时间序列分析（TIME SERIES ANALYSIS）

主讲：吴尚
复旦大学管理学院统计与数据科学系

季节模型

- 基本的季节模型SARMA(P,Q)s
- 乘法季节模型SARMA(p,q)(P,Q)s
- 非平稳的季节模型SARIMA(p,d,q)(P,D,Q)s

MA(1)12

- $Y_t = e_t - \theta e_{t-12}$
- $Cov(Y_t, Y_{t-1}) = Cov(e_t - \Theta e_{t-12}, e_{t-1} - \Theta e_{t-13}) = 0$
- $Cov(Y_t, Y_{t-12}) = Cov(e_t - \Theta e_{t-12}, e_{t-12} - \Theta e_{t-24}) =$
$= -\Theta \sigma_e^2$

MA(Q)s

- $Y_t = e_t - \Theta_1 e_{t-s} - \Theta_2 e_{t-2s} - \cdots - \Theta_\mathrm{Q} e_{t-Qs}$
- $\gamma_{ks} = Cov(Y_t, Y_{t-ks}) = Cov(e_t - \Theta_1 e_{t-s} - \cdots - \Theta_\mathrm{Q} e_{t-Qs}, e_{t-ks} - \Theta_1 e_{t-(k+1)s} - \cdots - \Theta_\mathrm{Q} e_{t-(k+Q)s})$
- $\gamma_{ks} = \begin{cases} (1 + \Theta_1^2 + \cdots + \Theta_Q^2)\sigma_e^2, & k = 0 \\ (-\Theta_k + \sum_{i=1}^{Q-k} \Theta_i \Theta_{k+i})\sigma_e^2, & 1 \le k \le Q \\ 0, & k > Q \end{cases}$

MA(Q)s

- $\rho_{ks} = \begin{cases} 1, & k = 0 \\ \frac{-\Theta_k + \sum_{i=1}^{Q-k} \Theta_i \Theta_{k+i}}{1+\Theta_1^2+\cdots+\Theta_Q^2}, & 1 \le k \le Q \\ 0, & k > Q \end{cases}$

- MA特征多项式:
  $$\Theta(x) = 1 - \Theta_1 x^s - \Theta_2 x^{2s} - \cdots - \Theta_{\mathrm{Q}} x^{Qs}$$

AR(1)12

- $Y_t = \Phi Y_{t-12} + e_t, \qquad |\Phi| < 1$
- $\gamma_k = Cov(Y_t, Y_{t-k}) = E[(\Phi Y_{t-12} + e_t)Y_{t-k}] = \Phi\gamma_{k-12}$
- $\rho_k = \Phi\rho_{k-12} \quad k = 1,2, \dots$
- $\rho_{12k} = \Phi^k \quad k = 1,2, \dots$

AR(1)12

- $\rho_k = \Phi\rho_{k-12} \quad k = 1,2,\ldots$
- $\rho_1 = \Phi\rho_{-11} = \Phi\rho_{11}$, $\rho_{11} = \Phi\rho_{-1} = \Phi\rho_1$
- $\rho_{11} = \rho_1 = 0$
- $\rho_k = 0 \quad k \neq 12,24, \ldots, 12m$

- AR特征多项式:

  $$\Phi(x) = 1 - \Phi_1 x^s - \Phi_2 x^{2s} - \cdots - \Phi_\mathrm{P} x^{Ps}$$

- 平稳条件是$\Phi(x) = 0$的所有根的模长大于1

SARMA(P,Q)s

- $\Phi(B^s)Y_t = \Theta(B^s)e_t$
- 其中:
- $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_{\mathrm{P}} B^{Ps}$
- $\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_{\mathrm{Q}} B^{Qs}$

SARMA(P,Q)s模型相关性特征

模型	自相关系数 ACF	偏自相关系数 PACF
AR(P)s	拖尾 (在滞后ks步，k=1,2...)	Ps阶截尾
MA(Q)s	Qs阶截尾	拖尾 (在滞后ks步，k=1,2...)
ARMA(P,Q)s	拖尾 (在滞后ks步，k=1,2...)	拖尾 (在滞后ks步，k=1,2...)

ACF或PACF的值在滞后不是ks时均为0 (k=1,2,···)

乘法季节模型 SARMA(p,q)(P,Q)s

- $\Phi(B^s)\phi(B)Y_t = \Theta(B^s)\theta(B)e_t$

- 其中，
- $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_{\mathrm{p}} B^p$
- $\Phi(B^s) = 1 - \Phi_1 B^s - \Phi_2 B^{2s} - \cdots - \Phi_{\mathrm{P}} B^{Ps}$
- $\theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_{\mathrm{q}} B^q$
- $\Theta(B^s) = 1 - \Theta_1 B^s - \Theta_2 B^{2s} - \cdots - \Theta_{\mathrm{Q}} B^{Qs}$

SARMA(p,q)(P,Q)s示例

- 例一： SARMA(0,1)(0,1)12
- $Y_t = (1-\theta B)(1-\Theta B^{12})e_t$， $|\Theta| < 1,\quad |\theta| < 1$
- $Y_t = e_t - \theta e_{t-1} - \Theta e_{t-12} + \theta\Theta e_{t-13}$

