

# 第三章 平稳时间序列的建模和预测



# 学习目标与要求

- 了解平稳时间序列的建模过程.
- 掌握模型识别的方法.
- 掌握自回归模型、移动平均模型和自回归移动平均模型中未知参数的常用估计方法.
- 理解自回归模型、移动平均模型和自回归移动平均模型的检验和优化的思想.
- 理解模型预测的准则, 并掌握平稳序列的预测方法.



# 本章结构

1. 自回归移动平均模型的识别

2. 参数估计

3. 模型的检验与优化

4. 序列的预测



# ARMA 模型的建模步骤

- 前面我们学习了 ARMA 模型的统计性质, 应用这些统计性质可以对观察值序列进行建模. 如果经过数据的预处理判别该序列为平稳非白噪声序列, 那么我们就可以按照 ARMA 模型的统计性质对该序列建模. 建模应该遵循第 1 章所述的建模步骤.

## 建模步骤:

- (1) 根据样本观察值, 计算自相关函数和偏自相关函数的估计值;



# ARMA 模型的建模步骤

- (2) 根据自相关函数和偏自相关函数的估计值的性质, 对  $ARMA(p,q)$  模型进行定阶, 给出  $p, q$  的值;
- (3) 对模型中的未知参数进行估计;
- (4) 对模型进行检验. 如果拟合模型未通过检验, 那么返回到第二步重新定阶, 再次选择拟合模型;
- (5) 模型优化. 如果有多个拟合模型通过了检验, 那么需要从这些模型中选择最优的拟合模型;
- (6) 利用优化后的拟合模型预测序列未来的走势.



# 自相关函数和偏自相关函数的估计

- 设  $x_1, x_2, \dots, x_n$  是平稳序列  $\{x_t, t \in T\}$  的一个样本, 则我们可以根据如下公式估计出该序列的自相关函数

$$\hat{\rho}(k) = \frac{\sum_{t=1}^{n-k} (x_t - \bar{x})(x_{t+k} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}, \quad \forall 0 < k < n.$$

- 将样本的自相关函数代入 Yule-Walker 方程

$$\begin{bmatrix} 1 & \hat{\rho}(1) & \dots & \hat{\rho}(k-1) \\ \hat{\rho}(1) & 1 & \dots & \hat{\rho}(k-2) \\ \vdots & \vdots & & \vdots \\ \hat{\rho}(k-1) & \hat{\rho}(k-2) & \dots & 1 \end{bmatrix} \begin{pmatrix} \phi_{k1} \\ \phi_{k2} \\ \vdots \\ \phi_{kk} \end{pmatrix} = \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(k) \end{pmatrix}. \quad (3.1)$$

# 自相关函数和偏自相关函数的估计

- 在方程 (3.1) 中, 依次取  $k = 1, 2, \dots, n$ , 并利用如下公式, 求得偏自相关函数的估计值

$$\hat{\phi}_{kk} = \frac{\hat{D}_k}{\hat{D}},$$

其中

$$\hat{D} = \begin{vmatrix} 1 & \hat{\rho}(1) & \dots & \hat{\rho}(k-1) \\ \hat{\rho}(1) & 1 & \dots & \hat{\rho}(k-2) \\ \vdots & \vdots & & \vdots \\ \hat{\rho}(k-1) & \hat{\rho}(k-2) & \dots & 1 \end{vmatrix}, D_k = \begin{vmatrix} 1 & \hat{\rho}(1) & \dots & \hat{\rho}(1) \\ \hat{\rho}(1) & 1 & \dots & \hat{\rho}(2) \\ \vdots & \vdots & & \vdots \\ \hat{\rho}(k-1) & \hat{\rho}(k-2) & \dots & \hat{\rho}(k) \end{vmatrix}$$

# 模型识别的方法

- 当估计出模型的自相关函数和偏自相关函数后, 我们可以根据估计值表现出的拖尾和截尾性质, 估计出合适的自相关阶数  $\hat{p}$  和移动平均阶数  $\hat{q}$ , 从而选择出适当的 ARMA 模型拟合观察值序列. 我们将这个过程称为**模型识别**. 可见, 此时模型识别过程就是模型定阶的过程.
- 自相关函数和偏自相关函数的截尾意味着从某步之后的自相关函数和偏自相关函数为零, 但是由于样本的随机性, 样本自相关函数和样本偏自相关函数不可能呈现出完美的截尾情况, 而只可能在零附近区域随机波动. 另一方面, 平稳时间序列具有短期相关性, 即随着延迟阶数  $k$  的增大, 样本自相关函数 和样本偏自相关函数迅速衰减至零附近波动.





# 模型识别的方法

因此,我们在定阶时必须考虑,随着延迟阶数的增大,样本自相关函数和样本偏自相关函数衰减到零附近波动时,何时可看做样本自相关函数或样本偏自相关函数的截尾,何时可看做正常衰减至零值附近的拖尾.但是这实际上没有绝对的标准,很大程度上依靠分析人员的主观经验.尽管如此,我们还是可以提供一些理论依据,帮助人们来做合理分析.

- 根据  $ARMA(p,q)$  模型的统计性质,我们可得到如下定阶原则

模 型	AR(p)	MA(q)	ARMA(p,q)
$\hat{\rho}(k)$	拖 尾	截 尾	拖 尾
$\hat{\phi}_{kk}$	截 尾	拖 尾	拖 尾

# 模型识别的方法

- 尽管样本自相关函数和样本偏自相关函数的截尾指的是，它们的值在零附近区域做小幅波动，而不是像总体自相关函数和总体偏自相关函数那样具有严格的截尾。不过，我们可以通过研究样本自相关函数和样本偏自相关函数的近似分布，来选取适当的阶数。
- 研究表明，当样本容量  $n$  充分大时，样本自相关函数  $\hat{\rho}(k)$  近似服从正态分布：

$$\hat{\rho}(k) \sim N(0, 1/n).$$

而样本偏自相关函数  $\hat{\phi}_{kk}$  也近似服从正态分布：

$$\hat{\phi}_{kk} \sim N(0, 1/n).$$

# 模型识别的方法

- 根据正态分布的性质, 得

$$P(|\hat{\rho}(k)| \leq 2/\sqrt{n}) \approx 95.5\% \quad \text{且} \quad P(|\hat{\phi}_{kk}| \leq 2/\sqrt{n}) \approx 95.5\%$$

因此, 若满足不等式

$$|\hat{\rho}(k)| \leq 2/\sqrt{n}$$

的比例达到了 95.5%, 则可以认为  $\hat{\rho}(k)$  截尾; 同样地, 若满足不等式  $|\hat{\phi}_{kk}| \leq 2/\sqrt{n}$  的比例达到了 95.5%, 则可以认为  $\hat{\phi}_{kk}$  截尾.

在实际应用中, 一般按照 2 倍标准差作为截尾标准, 即如果样本自相关函数或样本偏自相关函数在最初的  $l$  阶明显超出 2 倍标准差范围, 而之后几乎 95.5% 的值都落在 2 倍标准差范围内, 而且衰减到 2 倍标准差范围内的速度很快, 则通常可以认为  $l$  阶截尾.

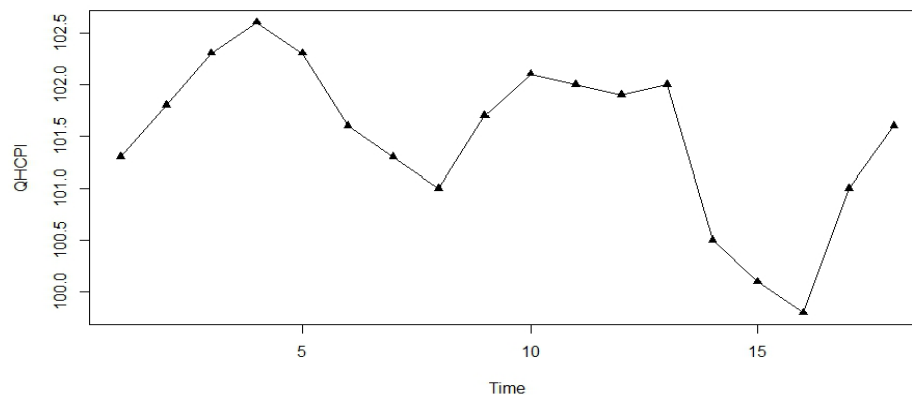


# 模型识别的方法

- **例 3.1** 选择合适的模型拟合 2016 年 1 月至 2017 年 6 月青海省居民消费指数.

**解：** 首先，读入数据并绘制时序图.

```
> x <- read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)
> QHCPI <- ts(x$QHCPI,start=1)
> plot(QHCPI,type="o",pch=17)
```



# 模型识别的方法

- 时序图显示序列具有平稳特征. 然后, 我们做白噪声检验.

```
> for (i in 1:2) print(Box.test(QHCPI,type="Ljung-Box",lag=6*i))
```

