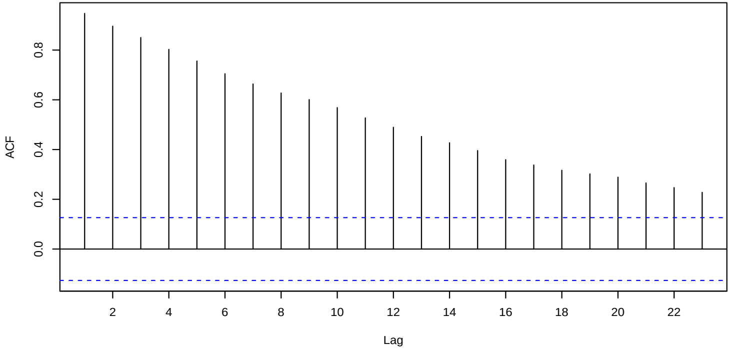
- 通过时间序列趋势图判断
- 通过ACF判断
- 单位根检验

非平稳时间序列案例：原油价格

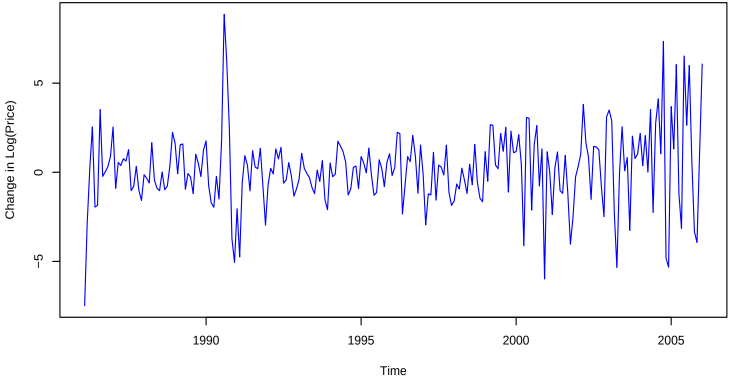
例 6.7：1986 年 1 月到 2006 年 1 月，每桶原油的月度价格时间序列图



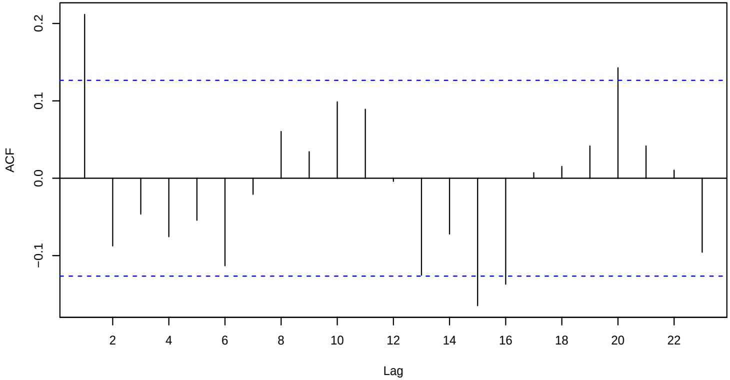
例 6.7（续 1）：每桶原油的月度价格的样本 ACF



例 6.7（续2）：每桶原油的月度价格取对数差分的时序图



例 6.7（续3）：原油价格序列对数差分的样本 ACF



Outline

1. 非平稳性
    - 平稳序列与非平稳序列
    - 过度差分
    - 单位根检验
2. 其他识别方法

过度差分：平稳序列

平稳序列的差分仍为平稳的：

- 假设序列 $\{Y_t\}$ 平稳，均值记为 μ，自协方差记为 γ_k
- 一阶差分 $W_t = Y_t - Y_{t-1}$：
  - 均值：
  $$E(W_t) = E(Y_t - Y_{t-1}) = \mu - \mu = 0$$
  - 自协方差：
  $$Cov(W_t, W_{t-k}) = Cov(Y_t - Y_{t-1}, Y_{t-k} - Y_{t-k-1}) = 2\gamma_k - \gamma_{k-1} - \gamma_{k+1}$$
- d 阶差分仍然平稳