- 例二：SARMA(0,1)(1,0)12
- $(1-\Phi B^{12})Y_t = (1-\theta B)e_t,\quad |\Phi| < 1,\quad |\theta| < 1$
- 即： $Y_t = \Phi Y_{t-12} + e_t - \theta e_{t-1}$

非平稳SARIMA(p,d,q)(P,D,Q)s

- $\Phi(B^s)\phi(B)\nabla^d\nabla_s^D Y_t = \Theta(B^s)\theta(B)e_t$
- 其中$\nabla_s^D = (1-B^s)^D, \nabla^d = (1-B)^d$

- 例： ARIMA(0,1,1)(0,1,1)12
- $\nabla_{12}\nabla Y_t = \Theta(B^{12})\theta(B)e_t$
- $(1-B^{12})(1-B)Y_t = (1-\Theta B^{12})(1-\theta B)e_t$
- $(1-B-B^{12}+B^{13})Y_t = (1-\Theta B^{12}-\theta B+\theta\Theta B^{13})e_t$

模型识别

- Step 1: 作数据的时间序列图。从图上或者数据的背景查看是否有趋势性、季节性;
- Step 2: 进行必要的差分，一般原则是:
  - 如果是只有季节性，没有趋势性，只做季节差分;
  - 如果有线性趋势，无明显的季节性，则做1阶差分;
  - 如果是非线性趋势，可考虑先对数据进行变换后再做1阶差分;
  - 如果同时有线性趋势和季节性，则要同时进行季节差分和1阶差分;
  - 如果既没有线性趋势，也无明显的季节性，则不需做任何差分。

模型识别

- Step 3: 对差分后（如果已进行差分）的数据序列，画出ACF 和PACF，并通过ACF 和PACF来进行模型的初步识别，具体如下:
  - 非季节项的确定：检验最初几步的ACF和PACF，如果ACF明显在置信区间之外，则指示应该有非季节的MA项；如果PACF明显在置信区间之外，则指示应该有非季节的AR项。
  - 季节项：与非季节项类似，只不过这时要看ACF和PACF在季节周期的倍数上的表现，如，对于月度数据，应该看滞后12, 24, 36, 等等。

模型拟合、模型诊断

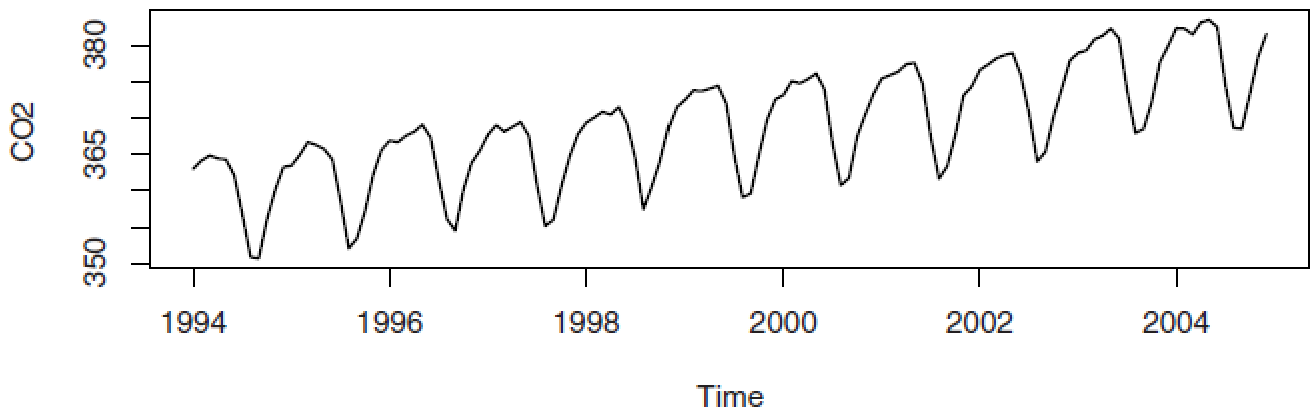
- Step 4: 对识别的模型进行拟合估计。
- Step 5: 检验残差： ACF, 散点图，QQ图，白噪声检验;
  如果有多个模型，需要计算比较AIC或BIC 。
- 如果模型不够理想，则重新识别（回到Step 3或Step 2）

季节项识别

- 一般只用一次季节差分;
- 应该避免多于一个季节项或SAR+SMA同时出现在同一模型中，这往往会导致模型的过度拟合或出现参数估计的问题。
- 如果从ACF和PACF的图形中较难判断时，可考虑利用信息准则来进行模型识别。