Box-Ljung test

data: QHCPI

X-squared = 15.53, df = 6, p-value = 0.01651

Box-Ljung test

data: QHCPI

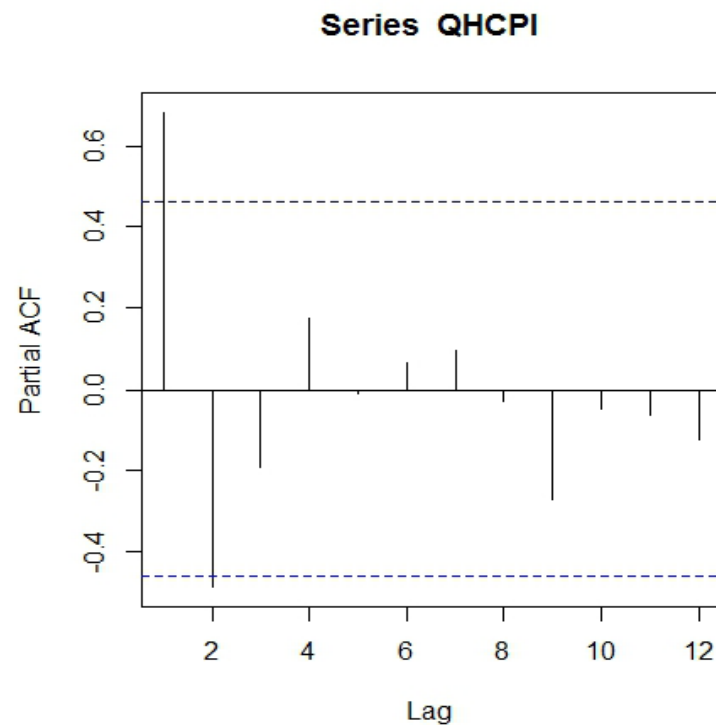
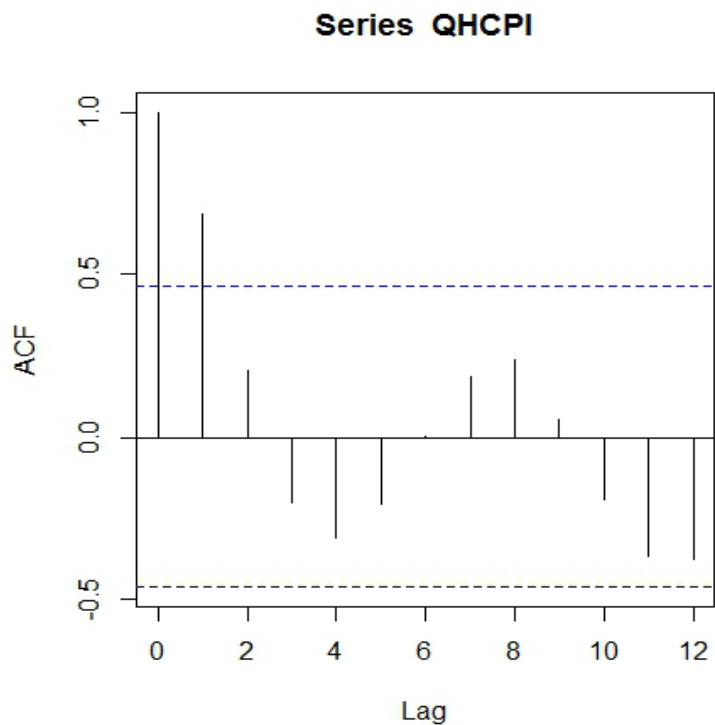
X-squared = 35.801, df = 12, p-value = 0.0003488

- 做延迟 6 阶和延迟 12 阶的白噪声检验, 表明该序列为非白噪声序列. 最后, 根据自相关函数图和偏自相关函数图定阶.



# 模型识别的方法

```
> par(mfrow=c(1,2))  
> acf(QHCPI)  
> pacf(QHCPI)
```



# 模型识别的方法

- 可见, 一方面, 从样本自相关函数图来看, 自相关函数延迟 1 阶之后, 衰减到 2 倍标准差范围内 (自相关图中的虚线); 而样本偏自相关函数图也表明偏自相关函数延迟 2 阶之后, 完全衰减到 2 倍标准差范围内 (偏自相关图中的虚线). 这些进一步说明该序列具有短期相关性, 显示序列的平稳的特征.
- 另一方面, 样本自相关函数图衰减到 2 倍标准差范围内值呈现“伪正弦波动”, 说明自相关函数呈现拖尾现象; 偏自相关函数图呈现了 2 阶截尾特征. 因此, 我们可以初步确定拟合模型为AR(2).

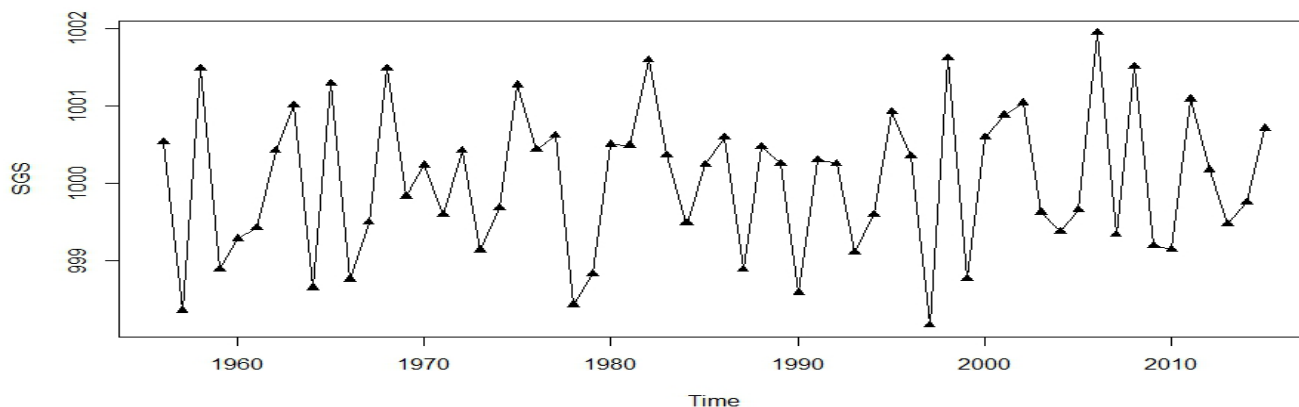


# 模型识别的方法

- 例 3.2 选择合适的模型拟合 1956 年至 2016 年某城市各月的交通事故数.

**解：** 读入数据，并绘制时序图，观察序列走势。

```
> x <-  
read.table("E:/DATA/CHAP3/SGS.csv",sep="," ,header=T)  
> SGS <- ts(x$JTSGS,start=1956)  
> plot(SGS,type="o",pch=17)
```





# 模型识别的方法

- 可见, 序列时序图呈现平稳特征. 下面进行延迟 6 阶和延迟 12 阶的白噪声检验.

```
> for (i in 1:2) print(Box.test(SGS,type="Ljung-Box",lag=6*i))
```

Box-Ljung test

data: SGS

X-squared = 8.8243, df = 6, p-value = 0.0183

Box-Ljung test

data: SGS

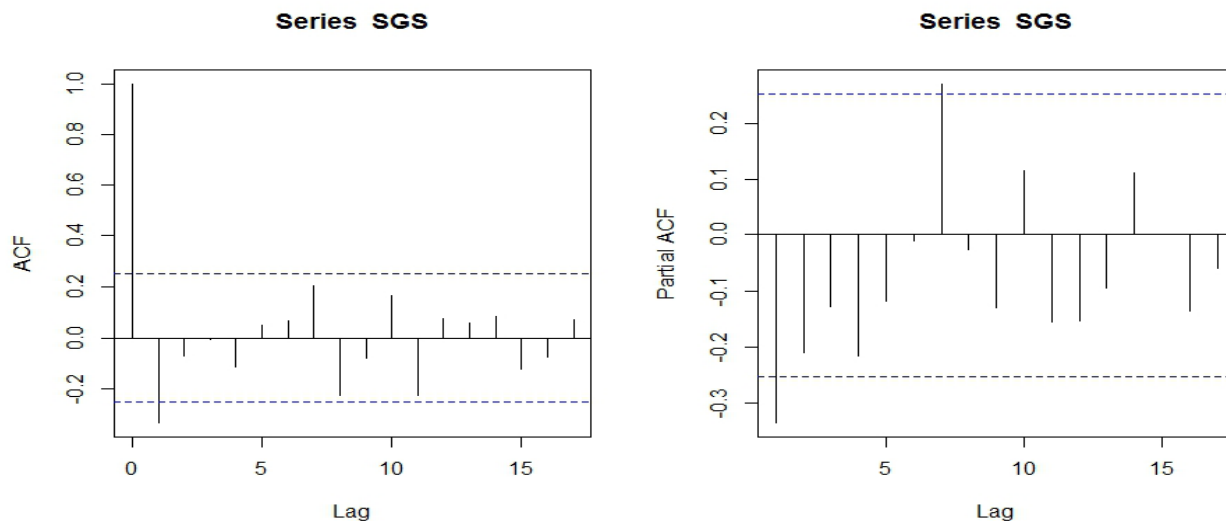
X-squared = 22.44, df = 12, p-value = 0.00328

检验表明, 该序列为非白噪声序列. 下面绘制自相关函数图和偏自相关函数图来进行模型初步识别.



# 模型识别的方法

- `> par(mfrow=c(1,2)); acf(SGS); pacf(SGS)`



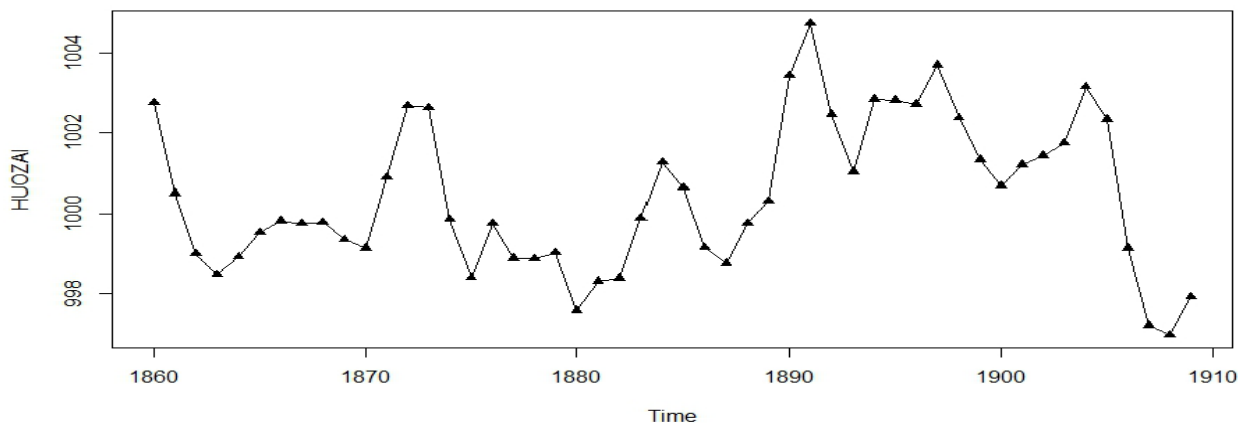
- 可见，自相关函数和偏自相关函数具有短期相关性，而且自相关函数延迟 1 阶之后，呈现明显的截尾特征，偏自相关函数却表现出拖尾形态，因此，初步选定拟合模型为 MA(1).

# 模型识别的方法

- 例 3.3 选择合适的模型拟合 1860 年至 1909 年国外某城市火灾发生数.