过度差分：非平稳序列

例 6.8：$\{Y_t\}$ 为随机游动序列

$$Y_t = Y_{t-1} + e_t$$

- 一阶差分得到白噪声模型

$$\nabla Y_t = Y_t - Y_{t-1} = e_t$$

- 二阶差分得到 $\theta = 1$ 的 MA(1) 模型

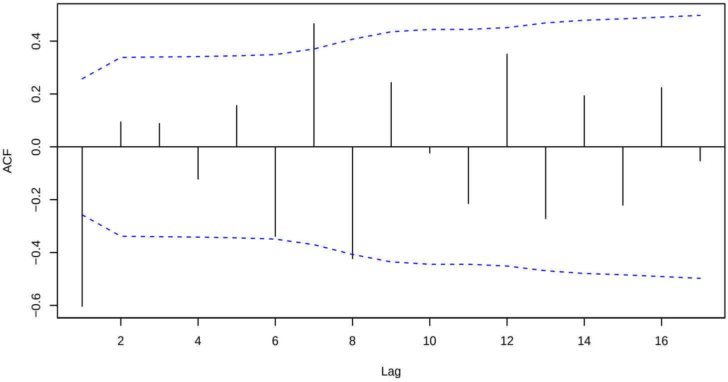
$$\nabla^2 Y_t = e_t - e_{t-1}$$

注：模型选取遵从简约原则

过度差分：非平稳序列

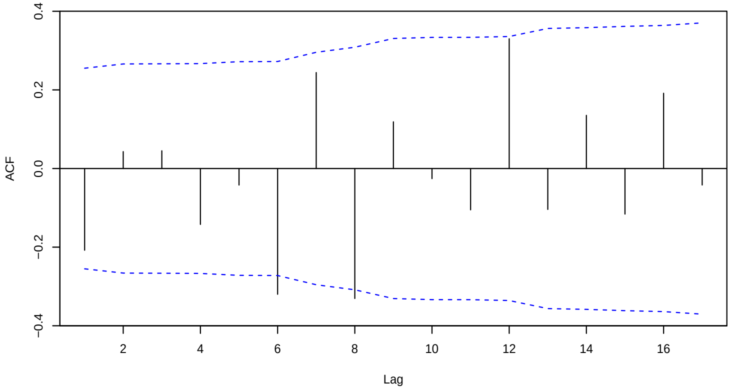
例 6.8（续 1）：

过度差分随机游动的的样本 ACF



例 6.8（续 2）：

正确差分随机游动的的样本 ACF



Outline

1. 非平稳性
   - 平稳序列与非平稳序列
   - 过度差分
   - 单位根检验
2. 其他识别方法

单位根检验：DF 检验

模型：

$$Y_t = \alpha Y_{t-1} + X_t$$

其中 $\{X_t\}$ 为平稳序列。

假设 $\{X_t\}$ 为 AR(k) 过程：

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_k X_{t-k} + e_t$$

$\{X_t\}$ 序列的 AR 特征方程：

$$\Phi(x) = 1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_k x^k = 0$$

将 X_t 带入模型中：

$$\begin{aligned} Y_t &= \alpha Y_{t-1} + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_k X_{t-k} + e_t \\ &= \alpha Y_{t-1} + \phi_1 (Y_{t-1} - \alpha Y_{t-2}) + \phi_2 (Y_{t-2} - \alpha Y_{t-3}) + \cdots \\ &\quad + \phi_k (Y_{t-k} - \alpha Y_{t-k-1}) + e_t \end{aligned}$$

AR 特征方程：

$$1 - \alpha x - \phi_1(x - \alpha x^2) - \phi_2(x^2 - \alpha x^3) - \cdots - \phi_k(x^k - \alpha x^{k+1}) = 0$$

即，

$$(1 - \alpha x)(1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_k x^k) = 0$$

单位根检验：DF 检验

AR 特征方程：

$$(1-\alpha x)(1-\phi_1 x-\phi_2 x^2-\cdots-\phi_k x^k)=0$$

- $\alpha=1$ 时，序列 $\{Y_t\}$ 非平稳，但一阶差分之后平稳。
- $|\alpha|<1$ 时，序列 $\{Y_t\}$ 平稳。
- $|\alpha|>1$ 时，序列 $\{Y_t\}$ 非平稳。