例一：CO_2含量

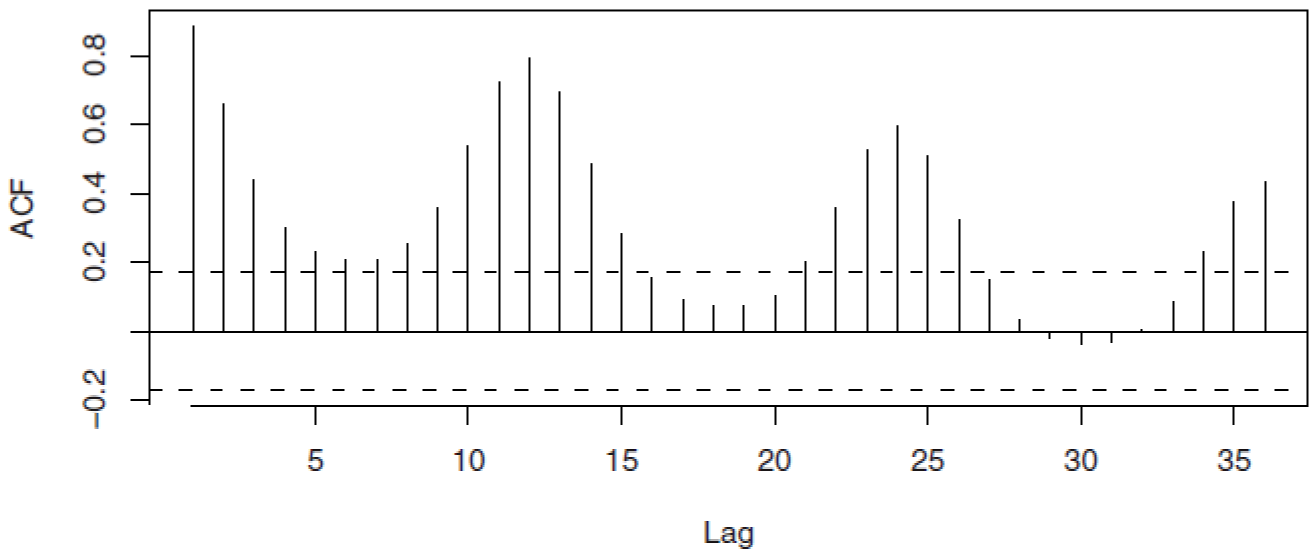
Exhibit 10.1 Monthly Carbon Dioxide Levels at Alert, NWT, Canada




```
> data(co2)
> win.graph(width=4.875,height=3,pointsize=8)
> plot(co2,ylab='CO2')
```

例一：CO_2含量

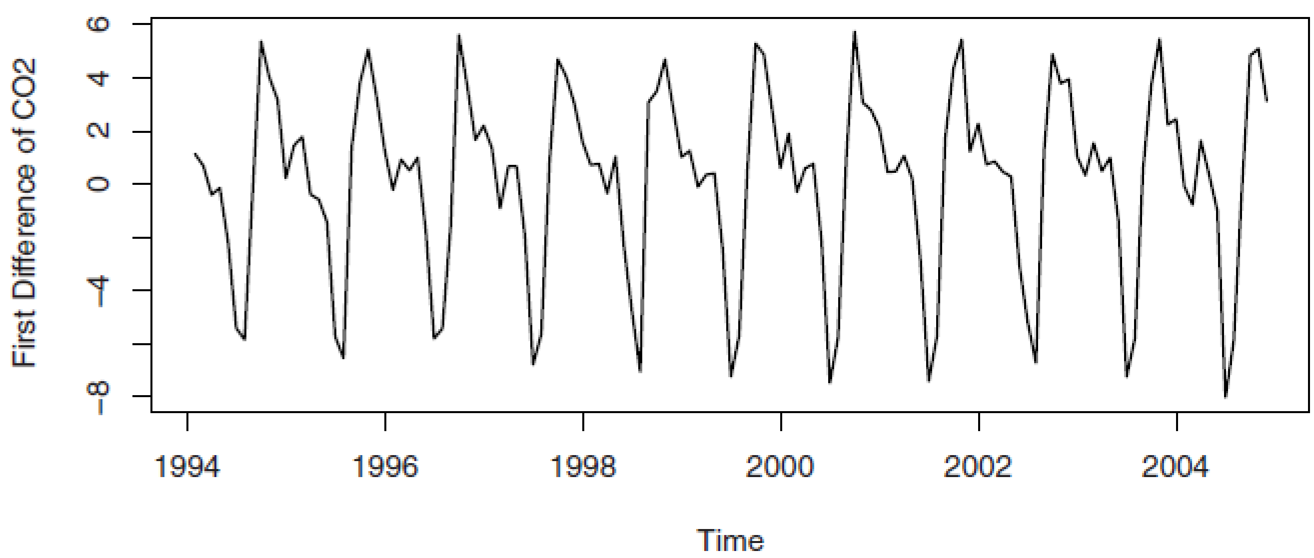
Exhibit 10.5 Sample ACF of CO_2 Levels




```
> acf(as.vector(co2),lag.max=36)
```