**解** 读入数据, 并绘制时序图, 观察序列走势.

```
> x <- read.table("E:/DATA/CHAP3/2.csv",sep=";",header=T)
> HUOZAI <- ts(x$X,start=1860)
> plot(HUOZAI,type="o",pch=17)
```



# 模型识别的方法

- 可以看出, 序列呈现平稳特征. 下面进行延迟 5 阶和延迟 10 阶的白噪声检验.

```
> for (i in 1:2) print(Box.test(HUOZAI,type="Ljung-Box",lag=5*i))
```

Box-Ljung test

data: HUOZAI

X-squared = 36.859, df = 5, p-value = 6.393e-07

Box-Ljung test

data: HUOZAI

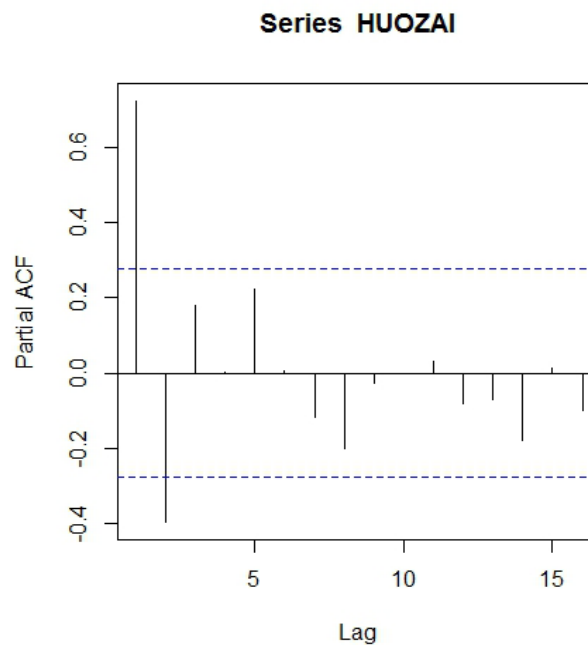
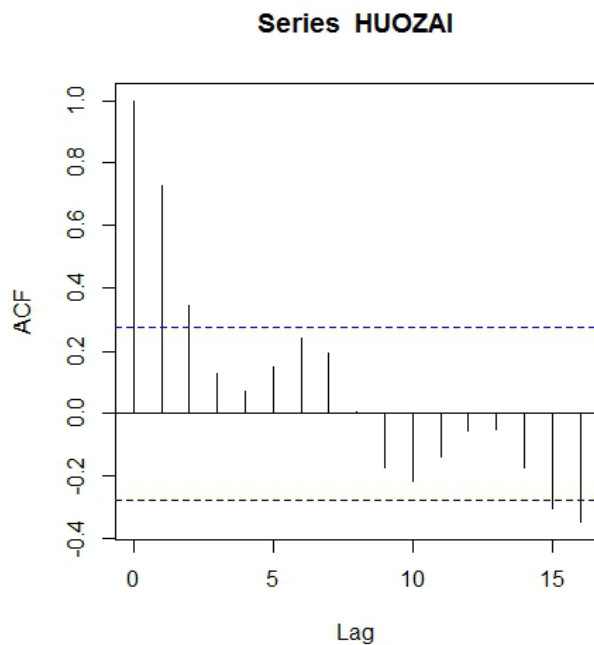
X-squared = 47.515, df = 10, p-value = 7.608e-07

- 检验表明,  $p$  值远远小于 0.05, 该序列为非白噪声序列. 下面绘制自相关函数图和偏自相关函数图来进行模型初步识别.



# 模型识别的方法

```
> par(mfrow=c(1,2));acf(HUOZAI);pacf(HUOZAI)
```



# 模型识别的方法

- 可见, 自相关函数和偏自相关函数具有短期相关性, 同时自相关函数和偏自相关函数都表现出明显的拖尾形态, 因此, 初步选定拟合模型为  $ARMA(2,1)$ .
- 这里需要说明的是, 在实际建模时, 由于  $p$  和  $q$  通常较低, 故在自相关函数和偏自相关函数都表现出拖尾时, 可以由低阶到高阶逐步尝试.



# 本章结构

1. 自回归移动平均模型的识别

2. 参数估计

3. 模型的检验与优化

4. 序列的预测



# 矩估计法

- 本节主要论述如何基于序列观察值对 ARMA(p,q) 模型的未知参数进行估计. 在这节中, 我们假定已经确定了序列是平稳的时间序列, 并且已经进行了模型识别, 即确定了自回归阶数  $p$  和移动平均阶数  $q$ . 本节主要介绍利用矩估计法、最小二乘估计法和极大似然估计法估计 ARMA(p,q) 模型中的未知参数  $\mu$ ,  $\sigma_\varepsilon^2$ ,  $\theta_i$ ,  $\phi_k$ .

- **矩估计法**

所谓**矩估计**, 就是令样本矩等于相应的总体矩, 通过求解所得方程, 得到未知参数的估计方法.





# 矩估计法

- **特点：**（1）矩估计法具有简单直观，计算量相对较小，且不需要假设总体分布等优点；（2）矩估计法忽略了观察值序列的其他信息，因而导致其估计精度不高。在实际中，它常被用来做初始估计，以确定最小二乘估计或极大似然估计中迭代计算的初值。

- **AR(p) 模型的矩估计**

在AR(1) 模型：  $x_t = \phi_1 x_{t-1} + \varepsilon_t$  中，未知参数为  $\phi_1$ 。

由于  $\rho(1) = \phi_1$ ，所以用样本自相关函数  $\hat{\rho}(1)$  替代总体自相关函数  $\rho(1)$  后，得  $\phi_1$  的估计  $\hat{\phi}_1$ ：  $\hat{\phi}_1 = \hat{\rho}(1)$ 。



# 矩估计法

在AR(2) 模型:  $x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \varepsilon_t$  中, 未知参数为  $\phi_1, \phi_2$ . 根据 Yule-Walker 方程, 得

$$\begin{cases} \rho(1) = \phi_1 + \rho(1)\phi_2, \\ \rho(2) = \rho(1)\phi_1 + \phi_2. \end{cases}$$

按照矩估计法的思想, 分别用延迟 1 阶和延迟 2 阶的样本自相关函数代替相应的总体自相关函数, 得

$$\begin{cases} \hat{\rho}(1) = \phi_1 + \hat{\rho}(1)\phi_2, \\ \hat{\rho}(2) = \hat{\rho}(1)\phi_1 + \phi_2. \end{cases}$$

求解后得到未知参数  $\phi_1, \phi_2$  的矩估计  $\hat{\phi}_1, \hat{\phi}_2$ :

$$\hat{\phi}_1 = \hat{\rho}(1) \frac{1 - \hat{\rho}(2)}{1 - [\hat{\rho}(1)]^2}, \quad \hat{\phi}_2 = \frac{\hat{\rho}(2) - [\hat{\rho}(1)]^2}{1 - [\hat{\rho}(1)]^2}$$



# 矩估计法

在AR(p) 模型:

$$x_t = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \cdots + \phi_p x_{t-p} + \varepsilon_t$$

中, 未知参数为  $\phi_1, \phi_2, \dots, \phi_p$ . 在 Yule-Walker 方程中, 分别用延迟  $k$  ( $1 \leq k \leq p$ ) 阶的样本自相关函数  $\hat{\rho}(k)$  代替总体自相关函数  $\rho(k)$ , 得到**样本 Yule-Walker 方程**

$$\begin{bmatrix} 1 & \hat{\rho}(1) & \cdots & \hat{\rho}(k-1) \\ \hat{\rho}(1) & 1 & \cdots & \hat{\rho}(k-2) \\ \vdots & \vdots & & \vdots \\ \hat{\rho}(k-1) & \hat{\rho}(k-2) & \cdots & 1 \end{bmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \\ \vdots \\ \phi_k \end{pmatrix} = \begin{pmatrix} \hat{\rho}(1) \\ \hat{\rho}(2) \\ \vdots \\ \hat{\rho}(k) \end{pmatrix}. \quad (3.4)$$

求解线性方程组 (3.4), 得到估计  $\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p$ . 称这样得到的估计为**Yule-Walker 估计**.



# 矩估计法

- MA(q) 模型的矩估计

首先考虑 MA(1) 模型:  $x_t = \varepsilon_t - \theta_1 \varepsilon_{t-1}$ . 该模型的待估参数是  $\theta_1$ . 由例 2.11 知

$$\rho(1) = \frac{-\theta_1}{1 + \theta_1^2}$$

从而得

$$\rho(1)\theta_1^2 + \theta_1 + \rho(1) = 0.$$

用  $\hat{\rho}(1)$  替代上面一元二次方程中  $\rho(1)$ , 并解之, 得

$$\hat{\theta}_1 = \frac{-1 \pm \sqrt{1 - 4\hat{\rho}^2(1)}}{2\hat{\rho}(1)}.$$

考虑到 MA(1) 模型的可逆性条件为  $|\theta_1| < 1$ , 可得未知参数的估计



# 矩估计法

$$\hat{\theta}_1 = \frac{-1 + \sqrt{1 - 4\hat{\rho}^2(1)}}{2\hat{\rho}(1)}.$$

对于高阶的 MA(q) 模型:  $x_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \cdots - \theta_q\varepsilon_{t-q}$ , 其待估参数  $\theta_1, \theta_2, \dots, \theta_q$  的计算较为复杂. 将方程组

$$\rho(k) = \frac{-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{k+i}}{1 + \theta_1^2 + \cdots + \theta_q^2}, \quad 1 \leq k \leq q$$

中的  $\rho(k)$  用  $\hat{\rho}(k)$  代替, 求解上述非线性方程组, 就可得未知参数的矩估计  $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ . 解非线性方程组比较麻烦, 一般都是借助数值算法求得的, 所以我们不再做进一步探讨.



# 矩估计法

- ARMA(p,q) 模型的矩估计

由于一般的 ARMA(p,q) 模型的矩估计将更为复杂, 且估计精度较差, 所以我们仅以ARMA(1,1) 模型的矩估计为例来说明估计过程.

ARMA(1,1) 模型:  $x_t = \phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$  的待估参数为  $\phi_1, \theta_1$ , 故需要构造两个方程. 由 ARMA(p,q) 模型的自相关函数公式

$$\rho(k) = \frac{\sum_{i=0}^{\infty} G_i G_{i+k}}{\sum_{i=0}^{\infty} G_i^2}.$$

知, 我们需首先确定 ARMA(1,1) 模型的 Green 函数.



# 矩估计法

- ARMA(p,q) 模型的矩估计

由于一般的 ARMA(p,q) 模型的矩估计将更为复杂, 且估计精度较差, 所以我们仅以ARMA(1,1) 模型的矩估计为例来说明估计过程.

ARMA(1,1) 模型:  $x_t = \phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$  的待估参数为  $\phi_1, \theta_1$ , 故需要构造两个方程. 由 ARMA(p,q) 模型的自相关函数公式

$$\rho(k) = \frac{\sum_{i=0}^{\infty} G_i G_{i+k}}{\sum_{i=0}^{\infty} G_i^2}. \quad (3.5)$$

知, 我们需首先确定 ARMA(1,1) 模型的 Green 函数.



