中國計量大學
CHINA JILIANG UNIVERSITY

数学建模

中国计量大学

第6章 时间序列分析模型

6.1 时间序列的基本概念

6.1.1 时间序列的定义

6.1.2 时间序列的特征统计量

6.1.3 时间序列的平稳性

6.1.4 时间序列分析模型的分类

6.1 时间序列的基本概念

6.1.1 时间序列的定义

- 自然界以及社会经济生活中存在着大量的指标都按照年、季、月或日等进行统计，随着时间的推移，就形成了相应的**时间序列**。
- **时间序列** 是某一统计指标长期变动的数量表现。
- **时间序列分析** 就是分析和研究时间序列在长期变动过程中所存在的统计规律性。

6.1 时间序列的基本概念

6.1.1 时间序列的定义

- **时间序列**：按时间顺序排列的一组随机变量

$$\ldots, X_1, X_2, \ldots, X_t, \ldots$$

简记为 $\{X_t, t \in T\}$ 或 $\{X_t\}$

- **观测值序列**: 时间序列的 n 个有序观测值，称之为长度为 n 的观测值序列

$$x_1, x_2, \cdots, x_n \quad \text{或} \quad \{x_t : t = 1, 2, \cdots, n\}$$

时间序列的例子

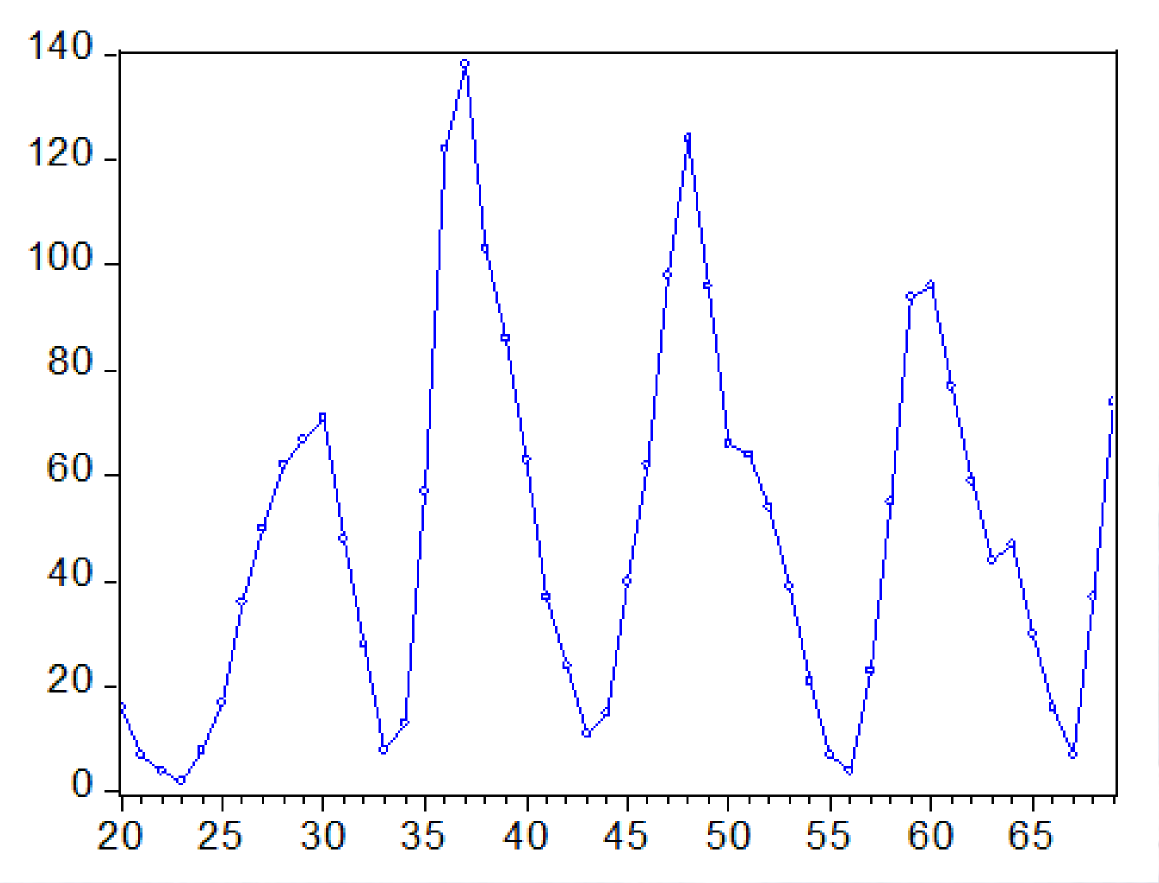
- **最早的时间序列分析可追溯到7000年前的古埃及**

古埃及人把尼罗河涨落情况逐天记录下来，得到所谓的时间序列。对这个时间序列长期的观察使他们发现尼罗河的涨落非常有规律。由于掌握了尼罗河泛滥的规律，使得古埃及的农业迅速发展，从而创建了埃及灿烂的史前文明。

时间序列的例子

- 1820年—1869年的太阳黑子数依时间变化图

  - 横轴是时间指标 *t* (这里 *t* 以年为单位)
  - 纵轴表示在时间 *t* 内太阳黑子个数的观测值
  - 这种图称为时间序列图

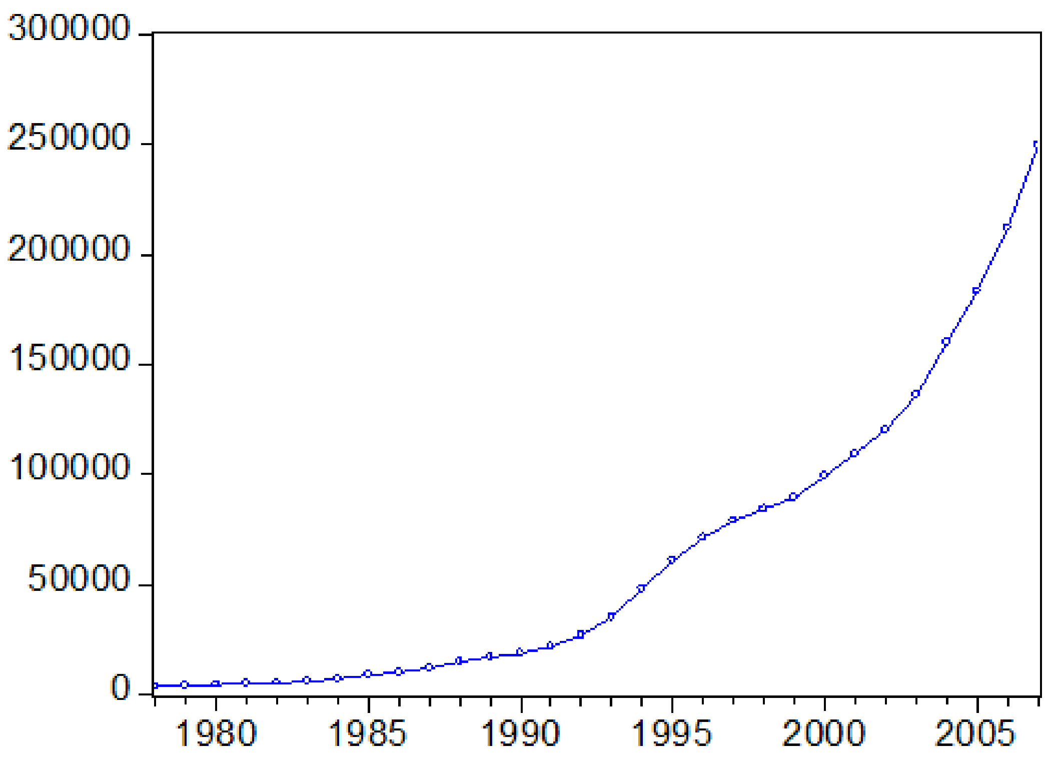


时间序列的例子

- 我国GDP数据（1978年-2007年）的时间序列图

  - GDP：国内生产总值
  - 常常被看成反映一个国家（地区）经济状况的重要指标

(单位:亿元)

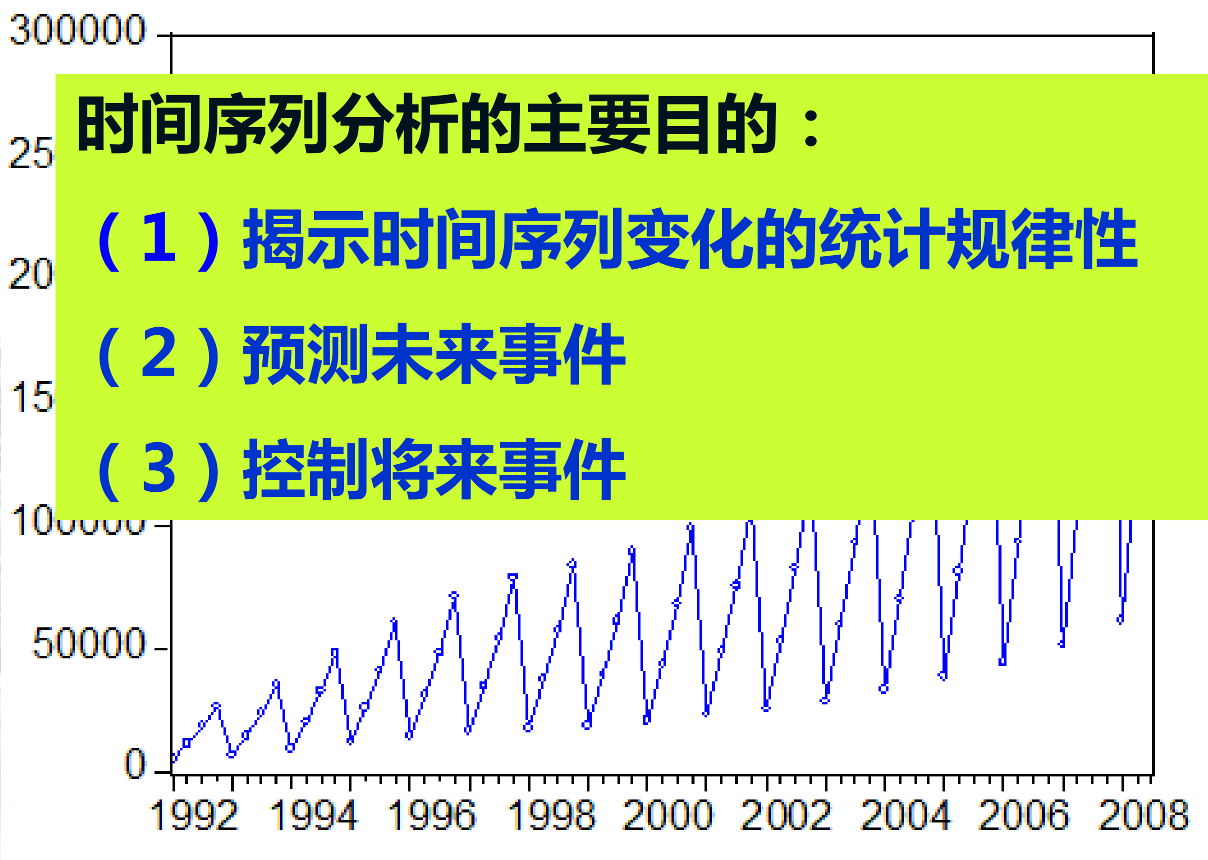


精思國計 細量民生

时间序列的例子

- **我国GDP季度数据（1992年第一季度至2008年第三季度）的时间序列**

(单位:亿元)



时间序列分析的主要目的：

（1）揭示时间序列变化的统计规律性

（2）预测未来事件

（3）控制将来事件

6.1 时间序列的基本概念

6.1.2 时间序列的特征统计量

- **均值** $\mu_X(t) = E[X_t], \quad t \in T$
- **方差** $D_X(t) = \gamma_X(t,t) = E[X_t - \mu_X(t)]^2, \quad t \in T$
- **自协方差**

$$\gamma_X(t,s) = E[(X_t - \mu_X(t))(X_s - \mu_X(s))], \quad t,s \in T$$

- **自相关系数**

$$\rho_X(t,s) = \frac{\gamma_X(t,s)}{\sqrt{D_X(t) \cdot D_X(s)}}, \quad t,s \in T$$

特征统计量的估计：

对于时间序列 $\{X_t, t \in T\}$ 在 n 个时刻的观测值 $x_1, \cdots, x_n$

- 样本均值 $\hat{\mu} = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- 样本方差 $\hat{\gamma}(0) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$
- 样本延迟k自协方差 $\hat{\gamma}(k) = \frac{1}{n}\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})$
- 样本延迟k自相关系数 $\hat{\rho}(k) = \frac{\hat{\gamma}(k)}{\hat{\gamma}(0)} = \frac{\sum_{i=1}^{n-k}(x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$

精思国计 细量民生

6.1 时间序列的基本概念

6.1.3 时间序列的平稳性

时间序列过程处于一种平稳状态：其主要性质与所考察的起始点无关。

- 严平稳
- 宽平稳

6.1 时间序列的基本概念

6.1.3 时间序列的平稳性

- **严平稳**

如果时间序列 $\{X_t, t \in T\}$ 的概率分布与时间 t 无关，则称该时间序列为**严平稳**的.

严平稳要求时间序列的**所有统计性质**都不会随着时间的推移而发生变化，这个要求比较苛刻。

6.1 时间序列的基本概念

6.1.3 时间序列的平稳性

- **宽平稳**

如果时间序列 $\{X_t, t\in T\}$ 满足

1) $E[X_t]=\mu$, μ为常数， $\forall t\in T$

2) $E[X_t^2]$为常数， $\forall t\in T$

3) $\gamma_X(t,s)=\gamma_X(k+t,k+s)$， $\forall t,s,k$ 且 $k+s,k+t\in T$

则称 $\{X_t, t\in T\}$ 为**宽平稳**的.

宽平稳只要求时间序列的低阶矩（一阶、二阶）都不会随着时间的推移而发生变化。

6.1 时间序列的基本概念

6.1.4 时间序列分析模型的分类

- 确定型时间序列模型

该方法将时间序列看作四部分组成：

- 长期变动趋势（T_t）：随时间推移逐渐增加或减少的长期变化趋势
- 季节变动趋势（S_t）：依一固定周期重复性的波动趋势
- 循环变动趋势（C_t）：周期长度不固定的波动趋势
- 不规则变动（R_t）：由许多不确定因素引起的变化（不可预测性）

6.1 时间序列的基本概念

6.1.4 时间序列分析模型的分类

- 确定型时间序列模型

（1）加法模型： $X_t = T_t + S_t + C_t + R_t$

（2）乘法模型： $X_t = T_t \cdot S_t \cdot C_t \cdot R_t$

（3）混合模型： $X_t = T_t \cdot S_t + R_t$； $X_t = S_t + T_t \cdot C_t \cdot R_t$

一般地，$E(R_t) = 0, D(R_t) = \sigma^2$

6.1 时间序列的基本概念

6.1.4 时间序列分析模型的分类

- 线性时间序列模型

  （1）自回归滑动平均（ARMA）模型

  （2）自回归综合滑动平均（ARIMA）模型

  （3）季节ARIMA(SARIMA)模型

- 非线性时间序列模型

  （1）自激励门限自回归（SETAR）模型

  （2）双线性（BL）模型

  （3）指数自回归(EAR)模型

6.1 时间序列的基本概念

- **时间序列建模的Box-Jenkins (B-J) 方法**

Box & Jenkins (1970)：

《Time Series Analysis: Forecasting and Control》

提出的一种时间序列建模方法

B-J 方法：是对时间序列进行模型识别、估计和诊断的系统方法，主要适用于单变量、同方差场合的线性模型，是线性时间序列模型的一种经典建模方法。

精思國計 細量民生

Thanks

6.2 ARMA模型

6.2.1 AR模型

6.2.2 MA模型

6.2.3 ARMA模型

引言

- 在时间序列的统计分析中, 平稳时间序列是一类重要的随机序列
- 平稳时间序列的分析和建模最常见的就是ARMA(Auto Regressive Moving Average)模型
- ARMA模型在实际应用中有许多优点:
  - 便于确定模型，由于是线性模型, 需要确定的参数少
  - 便于分析数据的结构和内在性质
  - 便于在最小方差意义下进行最佳预测和控制

引言

一般地，ARMA模型可以分为三种情形：

- AR（Auto Regressive）模型
- MA（Moving Average）模型
- ARMA（Auto Regressive Moving Average）模型

6.2.2 AR模型

- 由于许多实际系统惯性的作用，时间序列往往存在着前后依存关系
- 如果变量 X_t 当前的取值与过去直到 $t-p$ 期的取值相关，可建立：

p 阶自回归模型 —— AR(p)

（1）AR（p）模型的定义---p 阶自回归模型

AR(p) 模型的一般形式为

$$\begin{cases} X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t, \quad \phi_p \neq 0 \\ \varepsilon_t \sim WN(0, \sigma^2) \\ \forall s < t, \quad E[X_s \varepsilon_t] = 0 \end{cases}$$

特别当 $\phi_0 = 0$ 时，称为**中心化 AR(p) 模型.**

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t$$

AR(p)序列中心化变换

令

$$Y_t = X_t - \mu$$

其中，$\mu = \dfrac{\phi_0}{1-\phi_1-\cdots-\phi_p}$

则，$\{Y_t\}$ 为 $\{X_t\}$ 的中心化序列，即

$$Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + \varepsilon_t$$

AR(p) — 自回归系数多项式

引进延迟算子B, 中心化 AR(p) 模型可简记为

$$\Phi(B)X_t = \varepsilon_t$$

其中，自回归系数多项式

$$\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \cdots - \phi_p B^p$$

（2）AR(p) 模型的平稳性判别

判别原因：

AR模型是平稳时间序列的拟合模型，但AR模型本身不一定是平稳的。

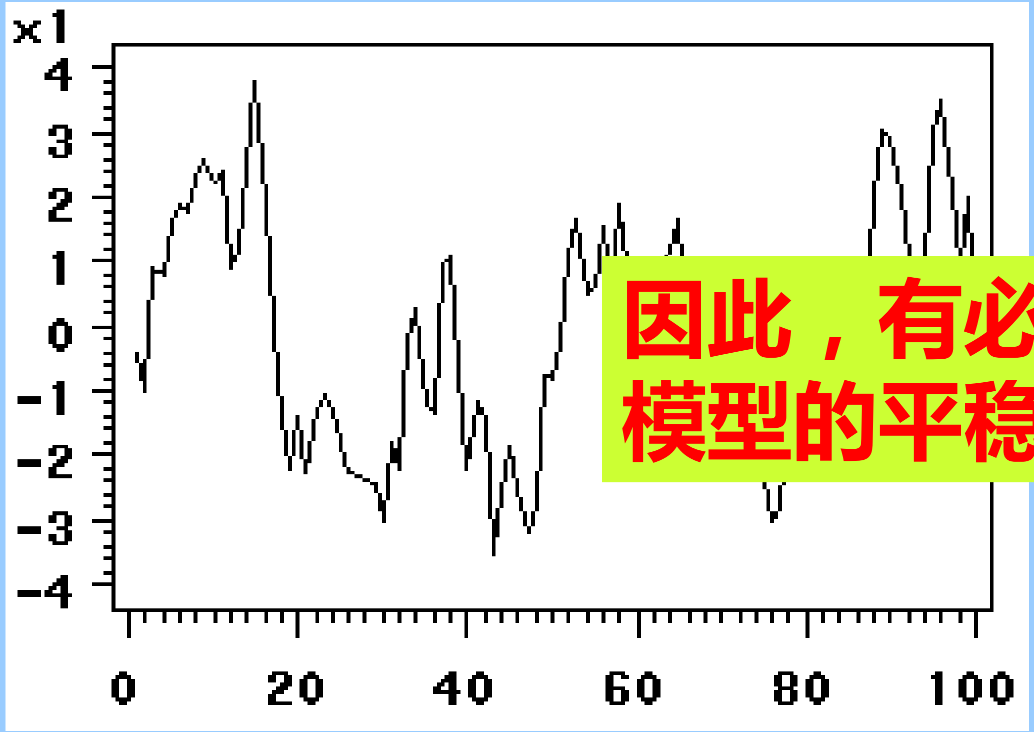
例：作图观察以下2个AR模型的平稳性 $\varepsilon_t \sim WN(0,1)$