例一：CO_2含量

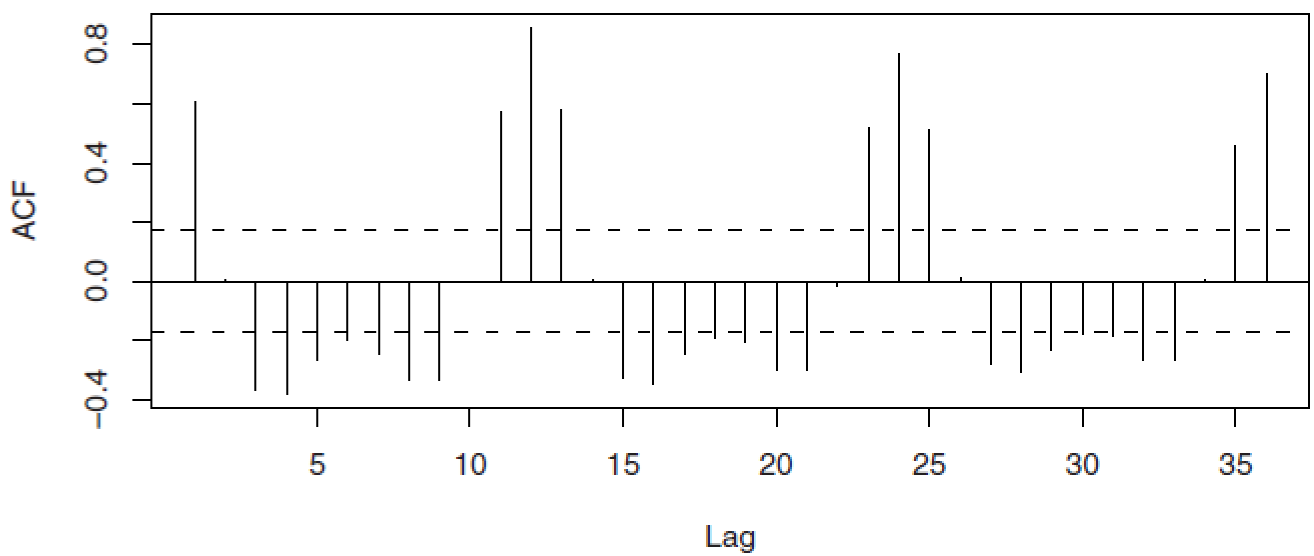
Exhibit 10.6 Time Series Plot of the First Differences of CO_2 Levels




```
> plot(diff(co2),ylab='First Difference of CO2',xlab='Time')
```

例一：CO_2含量

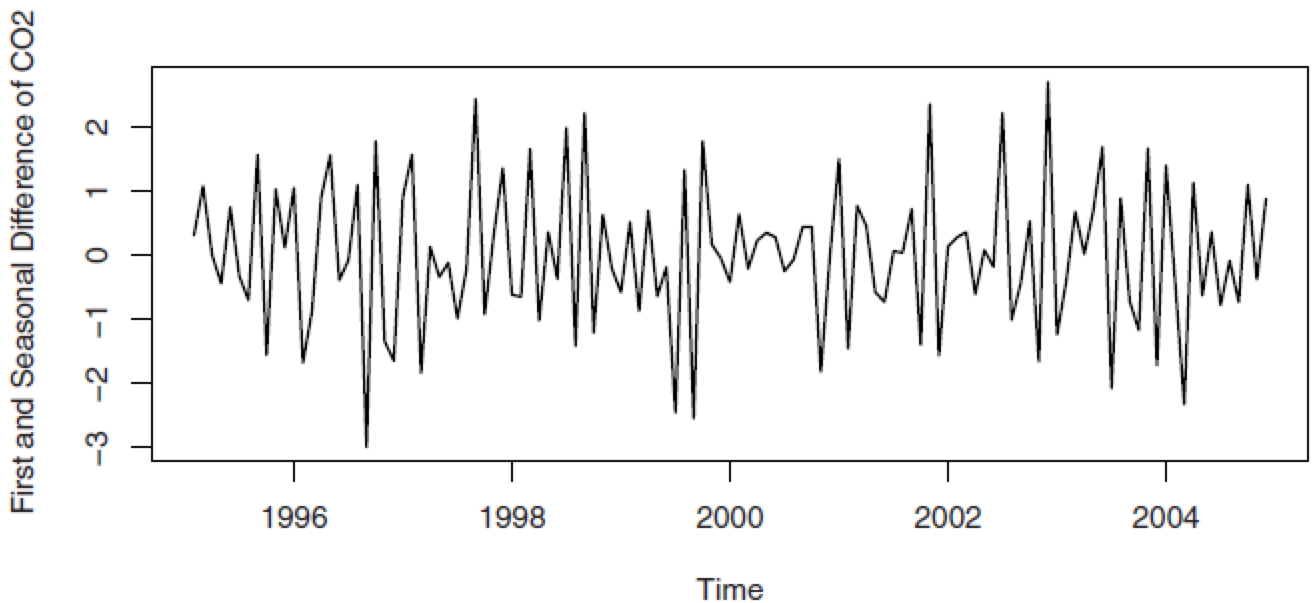
Exhibit 10.7 **Sample ACF of First Differences of CO_2 Levels**




```
> acf(as.vector(diff(co2)),lag.max=36)
```