# 矩估计法

根据 ARMA 模型 Green 函数的递推公式, 可推得 ARMA(1,1) 模型的 Green 函数为

$$\begin{cases} G_0 = 1; \\ G_i = (\phi_1 - \theta_1)\phi_1^{i-1}, i=1,2,\dots \end{cases} \quad (3.6)$$

在 (3.5) 式中分别取  $k=1,2$ , 并将 (3.6) 式代入, 得

$$\begin{cases} \rho(1) = \frac{(\phi_1 - \theta_1)(1 - \theta_1\phi_1)}{1 + \theta_1^2 - 2\theta_1\phi_1}; \\ \rho(2) = \phi_1\rho(1). \end{cases} \quad (3.7)$$





# 矩估计法

在(3.7)中,分别用  $\hat{\rho}(1)$  和  $\hat{\rho}(2)$  替换  $\rho(1)$  和  $\rho(2)$ ,并求解关于  $\phi_1, \theta_1$  的方程组,结合可逆性条件  $|\theta_1| < 1$ ,得到矩估计的唯一解

$$\begin{cases} \hat{\phi}_1 = \frac{\hat{\rho}(2)}{\hat{\rho}(1)}; \\ \hat{\theta}_1 = \begin{cases} \frac{c + \sqrt{c^2 - 4}}{2}, & c \leq -2; \\ \frac{c - \sqrt{c^2 - 4}}{2}, & c \geq 2, \end{cases} \end{cases} \quad \text{其中, } c = \frac{1 + \hat{\phi}_1^2 - 2\hat{\rho}(2)}{\hat{\phi}_1 - \hat{\rho}(1)}.$$

# 矩估计法

- 噪声方差  $\sigma_\varepsilon^2$  的矩估计

回顾序列  $\{x_t, t \in T\}$  均值  $\mu$  的估计为

$$\hat{\mu} = \bar{x} = \frac{1}{n} \sum_{t=1}^n x_t.$$

方差  $\sigma_x^2 = \gamma(0)$  的估计为

$$\sigma_x^2 = \frac{1}{n-1} \sum_{t=1}^n (x_t - \bar{x})^2$$

(1) 对于AR(p) 模型而言, 在等式  $x_t = \phi_0 + \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p} + \varepsilon_t$  的两边同时乘以  $x_t$ , 并求期望, 得

$$\gamma(0) = \phi_1 \gamma(1) + \phi_2 \gamma(2) + \dots + \phi_p \gamma(p) + \sigma_\varepsilon^2 \quad (3.8)$$



# 矩估计法

将  $\gamma(k) = \gamma(0)\rho(k)$  代入 (3.8) 式, 整理得

$$\gamma(0) = \frac{\sigma_{\varepsilon}^2}{1 - \phi_1\rho(1) - \phi_2\rho(2) - \dots - \phi_p\rho(p)}.$$

由 (3.9) 式得  $\sigma_{\varepsilon}^2$  的矩估计为

$$\hat{\sigma}_{\varepsilon}^2 = (1 - \hat{\phi}_1\hat{\rho}(1) - \hat{\phi}_2\hat{\rho}(2) - \dots - \hat{\phi}_p\hat{\rho}(p))\hat{\sigma}_x^2$$

特别地, 对于 AR(1) 模型, 因为  $\hat{\phi}_1 = \hat{\rho}(1)$ , 所以

$$\hat{\sigma}_{\varepsilon}^2 = (1 - \hat{\rho}^2(1))\hat{\sigma}_x^2$$

(2) 对于 MA(q) 模型, 使用  $\gamma(0) = (1 + \theta_1^2 + \dots + \theta_q^2)\sigma_{\varepsilon}^2$  得到

$$\hat{\sigma}_{\varepsilon}^2 = \frac{\hat{\sigma}_x^2}{1 + \hat{\theta}_1^2 + \dots + \hat{\theta}_q^2}.$$

# 矩估计法

(3)对 ARMA(p,q) 模型, 我们仅以 ARMA(1,1) 模型:

$$x_t = \phi_1 x_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}$$

为例讨论. 由

$$\begin{cases} E(\varepsilon_t x_t) = \sigma_\varepsilon^2 \\ E(\varepsilon_{t-1} x_t) = (\phi_1 - \theta_1) \sigma_\varepsilon^2 \end{cases}$$

得

$$\begin{cases} \gamma(0) = E(x_t^2) = \phi_1 \gamma(1) + [1 - \theta_1(\phi_1 - \theta_1)] \sigma_\varepsilon^2 \\ \gamma(1) = E(x_{t-1} x_t) = \phi_1 \gamma(0) - \theta_1 \sigma_\varepsilon^2 \end{cases}$$

于是

$$\sigma_x^2 = \gamma(0) = \frac{1 - 2\phi_1\theta_1 + \theta_1^2}{1 - \phi_1^2} \sigma_\varepsilon^2$$

从而

$$\hat{\sigma}_\varepsilon^2 = \frac{1 - \hat{\phi}_1^2}{1 - 2\hat{\phi}_1\hat{\theta}_1 + \hat{\theta}_1^2} \hat{\sigma}_x^2$$



# 最小二乘估计

- 所谓的**最小二乘估计** (least squares estimation), 就是使得残差平方和达到最小的未知参数值. 下面分情况来讨论.
- **AR(p) 模型的最小二乘估计**

AR(p) 模型的待估参数为  $\Phi = (\phi_1, \phi_2, \dots, \phi_p)$ . 记

$$F_t(\Phi) = \phi_1 x_{t-1} + \phi_2 x_{t-2} + \dots + \phi_p x_{t-p}$$

则残差项为

$$\varepsilon_t = x_t - F_t(\Phi)$$

条件残差平方和  $Q(\Phi)$  为

$$Q(\Phi) = \sum_{t=p+1}^n \varepsilon_t^2 = \sum_{t=p+1}^n (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \dots - \phi_p x_{t-p})^2$$



# 最小二乘估计

按照最小二乘估计的思想, 使得条件残差平方和  $Q(\Phi)$  达到达到最小的  $\Phi = (\phi_1, \phi_2, \dots, \phi_p)$  的  $\Phi$  的取值  $\hat{\Phi} = (\hat{\phi}_1, \hat{\phi}_2, \dots, \hat{\phi}_p)$  就是待估参数的最小二乘估计值.

对于 AR(1) 模型而言, 有

$$Q(\Phi) = \sum_{t=2}^n \varepsilon_t^2 = \sum_{t=2}^n (x_t - \phi_1 x_{t-1})^2$$

根据极值原理, 令  $\frac{dQ(\Phi)}{d\phi_1} = 0$ , 解之得  $\hat{\phi}_1 = \frac{\sum_{t=2}^n x_t x_{t-1}}{\sum_{t=2}^n x_{t-1}^2}$

由于总体均值为零, 所以样本均值也可近似视为零. 将上式与  $\hat{\rho}(1)$  的估计式

$$\hat{\rho}(1) = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$



# 最小二乘估计

由于总体均值为零, 所以样本均值也可近似视为零. 将上式与  $\hat{\rho}(1)$  的估计式

$$\hat{\rho}(1) = \frac{\sum_{t=1}^{n-1} (x_t - \bar{x})(x_{t+1} - \bar{x})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

相对照, 仅仅分母中缺少了一项  $x_n^2$ . 而对于平稳序列来说,  $n$  较大时, 这个缺项可忽略. 因此, 可得  $\hat{\phi}_1 = \hat{\rho}(1)$ .

可见对大样本而言,  $\phi_1$  的最小二乘估计与矩估计是一致的.

对于AR(2) 模型, 我们有

$$Q(\Phi) = \sum_{t=3}^n \varepsilon_t^2 = \sum_{t=3}^n (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2})^2$$



# 最小二乘估计

令

$$Q(\Phi) = \sum_{t=3}^n \varepsilon_t^2 = \sum_{t=3}^n (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2})^2$$

$$\frac{\partial Q(\Phi)}{\partial \phi_1} = -2 \sum_{t=3}^n (x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2}) x_{t-1} = 0.$$

将上式写成  $\sum_{t=3}^n x_t x_{t-1} = \phi_1 \sum_{t=3}^n x_{t-1}^2 + \phi_2 \sum_{t=3}^n x_{t-1} x_{t-2}$

上式两边同时除以  $\sum_{t=3}^n x_t^2$ , 得

$$\frac{\sum_{t=3}^n x_t x_{t-1}}{\sum_{t=3}^n x_t^2} = \phi_1 \frac{\sum_{t=3}^n x_{t-1}^2}{\sum_{t=3}^n x_t^2} + \phi_2 \frac{\sum_{t=3}^n x_{t-1} x_{t-2}}{\sum_{t=3}^n x_t^2} \quad (*)$$



# 最小二乘估计

$$\frac{\sum_{t=3}^n x_t x_{t-1}}{\sum_{t=3}^n x_t^2} = \phi_1 \frac{\sum_{t=3}^n x_{t-1}^2}{\sum_{t=3}^n x_t^2} + \phi_2 \frac{\sum_{t=3}^n x_{t-1} x_{t-2}}{\sum_{t=3}^n x_t^2} \quad (*)$$

在样本容量较大,且平稳假设下, (\*) 式中三个分数非常接近  $\hat{\rho}(1)$ . 于是, 可近似得到如下等式

$$\hat{\rho}(1) = \phi_1 + \hat{\rho}(1)\phi_2 \quad (1)$$

按照同样的思路, 由  $\frac{\partial Q(\Phi)}{\partial \phi_2} = 0$ , 可推得

$$\hat{\rho}(2) = \hat{\rho}(1)\phi_1 + \phi_2 \quad (2)$$

可见, (1) 式和 (2) 式恰好是 AR(2) 模型的样本 Yule-Walker 方程. 求解后得未知参数的估计. 它与矩估计一致.



# 最小二乘估计

- 可以证明, 在一般的平稳 AR(p) 模型情况下, 可得出完全类似结论: 未知参数  $\Phi$  的条件最小二乘估计与其矩估计一样可以通过求解样本 Yule-Walker 方程 (3.4) 得到.
- **MA(q) 模型和 ARMA(p,q) 模型的最小二乘估计**

由于随机干扰项是不可观测的, 所以对于 MA(q) 模型和 ARMA(p,q) 模型不能直接实行最小二乘估计法. 下面我们简述其思想.

从 ARMA(p,q) 模型

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q},$$



# 最小二乘估计

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q}, \quad (3.20)$$

得到

$$\varepsilon_t = x_t - \sum_{i=1}^p \phi_i x_{t-i} + \sum_{k=1}^q \theta_k \varepsilon_{t-k} \quad (3.21)$$

利用模型 (3.20) 的逆转形式  $\varepsilon_t = \sum_{i=0}^{\infty} I_i x_{t-i}$

将  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$  代入 (3.21) 式中, 得

$$\varepsilon_t = x_t - \sum_{i=1}^p \phi_i x_{t-i} + \sum_{k=1}^q \theta_k \sum_{i=0}^{\infty} I_i x_{t-k-i} \quad (3.22)$$

通过数值优化算法, 即可得到相应估计.