(1) $X_t = 0.8X_{t-1} + \varepsilon_t$　　(2) $X_t = -1.1X_{t-1} + \varepsilon_t$

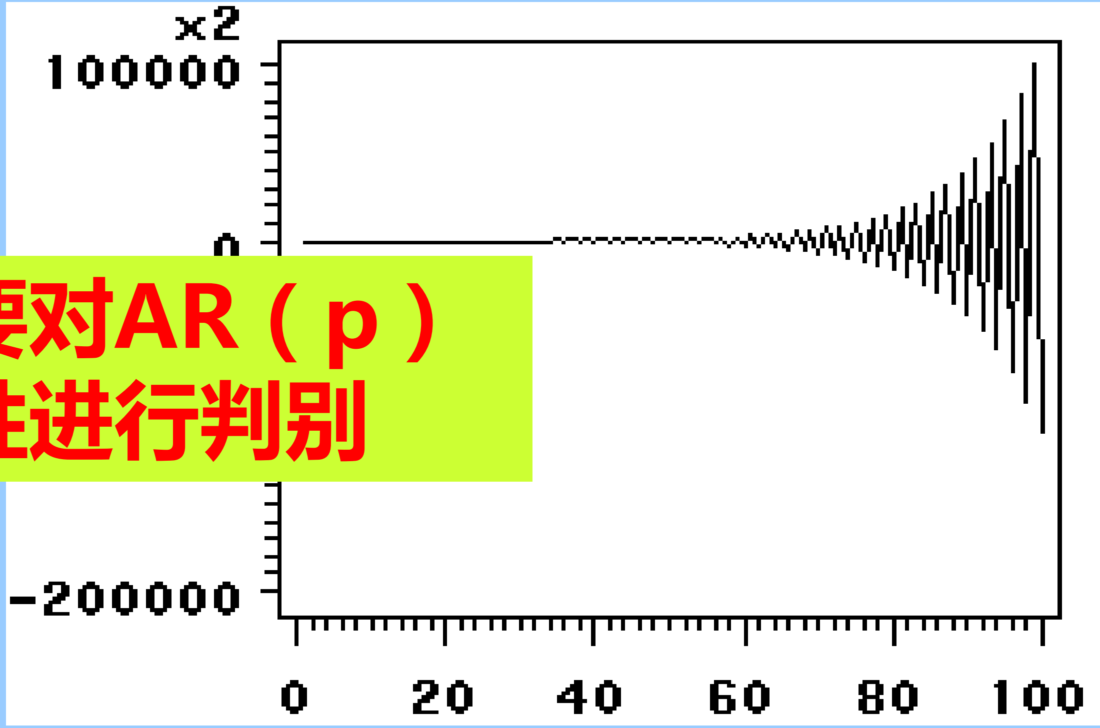
Matlab代码：

```
mdl = arima('Constant', 0, 'ARLags', [1], 'AR', {0.8}, 'Variance', 1)
x = simulate(mdl,1000);
plot(x)
```

模拟产生的AR(p)序列时序图



(1) $X_t = 0.8X_{t-1} + \varepsilon_t$



(2) $X_t = -1.1X_{t-1} + \varepsilon_t$

因此，有必要对AR（p）模型的平稳性进行判别

AR(p) 模型的平稳性判别

特征根判别法：

AR(p)模型平稳的充要条件是其特征方程的p个特征根都在单位圆内。

AR(1) 模型的平稳性判别

$$X_t = \phi X_{t-1} + \varepsilon_t$$

特征方程：$\lambda - \phi = 0$

特征根：$\lambda = \phi$

AR(1)模型平稳的充要条件是：$|\lambda| = |\phi| < 1$

AR(2) 模型的平稳性判别

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \varepsilon_t$$

特征方程

$$\lambda^2 - \phi_1 \lambda - \phi_2 = 0$$

平稳条件

$$|\lambda_1| < 1 \text{且} |\lambda_2| < 1$$

特征根

$$\lambda_{1,2} = \frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2}$$

精思国计 细量民生

例. 判别如下AR模型的平稳性

(1) $X_t = 0.8X_{t-1} + \varepsilon_t$ (2) $X_t = -1.1X_{t-1} + \varepsilon_t$

(3) $X_t = X_{t-1} - 0.5X_{t-2} + \varepsilon_t$ (4) $X_t = X_{t-1} + 0.5X_{t-2} + \varepsilon_t$

模型	特征根判别	结论
(1)	$\lambda_1 = 0.8$	平稳
(2)	$\lambda_1 = -1.1$	非平稳
(3)	$\lambda_1 = \frac{1+i}{2}$ $\lambda_2 = \frac{1-i}{2}$	平稳
(4)	$\lambda_1 = \frac{1+\sqrt{3}}{2}$ $\lambda_2 = \frac{1-\sqrt{3}}{2}$	非平稳

（3）平稳AR(p) 模型自相关系数的性质

定义

滞后k自相关系数：随机变量 X_{t-k} 与 X_t 的相关系数。

$$\rho_k=\rho_{X_t,X_{t-k}}=\frac{E[(X_t-E[X_t])(X_{t-k}-E[X_{t-k}])]}{E[(X_{t-k}-E[X_{t-k}])^2]}$$

平稳AR(p) 模型自相关系数的性质

$$\rho(k)=\sum_{i=1}^{p}c_i\lambda_i^k,\quad c_1,c_2,\cdots,c_p\text{不能恒等于零}$$

由 $|\lambda_i|<1\ (i=1,\ldots,p)$

$$\rho(k)=\sum_{i=1}^{p}c_i\lambda_i^k\to 0\quad (k\to\infty)$$

ρ_k 呈负指数衰减，且不会在有限阶后恒为零

AR(p) 模型自相关系数具有**拖尾性**

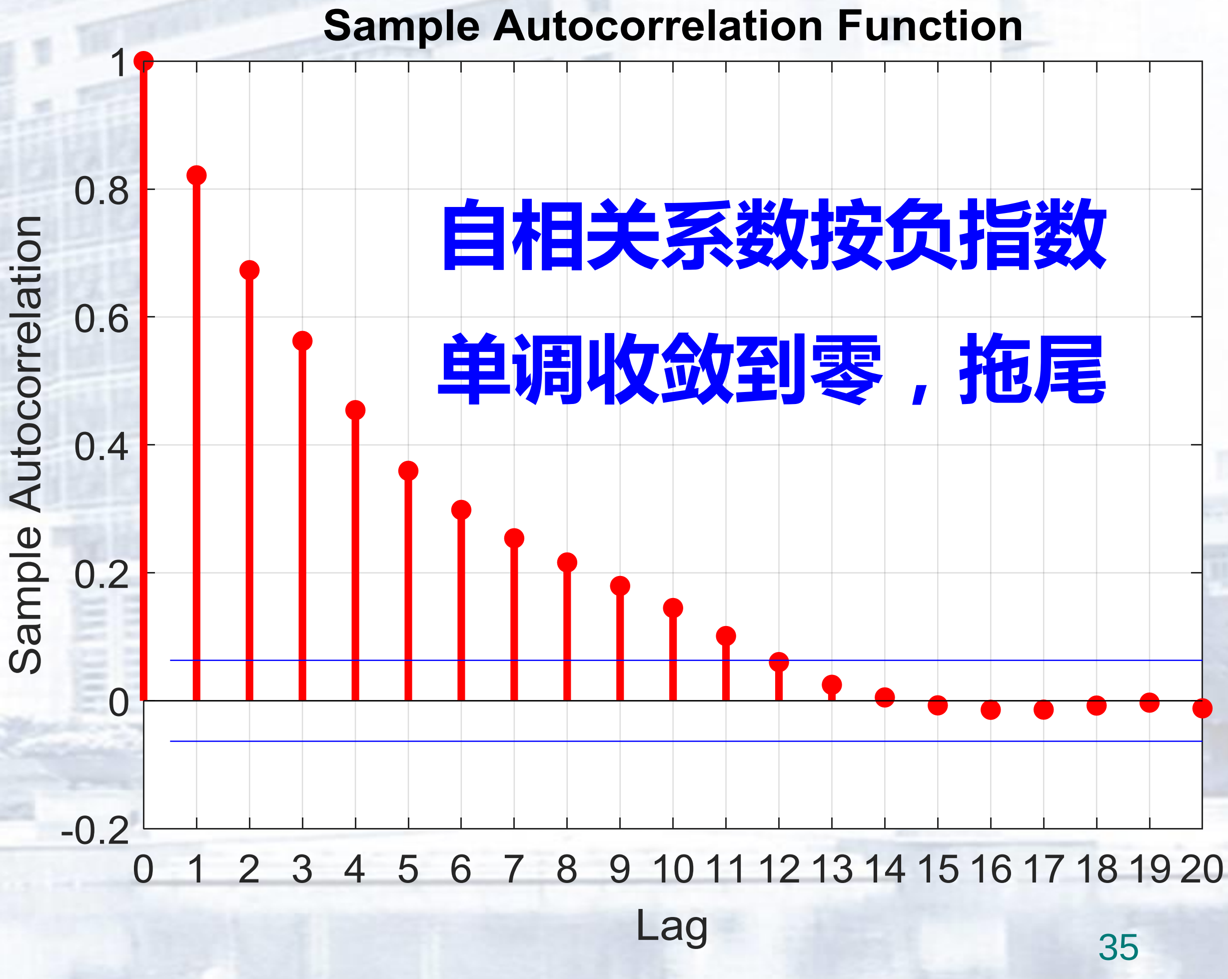
例. 考察如下 AR 模型的自相关图

$$\varepsilon_t \sim WN(0,1)$$

$$(1)\quad X_t = 0.8X_{t-1} + \varepsilon_t$$

Matlab代码：

```
mdl = arima('Constant', 0, ...
                    'ARLags', [1], ...
                    'AR', {0.8}, ...
  'Distribution', 'Gaussian', ...
  'Variance', 1)
x = simulate(mdl,1000);
autocorr(x)
```



精思國計 細量民生

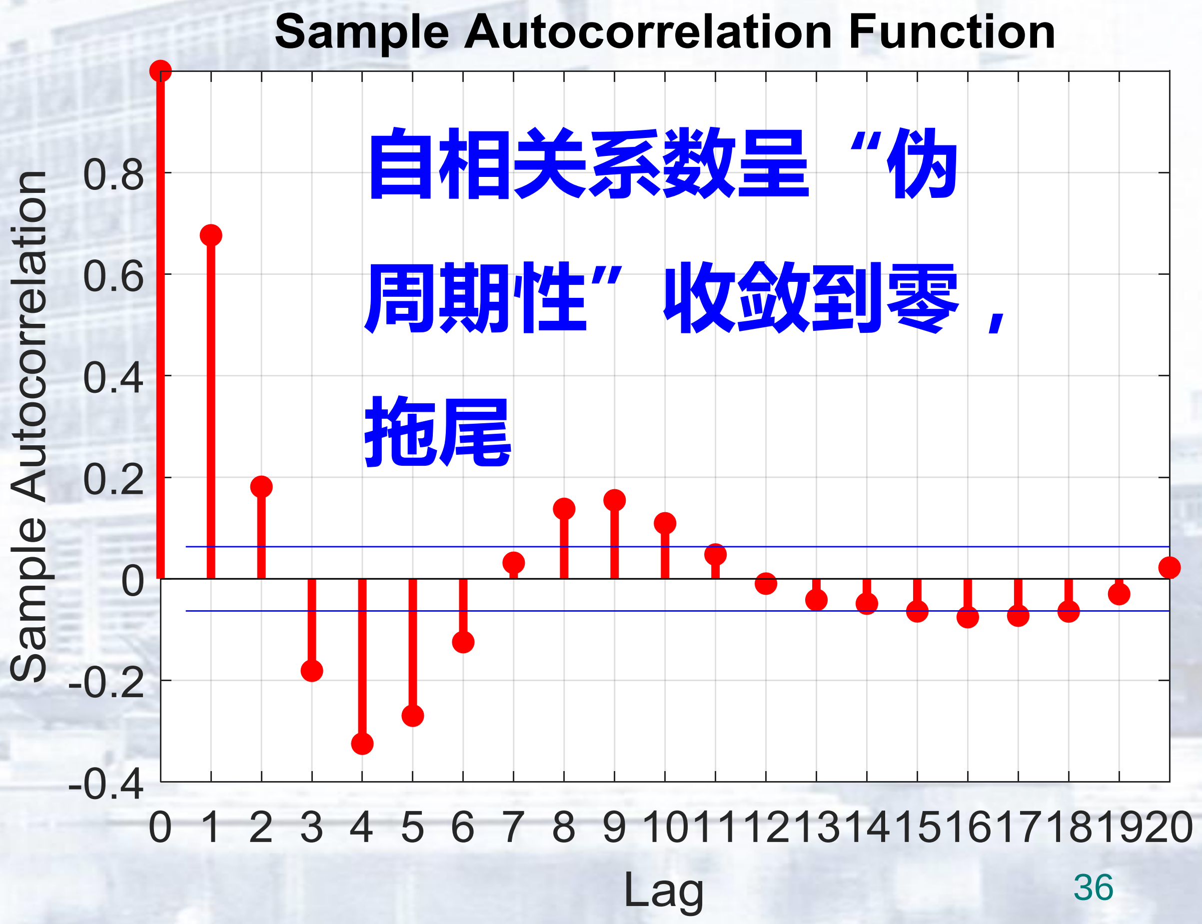
例. 考察如下 AR 模型的自相关图

$$(2)\quad X_t = X_{t-1} - 0.5X_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0,1)$$

Matlab代码：

```
mdl = arima('Constant', 0, ...
            'ARLags', [1, 2], ...
            'AR', {1, -0.5}, ...
  'Distribution', 'Gaussian', ...
  'Variance', 1)
x = simulate(mdl,1000);
autocorr(x)
```



（4）AR(p) 模型偏自相关系数的性质

定义

滞后k偏自相关系数：在给定中间 k-1 个随机变量 X_{t-1} ，…，X_{t-k+1} 的条件下，X_{t-k} 与 X_t 的相关系数。

$$\phi_{kk} = \rho_{X_t, X_{t-k} | X_{t-1}, \cdots, X_{t-k+1}}$$

滞后k偏自相关系数可以理解为：在剔除了中间 k-1 个随机变量的干扰之后，X_{t-k} 对 X_t 影响的相关系数。

AR(p) 模型偏自相关系数的性质

计算

性质： 滞后 k 偏自相关系数就等于如下 k 阶自回归模型第 k 个回归系数的值，

$$\begin{cases} \rho_1 = \phi_{k1}\rho_0 + \phi_{k2}\rho_1 + \cdots + \phi_{kk}\rho_{k-1} \\ \rho_2 = \phi_{k1}\rho_1 + \phi_{k2}\rho_0 + \cdots + \phi_{kk}\rho_{k-2} \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ \rho_k = \phi_{k1}\rho_{k-1} + \phi_{k2}\rho_{k-2} + \cdots + \phi_{kk}\rho_0 \end{cases} \Rightarrow \phi_{kk} = \frac{D_k}{D}$$

精思国计 细量民生

AR(p) 模型偏自相关系数的性质

- ***p*阶截尾性**

平稳AR(*p*)模型的偏自相关系数具有*p*阶截尾性

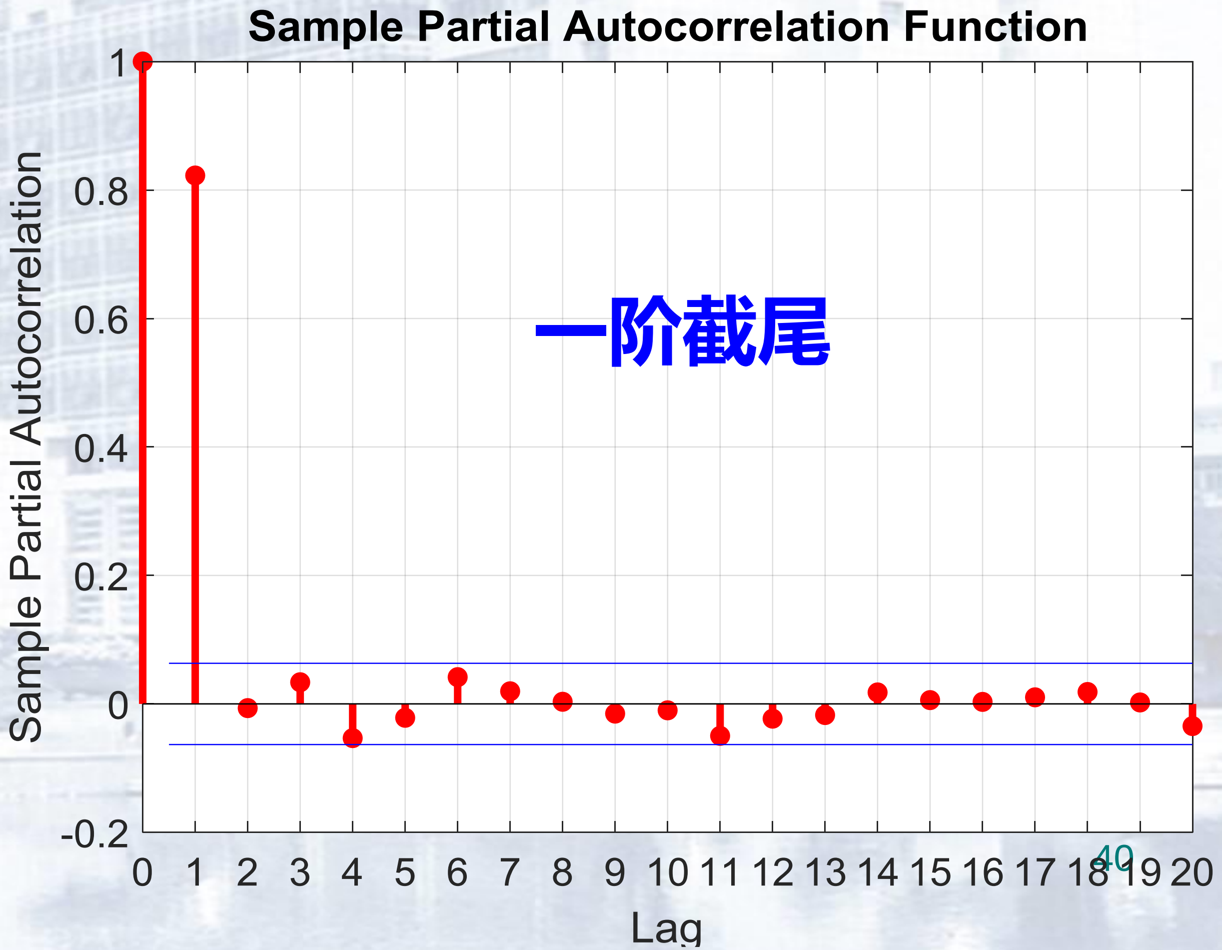
$$\phi_{kk}=0, \quad k>p$$

例. 考察如下 AR 模型的偏自相关图

$$(1)\quad X_t = 0.8X_{t-1} + \varepsilon_t,\quad \varepsilon_t \sim WN(0,1)$$

Matlab代码：

```
mdl = arima('Constant', 0, ...
            'ARLags', [1], ...
            'AR', {0.8}, ...
  'Distribution', 'Gaussian', ...
  'Variance', 1)
x = simulate(mdl,1000);
parcorr(x)
```