例一：CO_2含量

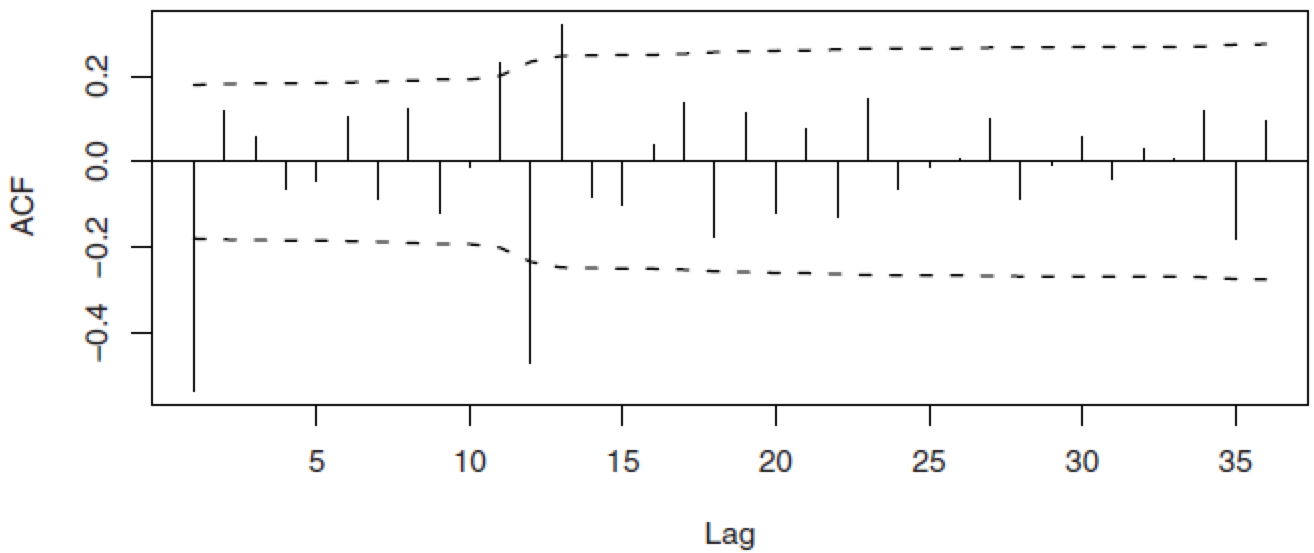
Exhibit 10.8 **Time Series Plot of First and Seasonal Differences of CO_2**




```
> plot(diff(diff(co2),lag=12),xlab='Time',
   ylab='First and Seasonal Difference of CO2')
```

例一：CO_2含量

Exhibit 10.9 Sample ACF of First and Seasonal Differences of CO_2




```
> acf(as.vector(diff(diff(co2),lag=12)),lag.max=36,ci.type='ma')
```

例一：CO_2含量

- ARIMA(0,1,1)(0,1,1)12
- $(1-B^{12})(1-B)Y_t = (1-\Theta B^{12})(1-\theta B)e_t$

Exhibit 10.10 Parameter Estimates for the CO_2 Model

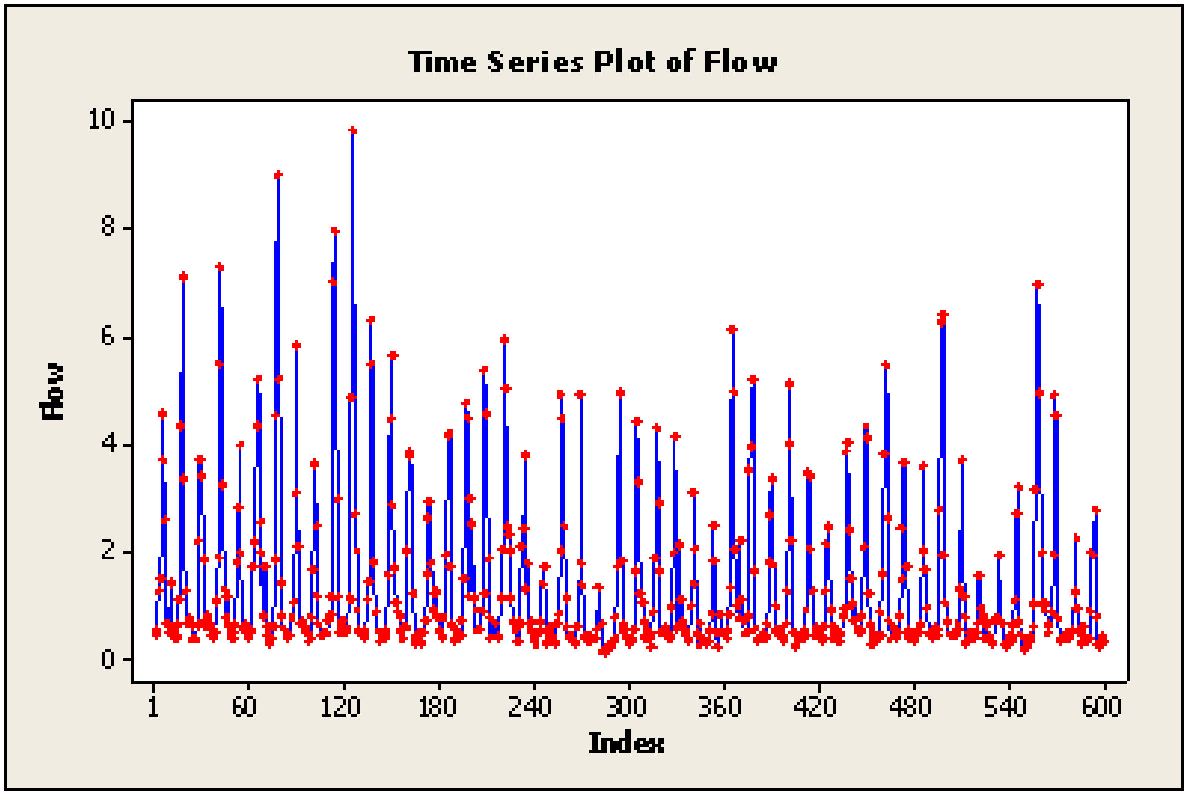
Coefficient	θ	Θ
Estimate	0.5792	0.8206
Standard error	0.0791	0.1137