DF 检验：

- 原假设 H_0：$\alpha=1$；备择假设 H_1：$|\alpha|<1$
- 只适用于 AR(1) 模型

单位根检验：DF 检验

DF 检验的三种形式

- 无常数项的回归过程：

$$Y_t = \alpha Y_{t-1} + X_t$$

- 含常数项的回归过程：

$$Y_t = \mu + \alpha Y_{t-1} + X_t$$

- 有线性趋势的回归过程：

$$Y_t = \mu + \beta t + \alpha Y_{t-1} + X_t$$

单位根检验：ADF 检验

模型：

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + X_t$$

其中 $\{X_t\}$ 为平稳序列，不妨假设为 AR(k) 过程。

AR 特征方程：

$$(1 - \alpha_1 x - \alpha_2 x^2 - \cdots - \alpha_p x^p)(1 - \phi_1 x - \phi_2 x^2 - \cdots - \phi_k x^k) = 0$$

- $|x_i| > 1$ 时，序列 $\{Y_t\}$ 平稳
- 若有一个单位根存在，则 $1 - \alpha_1 - \alpha_2 - \cdots - \alpha_p = 0$，且序列 $\{Y_t\}$ 非平稳。

单位根检验：ADF 检验

AR 特征方程：

$$(1-\alpha_1 x-\alpha_2 x^2-\cdots-\alpha_p x^p)(1-\phi_1 x-\phi_2 x^2-\cdots-\phi_k x^k)=0$$

- 若序列 $\{Y_t\}$ 平稳，

$$\alpha_1+\alpha_2+\cdots+\alpha_p<1$$

- 若序列 $\{Y_t\}$ 非平稳，则至少存在一个单位根，即

$$1-\alpha_1-\alpha_2-\cdots-\alpha_p=0$$

ADF 检验：

- 原假设 H_0：$a=0$；备择假设 H_1：$a<0$，
  其中 $a=\alpha_1+\alpha_2+\cdots+\alpha_p-1$。
- 适用于 AR(p) 模型

ADF 检验的三种形式

- 无常数项的回归过程：

$$Y_t = \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + X_t$$

- 含常数项的回归过程：

$$Y_t = \mu + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + X_t$$

- 有线性趋势的回归过程：

$$Y_t = \mu + \beta t + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \cdots + \alpha_p Y_{t-p} + X_t$$

注：有限样本情况下或者序列含有异常点时，易出现接受原假设的现象。

Outline

1. 非平稳性
2. 其他识别方法

其他识别方法

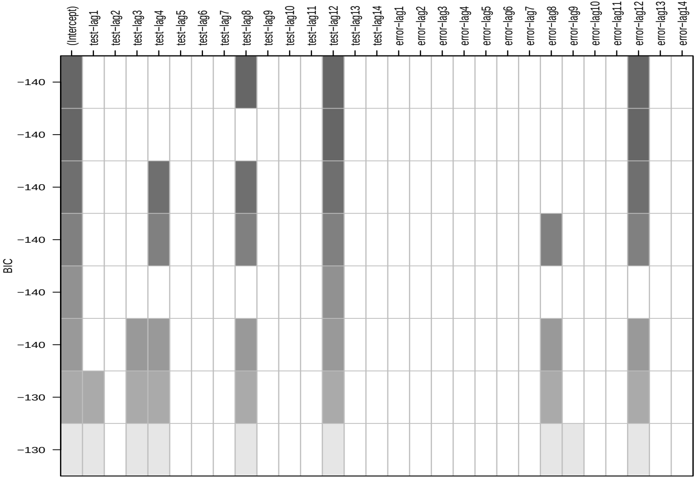
两种常用的准则

- AIC：
$$AIC = -2\log(\text{极大似然估计}) + 2k$$
- BIC：
$$BIC = -2\log(\text{极大似然估计}) + k\log(n)$$