# 极大似然估计

- 所谓**极大似然估计**,指的是建立在极大似然准则基础上的估计方法. **极大似然准则**认为,样本来自使得该样本出现概率最大的总体. 因此,未知参数的极大似然估计就是使得似然函数 (即联合密度函数) 达到最大的参数值.
- 使用极大似然估计必须提前知道总体的分布结构. 在时间序列分析中,序列总体的具体分布通常未知,为了便于分析和计算,一般假定序列服从多元正态分布.
- 考虑中心化 ARMA(p,q) 模型 (3.20).

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q},$$



# 极大似然估计

$$x_t = \phi_1 x_{t-1} + \dots + \phi_p x_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \dots - \theta_q \varepsilon_{t-q},$$

记  $\Theta = (\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q)^T$ ;

$$x = (x_1, x_2, \dots, x_n)^T;$$

$$\Sigma_n = E(xx^T) = \Omega \sigma_\varepsilon^2,$$

其中,

$$\Omega = \begin{pmatrix} \sum_{i=0}^{\infty} G_i^2 & \cdots & \sum_{i=0}^{\infty} G_i G_{i+n-1} \\ \vdots & \ddots & \vdots \\ \sum_{i=0}^{\infty} G_i G_{i+n-1} & \cdots & \sum_{i=0}^{\infty} G_i^2 \end{pmatrix}$$

# 极大似然估计

则似然函数为

$$\begin{aligned} L(\Theta; x) &= p(x_1, \dots, x_n; \Theta) = (2\pi)^{-\frac{n}{2}} |\Sigma_n|^{-\frac{1}{2}} \exp \left\{ -\frac{x^T \Sigma_n^{-1} x}{2} \right\} \\ &= (2\pi)^{-\frac{n}{2}} (\sigma_\varepsilon^2)^{-\frac{n}{2}} |\Omega|^{-\frac{1}{2}} \exp \left\{ -\frac{x^T \Omega^{-1} x}{2\sigma_\varepsilon^2} \right\}. \end{aligned}$$

令  $G(\Theta) = x^T \Omega^{-1} x$ , 则对数似然函数为

$$l(\Theta; x) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma_\varepsilon^2} G(\Theta)$$

根据极值原理, 对对数似然函数中的未知参数求偏导, 得到似然方程组



# 极大似然估计

令  $G(\Theta) = x^T \Omega^{-1} x$  ,则对数似然函数为

$$l(\Theta; x) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma_\varepsilon^2} G(\Theta)$$

根据极值原理, 对对数似然函数中的未知参数求偏导, 得到似然方程组

$$\begin{cases} \frac{\partial l(\Theta; x)}{\partial \sigma_\varepsilon^2} = -\frac{n}{2\sigma_\varepsilon^2} + \frac{G(\Theta)}{2\sigma_\varepsilon^4}; \\ \frac{\partial l(\Theta; x)}{\partial \Theta} = -\frac{1}{2} \frac{\partial \ln |\Omega|}{\partial \Theta} - \frac{1}{2\sigma_\varepsilon^2} \frac{\partial G(\Theta)}{\partial \Theta} \end{cases}$$

求解上述似然方程, 就可得到未知参数的极大似然估计.



# 实例

- 在 R 语言中, 参数估计可通过调用函数 `arima` 完成, 该函数的命令格式如下:

```
arima(x, order=, include.mean=, method= )
```

参数说明:

- `x`: 要进行模型拟合的序列名.
- `order`: 指定模型阶数. `order=c(p,d,q)`, 其中 `p` 为自回归阶数; `q` 为移动平均阶数; `d` 为差分阶数, 差分阶数后面章节才会用到, 在本章取 `d = 0`.



# 实例

`arima(x, order=, include.mean=, method= )`

- `include.mean`: 决定是否包含常数项. 如果 `include.mean=T`, 那么需要拟合常数项, 这是系统默认设置; 如果 `include.mean=F`, 那么意味着不需要拟合常数项.
- `method`: 指定参数估计方法. 如果 `method=“CSS-ML”`, 那么指定参数估计方法是条件最小二乘和极大似然估计混合方法, 这是系统默认设置; 如果 `method=“ML”`, 那么指定参数估计方法是极大似然估计法; 如果 `method=“CSS”`, 那么指定参数估计方法是条件最小二乘估计法.



# 实例

**例 3.4** 确定 2016 年 1 月至 2017 年 6 月青海省居民消费价格指数序列拟合模型的口径(即对该序列的未知参数进行估计).

**解** 根据例 3.1, 我们已经将模型识别为 AR(2). 现采用极大似然估计法估计未知参数. 具体命令及运行结果如下:

```
> x <-  
read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)  
  
> QHCPI <- ts(x$QHCPI,start=1)
```



# 实例

```
> QHCPI.fix <-  
arima(QHCPI,order=c(2,0,0),method="ML")
```

```
> QHCPI.fix
```

Call:

```
arima(x = QHCPI, order = c(2, 0, 0), method = "ML")
```

Coefficients:

	ar1	ar2	intercept
	1.0233	-0.5067	101.5002
s.e.	0.1975	0.1928	0.2245



# 实例

Coefficients:

	ar1	ar2	intercept
	1.0233	-0.5067	101.5002
s.e.	0.1975	0.1928	0.2245

sigma^2 estimated as 0.2123: log likelihood = -12.2, aic = 32.4

根据估计结果, 确定该模型的口径为

$$x_t - 101.5002 = \frac{\varepsilon_t}{1 - 1.0233B + 0.5067B^2}, \sigma_{\varepsilon}^2 = 0.2123.$$



# 实例

根据估计结果, 确定该模型的口径为

$$x_t - 101.5002 = \frac{\varepsilon_t}{1 - 1.0233B + 0.5067B^2}, \sigma_\varepsilon^2 = 0.2123.$$

也可写成如下形式:

$$x_t = 49.0652 + 1.0233x_{t-1} - 0.5067x_{t-2} + \varepsilon_t, \sigma_\varepsilon^2 = 0.2123.$$

**例 3.5** 确定 1956 年至 2016 年某城市各月份交通事故数序列拟合模型的口径.

**解** 根据例 3.2, 我们已经将模型识别为 MA(1). 现采用条件最小二乘法估计未知参数. 具体命令及运行结果如下:



# 实例

```
> x <-  
read.table("E:/DATA/CHAP3/SGS.csv",sep=";",hea  
der=T)  
  
> SGS <- ts(x$JTSGS,start=1956)  
  
> SGS.fix <-  
arima(SGS,order=c(0,0,1),method="CSS")  
  
> SGS.fix
```



# 实例

Call:

```
arima(x = SGS, order = c(0, 0, 1), method = "CSS")
```

Coefficients:

	ma1	intercept
	-0.5446	1000.0257
s.e.	0.1234	0.0508

sigma<sup>2</sup> estimated as 0.7223: part log likelihood = -75.38



# 实例

根据估计结果, 确定该模型的口径为

$$x_t = 1000.0257 + \varepsilon_t - 0.5446\varepsilon_{t-1}, \sigma_\varepsilon^2 = 0.7223$$

**例 3.6** 确定 1860 年至 1909 年国外某城市火灾发生数序列拟合模型的口径.

**解** 根据例 3.3, 我们已经将模型识别为 ARMA(2,1). 现采用条件最小二乘法估计未知参数. 具体命令及运行结果如下:

```
> x <-  
read.table("E:/DATA/CHAP3/2.csv",sep=";",header  
=T)
```





# 实例

```
> HUOZAI <- ts(x$X,start=1860)
```

```
> HUOZAI.fix <-  
arima(HUOZAI,order=c(2,0,1),method="CSS")
```

```
> HUOZAI.fix
```

Call:

```
arima(x = HUOZAI, order = c(2, 0, 1), method =  
"CSS")
```



# 实例

Coefficients:

	ar1	ar2	ma1	intercept
	0.5171	-0.0858	0.9268	1000.1156
s.e.	0.1449	0.1481	0.0857	0.4944

sigma^2 estimated as 0.9821: part log likelihood = -70.5

根据估计结果, 确定该模型的口径为

$$x_t = 1000.1156 + 0.5171x_{t-1} - 0.0858x_{t-2} + \varepsilon_t + 0.9268\varepsilon_{t-1}, \sigma_\varepsilon^2 = 0.9821.$$



# 本章结构

1. 自回归移动平均模型的识别

2. 参数估计

3. 模型的检验与优化

4. 序列的预测



# 最基本的特征

经过模型识别和参数估计之后, 接下来我们要对模型进行诊断性检验, 即检测已知观测数据在用既定模型拟合时的合理性. 一个好的拟合模型应该具备如下两个最基本的特征:

- (1) 拟合模型应该提取了观察值序列的几乎全部相关信息, 因而拟合残差项中将不再蕴含相关信息, 也即残差序列应该为白噪声序列.
- (2) 拟合模型应该是最精简的模型. 换句话说, 拟合模型不再含有任何冗余参数, 因为参数个数过多必然影响估计的精度.

在本节中, 我们就基于以上两点来讨论模型诊断方法. 然后, 利用模型诊断的结论, 提出改进模型的方法.



# 残差的检验

如果模型被正确识别, 参数估计足够精确, 那么残差应该具有白噪声的性质, 即残差序列应表现出独立、同分布、零均值和相同标准差的性质. 反之, 如果残差序列为非白噪声序列, 那就意味着残差序列中还残留着相关信息未被提取, 说明拟合模型不够有效, 需要重新选择其他模型进行拟合. 因此, 残差的检验指的就是残差序列的白噪声检验.

最简单的残差检验就是观察残差序列的时序图. 如果残差序列的时序图围绕横轴波动, 且波动范围有界, 但是波动既无趋势性, 也无周期性, 表现出较明显的随机性, 那么残差序列就可能为白噪声序列.

但是较为可靠的检验还是 1.5 节引入的白噪声检验.