$\hat{\sigma}_e^2 = 0.5446$: log-likelihood $= -139.54$, AIC $= 283.08$

```
> m1.co2=arima(co2,order=c(0,1,1),seasonal=list(order=c(0,1,1),
   period=12))
> m1.co2
```

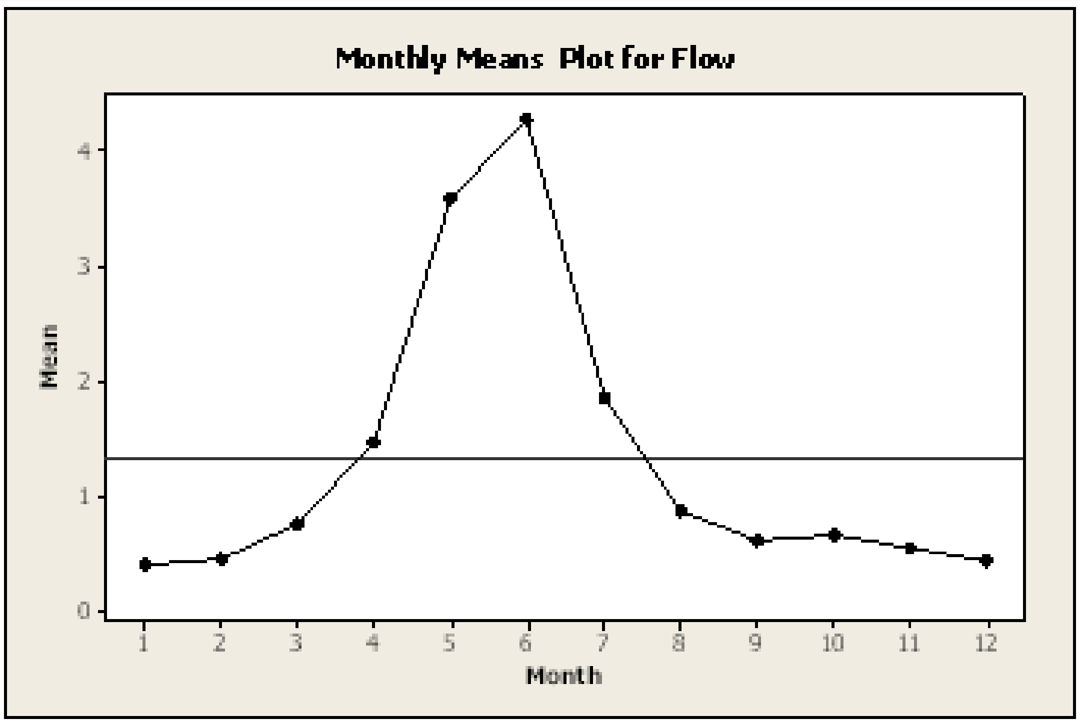
例二：月度流量数据

科罗拉多河在某个固定观测点的月度流量数据，有600个观测值。



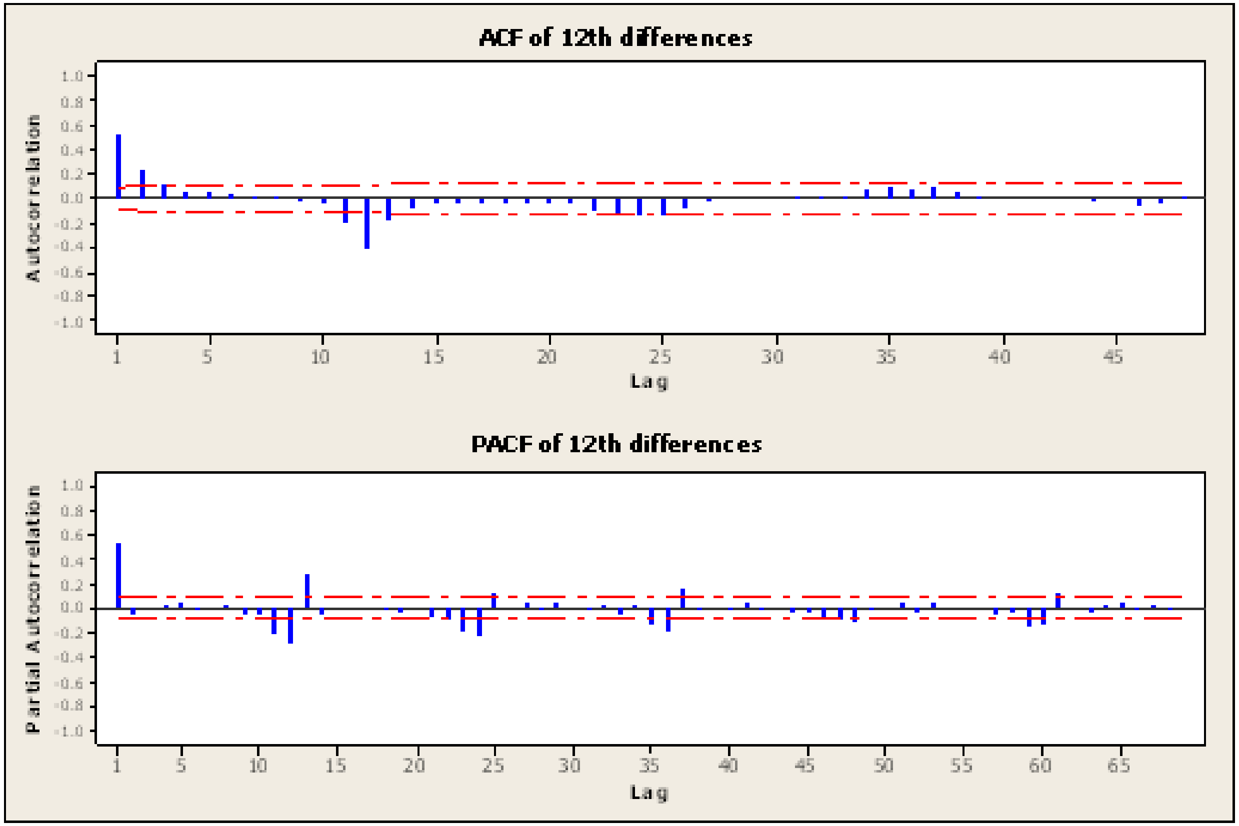