精思国计 细量民生

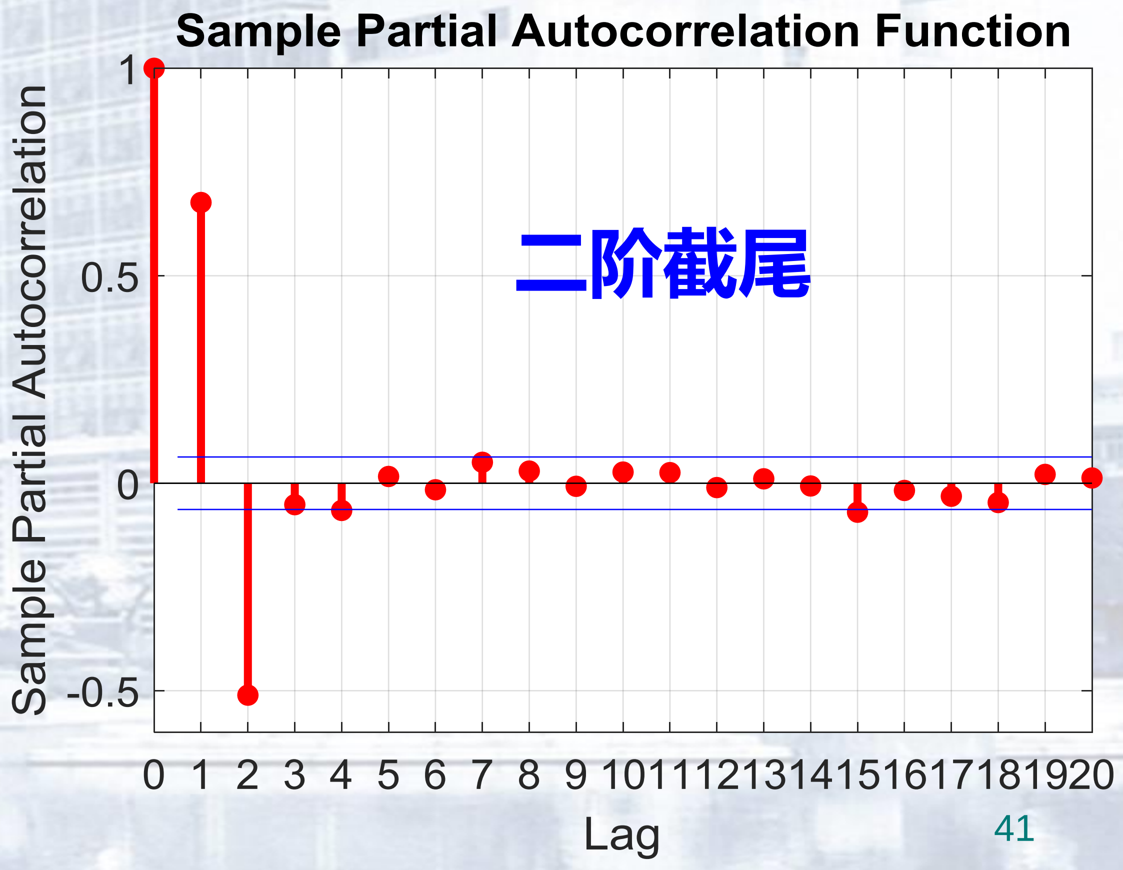
例. 考察如下 AR 模型的偏自相关图

$$(2)\quad X_t = X_{t-1} - 0.5X_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0,1)$$

Matlab代码：

```
mdl = arima('Constant', 0, ...
            'ARLags', [1, 2], ...
              'AR', {1, -0.5}, ...
  'Distribution', 'Gaussian', ...
  'Variance', 1)
x = simulate(mdl,1000);
parcorr(x)
```



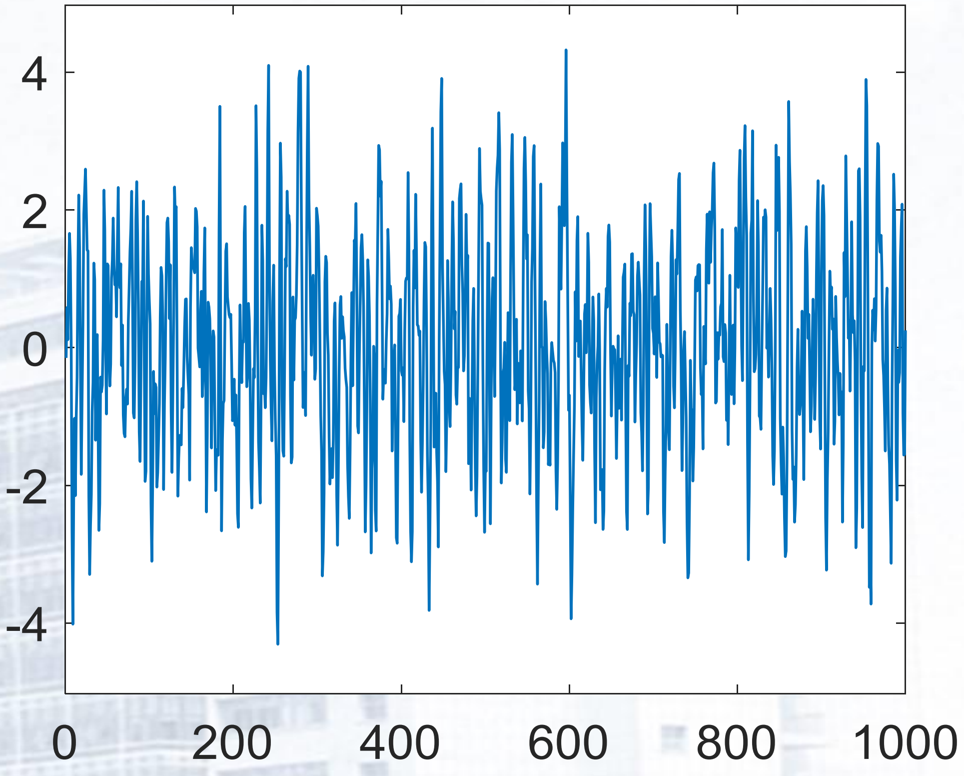
（5）AR(p) 模型的判据

- 自相关系数的拖尾
- 偏自相关系数阶p阶截尾

例.（AR模型实例）基于给定的平稳时间序列数据建立AR模型

Matlab代码：

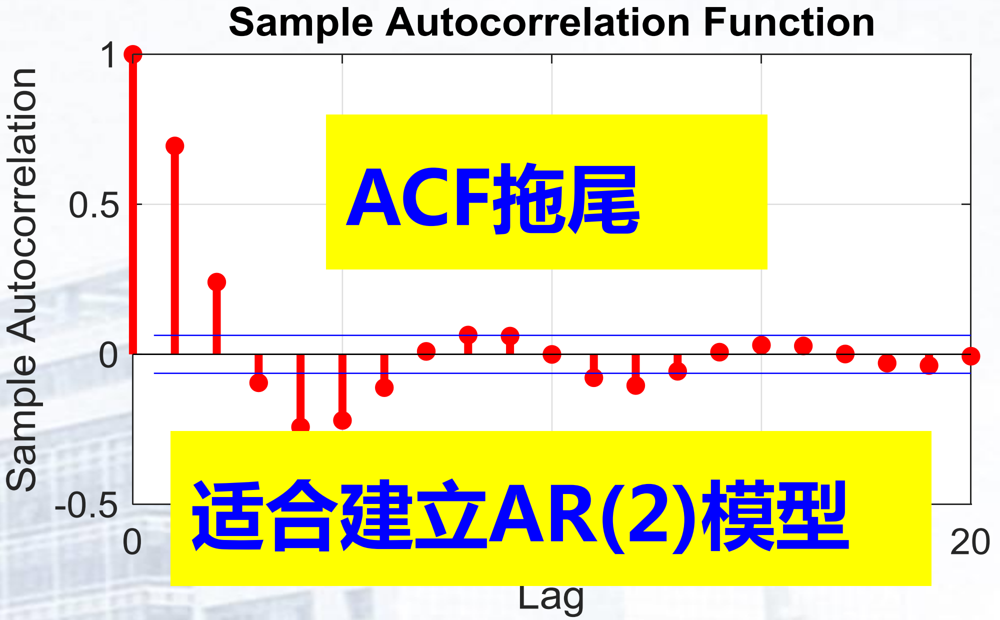
```
load data1.mat x;  %x为某平稳序列
figure(1), autocorr(x)
figure(2), parcorr(x)
```

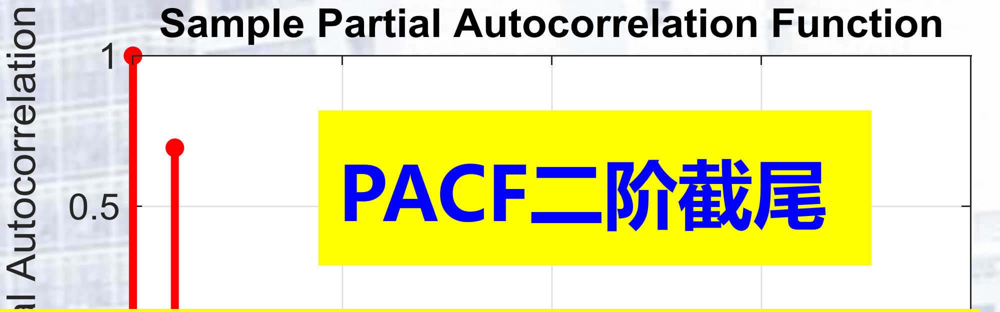


例.（AR模型实例）基于给定的平稳时间序列数据建立AR模型

Matlab代码：

```
load data1.mat x;  %x为某平稳序列
figure(1), autocorr(x)
figure(2), parcorr(x)
mdl = arima(2, 0, 0)    %创建AR(2)模型
EstMdl = estimate(mdl,x);
```





ARIMA(2,0,0) Model:

Constant	0.024
AR{1}	1.02
AR{2}	-0.47
Variance	0.93

$$X_t = 0.024 + 1.02X_{t-1} - 0.47X_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, 0.93)$$

精思國計 細量民生

Thanks

6.2.2 MA模型

- 时间序列{Xt }的记忆是关于过去外部干扰的记忆
- 时间序列{Xt }可以表示成现在干扰值和过去直到t-q期干扰值的线性组合

q阶滑动平均模型 —— MA(q)

精思国计 细量民生

（1）MA(q)模型的定义 — q阶移动平均模型

MA(q) 模型的一般形式为

$$\begin{cases} X_t = \mu + \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q} \\ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \quad Var(\varepsilon_t) = \sigma^2, E(\varepsilon_t\varepsilon_s) = 0, s \neq t \end{cases}$$

特别当 $\mu = 0$ 时，称为**中心化 MA(q) 模型.**

$$X_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$$

MA(q) — 移动平均系数多项式

引进延迟算子B, 中心化 MA(q) 模型可简记为

$$X_t = \Theta(B)\varepsilon_t$$

其中，滑动平均系数多项式

$$\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_q B^q$$

（2）MA(q)模型自相关系数的性质

$$\rho_k = \begin{cases} 1, & k = 0 \\ \dfrac{-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{k+i}}{1+\theta_1^2+\cdots+\theta_q^2}, & 1 \le k \le q \\ 0, & k > q \end{cases}$$

- **自相关系数q阶截尾**

（3）MA(q) 模型偏自相关系数的性质

$$\phi_{kk}=\frac{E[X_t X_{t-k}\mid X_{t-1},\ldots,X_{t-k+1}]}{Var[X_{t-k}\mid X_{t-1},\ldots,X_{t-k+1}]}$$

$$=(-\theta_1\varepsilon_{t-1}-\cdots-\theta_q\varepsilon_{t-q})(-\theta_1\varepsilon_{t-k-1}-\cdots-\theta_q\varepsilon_{t-k-q+1})\sigma^{-2}$$

$\theta_1,\cdots\theta_q$不全为零 → ϕ_{kk} 不会在有限阶后恒为零

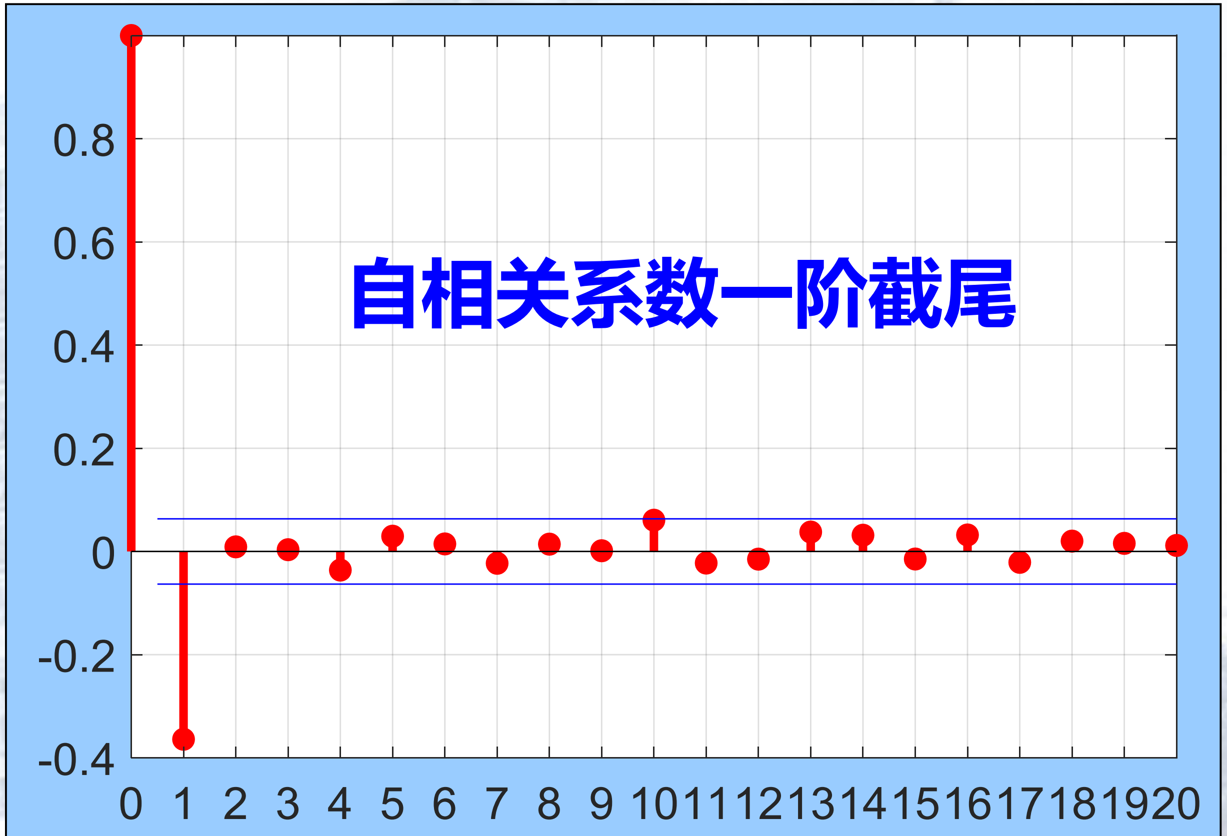
→ MA(q) 模型偏自相关系数具有**拖尾性**

例. MA(1)模型自相关系数一阶截尾

Matlab代码：

```
mdl = ...
arima('Constant', 0, ...
         'MALags', [1], ...
         'MA', {-0.4}, ...
         'Variance', 1)
x = simulate(mdl, 1000);
autocorr(x)
```

样本自相关图



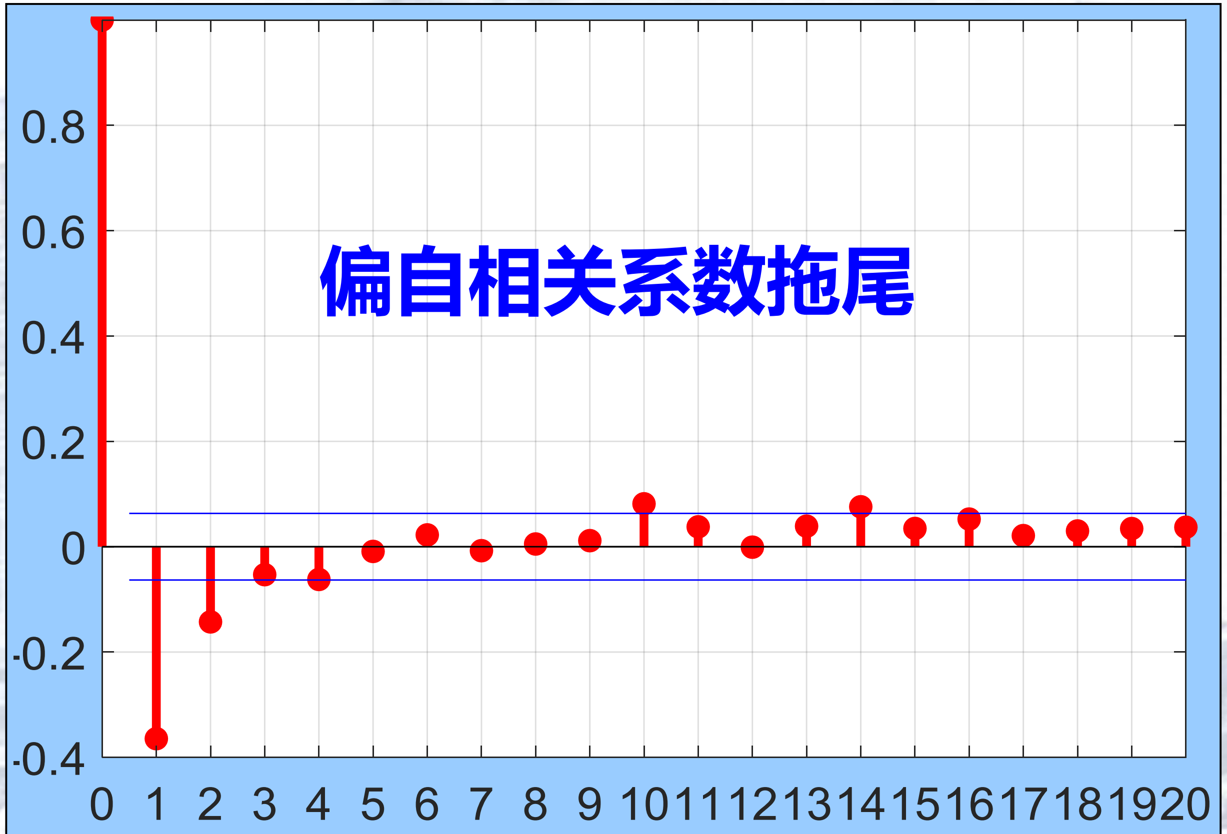
$$X_t = \varepsilon_t - 0.4\varepsilon_{t-1}$$