其他识别方法

模型：

$$Y_t = 0.8Y_{t-12} + e_t + 0.7e_{t-12}$$



时间序列分析与R 语言实战
第六章：模型识别（五）：真实时间序列案例

Outline

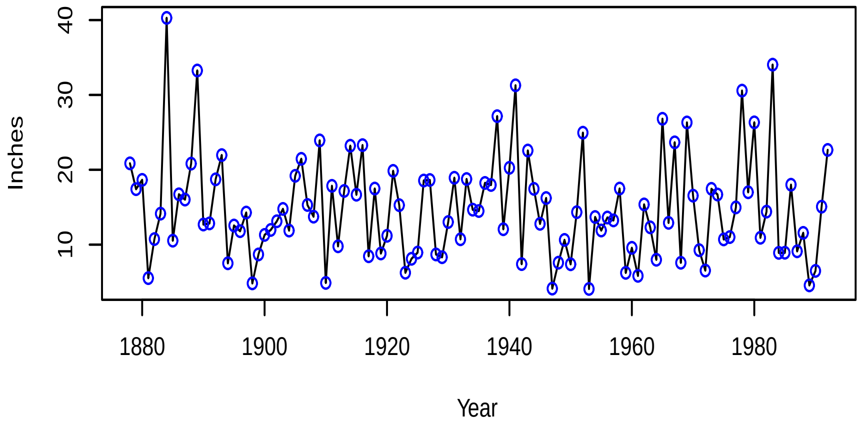
1. 洛杉矶降雨量
2. 化工颜色属性
3. 加拿大野兔丰度
4. 石油价格

Outline

1. 洛杉矶降雨量
2. 化工颜色属性
3. 加拿大野兔丰度
4. 石油价格

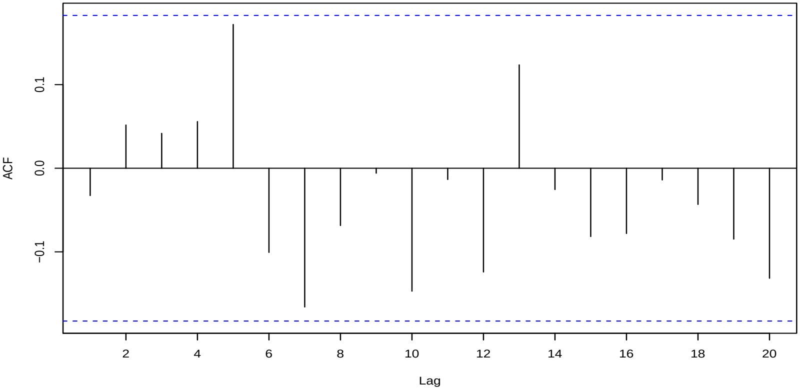
洛杉矶降雨量

LA 1878-1992期间年降水量的时间序列图

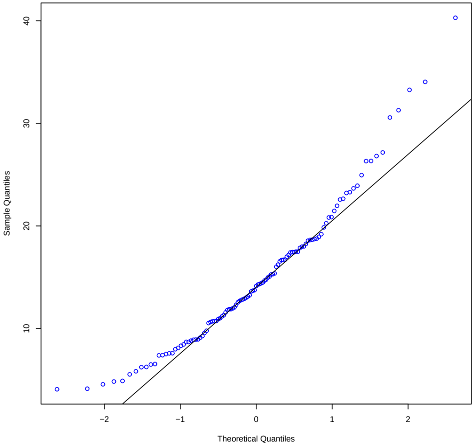


洛杉矶降雨量

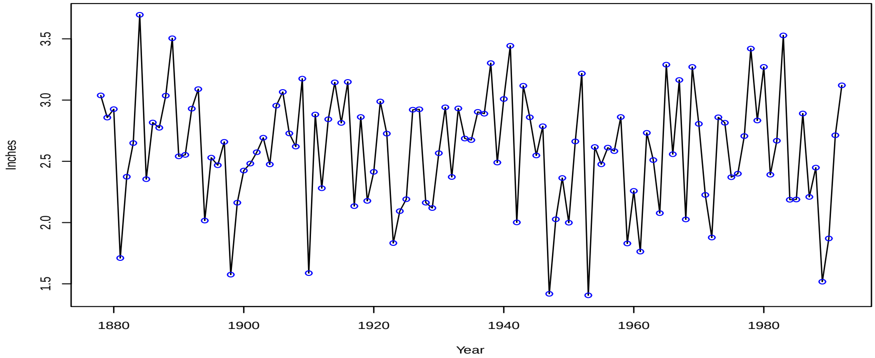
LA 年降水量的样本 ACF