例二：月度流量数据

根据数据的实际背景，猜测数据可能会有季节性因素，但数据太多，从时间序列图中很难看出，因此，计算月度平均有：

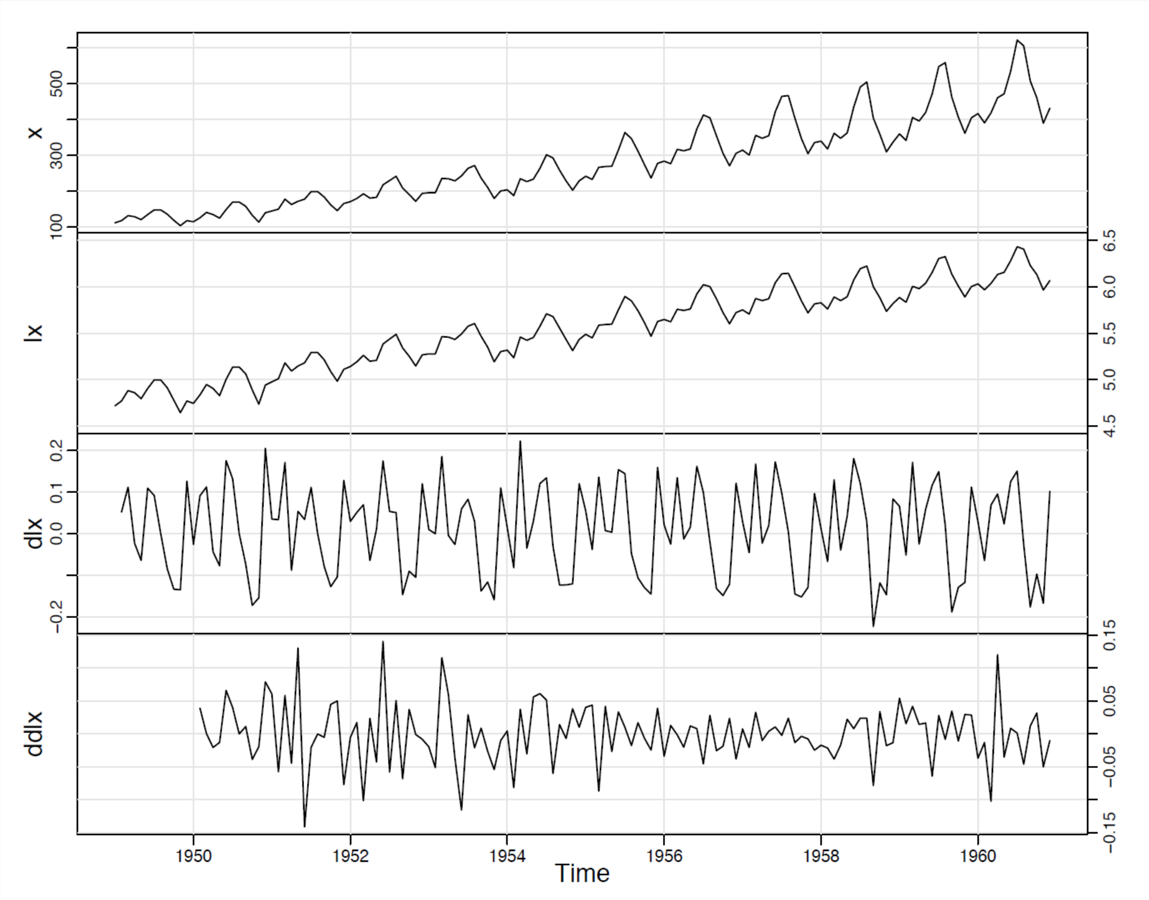


例二：月度流量数据

尝试对原时间序列进行滞后12步差分（季节差分），并画出差分后序列的ACF、PACF



例三：国际航空旅客数据


x
100
300
500
lx
4.5
5.0
5.5
6.0
6.5
dlx
−0.2
0.0
0.1
0.2
ddlx
−0.15
−0.05
0.05
0.15
1950
1952
1954
1956
1958
1960
Time