例. MA(1)模型偏自相关系数拖尾

Matlab代码：

```
mdl = ...
arima('Constant', 0, ...
         'MALags', [1], ...
         'MA', {-0.4}, ...
         'Variance', 1)
x = simulate(mdl,1000);
parcorr(x)
```

样本偏自相关图



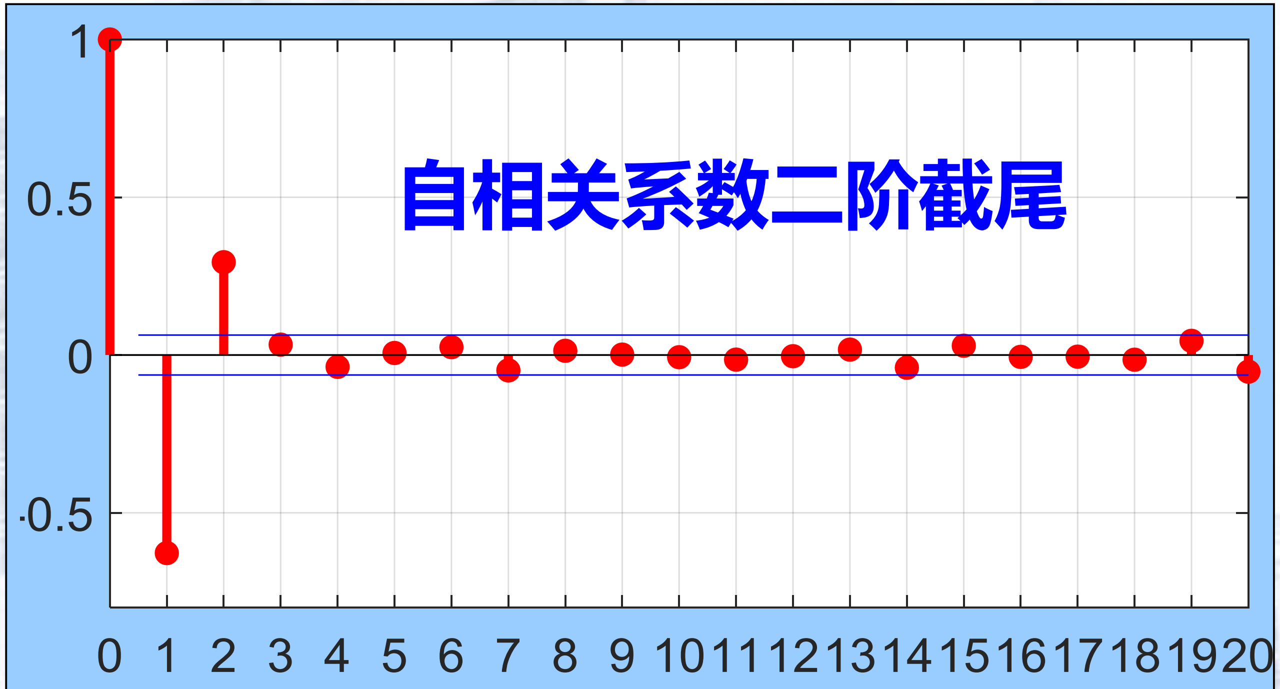
$$X_t = \varepsilon_t - 0.4\varepsilon_{t-1}$$

例. MA(2)模型自相关系数二阶截尾

Matlab代码：

```
mdl = ...
arima('Constant', 0, ...
        'MALags', [1,2], ...
        'MA', {-4/5, 16/25}, ...
        'Variance', 1)
x = simulate(mdl,1000);
autocorr(x)
```

样本自相关图



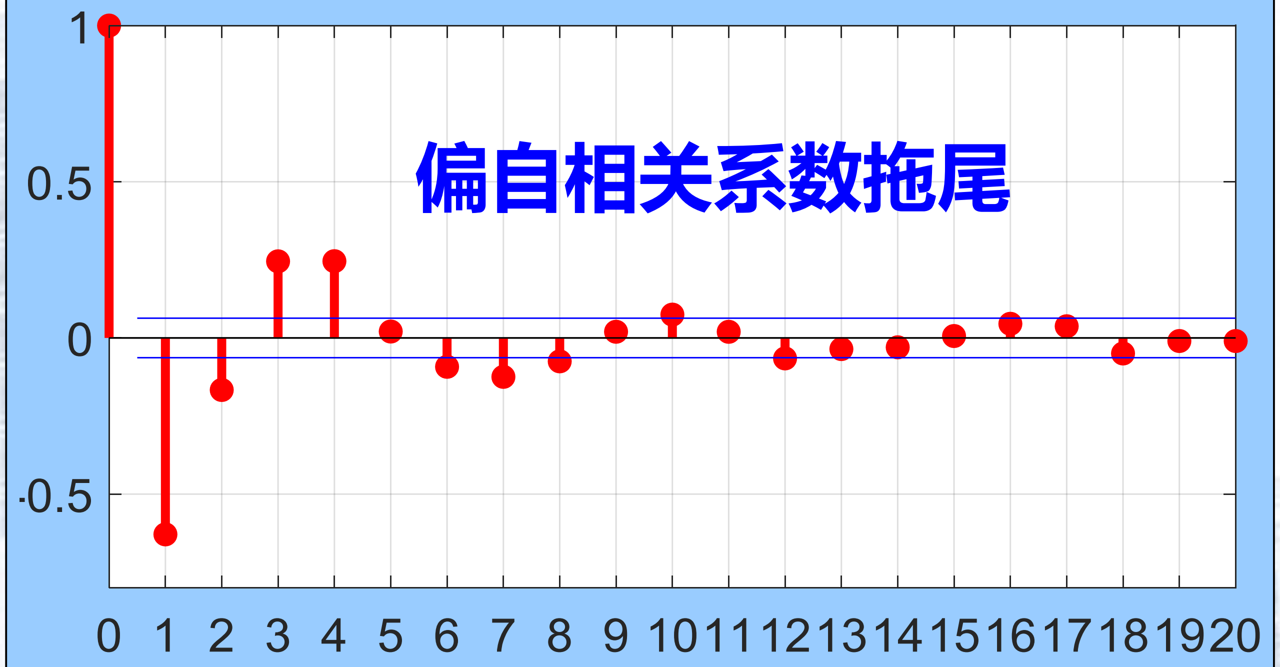
$$X_t = \varepsilon_t - 4/5\varepsilon_{t-1} + 16/25\varepsilon_{t-2}$$

例. MA(2)模型偏自相关系数拖尾

Matlab代码：

```
mdl = ...
arima('Constant', 0, ...
          'MALags', [1,2], ...
          'MA', {-4/5, 16/25}, ...
          'Variance', 1)
x = simulate(mdl,1000);
parcorr(x)
```

样本偏自相关图



$$X_t = \varepsilon_t - 4/5\varepsilon_{t-1} + 16/25\varepsilon_{t-2}$$

（4）MA(q)模型的判据

- 自相关系数q阶截尾
- 偏自相关系数拖尾

但是：

通过这些特征选择的MA（q）模型可能对应多种(偏)自相关系数特征，因此需要保证两者的一一对应。

这就需要保证MA模型的可逆性

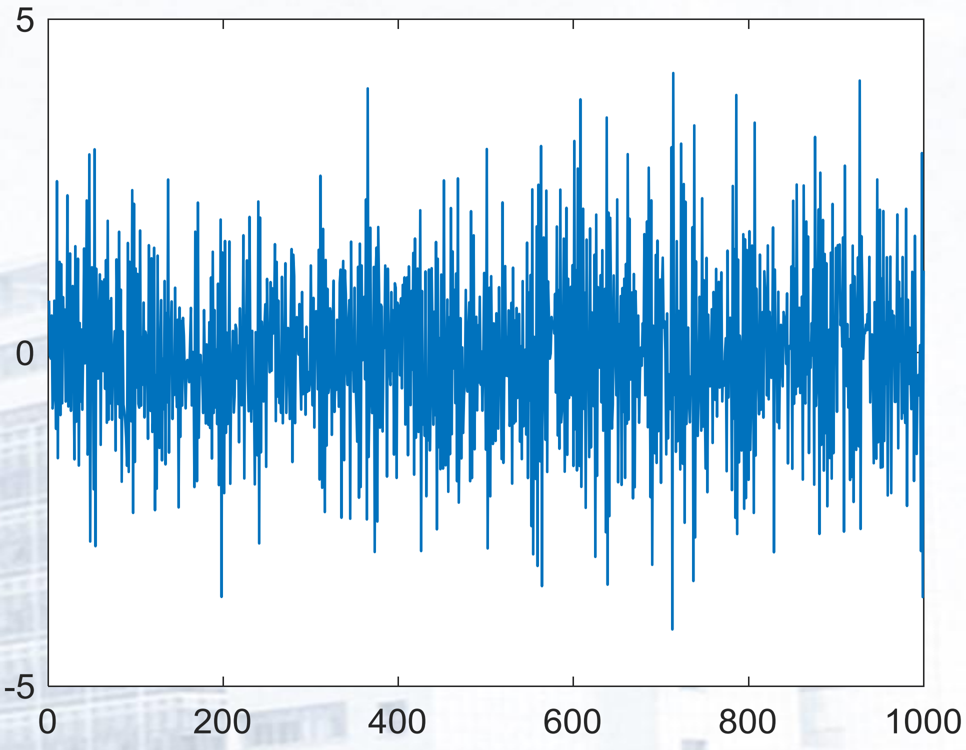
MA(q) 模型的可逆性判别

特征根判别法：

MA(q)可逆等价于其特征方程的特征根都在单位圆内

$$|\lambda_i| < 1 \quad i = 1, \cdots, q$$

例. （MA模型实例）基于给定的平稳时间序列数据建立MA模型



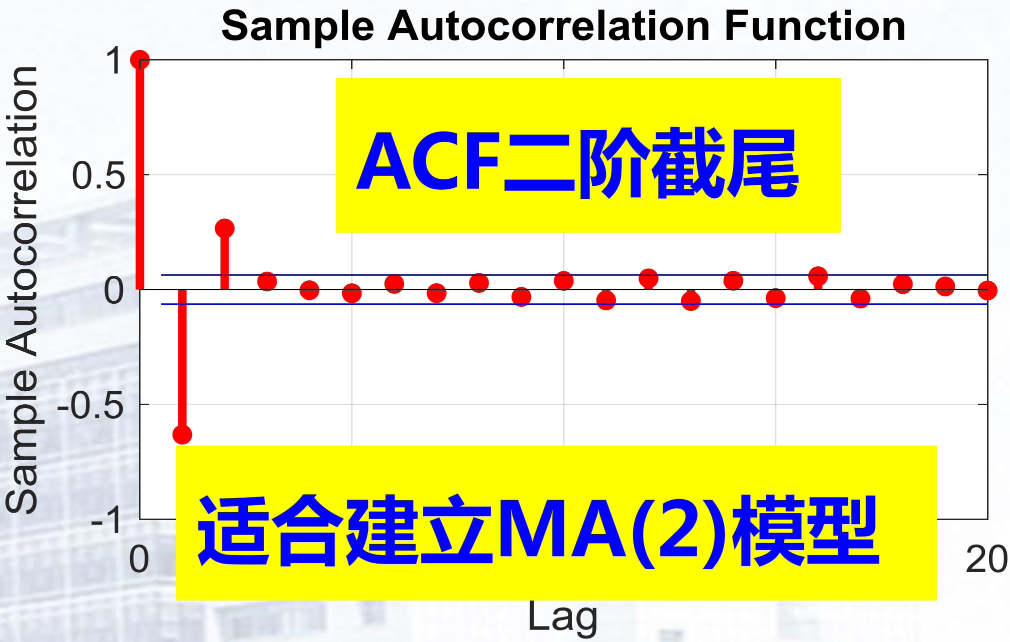
Matlab代码：

```
load data2.mat x;  %x为某平稳序列
figure(1), autocorr(x)
figure(2), parcorr(x)
```

例.（MA模型实例）基于给定的平稳时间序列数据建立MA模型

Matlab代码：

```
load data2.mat x;  %x为某平稳序列
figure(1), autocorr(x)
figure(2), parcorr(x)
mdl = arima(0, 0, 2)    %创建MA(2)模型
EstMdl = estimate(mdl, x);
```



ARIMA(0,0,2) Model:

Constant	-0.024
MA{1}	-0.83
MA{2}	0.64
Variance	0.89

$$X_t = -0.024 - 0.83\varepsilon_{t-1} + 0.64\varepsilon_{t-2} + \varepsilon_t$$

$$\varepsilon_t \sim WN(0, 0.89)$$

精思國計 細量民生

Thanks

6.2.3 ARMA模型

- 时间序列{Xt }的当前值不仅与自身的过去值有关，而且还与其以前的外部干扰有关
- 时间序列{Xt }模型中既包括自身的滞后项，也包括过去的外部干扰

自回归滑动平均模型 —— ARMA(p,q)

ARMA(p,q) — 自回归移动平均模型

ARMA(p,q) 模型的一般形式为

$$\begin{cases} X_t = \phi_0 + \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \phi_p \neq 0, \quad \theta_q \neq 0 \\ \varepsilon_t \sim WN(0, \sigma^2) \\ \forall s < t, \quad E[X_s \varepsilon_t] = 0 \end{cases}$$

特别当 $\phi_0 = 0$ 时，称为**中心化 ARMA(p,q) 模型.**

$$X_t = \phi_1 X_{t-1} + \cdots + \phi_p X_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q}$$

ARMA(p,q) —系数多项式

引进延迟算子B, 中心化 ARMA(p,q) 模型可简记为

$$\Phi(B)X_t=\Theta(B)\varepsilon_t$$

其中，p阶自回归系数多项式

$$\Phi(B)=1-\varphi_1 B-\varphi_2 B^2-\cdots-\varphi_p B^p$$

其中，q阶滑动平均系数多项式

$$\Theta(B)=1-\theta_1 B-\theta_2 B^2-\cdots-\theta_q B^q$$

ARMA(p,q) 模型的平稳性和可逆性

- ARMA(p,q) 模型的平稳性
  - ARMA(p,q)模型的平稳性完全由其自回归部分AR(p)的平稳性决定
  - 对应AR(p)模型的特征根都在单位圆内
- ARMA(p,q) 模型的可逆性
  - ARMA(p,q)模型的可逆性完全由其滑动平均部分MA(q)的可逆性决定
  - MA(q)模型的特征根都在单位圆内

ARMA(p,q) 模型（偏）自相关系数的性质

- 自相关系数拖尾（k=q阶后指数衰减）
- 偏自相关系数拖尾（k=p阶后指数衰减）

例. 考察如下ARMA(1,1)模型：

$$X_t = 0.5X_{t-1} + \varepsilon_t - 0.8\varepsilon_{t-1} \quad \varepsilon_t \sim WN(0,1)$$

自相关系数和偏自相关系数的性质。

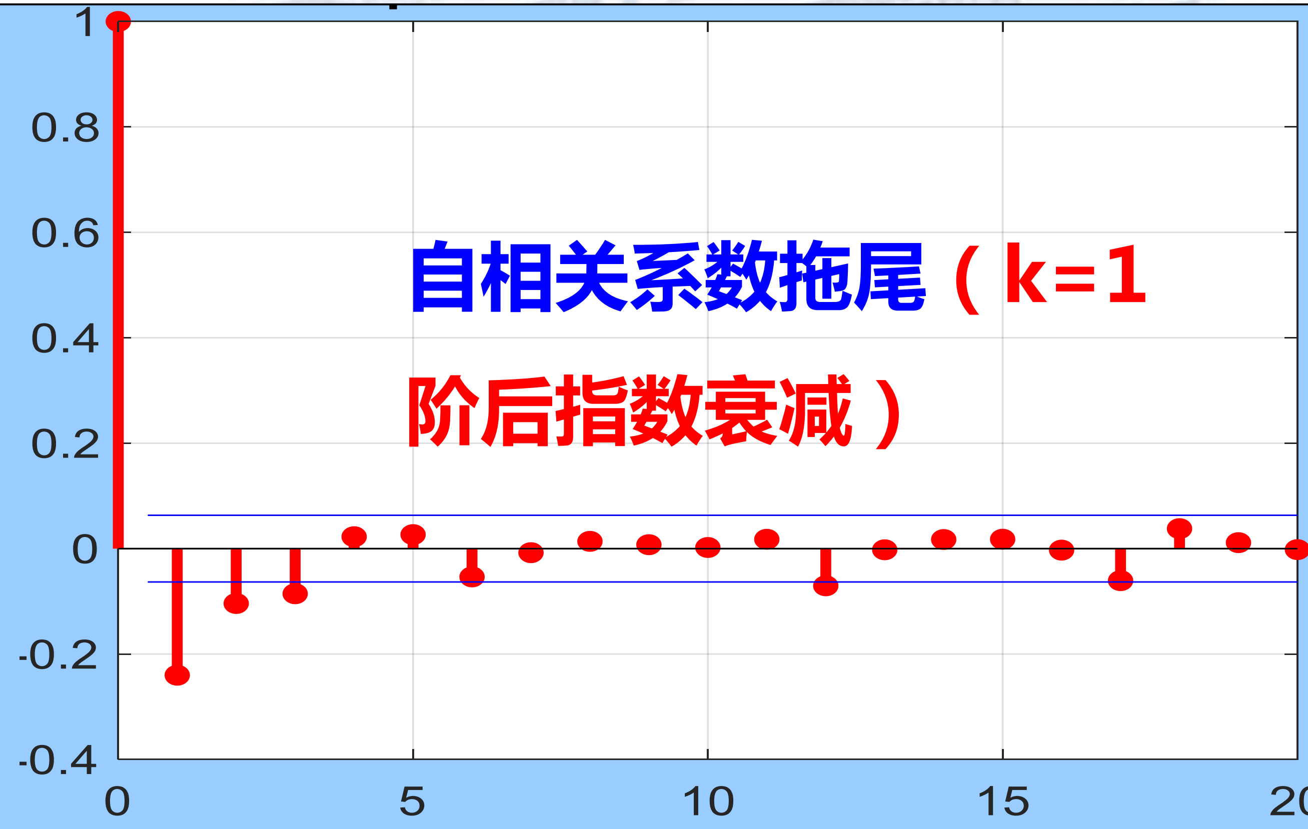
Matlab代码：

```
mdl = arima('Constant', 0, 'AR', {0.5}, 'MA', {-0.8}, 'Variance', 1)
x = simulate(mdl,1000);
figure(1), autocorr(x)
figure(2), parcorr(x)
```