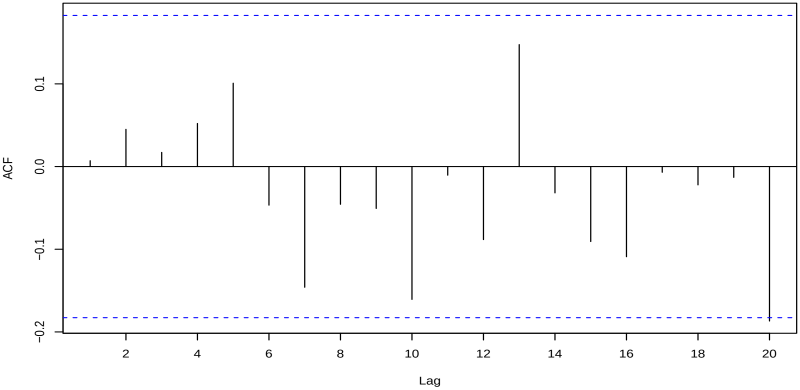
LA 年降水量的 QQ 正态图



LA 年降水量对数的时间序列图

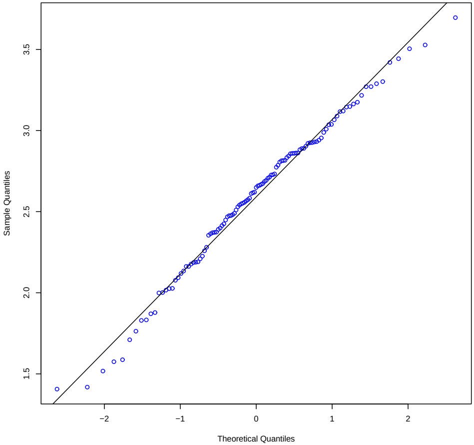


LA 年降水量对数的样本 ACF



洛杉矶降雨量

LA 年降水量对数的 QQ 正态图：$N(2.593, 0.477)$

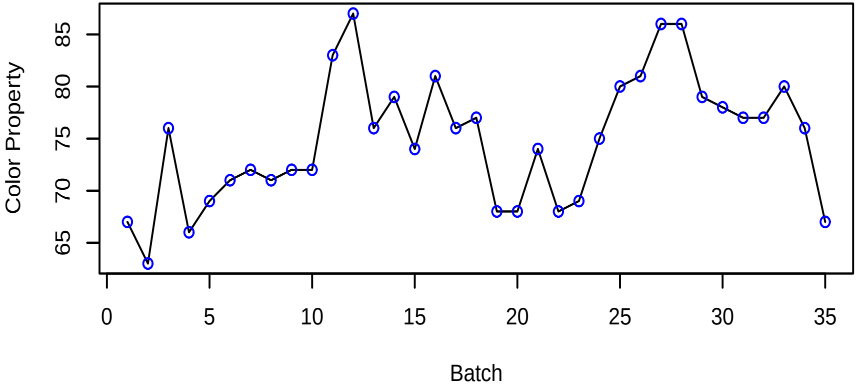


Outline

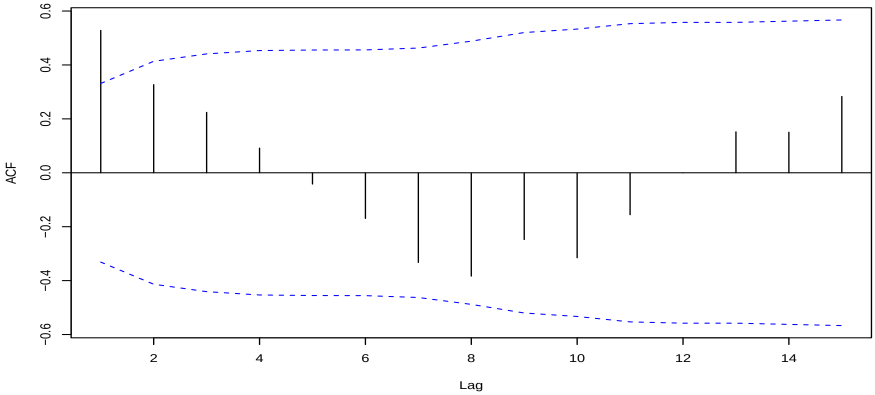
1. 洛杉矶降雨量
2. 化工颜色属性
3. 加拿大野兔丰度
4. 石油价格

化工颜色属性

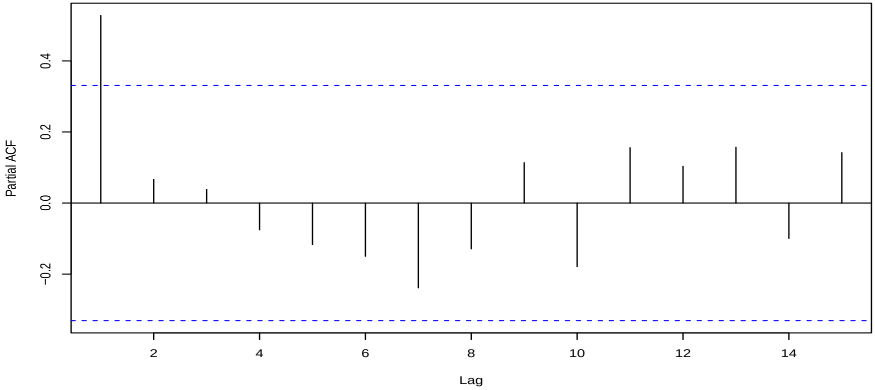
某化工过程中颜色属性的时间序列图



化工颜色属性的样本 ACF