例三：国际航空旅客数据

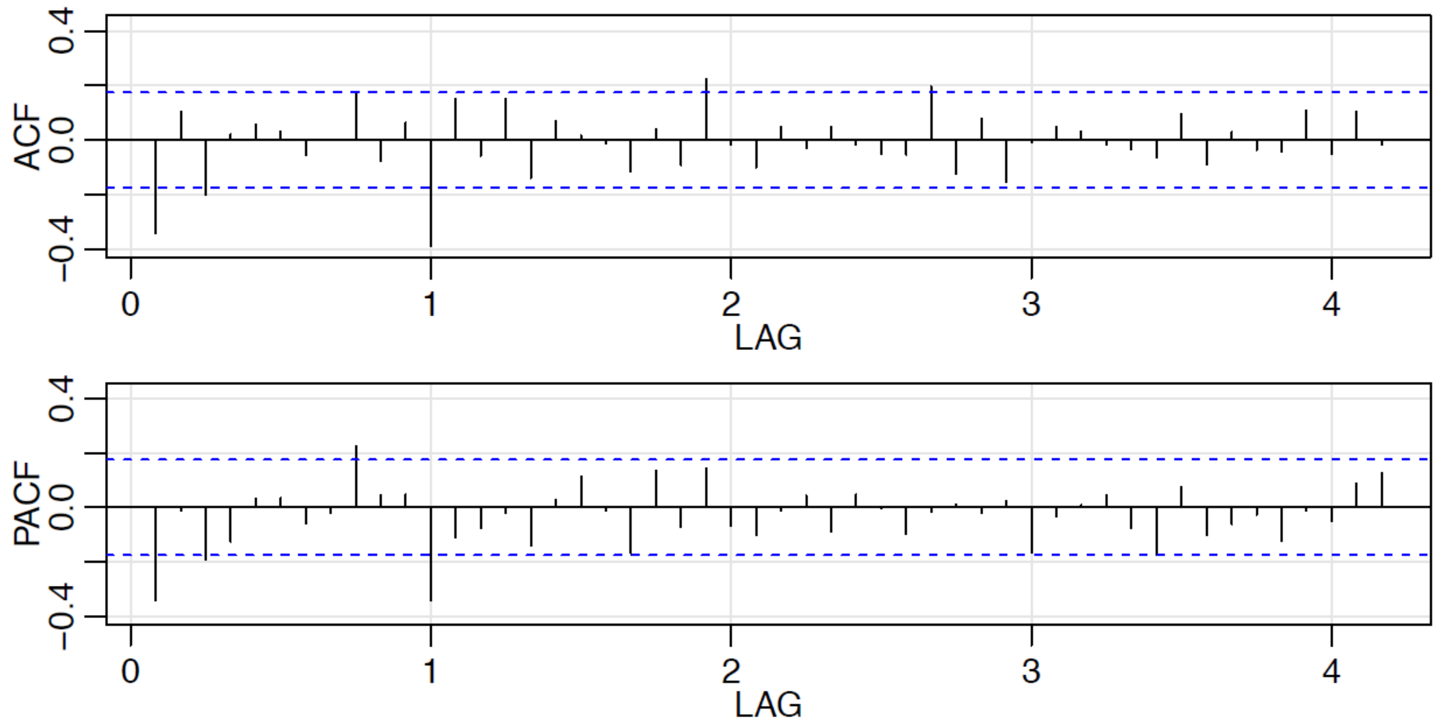


Fig. 3.23. *Sample ACF and PACF of* `ddlx` ($\nabla_{12}\nabla \log x_t$).

例三：国际航空旅客数据

- 由ACF和PACF可以初步判断，

季节项：1s(s=12)阶截尾，PACF在1s,2s,3s,…拖尾，于是P=0, Q=1,s=12

非季节项：ARMA(1,1)

- 尝试建模SARIMA(1,1,1)(0,1,1)12

```
sarima(lx, 1,1,1, 0,1,1,12)
 Coefficients:
          ar1      ma1      sma1
       0.1960  -0.5784   -0.5643
 s.e.  0.2475   0.2132    0.0747
 sigma^2 estimated as 0.001341
 $AIC -5.5726  $AICc -5.556713  $BIC -6.510729
```

例三：国际航空旅客数据

- 显然，非季节项AR系数可能为零，因此尝试SARIMA(0,1,1)(0,1,1)12或SARIMA(1,1,0)(0,1,1)12

```
sarima(lx, 0,1,1, 0,1,1,12)
 Coefficients:
           ma1      sma1
       -0.4018   -0.5569
 s.e.   0.0896    0.0731
 sigma^2 estimated as 0.001348
 $AIC -5.58133  $AICc -5.56625  $BIC -6.540082
sarima(lx, 1,1,0, 0,1,1,12)
 Coefficients:
           ar1      sma1
       -0.3395   -0.5619
 s.e.   0.0822    0.0748
 sigma^2 estimated as 0.001367
 $AIC -5.567081  $AICc -5.552002  $BIC -6.525834
```

例三：国际航空旅客数据

- SARIMA(0,1,1)(0,1,1)12的模型检验