例. ARMA(1,1) 模型

自相关系数和偏自相关系数拖尾性

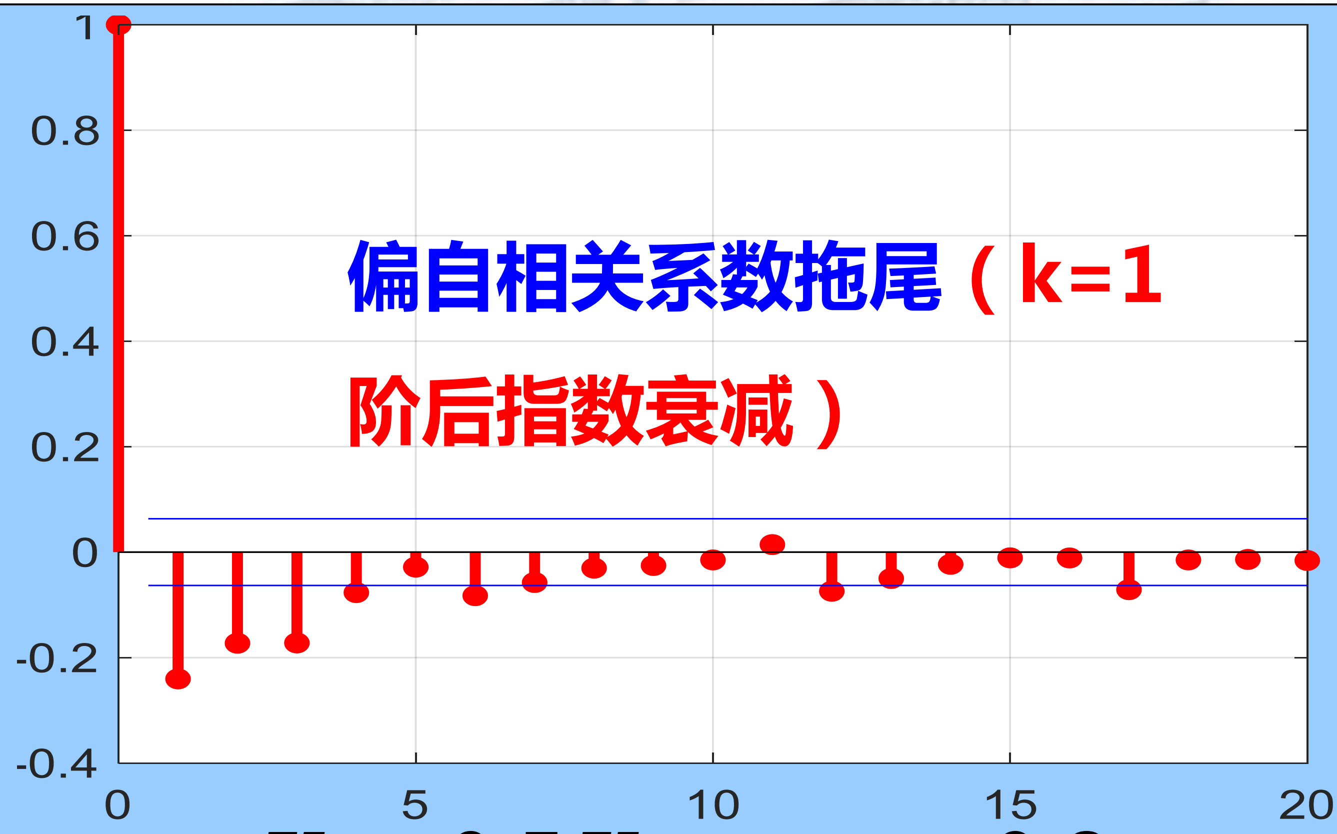
样本自相关图



$$X_t - 0.5X_{t-1} = \varepsilon_t - 0.8\varepsilon_{t-1}$$

例. ARMA(1,1) 模型
自相关系数和偏自相关系数拖尾性

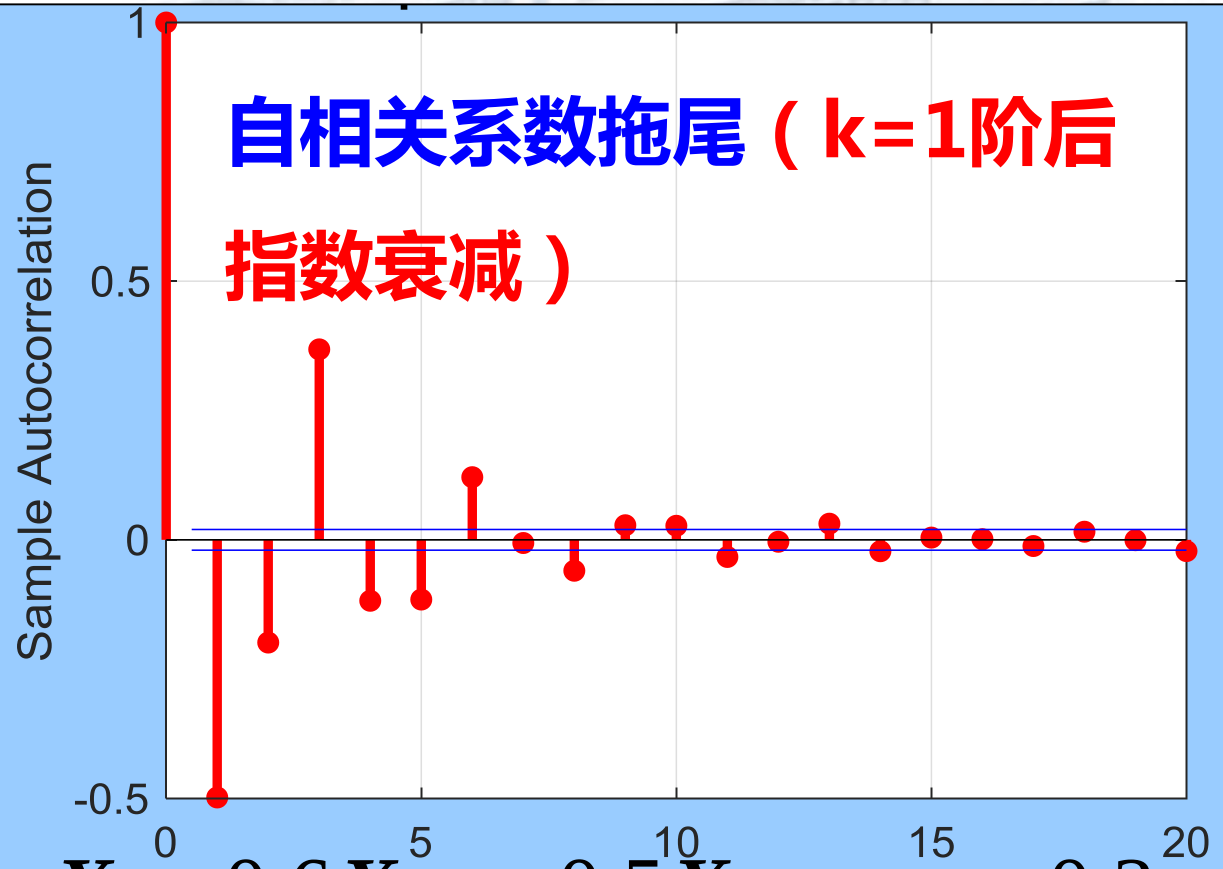
样本偏自相关图



$$X_t - 0.5X_{t-1} = \varepsilon_t - 0.8\varepsilon_{t-1}$$

例. ARMA(2,1) 模型
自相关系数和偏自相关系数拖尾性

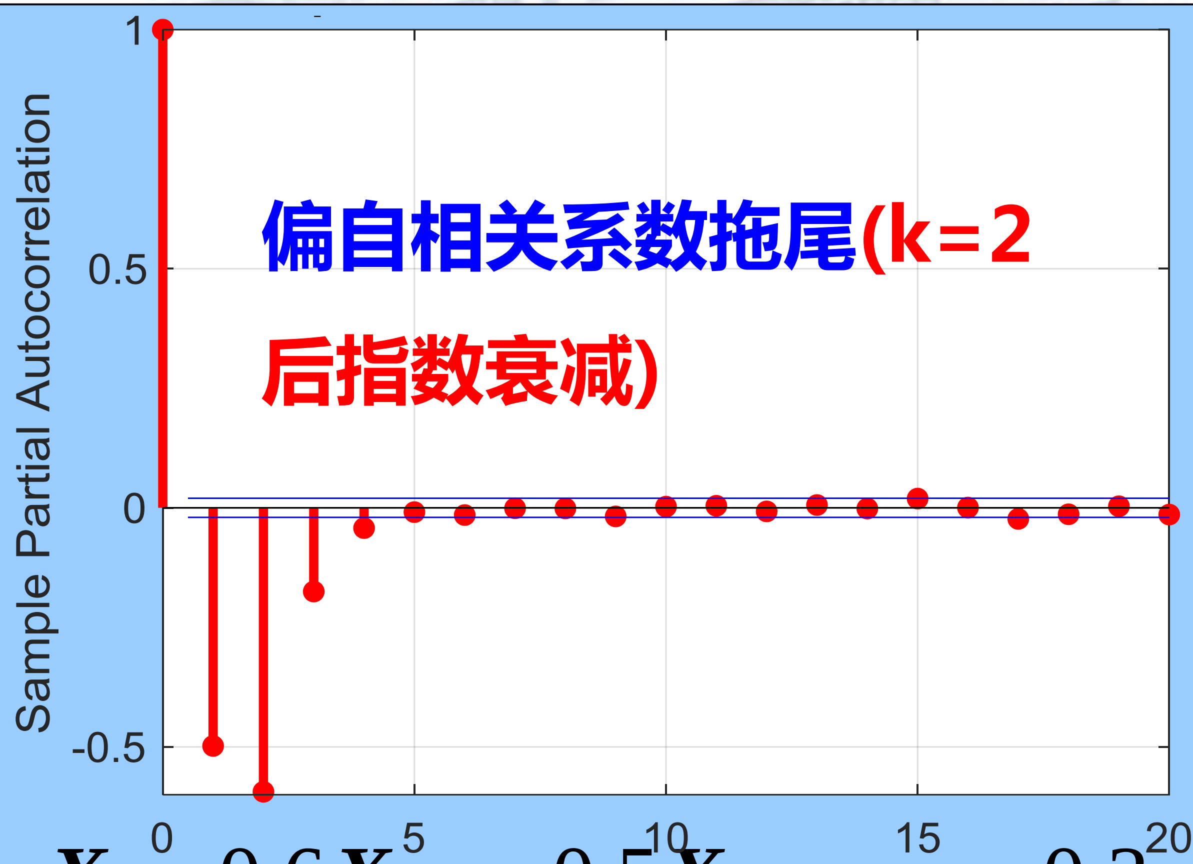
样本自相关图



$$X_t - 0.6X_{t-1} - 0.5X_{t-1} = \varepsilon_t - 0.3\varepsilon_{t-1}$$

例. ARMA(2,1) 模型
自相关系数和偏自相关系数拖尾性

样本偏自相关图



$$X_t - 0.6X_{t-1} - 0.5X_{t-1} = \varepsilon_t - 0.3\varepsilon_{t-1}$$

ARMA(p,q) 模型的判据

模　型	自相关系数	偏自相关系数
ARMA(p,q)	拖尾 (k=q后指数衰减)	拖尾 (k=p后指数衰减)

ARMA(p,q) 模型的识别判据

模　型	自相关系数	偏自相关系数
AR(p)	拖尾	p阶截尾
MA(q)	q阶截尾	拖尾
ARMA(p,q)	拖尾	拖尾

在实际建模过程中，该识别方法有如下几个缺点：

(1) 由于样本随机性，(P)ACF是截尾还是拖尾有时难以准确判断。

(2) ARMA(p,q)模型阶数p和q有时难以准确判断，非常依赖经验，主观性强。

精思國計 細量民生

Thanks

6.3 ARMA模型的识别（定阶）

6.3.1 自相关和偏自相关系数法

6.3.2 信息准则函数法

6.3.1 自相关和偏自相关系数定阶法

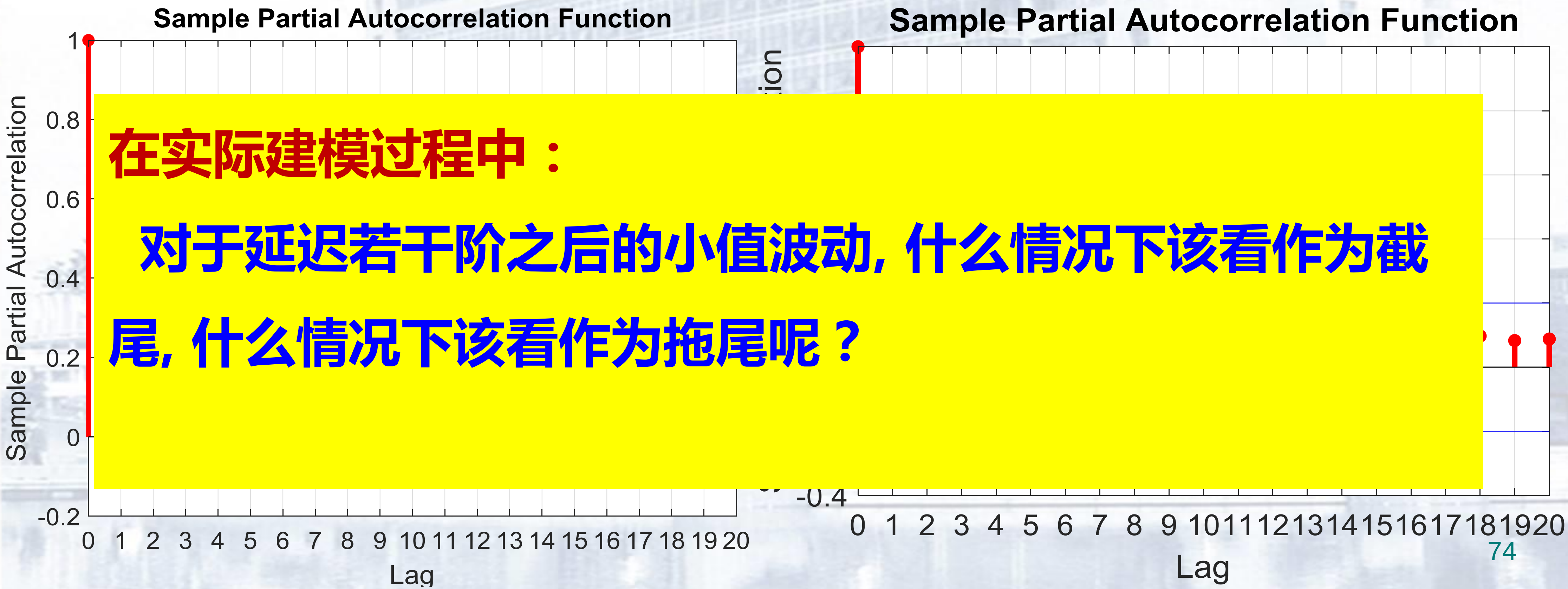
- 通过观察样本的自相关系数（ACF）和偏自相关系数（PACF），进行大体地判断

模　型	自相关系数	偏自相关系数
AR(p)	拖尾	p阶截尾
MA(q)	q阶截尾	拖尾
ARMA(p,q)	拖尾	拖尾

模型定阶的困难性

$$X_t = 0.8X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim WN(0,1)$$

- 由于样本的随机性，样本的相关系数不会呈现出完美的理论截尾情况，本应截尾的 $\hat{\rho}_k$ 或 $\hat{\varphi}_{kk}$ 仍会呈现出小值振荡的情况



在实际建模过程中：

对于延迟若干阶之后的小值波动，什么情况下该看作为截尾，什么情况下该看作为拖尾呢？

样本相关系数的近似分布

- 由两个相关系数的分布定理可知

$$\hat{\rho}_k \sim N(0,\frac{1}{n}) \text{ ， } n \to \infty \qquad \hat{\varphi}_{kk} \sim N(0,\frac{1}{n}) \text{ ， } n \to \infty$$

由正态分布的性质

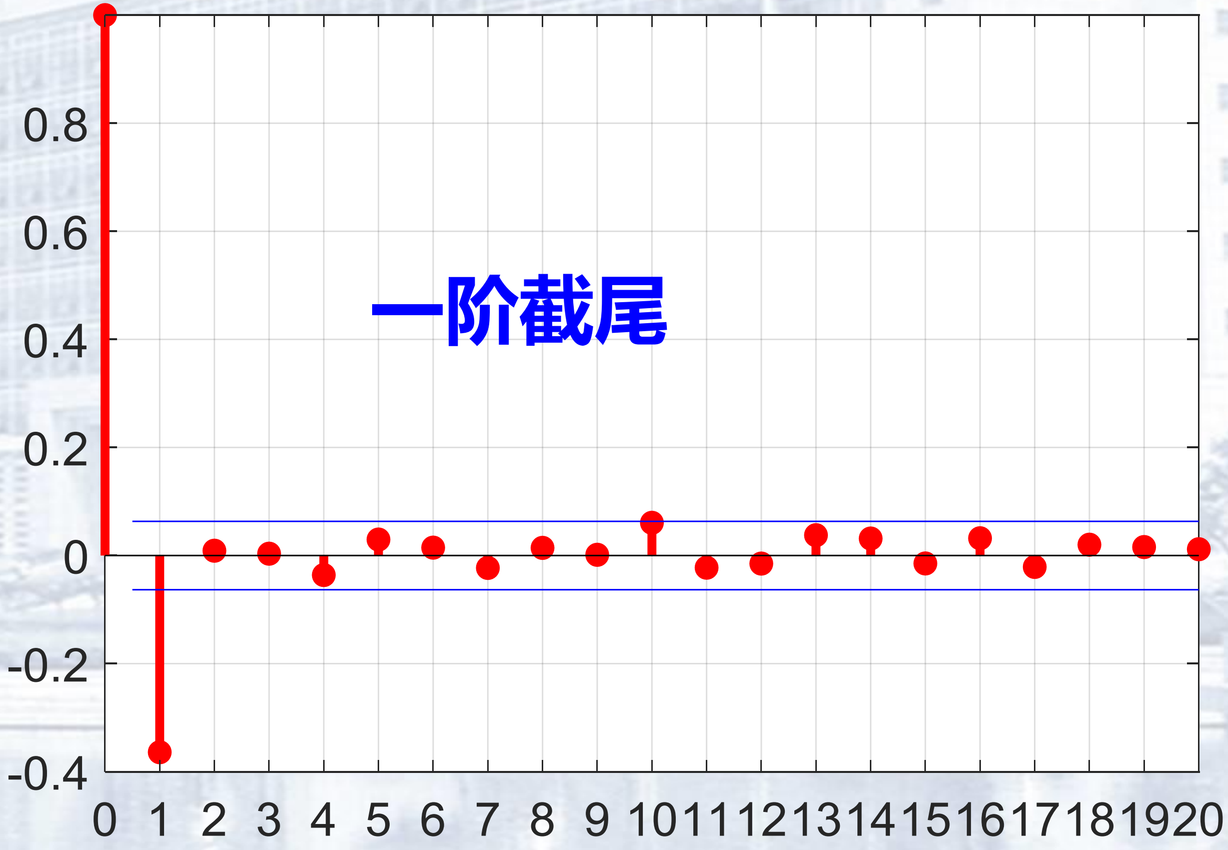
$$\Pr\left(|\hat{\rho}_k| \le \frac{2}{\sqrt{n}}\right) \ge 0.95 \qquad \Pr\left(|\hat{\varphi}_{kk}| \le \frac{2}{\sqrt{n}}\right) \ge 0.95$$