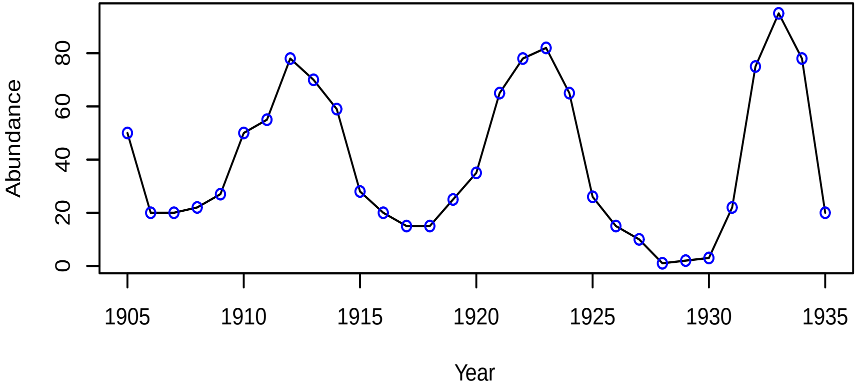
化工颜色属性的样本 PACF



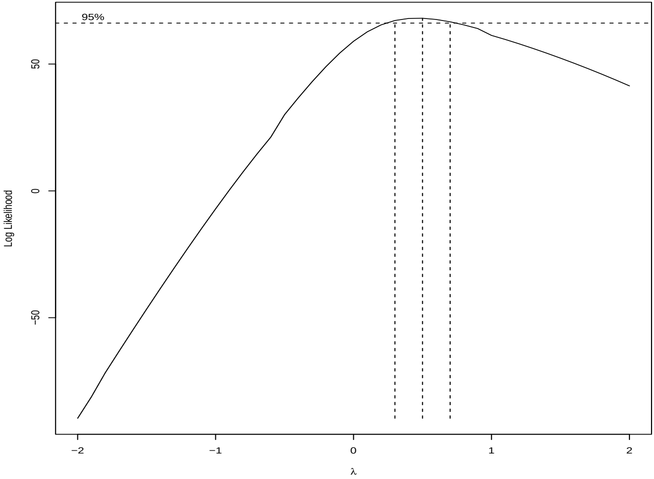
Outline

1. 洛杉矶降雨量
2. 化工颜色属性
3. 加拿大野兔丰度
4. 石油价格

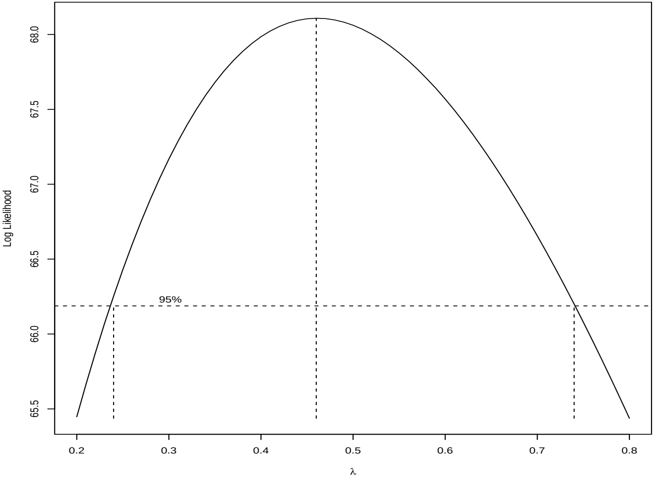
加拿大野兔时间序列图


Abundance
0
20
40
60
80
1905
1910
1915
1920
1925
1930
1935
Year

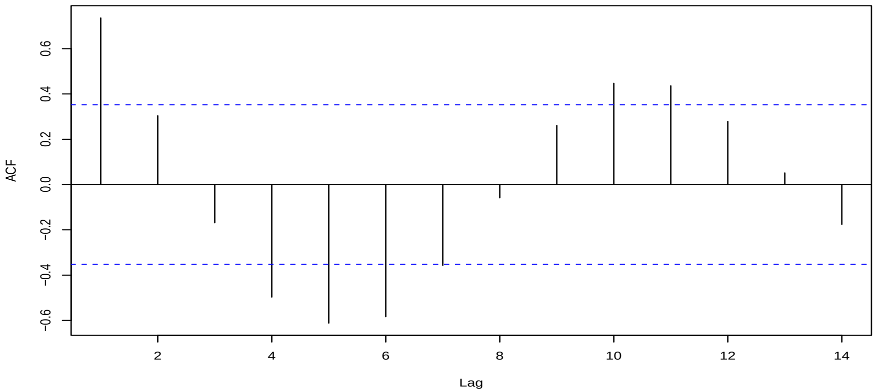
加拿大野兔丰度 Box-Cox 变换



加拿大野兔丰度 Box-Cox 变换

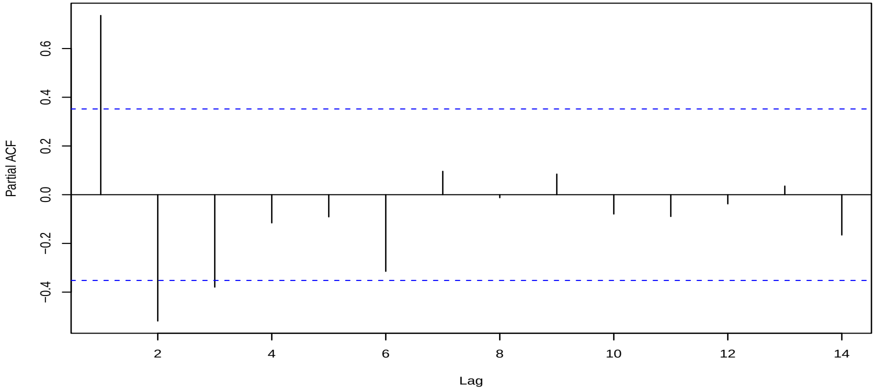


加拿大野兔丰度开方序列的的样本 ACF



加拿大野兔丰度