# 残差的检验

原假设和备择假设分别为：

**原假设**  $H_0 : \rho(1) = \rho(2) = \cdots = \rho(m) = 0, \forall m \geq 1;$

**备择假设**  $H_1 : \text{至少存在某个 } \rho(k) \neq 0, \forall m \geq 1, k \leq m.$

**检验统计量取为**  $Q_{LB}$ ：

$$Q_{LB} = n(n+2) \sum_{k=1}^m \frac{\hat{\rho}^2(k)}{n-k} \sim \chi^2(m), \forall m > 0.$$

一般来讲，当检验的  $p$  值显著大于显著性水平 0.05，我们就不能拒绝原假设，也就是有理由相信原假设成立，认为序列是白噪声序列；当检验的  $p$  值显著小于显著性水平 0.05，我们就拒绝原假设，从而有理由相信备择假设成立，认为序列值之间有相关关系，该序列是非白噪声序列。



# 残差的检验

**例 3.7** 对 2016 年 1 月至 2017 年 6 月青海省居民消费价格指数序列拟合模型的残差序列进行检验 (也称为模型的显著性检验).

**解** 残差序列检验的具体命令及运行结果如下:

```
> x <-  
read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)  
  
> QHCPI <- ts(x$QHCPI,start=1)  
  
> QHCPI.fix <-  
arima(QHCPI,order=c(2,0,0),method="ML")
```



# 残差的检验

```
> for(i in 1:2) print(Box.test(QHCPI.fix$residual,  
type="Ljung-Box",+lag=6*i))
```

Box-Ljung test

data: QHCPI.fix\$residual

X-squared = 4.311, df = 6, p-value = 0.6347

Box-Ljung test

data: QHCPI.fix\$residual

X-squared = 10.258, df = 12, p-value = 0.5933

可见, 延迟 6 阶和延迟 12 阶的统计量 Q LB 的 p 值显著大于显著性水平 0.05, 所以可以认为拟合模型的残差序列是白噪声序列.





# 过度拟合检验

- 在模型诊断中, 需要特别引起注意的问题是, 由过度拟合而产生的参数冗余问题.
- 举例: 对于ARMA(p,q) 模型:  $\Phi(B)x_t = \Theta(B)\varepsilon_t$ , 如果在该模型两边同时乘以  $1 - cB$ , 可得  $(1 - cB)\Phi(B)x_t = (1 - cB)\Theta(B)\varepsilon_t$ . 从数学角度看, 两个模型仍然等同, 但实际上模型的待估参数增加了, 产生了冗余参数. 由于样本量有限, 参数个数的增加必然导致估计精度的下降. 因此, 有必要舍弃冗余参数, 精简模型.
- 为了舍弃对模型影响不显著的参数, 我们做如下参数显著性假设:

原假设  $H_0: \alpha_i = 0 \Leftrightarrow$  备择假设  $H_1: \alpha_i \neq 0$

其中,  $\alpha_i$  是模型第  $i$  个参数值.



# 过度拟合检验

- 原假设  $H_0: \alpha_i = 0 \Leftrightarrow$  备择假设  $H_1: \alpha_i \neq 0$

其中,  $\alpha_i$  是模型第  $i$  个参数值.

可以证明, 第  $i$  个参数  $\alpha_i$  的  $t$  检验统计量可按如下方式给出:

$$t = \hat{\phi}_i / \sigma$$

其中,  $\hat{\phi}_i$  为该参数的估计值;  $\sigma$  为该参数估计值的标准差.

- R 语言没有提供参数的显著性检验结果, 因为默认输出参数均显著非零. 如果想要得到  $t$  检验统计量的值和检验的  $p$  值, 就得动手算出  $t$  统计量的值, 并调用  $t$  分布的  $p$  值函数  $pt()$ . 该函数的命令格式如下:

```
pt(t, df=, lower.tail= )
```



# 过度拟合检验

- **例 3.8** 对 2016 年 1 月至 2017 年 6 月青海省居民消费价格指数序列拟合模型的参数进行显著性检验.

**解:** 参数的显著性检验的具体命令及运行结果如下:

```
> x <- read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)

> QHCPI <- ts(x$QHCPI,start=1)

> QHCPI.fix <- arima(QHCPI,order=c(2,0,0),method="ML")

> QHCPI.fix
```



# 过度拟合检验

- **例 3.8** 对 2016 年 1 月至 2017 年 6 月青海省居民消费价格指数序列拟合模型的参数进行显著性检验.

**解:** 参数的显著性检验的具体命令及运行结果如下:

```
> x <- read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)

> QHCPI <- ts(x$QHCPI,start=1)

> QHCPI.fix <- arima(QHCPI,order=c(2,0,0),method="ML")

> QHCPI.fix
```



# 过度拟合检验

- Call:

```
arima(x = QHCPI, order = c(2, 0, 0), method = "ML")
```

Coefficients:

ar1	ar2	intercept
1.0233	-0.5067	101.5002
s.e. 0.1975	0.1928	0.2245

sigma<sup>2</sup> estimated as 0.2123: log likelihood = -12.2, aic = 32.4

```
> # ar1 系数的显著性检验
```

```
> t1 <- 1.0233/0.1975
```



# 过度拟合检验

> # ar1 系数的显著性检验

> t1 <- 1.0233/0.1975; pt(t1,df=15,lower.tail=F)

[1] 5.584387e-05

> # ar2 系数的显著性检验

> t2 <- -0.5067/0.1928; pt(t2,df=15,lower.tail=T)

[1] 0.009501987

> # 常数的显著性检验

> t0 <- 101.5002/0.2245; pt(t0,df=15,lower.tail=F)

[1] 9.938772e-33



# 模型优化

三个系数检验的  $p$  值均小于 0.05, 故三个系数均显著非零.

## 模型优化

- 由于样本的随机性和定阶过程很大程度上依赖于分析人员的主观判断, 所以在模型识别时, 可能就会有若干个备选模型符合条件, 而且有时会出现多个模型都通过了检验, 那么现在的问题是, 到底哪个模型最有效呢? 下面, 我们介绍几个选择模型的方法.
- 信息准则法
  - (1) AIC 准则



# 模型优化

- 信息准则法

- (1) AIC 准则

AIC (Akaike information criterion) 准则是由日本统计学家 Akaike 于 1973 年提出的,它是基于最小信息量思想的准则.

从统计的观点来看,一个事件的发生如果给人们带来了信息,那么就应该认为该事件是一个随机事件.显然,一件为人们所完全预料的事件,不会给人们带来信息.假定 A 和 B 是两个随机事件,且  $P(A) > P(B)$ , 那么概率小的事件带给人们更多信息.因此,事件 B 的信息比事件 A 的信息多.





# 模型优化

- 一般地, 对于一个事件  $A$ , 我们可以用  $-\ln(p(A))$  来刻画一个随机事件  $A$  的信息量. 这里指出, 对于随机事件  $A$  来说,  $p(A)$  是其概率; 而对随机变量来说,  $p(\cdot)$  是该随机变量的概率密度函数.
- 基于上述信息量的考虑, AIC 准则建议评判一个拟合模型的优劣可以从如下两方面考察:
  - ① 似然函数值的大小. 似然函数值越大说明模型拟合的效果越好.
  - ② 模型中未知参数的个数. 模型中未知参数越多, 估计的难度就越大, 相应地, 估计的精度就越差.



# 模型优化

- 一个拟合模型应该是在兼顾考虑拟合精度和未知参数的个数下, 从中选择最优的配置. 基于此, AIC 函数被提出, 它是拟合精度和参数个数的加权函数:

$$\text{AIC} = -2\ln(\text{模型的极大似然函数值}) + 2(\text{模型中未知参数的个数})$$

- AIC 准则认为, 使得 AIC 函数达到最小的模型是最优模型.
- 考虑 ARMA(p,q) 模型. 从其对数似然函数

$$l(\Theta; x) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma_\varepsilon^2) - \frac{1}{2}\ln|\Omega| - \frac{1}{2\sigma_\varepsilon^2}G(\Theta)$$

# 模型优化

- 考虑 ARMA(p,q) 模型. 从其对数似然函数

$$l(\Theta; x) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma_\varepsilon^2) - \frac{1}{2} \ln |\Omega| - \frac{1}{2\sigma_\varepsilon^2} G(\Theta)$$

可以证明,  $l(\Theta; x) \propto -\frac{n}{2} \ln(\sigma_\varepsilon^2)$ . 因为中心化 ARMA(p,q) 模型的未知参数的个数为  $p + q + 1$ , 非中心化 ARMA(p,q) 模型的未知参数的个数为  $p+q + 2$ , 所以可得, 中心化 ARMA(p,q) 模型的 AIC 函数为

$$\text{AIC} = n \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q + 1).$$

非中心化 ARMA(p,q) 模型的 AIC 函数为

$$\text{AIC} = n \ln(\hat{\sigma}_\varepsilon^2) + 2(p + q + 2).$$



# 模型优化

## (2) BIC 准则

AIC 准则为模型选择提供了重要依据, 但是 AIC 准则也有不足之处. 理论上已经证明, AIC 准则不能给出模型阶的相合估计, 即当样本容量  $n$  趋于无穷大时, 由 AIC 准则确定的模型的阶不能收敛到模型的真实阶, 而是比模型真实阶偏高. 为了弥补 AIC 准则的不足, Akaike 于 1976 年提出了 BIC 准则. 他定义了如下 BIC 函数:

$$\text{BIC} = -2\ln(\text{模型的极大似然函数值}) \\ + \ln(n)(\text{模型中未知参数的个数})$$

- BIC 准则规定, 使得 BIC 函数达到最小的模型是最优模型.



# 模型优化

- BIC 函数与 AIC 函数相比较不同的地方是, BIC 函数将 AIC 函数中未知参数个数的权重由常数 2 变成了样本容量的对数  $\ln(n)$ . 理论上已经证明, BIC 准则确定的最优模型是真实阶数的相合估计.
- 根据 BIC 函数的定义, 容易得到中心化 ARMA(p,q) 模型的 BIC 函数为

$$\text{BIC} = n \ln(\hat{\sigma}_{\varepsilon}^2) + \ln(n)(p + q + 1).$$

非中心化 ARMA(p,q) 模型的 BIC 函数为

$$\text{BIC} = n \ln(\hat{\sigma}_{\varepsilon}^2) + \ln(n)(p + q + 2).$$

# 模型优化

- 在实际应用中, 我们总是从所有通过检验的模型中选择使得 AIC 或 BIC 函数达到最小的模型为相对最优模型. 这一过程将伴随模型定阶过程反复进行, 直至满意为止.
- 在 R 语言中, 通过调用程序包 forecast 中的函数 Airmo( ) 也可以进行参数估计, 其命令格式和参数使用方式与函数 airmo( ) 的一样. 所不同的是, 函数 Arima( ) 的运行结果会同时给出 AIC、BIC 和 AICc 的值. 例如: 在例 3.4 中, 如果用函数 Arima( ) 替换函数 arima( ) 进行参数估计, 那么可从运行结果中提取 AIC 的值为 32.39511, BIC 的值为 35.95659.