因此，在实际建模过程中：

如果至少95%的（偏）自相关系数落在2个标准差之内，则可近似认为（偏）自相关系数是收敛到0的状态。

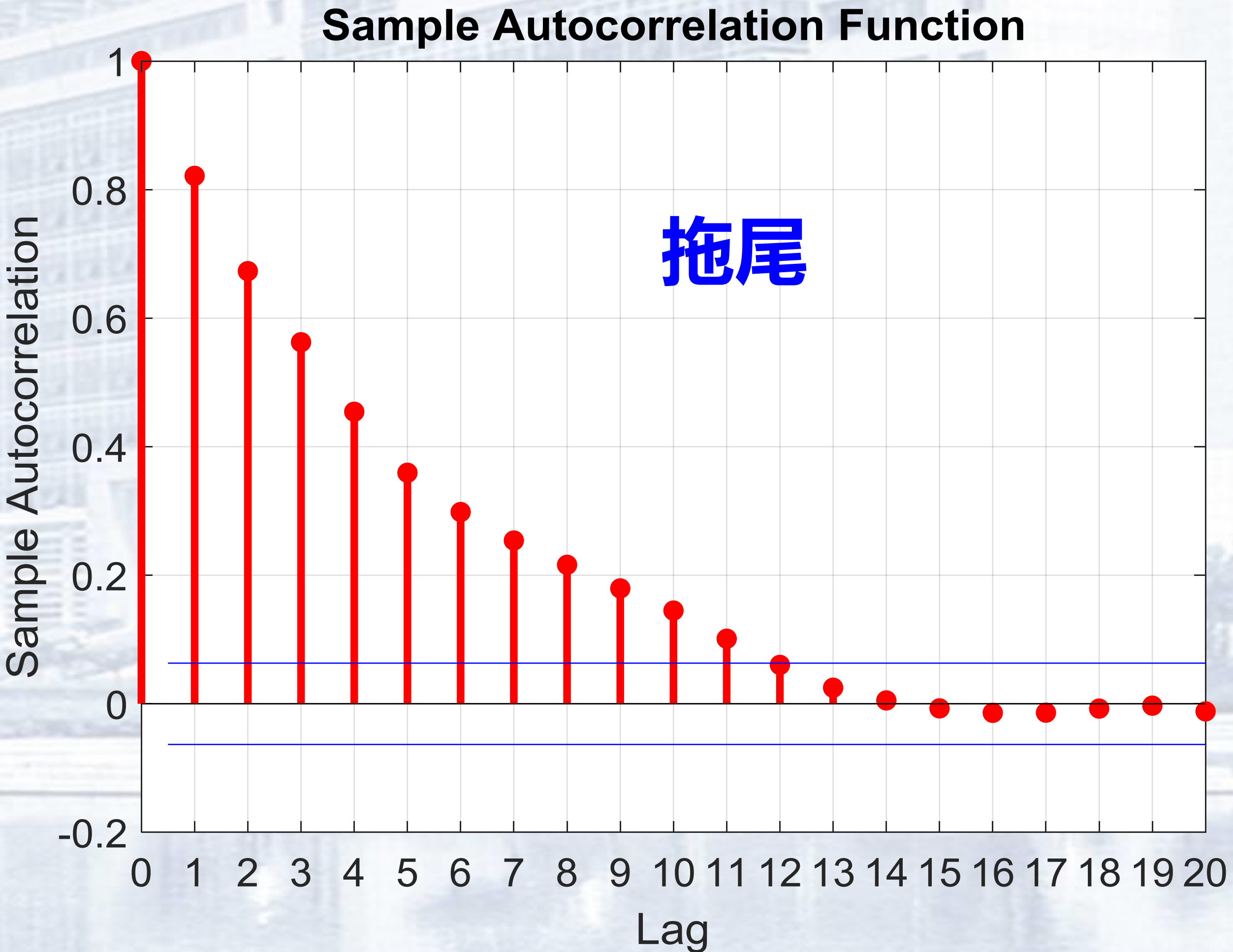
模型定阶的经验方法

- **截尾**：如果最初的 d 阶样本(偏)自相关系数明显在2倍标准差范围外，而后几乎95%的(偏)自相关系数都落在2倍标准差的范围以内，而且由非零自相关系数衰减为小值波动的过程非常突然. **这时，通常视为d阶截尾.**



模型定阶的经验方法

- **拖尾**：如果有超过5%的相关系数落在2倍标准差范围之外，或者是由显著非零的相关系数衰减为小值波动的过程比较缓慢或者非常连续，**通常视为相关系数拖尾.**



6.3.2 信息准则函数定阶法

- 基于相关系数图形的定阶法具有很强的主观性，是一种较为粗略的方法

- 信息准则函数定阶法则可以帮助我们在一些模型中选择相对最优的模型

6.3.2 准则函数定阶法

- FPE准则法
- AIC准则法
- BIC准则法

（1）FPE准则法

- 基本思想:

用模型一步预测误差的方差来判定模型的好坏，方差越小，认为模型拟合得越好（适用于AR(p)模型）

FPE准则函数：

$$FPE_p = \frac{n+p}{n-p}\hat{\sigma}^2$$

其中$\hat{\sigma}^2$是模型误差的方差，n样本容量，p为模型阶数

一般认为FPE_p越小越好.

例. 基于某实测时间序列数据建立AR(p)模型。

p	残差的方差	FPE_p
0	1.7203	1.7203
1	0.5097	0.5202
2	0.4790	$\boxed{0.4989}$
3	0.4728	0.5027
4	0.4708	0.5109

- 模型取AR(2)较合适

（2）AIC准则法

- 基本思想:

  既考虑模型对原始数据的预测准确性，也考虑模型中所含待定参数的个数, 适用于ARMA模型的检验

AIC准则函数：（ARMA(p,q)模型）

$$\mathrm{AIC} = \ln(\hat{\sigma}^2) + \frac{2}{n}(p+q+1)$$

其中 $\hat{\sigma}^2$ 是模型误差的方差，n 样本容量，p, q 为模型阶数

- 说明 一般认为AIC越小越好.

  第一项: 体现了模型拟合的好坏, 它随着阶数的增大而减小

  第二项: 体现了模型本身的复杂程度, 它随着阶数的增大而变大

例. 基于某实测时间序列数据建立AR(p) 模型。

p	残差的方差	AIC_p
1	418.17	1074
2	243.92	$\boxed{975.4}$
3	242.94	976.7
4	241.80	977.9
5	241.80	979.9

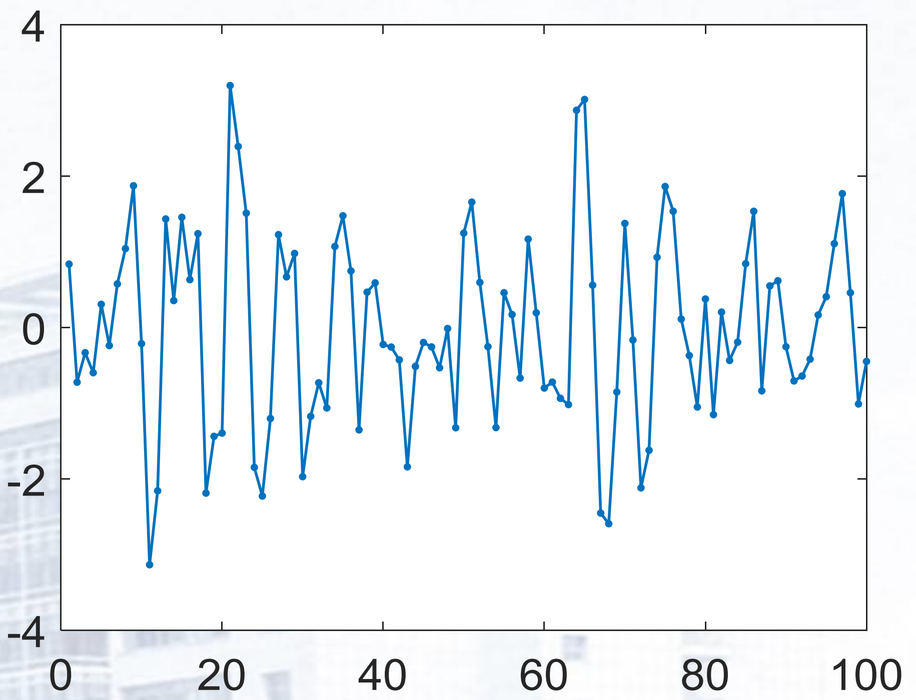
- **模型取AR(2)较合适**

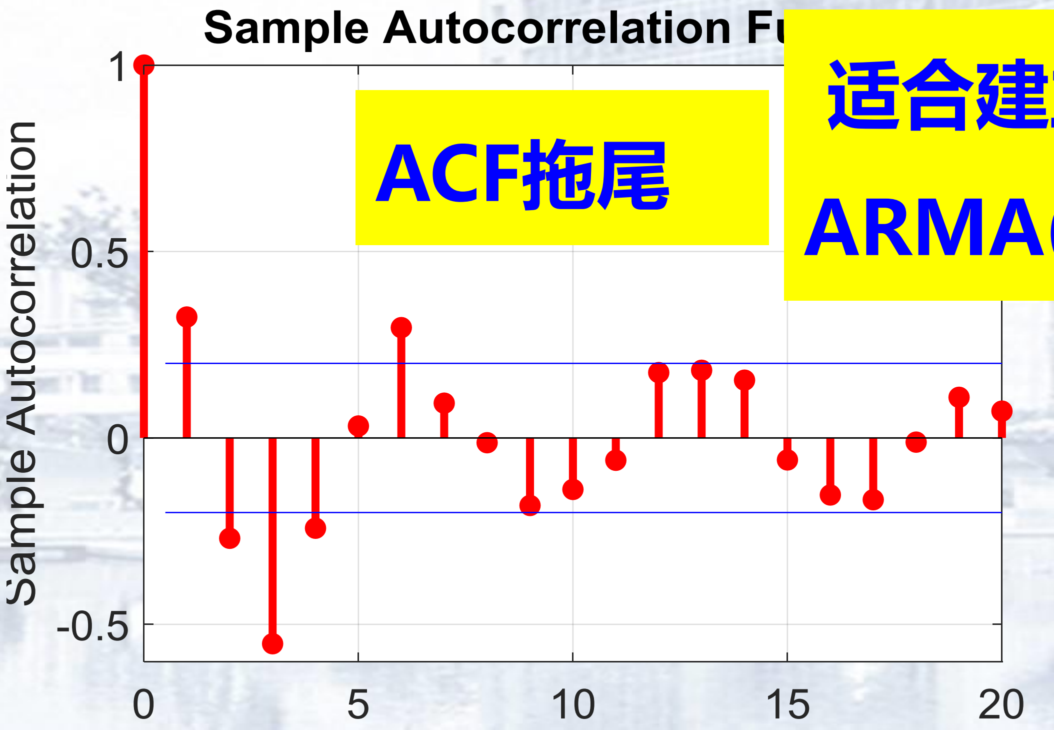
精思国计 细量民生

例. ARMA模型创建实例

Matlab代码：

```
load data3.mat x;  %x为某平稳序列数据
figure(1), autocorr(x); %自相关系数图
figure(2), parcorr(x);   %偏自相关系数图
```

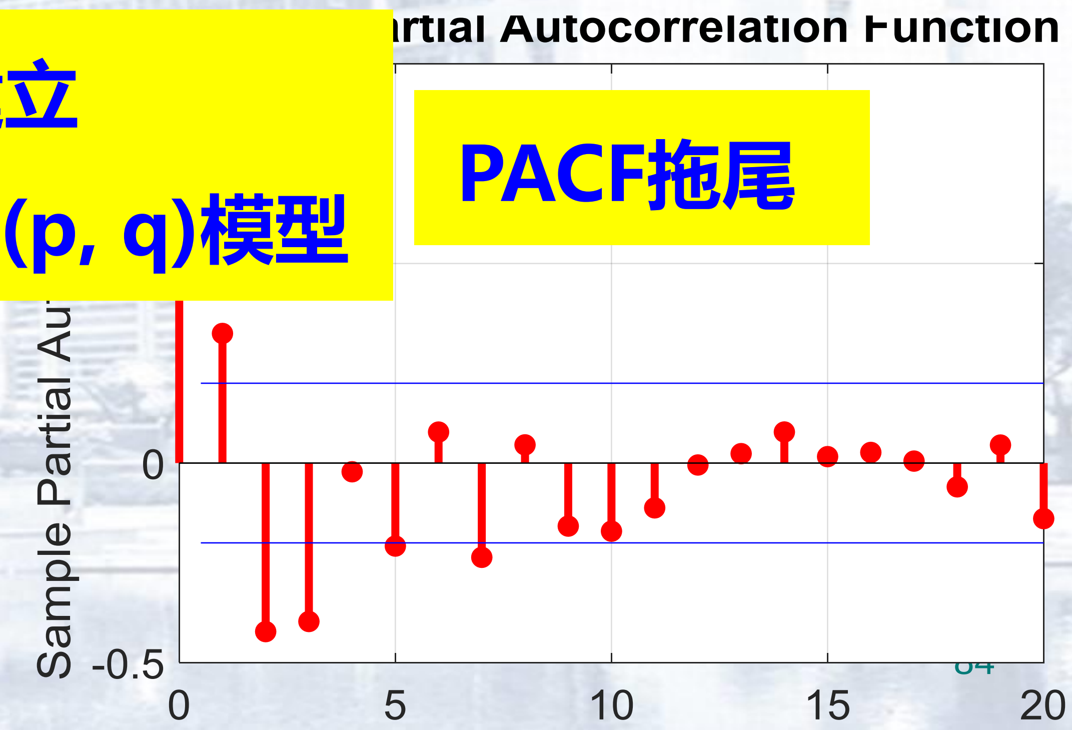




ACF拖尾

适合建立
ARMA(p, q)模型

PACF拖尾



例. ARMA模型创建实例

Matlab代码：(续)

```
%% AIC 定阶
maxLags = 5;
AICSet = zeros(maxLags, maxLags);
for i = 1:maxLags
for j = 1:maxLags
        mdl = arima('ARLags', [1:i], 'MALags', [1:j]);
        [EstMdl, EstParamCov, LogL, info] = estimate(mdl, x);
        AICSet(i, j) = aicbic(LogL, length(info.X));
end
end
```

例. ARMA模型创建实例

Matlab代码：(续)

```
% AIC结果可视化
figure(3),
heatmap(AICSet/1000)
xlabel('AR Lags')
ylabel('MA Lags')
title('AIC准则')
```

可以建立

ARMA(2, 4)模型

例. 创建ARMA模型实例

Matlab代码：(续)

```
% 建立最优ARMA(2,4)模型
mdl = arima(2, 0, 4)
EstMdl = estimate(mdl, x);
```

ARIMA(2,0,4) Model:

Constant	-0.0242
AR{1}	0.8233
AR{2}	-0.5201
MA{1}	-0.6005
MA{2}	0.0218
MA{3}	-0.2331
MA{4}	0.0222
Variance	0.9383

$$X_t = 0.024 + 0.823X_{t-1} - 0.520X_{t-2}$$

$$+\varepsilon_t + 0.601\varepsilon_{t-1} - 0.021\varepsilon_{t-2} + 0.233\varepsilon_{t-3} - 0.022\varepsilon_{t-4}$$

$$\varepsilon_t \sim N(0, 0.938)$$

精思國計 細量民生

Thanks

6.4 ARIMA模型

6.4.1 ARIMA模型的定义

6.4.2 ARIMA模型的B-J建模思想

6.4.1 ARIMA模型的定义

- 平稳时间序列：
  - 常数均值、常数方差，自协方差函数平稳
  - 建模：ARMA模型
- 非平稳时间序列：
  - 均值非平稳，方差和自协方差非平稳
  - 处理方法：差分运算，平稳化变换等
  - 建模：ARIMA模型，SARIMA模型

6.4.1 ARIMA模型的定义

实际中的很多非平稳时间序列都可以通过适当的差分运算，转化为平稳序列

- 蕴含着显著线性趋势的序列，一阶差分就可以实现趋势平稳；
- 蕴含着曲线趋势的序列，通常低阶(二阶或三阶)差分就可以提取出曲线趋势的影响，实现趋势平稳；
- 对于蕴含着固定周期的序列，通过步长为周期长度的差分，提取出周期的影响，实现趋势平稳。