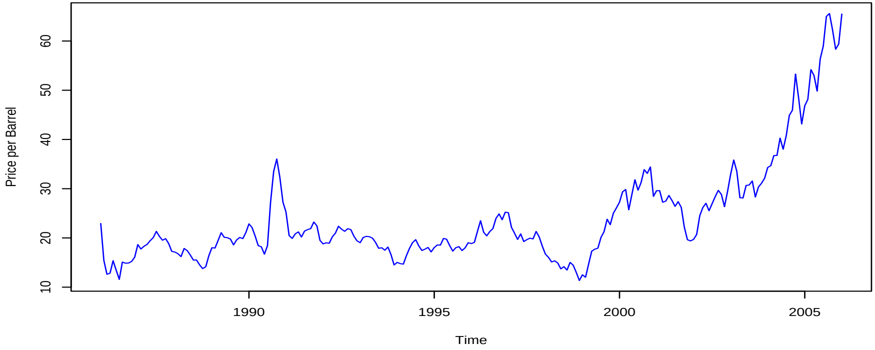
加拿大野兔丰度开方序列的的样本 PACF



Outline

石油价格

1986 年 1 月到 2006 年 1 月，每桶原油的月度价格时间序列图

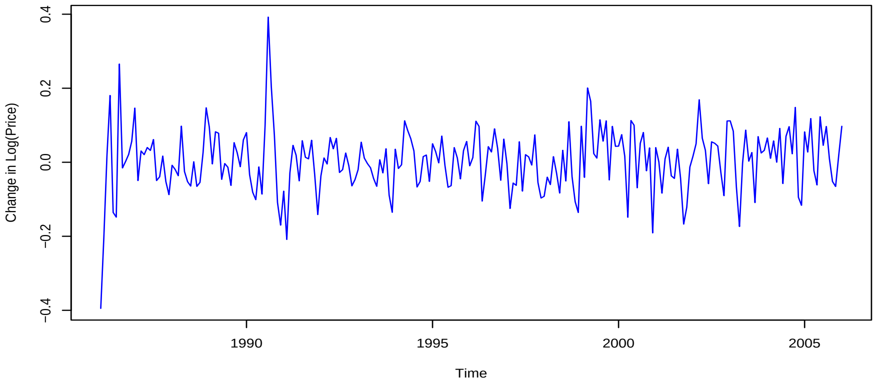


ADF 检验:

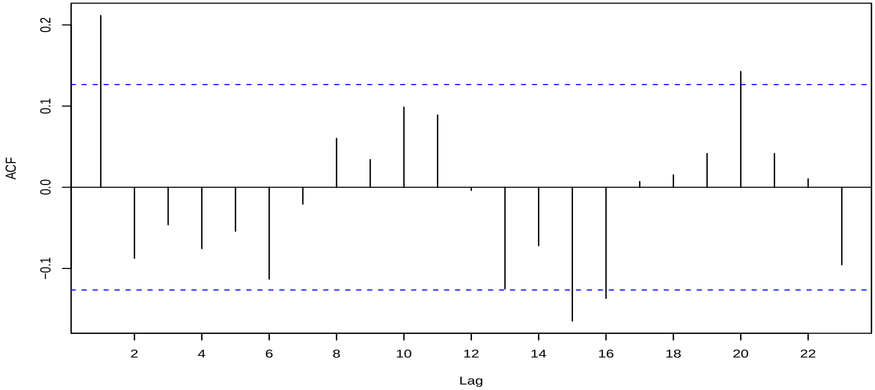
```
> adf.test(log(oil.price))

        Augmented Dickey-Fuller Test

data:  log(oil.price)
Dickey-Fuller = -1.1119, Lag order = 6, p-value = 0.9189
alternative hypothesis: stationary
```

石油价格

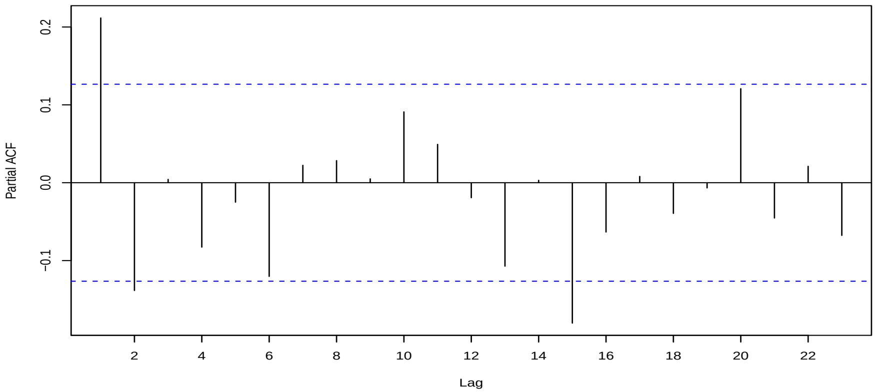
石油价格序列取对数之后再进行差分所得时间序列图



石油价格序列取对数差分的样本 ACF



石油价格序列取对数差分的样本 PACF



石油价格

石油价格序列取对数差分的 EACF

AR/MA	0	1	2	3	4	5	6	7	8	9	10	11	12	13
0	X	O	O	O	O	O	O	O	O	O	O	O	O	O
1	X	X	O	O	O	O	O	O	O	O	X	O	O	O
2	O	X	O	O	O	O	O	O	O	O	O	O	O	O
3	O	X	O	O	O	O	O	O	O	O	O	O	O	O
4	O	X	X	O	O	O	O	O	O	O	O	O	O	O
5	O	X	O	X	O	O	O	O	O	O	O	O	O	O
6	O	X	O	X	O	O	O	O	O	O	O	O	O	O
7	X	X	O	X	O	O	O	O	O	O	O	O	O	O

石油价格序列取对数差分的最优子集 ARMA 模型