# 模型优化

- **例 3.9** 在例 3.1 中, 通过对 2016 年 1 月至 2017 年 6 月青海省居民消费价格指数序列的自相关函数图和偏自相关函数图的观察, 初步将序列模型识别为  $AR(2)$ , 然后在例 3.4 中进行了未知参数的估计, 最后在例 3.7 和例 3.8 中对模型进行了检验, 并且所建模型通过了检验.
- 我们也可以将序列识别为  $ARMA(2,1)$  模型, 然后对该模型进行估计, 并对估计结果进行检验.

```
> x <- read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)
```

```
> QHCPI <- ts(x$QHCPI,start=1)
```

```
> QHCPI.fix2 <- Arima(QHCPI,order=c(2,0,1),method="ML")
```



# 模型优化

- **例 3.9** 在例 3.1 中, 通过对 2016 年 1 月至 2017 年 6 月青海省居民消费价格指数序列的自相关函数图和偏自相关函数图的观察, 初步将序列模型识别为  $AR(2)$ , 然后在例 3.4 中进行了未知参数的估计, 最后在例 3.7 和例 3.8 中对模型进行了检验, 并且所建模型通过了检验.
- 我们也可以将序列识别为  $ARMA(2,1)$  模型, 然后对该模型进行估计, 并对估计结果进行检验.

```
> x <- read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)
```

```
> QHCPI <- ts(x$QHCPI,start=1)
```

```
> QHCPI.fix2 <- Arima(QHCPI,order=c(2,0,1),method="ML")
```





# 模型优化

- ```
> for(i in 1:2) print(Box.test(QHCPI.fix2$residual,  
+ type="Ljung-Box", lag=6*i))
```

Box-Ljung test

data: QHCPI.fix2\$residual

X-squared = 3.2717, df = 6, p-value = 0.7741

Box-Ljung test

data: QHCPI.fix2\$residual

X-squared = 8.1046, df = 12, p-value = 0.7769



# 模型优化

- 检验结果表明, 用模型  $\text{ARMA}(2,1)$  拟合序列, 依然通过了模型检验. 进一步, 提取 AIC 函数和 BIC 函数的值.

```
> QHCPI.fix2$aic
```

```
[1] 32.79913
```

```
> QHCPI.fix2$bic
```

```
[1] 37.25099
```

- 由此可以看出, AIC 函数和 BIC 函数的值比用  $\text{AR}(2)$  拟合序列的  $\text{AIC}=32.39511$  和  $\text{BIC}=35.95659$  值都大, 因此, 选用模型  $\text{AR}(2)$  拟合序列的结果更优.



# 模型优化

- 2. F 检验法

所谓 **F 检验法**, 就是通过比较 ARMA(p,q) 模型和 ARMA(p-1,q-1) 模型的残差平方和, 并用 F 检验判定阶数降低后的模型与原来模型之间是否存在显著性差异的方法.

- 设 ARMA(p,q) 模型和 ARMA(p-1,q-1) 模型的残差平方和分别为  $R_0$  和  $R_1$ ; 自由度分别为  $d_0$  和  $d_1$ , 则检验的假设可表示为

原假设  $H_0: \phi_p = 0$  且  $\theta_q = 0 \leftrightarrow$  备择假设  $H_1: \phi_p \neq 0$  或  $\theta_q \neq 0$ .

- 构造如下检验统计量:



# 模型优化

- 构造如下检验统计量:

$$F = \frac{R_1 - R_0}{d_1 - d_0} \bigg/ \frac{R_0}{n - p - (p + q + 1)}$$

原假设为真时, 统计量  $F$  服从第一自由度为  $d_1 - d_0$ , 第二自由度为  $n - p - (p + q + 1)$  的  $F(d_1 - d_0, n - p - (p + q + 1))$  的分布. 对于显著性水平  $\alpha$ , 可得到临界值  $F_\alpha$ .

- 当  $F > F_\alpha$  时拒绝原假设, 模型选为 ARMA(p, q) 模型更好些;
- 当  $F \leq F_\alpha$  时接受原假设, 意味着两个模型的拟合精度没有显著性差异, 模型选阶数更低的 ARMA(p - 1, q - 1) 模型更好些. 当然, 也有可能选 ARMA(p, q - 1) 模型或选 ARMA(p - 1, q) 模型更好些.



# 模型优化

## ● 3. 残差方差图法

拟合模型与真实数据之间的差异越小, 拟合模型就越有效. 残差是描述拟合模型与真实数据差异的重要方法, 因而也成为判断拟合模型是否合适的重要标准. 在实际建模时, 我们当然希望模型残差方差尽量小些, 因为残差方差越小说明拟合模型越有效. 经过证明, 模型的残差方差可以用下式来估计:

$$\hat{\sigma}^2 = \frac{\text{条件残差平方和}}{\text{实际观测值个数}-\text{模型的参数个数}}.$$

- 假设样本容量为  $n$ . 对于  $AR(p)$  模型来说, 模型中有滞后阶数为  $1, 2, \dots, p$  阶的项, 所以第一个有效观测值应该从  $p + 1$  期开始, 模型有效的样本容量为  $n - p$ , 估计的参数为  $p + 1$  个, 因而  $AR(p)$  模型的残差方差为

# 模型优化

$$\hat{\sigma}^2 = \frac{\text{条件残差平方和}}{n-p-(p+1)}.$$

- 对于 MA(q) 模型来说, 模型的有效观测值仍为 n 个, 有 q + 1 个待估参数, 因此, MA(q) 模型的残差方差为

$$\hat{\sigma}^2 = \frac{\text{条件残差平方和}}{n-(q+1)}.$$

- 对于 ARAR(p,q) 模型来说, 第一个有效观测值仍是从 p + 1 期开始的, 模型的待估参数有 p + q + 1 个, 因此, ARMAR(p,q) 模型的残差方差为

$$\hat{\sigma}^2 = \frac{\text{条件残差平方和}}{(n-p)-(p+q+1)}.$$

# 模型优化

- 在进行模型选择时, 通常会选择残差方差小的模型. 虽然增加模型的阶数会减少残差方差, 但是也会导致自由度的损失, 估计精度下降. 所以, 在进行模型选择时, 除了要使得残差方差尽可能小之外, 还需要使得模型尽可能精简. 在模型残差相差不大的情况下, 尽量选择阶数低的模型, 以免损失过多的自由度.
- **4. R 语言自动拟合法**

R 语言的 `auto.arima()` 函数提供了自动定阶、参数估计和计算信息量的功能, 可以帮助数据分析人员进行参考分析. 该函数的命令格式如下:

```
auto.arima(x, max.p=, max.q=, ic= )
```



# 模型优化

```
auto.arima(x, max.p=, max.q=, ic= )
```

该函数的参数说明:

- x: 需要建模的序列名.
- max.p: 自相关系数的最高阶数, 不特殊指定的话, 系统默认值为 5.
- max.q: 移动平均系数的最高阶数, 不特殊指定的话, 系统默认值为 5.
- ic: 指定信息量准则. ic 有 “aicc”, “aic”, “bic” 三个选项. 系统默认选 aic.





# 模型优化

- 使用 `auto.arima()` 函数之前, 需要先下载安装 `zoo` 和 `forecast` 两个程序包, 并用 `library` 调用它们. 需要指出的是, 有时选择不同的信息量准则会导致模型的阶数不同; 有时系统指定的模型阶数会高于真实阶数.
- **例 3.10** 使用函数 `auto.arima()` 对 2016 年 1 月至 2017 年 6 月宁夏回族自治区居民消费价格指数序列进行自动拟合  
**解** 自动拟合的具体命令及运行结果如下:

```
> x <- read.table("E:/DATA/CHAP3/cpi.csv",sep=";",header=T)
> NXCPI<- ts(x$NXCPI,start=1)
> NXCPI.fix <- auto.arima(NXCPI)
> NXCPI.fix
```



# 模型优化

Series: NXCPI

ARIMA(0,0,1) with non-zero mean

Coefficients:

ma1

mean

0.5277

101.4093

s.e. 0.1712

0.2267

sigma^2 estimated as 0.4619: log likelihood=-17.69

AIC=41.38 AICc=43.1 BIC=44.06

自动拟合表明, 软件程序自动识别为 MA(1) 模型.



# 本章结构

1. 自回归移动平均模型的识别

2. 参数估计

3. 模型的检验与优化

4. 序列的预测



# 预测准则

- 时间序列分析的最终目的之一是预测序列未来的变化、发展. 所谓**预测** (forecast), 就是根据现在与过去的随机序列的样本取值, 对未来某个时刻序列值进行估计. 目前, 许多预测方法都是从线性预测的理论中发展而来的. 对于平稳序列来讲, 最常用的预测方法是线性最小方差预测. 线性是指预测值是观测值的线性函数, 最小方差是指预测的均方误差达到最小.

- 预测准则**

设  $x_t, x_{t-1}, \dots, x_{t-n}$  ( $n$  可以有限, 也可以无限) 是时间序列  $\{x_t\}$  的观测值, 也称为该序列的**历史信息**, 简记  $\Theta_t$ , 对将来某时刻  $t+l$  ( $l > 0$ ) 的序列值  $x_{t+l}$  进行估计, 称为序列的第  $l$  步预测, 预测值记为  $\hat{x}_{t+l}$ .



# 预测准则

- 显然, 预测值  $\hat{x}_{t+l}$  应该是关于历史信息  $\Theta_t$  的函数, 因此, 预测实质上就是求一个函数  $f$ , 使得

$$\hat{x}_{t+l} = f(\Theta_t) .$$

习惯上, 称上述函数为**预测函数**.

预测的准确程度是由预测误差

$$e_t(l) = x_{t+l} - \hat{x}_{t+l}$$

决定的, 因此预测函数的选取应该使得预测误差尽可能地小. 于是, 需要确定一种准则, 使得依据这种准则能够衡量采用某种预测函数所得的预测误差比采用其他预测函数所得的预测误差小.



# 预测准则

- 在统计上, 我们一般用均方误差来衡量一个估计量的好坏. 所谓**均方误差** (mean squared error), 就是误差平方的期望, 即

$$E[e_t(l)]^2 = E(x_{t+l} - \hat{x}_{t+l})^2 = E[x_{t+l} - f(\Theta_t)]^2 \quad (3.26)$$

一般地, 均方误差越小, 估计量越准确, 因此, (3.26) 式可选作我们的预测准则. 下面来求在这个预测准则下的最佳预测, 也称为**最小均方误差 (minimum mean square error) 预测**.