6.4.1 ARIMA模型的定义

ARIMA模型的定义：

如果时间序列 $\{X_t\}$ 的 d 阶差分 $Y_t = (1-B)^d X_t$ 是一个平稳的 ARMA(p, q) 序列，其中 $d \geq 1$ 是整数，则称 X_t 为具有阶 p，d 和 q 的ARIMA模型，记为

$$X_t \sim \text{ARIMA}(p,d,q).$$

ARIMA模型的结构

$$\begin{cases} \Phi(B)\nabla^d X_t = \Theta(B)\varepsilon_t \\ \varepsilon_t \sim WN(0,\sigma^2), \quad E[X_s\varepsilon_t] = 0, \forall s < t \end{cases}$$

$$\nabla^d = (1-B)^d$$

$\Phi(B) = 1 - \phi_1 B - \cdots - \phi_p B^p$为$ARMA(p,q)$的自回归系数多项式

$\Theta(B) = 1 - \theta_1 B - \cdots - \theta_q B^q$为$ARMA(p,q)$的滑动平均系数多项式

- **ARIMA模型的实质就是差分运算与ARMA模型的组合**

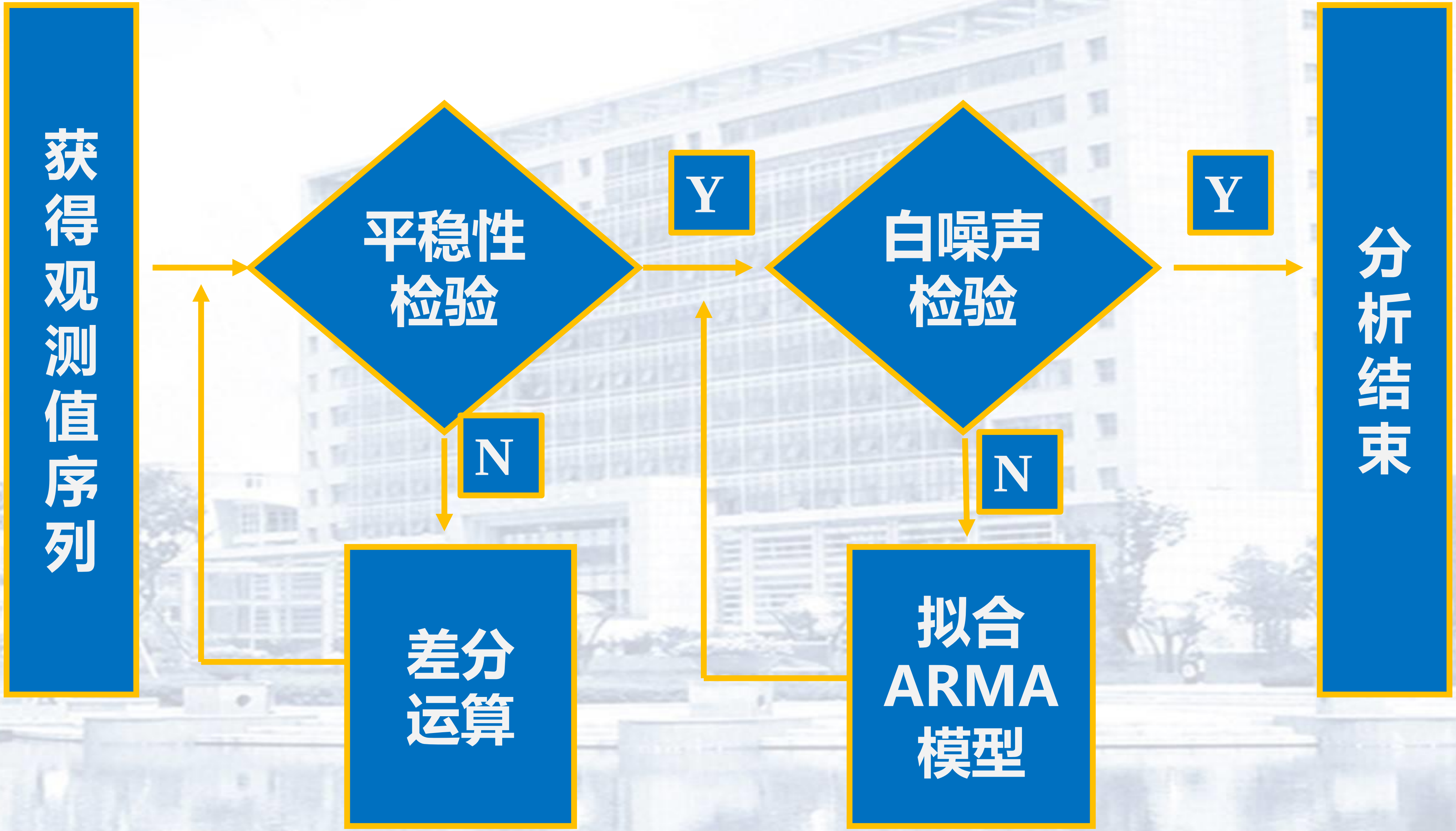
精思国计 细量民生

6.4.2 ARIMA模型的Box-Jenkins建模思想

Box-Jenkins建模思想可分为如下4个步骤：

- (1) 对原序列进行平稳性检验, 如果序列不平稳, 通过差分变换或者其他变换, 使序列平稳；
- (2) 通过相关系数法或准则函数法, 确定ARMA模型的阶数p和q；
- (3) 估计模型的未知参数, 并检验参数的显著性；
- (4) 进行诊断分析, 残差检验。

Box-Jenkins的ARIMA(p,d,q)建模思想



Box-Jenkins的ARIMA(p,d,q)建模思想

建模过程中需要一些必要的检验工作：

- (1) 时间序列数据的平稳性检验；
- (2) 残差序列的白噪声检验；
- (3) 模型参数的显著性检验；

Box-Jenkins的ARIMA(p,d,q)建模思想

❑ 时间序列数据的平稳性检验

标准方法是**单位根检验**，这里介绍两种实现方法：

（1）Augmented Dickey-Fuller (ADF) 检验

（2）Phillips-Perron (PP) 检验

```
[h, p] = adftest(X)
```

```
[h, p] = pptest(X)
```

- 检验时序数据 X 是否趋势平稳（原假设为：有单位根，不平稳）。
- h=0 表示接受原假设，即X不平稳；

  h =1 表示拒绝原假设，即可以认为X平稳；
- p 为统计量的相伴概率。

Box-Jenkins的ARIMA(p,d,q)建模思想

❑ 残差序列的白噪声检验

主要检验残差序列是否为白噪声序列，是否存在自相关性。这里介绍Ljung-Box Q方法：

```
[h, p] = lbqtest(res)
```

- 检验残差序列 res 是否存在自相关性（原假设为：不存在自相关性，是白噪声序列）。
- h=0 表示接受原假设，即认为res是白噪声序列；
  h =1 表示拒绝原假设，即认为res不是白噪声序列；
- p 为统计量的相伴概率。

精思國計 細量民生

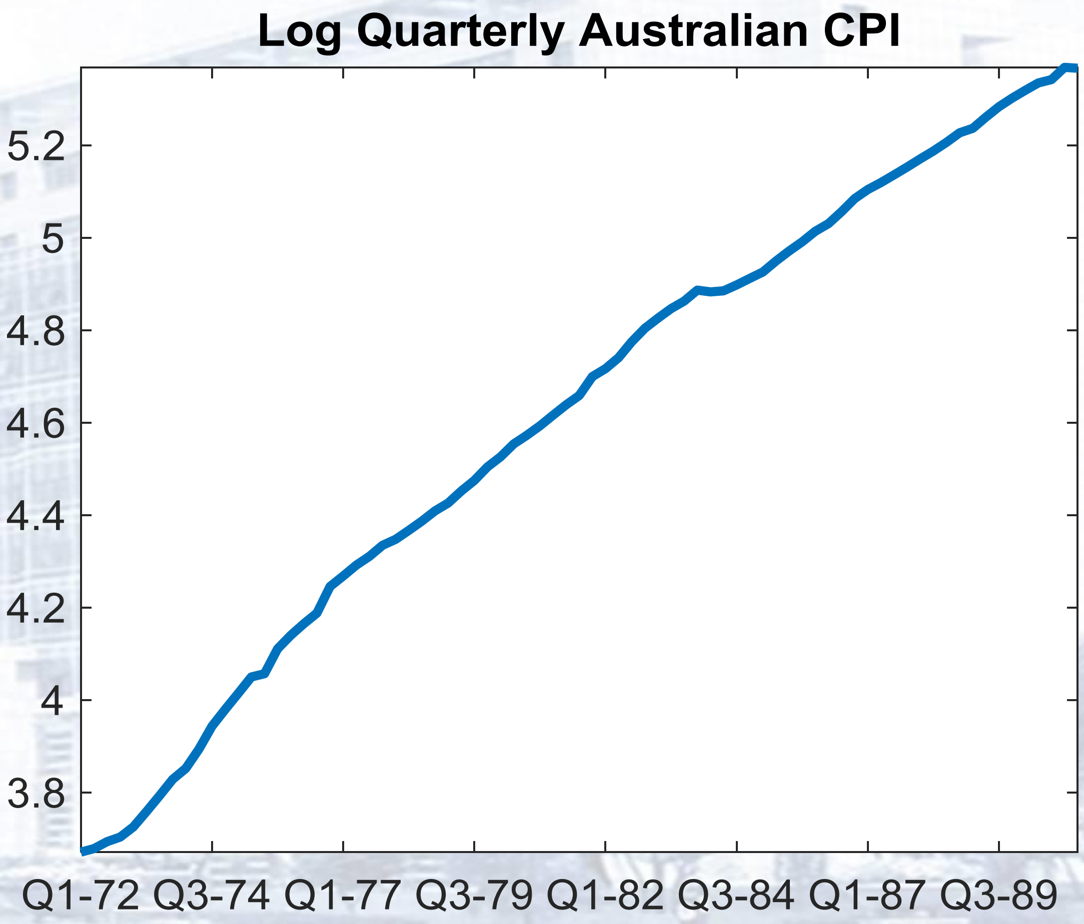
Thanks

6.5 ARIMA模型建模实例

例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
load Data_JAustralian
y = DataTable.PAU;
%% (1) 平稳性检验
figure(1), plot(y)
```



例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
load Data_JAustralian
y = DataTable.PAU;
%% (1) 平稳性检验
figure(1), plot(y)
% 平稳性的单位根检验
[h, p] = adftest(y);
```

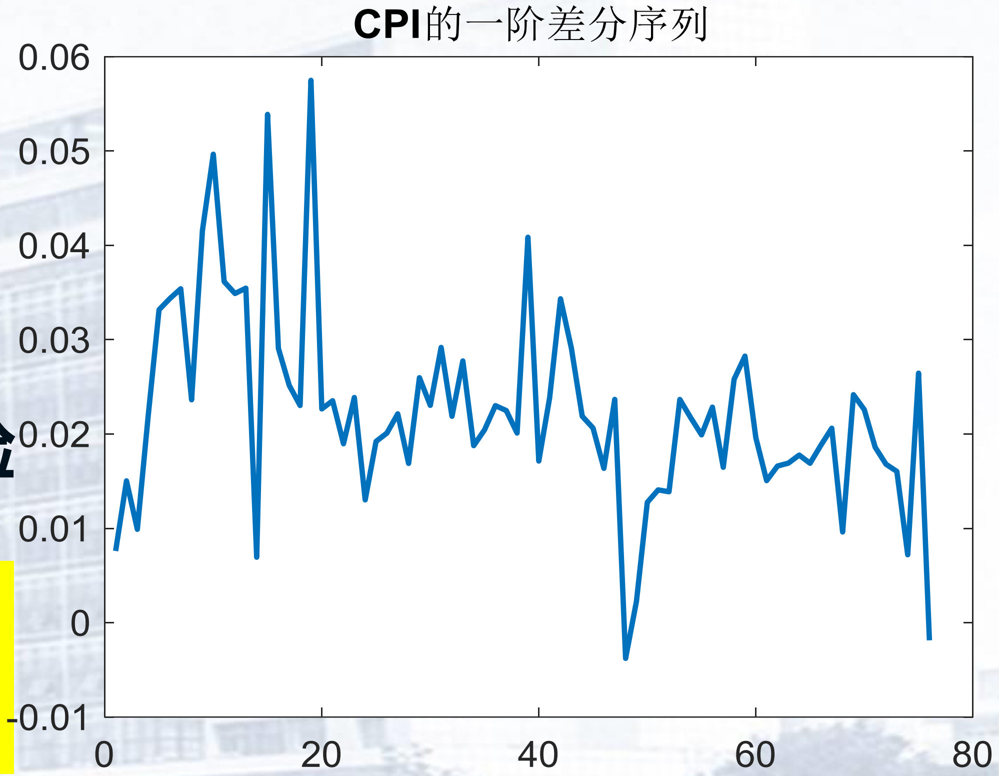
h = 0
p = 0.999

接受原假设，y序列非平稳

例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
%% (2) 差分运算
diff_x = diff(x); % 一阶差分运算
figure(4), plot(diff_x)
[h, p] = adftest(diff_x); %平稳性adf检验
```



h = 1 p = 0.0014	拒绝原假设，可以认为diff_x序列平稳

例1. 对消费指数CPI时间序列数据建立ARIMA模型

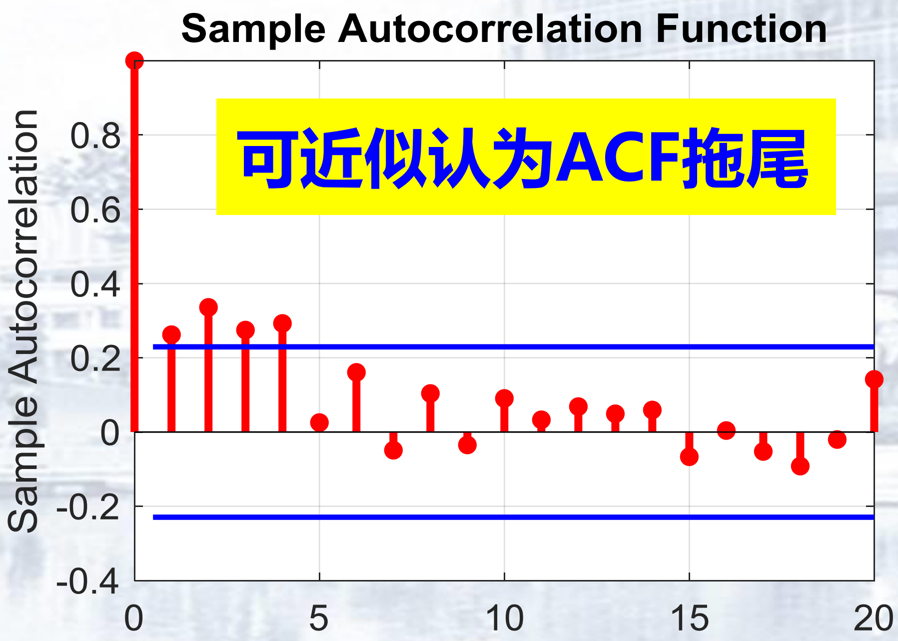
Matlab代码：

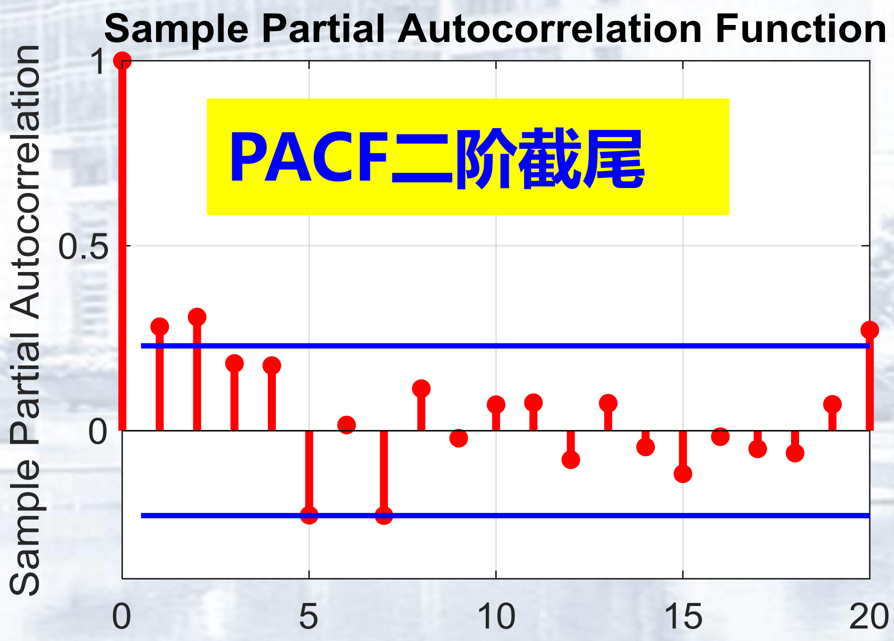
%% (3) 模型定阶

```
figure(4), autocorr(diff_x); %自相关系数图
figure(5), parcorr(diff_x);  %偏自相关系数图
```

一阶差分序列diff_x可考虑用AR(2)建模;

原序列x可用ARIMA(2,1,0)建模





例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
%% (4) 拟合ARIMA(2,1,0)模型
mdl = arima(2, 1, 0);
EstMdl = estimate(mdl, diff_y);

%% (5) 模型参数显著性检验
% 观察t统计量以及相伴概率
```

ARIMA(2,1,0) Model:

	Value	TStat	PValue
Constant	0.0101	3.071	0.002
AR{1}	0.212	2.222	0.026
AR{2}	0.337	3.250	0.001
Variance	9.230e-05	8.307	9.850e-17

模型参数显著性检验：

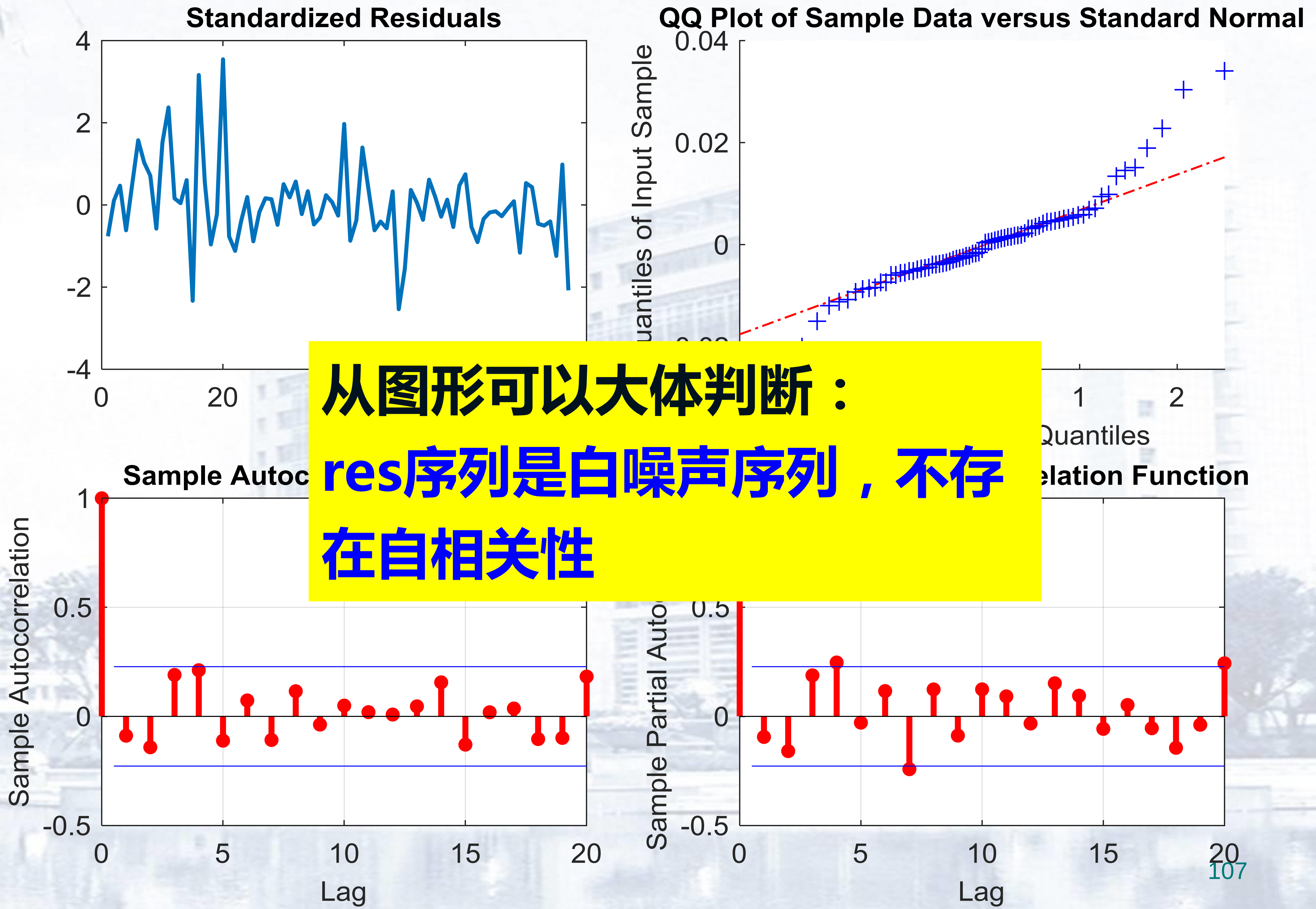
模型参数AR{1}、 AR{2}检验的t统计量相伴概率p值都小于0.05，模型参数显著。

例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
%% (6) 残差的白噪声检验
% 检验残差是否为白噪声序列、是否存在相关性
res = infer(EstMdl,y); %残差序列
% 可视化分析
figure(6),
subplot(2,2,1), plot(res./sqrt(EstMdl.Variance))
subplot(2,2,2), qqplot(res)
subplot(2,2,3), autocorr(res)
subplot(2,2,4), parcorr(res)
```

精思国计 细量民生



从图形可以大体判断：

res序列是白噪声序列，不存在自相关性

精思国计 细量民生

例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
%% (6) 模型合理性检验
% 检验残差是否为白噪声序列、是否存在自相关性
% Ljung-Box Q (lbq) 检验
[h, p] = lbqtest(res)
```

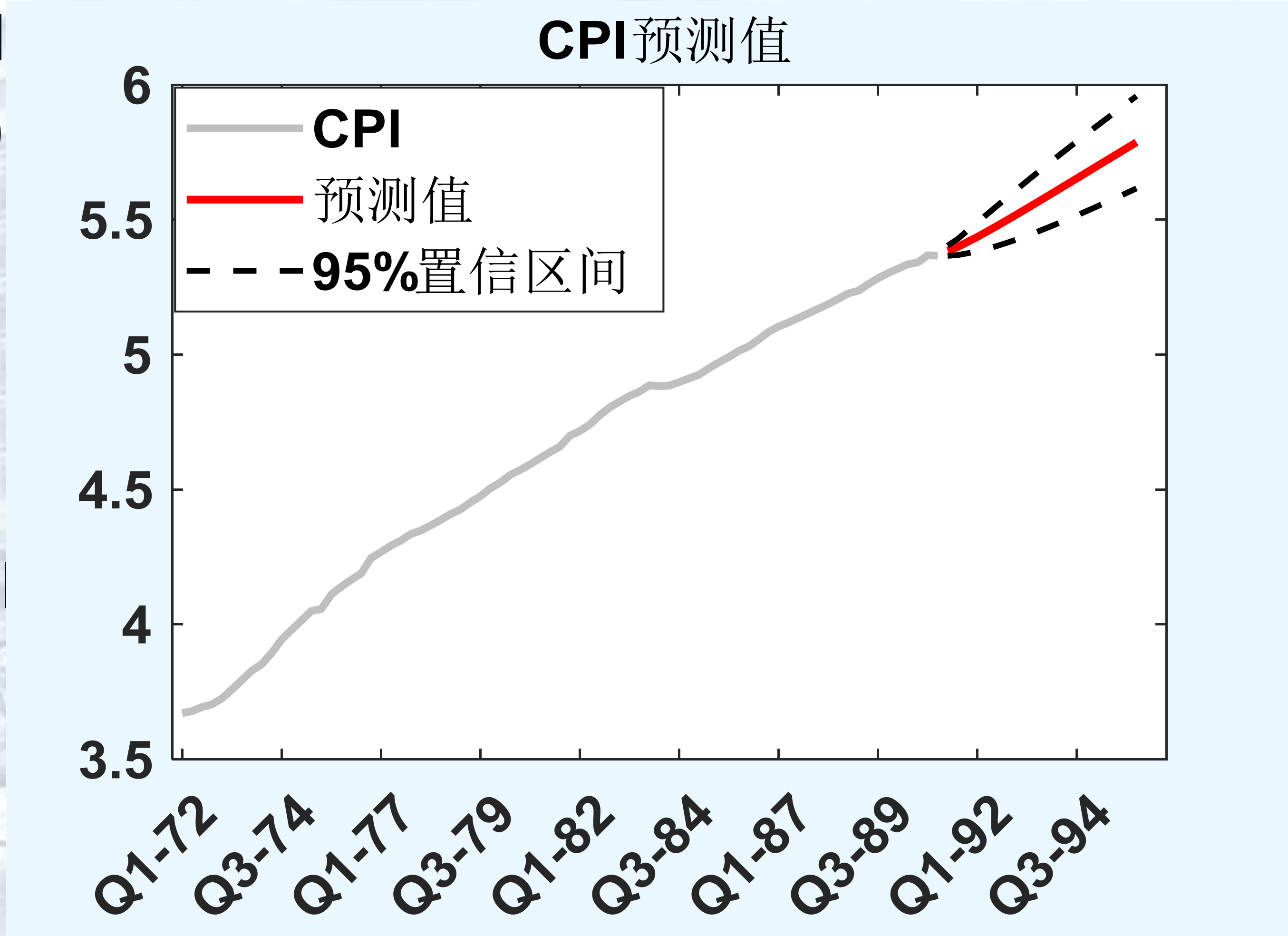
h = 0
p = 0.284

接受原假设，即可以认为res序列是白噪声序列，不存在自相关性

例1. 对消费指数CPI时间序列数据建立ARIMA模型

Matlab代码：

```
%% (7) 模型预测
[yF, yMSE] = forecast(EstM
UB = yF + 1.96*sqrt(yMSE)
LB = yF - 1.96*sqrt(yMSE);
figure(7), hold on
plot(y,'Color',[.75,.75,.75]);
plot(78:97,yF,'r');
plot(78:97,UB,'k--', 78:97,L
legend('CPI','预测值',...
   '95%置信区间')
title('CPI预测值')
```



例2.（季节ARIMA(SARIMA)模型）

对中国季度GDP数据建立SARIMA模型

Matlab代码：

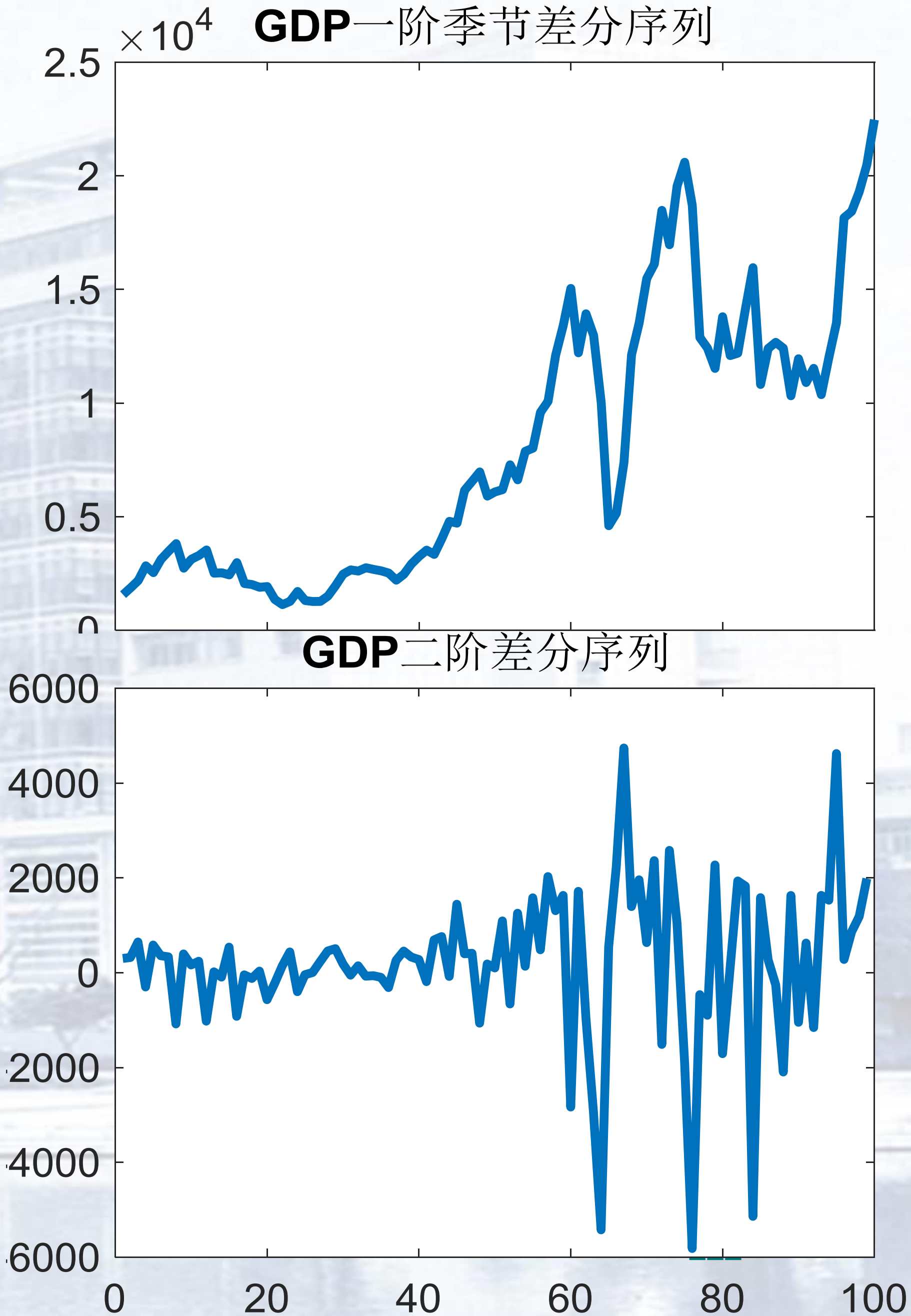
```
load data5.mat y;
%% (1) 平稳性检验
figure(1)
plot(y,'linewidth',3)
title('1992-2017年中国GDP')
```

例2.（**季节ARIMA(SARIMA)模型**）

对中国季度GDP数据建立SARIMA模型

Matlab代码：

```
%% (2) 差分运算
% 季节差分
diff1_y = y(5:end)-y(1:end-4);
figure(2),
plot(diff1_y,'linewidth',3)
title('GDP一阶季节差分序列')
% 增长趋势差分
diff2_y = diff(diff1_y);
figure(3),
plot(diff2_y,'linewidth',3)
title('GDP二阶差分序列')
```



精思国计 细量民生

例2.（季节ARIMA(SARIMA)模型）

对中国季度GDP数据建立SARIMA模型

Matlab代码：

```
%% (3) 平稳性检验
[h, p] = adftest(diff2_y); % adf单位根检验
```

h = 1
p = 0.001

拒绝原假设，可以认为diff2_y序列平稳

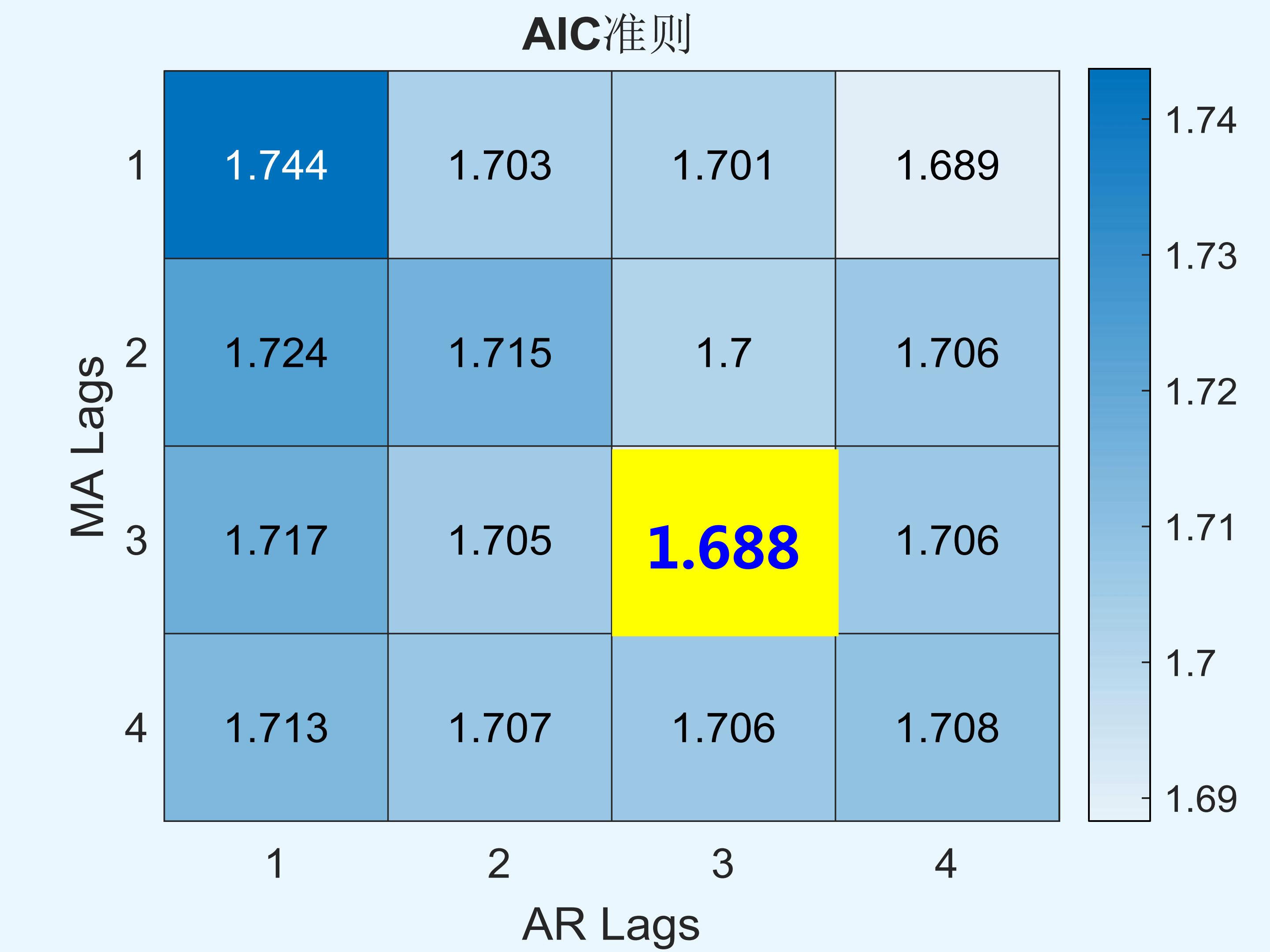
例2. （季节ARIMA(SARIMA)模型）

对中国季度GDP数据建立SARIMA模型

原序列x可用 SARIMA(3,1,3)建模

Matlab代码：

```
%% (4) 模型定阶
maxLags = 4;
AICSet = zeros(maxLag
for i = 1:maxLags
for j = 1:maxLags
        mdl = arima('AR
        [EstMdl, EstParar
        AICSet(i, j)= aicb
end
end
figure(4), heatmap(AI
```



例2. （季节ARIMA(SARIMA)模型）

对中国季度GDP数据建立SARIMA模型

Matlab代码：

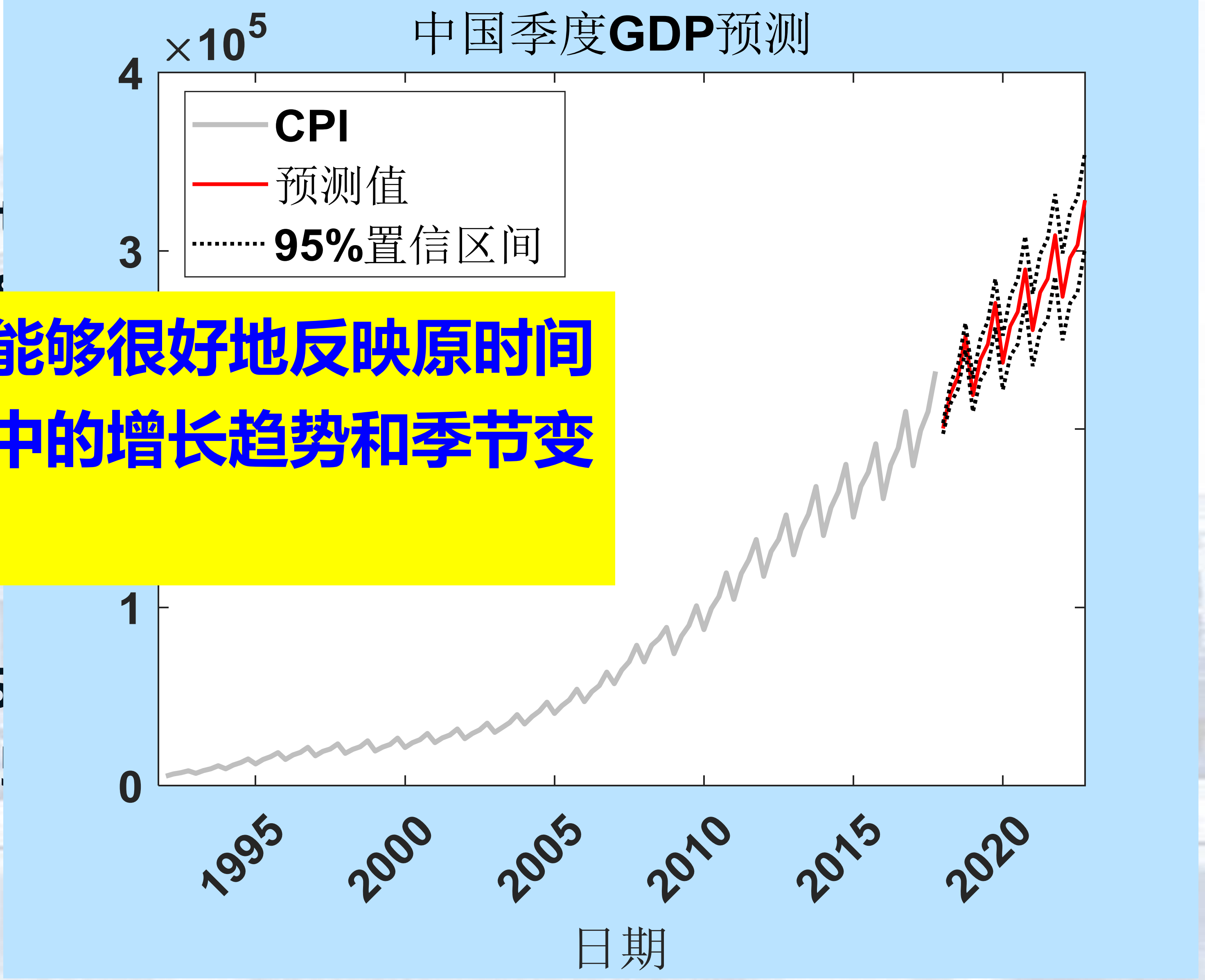
```
%% (5) 拟合SARIMA(3,1,3)模型
mdl = arima('Seasonality', 4, ...
            'D', 1, ...
            'ARLags', 1:3, ...
            'MALags', 1:3)
[EstMdl] = estimate(mdl, y');
```

例2. （季节ARIMA(SARIMA)模型）

对中国季度GDP数据建立SARIMA模型

Matlab代码：

```
%% (6) 模型预测
[yF, yMSE] = forecast(Est
UB = yF + 1.96*sqrt(yMS
LB = yF - 1.96
figure(5), hold
plot(y,'Color',
plot(105:124,yF,'r');
plot(105:124,UB,'k:', 105
legend('GDP','预测值', '9
title('GDP预测')
```



预测曲线能够很好地反映原时间序列数据中的增长趋势和季节变化趋势。

精思國計

細量民生

Thanks