- 设  $f(\Theta_t)$  为  $x_{t+l}$  的任一预测, 根据 (3.26) 式得

$$E[x_{t+l} - f(\Theta_t)]^2 = E[x_{t+l} - E(x_{t+l} | \Theta_t) + E(x_{t+l} | \Theta_t) - f(\Theta_t)]^2$$



# 预测准则

- 设  $f(\Theta_t)$  为  $x_{t+l}$  的任一预测, 根据 (3.26) 式得

$$\begin{aligned} E[x_{t+l} - f(\Theta_t)]^2 &= E[x_{t+l} - E(x_{t+l} | \Theta_t) + E(x_{t+l} | \Theta_t) - f(\Theta_t)]^2 \\ &= E[x_{t+l} - E(x_{t+l} | \Theta_t)]^2 \\ &\quad + 2E\{[x_{t+l} - E(x_{t+l} | \Theta_t)][E(x_{t+l} | \Theta_t) - f(\Theta_t)]\} \\ &\quad + E\{[E(x_{t+l} | \Theta_t) - f(\Theta_t)]^2\} \end{aligned} \quad (3.27)$$

记

$$\eta_{t+l} = [x_{t+l} - E(x_{t+l} | \Theta_t)][E(x_{t+l} | \Theta_t) - f(\Theta_t)]$$

则有  $E(\eta_{t+l} | \Theta_t) = [E(x_{t+l} | \Theta_t) - f(\Theta_t)]E\{[x_{t+l} - E(x_{t+l} | \Theta_t)] | \Theta_t\} = 0$ .  
(3.28)

进而得到  $E(\eta_{t+l}) = E[E(\eta_{t+l} | \Theta_t)] = 0$ .

将 (3.28) 式代入 (3.27) 式得



# 预测准则

$$\begin{aligned} E[x_{t+l} - f(\Theta_t)]^2 &= E[x_{t+l} - E(x_{t+l} | \Theta_t)]^2 + E\{[E(x_{t+l} | \Theta_t) - f(\Theta_t)]^2\} \\ &\geq E[x_{t+l} - E(x_{t+l} | \Theta_t)]^2 \end{aligned}$$

于是得到最小均方误差预测:

$$f(\Theta_t) = E(x_{t+l} | \Theta_t) \quad (3.29)$$

可见, 最小均方误差预测是  $x_{t+l}$  关于  $\Theta_t$  的条件期望. 这种预测具有许多优良的性质, 但是计算较为复杂.

事实上, 对于平稳序列来讲, 我们更感兴趣在  $\Theta_t$  的线性函数类中寻求  $x_{t+l}$  的最佳预测, 换句话说, 就是寻找  $x_t, x_{t-1}, \dots, x_{t-n}$  ( $n$  可以有限, 也可以无限) 的线性函数

$$\hat{x}_{t+l} = f(\Theta_t) = \alpha^T \Theta_t = \alpha_0 x_t + \alpha_1 x_{t-1} + \dots,$$





# 预测准则

使得 (3.26) 式达到最小, 也即在  $\Theta_t$  张成的线性空间  $\text{span}\{\Theta_t\}$  中, 寻找使得 (3.26) 式达到最小的  $\Theta_t$  的线性组合. 我们称这种预测为**线性最小方差 (linear minimum variance) 预测**.

现在我们来寻找  $x_{t+l}$  的线性最小方差预测. 为此, 首先引入投影的概念. 对于线性空间  $\text{span}\{\Theta_t\}$  中的元素  $\hat{x}'_{t+l} = \alpha^T \Theta_t$  来说, 如果满足

$$E[(x_{t+l} - \alpha^T \Theta_t) \Theta_t^T] = 0^T, \quad (3.30)$$

那么我们称  $\hat{x}'_{t+l}$  为  $x_{t+l}$  在  $\text{span}\{\Theta_t\}$  中的**投影 (projection)**.

下面我们证明,  $x_{t+l}$  在  $\text{span}\{\Theta_t\}$  中的投影  $\hat{x}'_{t+l}$  就是我们想要寻找的  $x_{t+l}$  的线性最小方差预测.



# 预测准则

- 设  $\beta^T \Theta_t \in \text{span}\{\Theta_t\}$  是  $x_{t+l}$  的任意一个预测, 则

$$\begin{aligned} E(x_{t+l} - \beta^T \Theta_t)^2 &= E(x_{t+l} - \alpha^T \Theta_t + \alpha^T \Theta_t - \beta^T \Theta_t)^2 \\ &= E(x_{t+l} - \alpha^T \Theta_t)^2 + 2E[(x_{t+l} - \alpha^T \Theta_t)(\alpha^T \Theta_t - \beta^T \Theta_t)] \\ &\quad + E(\alpha^T \Theta_t - \beta^T \Theta_t)^2 \end{aligned} \quad (3.31)$$

(3.31) 式右边中间项为

$$E[(x_{t+l} - \alpha^T \Theta_t)(\alpha - \beta)^T \Theta_t] = E\{[(x_{t+l} - \alpha^T \Theta_t)\Theta_t^T](\alpha - \beta)\} = 0. \quad (3.32)$$

将(3.32)式代入(3.31)式中, 得

$$E(x_{t+l} - \beta^T \Theta_t)^2 = E(x_{t+l} - \alpha^T \Theta_t)^2 + E(\alpha^T \Theta_t - \beta^T \Theta_t)^2 \quad (3.33)$$

从(3.33)式可以看到, 当  $\alpha^T \Theta_t = \beta^T \Theta_t$  时,  $E(x_{t+l} - \beta^T \Theta_t)^2$  达到最小值, 所以  $x_{t+l}$  在  $\text{span}\{\Theta_t\}$  中的投影  $\hat{x}'_{t+l} = \alpha^T \Theta_t$  为  $x_{t+l}$  的线性最小方差预测.



# 自回归移动平均模型的预测

- 上节中, 我们讨论了一般序列预测的思想. 对于 ARMA(p,q) 模型来讲, 通过其传递形式和逆转形式可以更容易求出它的线性最小方差预测值.
- 1. 线性最小方差预测

设平稳可逆的 ARMA(p,q) 模型的传递形式为

$$x_t = \sum_{i=0}^{\infty} G_i \varepsilon_{t-i} \quad (3.34)$$

其中,  $\{G_i\}$  是 Green 函数. 假如我们具有一直到  $t$  期的  $\varepsilon$  观测值  $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$ , 则根据(3.29) 式知,  $x_{t+l}$  的线性最小方差预测  $\hat{x}_{t+l}$  具有如下形式:

$$\hat{x}_{t+l} = E(x_{t+l} | \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sum_{i=0}^{\infty} G_{l+i} \varepsilon_{t-i} \quad (3.35)$$



# 自回归移动平均模型的预测

- 因为预测误差

$$x_{t+l} - \hat{x}_{t+l} = x_{t+l} - E(x_{t+l} | \varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \varepsilon_{t+l} + G_1 \varepsilon_{t+l-1} + \dots + G_{l-1} \varepsilon_{t+1}$$

与  $\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  满足  $E[(x_{t+l} - \hat{x}_{t+l}) \varepsilon_{t-k}] = 0, k = 0, 1, 2, \dots$ , 所以由 (3.30) 式知  $\hat{x}_{t+l}$  是  $x_{t+l}$  在线性空间  $\text{span}\{\varepsilon_t, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots\}$  上的投影, 故  $\hat{x}_{t+l}$  是  $x_{t+l}$  的线性最小方差预测. 此时, 预测误差的均值和方差分别为

$$E[e_t(l)] = E(x_{t+l} - \hat{x}_{t+l}) = 0$$

和

$$E[e_t(l)]^2 = E(x_{t+l} - \hat{x}_{t+l})^2 = \sum_{i=0}^{l-1} G_i^2 \sigma_\varepsilon^2.$$

再设平稳可逆的 ARMA(p,q) 模型的逆转形式为



# 自回归移动平均模型的预测

- 再设平稳可逆的 ARMA(p,q) 模型的逆转形式为  $\varepsilon_t = \sum_{j=0}^{\infty} I_j x_{t-j}$   
其中,  $\{I_j\}$  是逆转函数. 于是

$$\hat{x}_{t+l} = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} G_{l+i} I_j x_{t-i-j}. \quad (3.37)$$

将 (3.37) 式简记为

$$\hat{x}_{t+l} = \sum_{k=0}^{\infty} C_{l+k} x_{t-k}.$$

由 (3.34) 式得

$$x_{t+l} = \sum_{i=0}^{\infty} G_i \varepsilon_{t+l-i} = \sum_{i=0}^{l-1} G_i \varepsilon_{t+l-i} + \sum_{i=l}^{\infty} G_i \varepsilon_{t+l-i} = e_t(l) + \hat{x}_{t+l}.$$

于是可推得

$$E(x_{t+l} | x_t, x_{t-1}, \dots) = E[e_t(l) | x_t, x_{t-1}, \dots] + E[\hat{x}_{t+l} | x_t, x_{t-1}, \dots] = \hat{x}_{t+l} \quad (3.38)$$

和 
$$\text{Var}(x_{t+l} | x_t, x_{t-1}, \dots) = \text{Var}(e_t(l)) \quad (3.39)$$



# 自回归移动平均模型的预测

(3.38) 式再次说明  $\hat{x}_{t+l}$  是  $x_{t+l}$  的线性最小方差预测. (3.39) 说明在此预测下的方差只与预测步长  $l$  有关, 而与预测起始点  $t$  无关. 预测步长  $l$  越大, 预测值的方差也越大, 因而为了保证预测精度, 时间序列数据只适合做短期预测.

在正态假设下, 有

$$x_{t+l} | x_t, x_{t-1}, \dots \sim N(\hat{x}_{t+l}, \text{Var}[e_t(l)]),$$

因而,  $x_{t+l} | x_t, x_{t-1}, \dots$  的置信水平为  $1-\alpha$  的置信区间为

$$(\hat{x}_{t+l} - z_{1-\alpha/2} \cdot \sigma_\varepsilon^2 \sum_{i=0}^{l-1} G_i^2, \hat{x}_{t+l} + z_{1-\alpha/2} \cdot \sigma_\varepsilon^2 \sum_{i=0}^{l-1} G_i^2)$$



# 自回归移动平均模型的预测

- 2. AR(p) 模型的预测

设  $\{x_t, t \in T\}$  是AR(p) 模型, 则由(3.38) 式得

$$\begin{aligned}\hat{x}_{t+l} &= E(x_{t+l} \mid x_t, x_{t-1}, \dots) \\ &= E(\phi_1 x_{t+l-1} + \dots + \phi_p x_{t+l-p} + \varepsilon_{t+l} \mid x_t, x_{t-1}, \dots) \\ &= \phi_1 \hat{x}_{t+l-1} + \dots + \phi_p \hat{x}_{t+l-p}\end{aligned}$$

其中

$$\hat{x}_{t+i} = \begin{cases} \hat{x}_{t+i}, & i \geq 1; \\ x_{t+i}, & i \leq 0. \end{cases}$$

预测方差为

$$\text{Var}[e_t(l)] = E[e_t(l)]^2 = E(x_{t+l} - \hat{x}_{t+l})^2 = \sum_{i=0}^{l-1} G_i^2 \sigma_\varepsilon^2. \quad l \geq 1$$



# 自回归移动平均模型的预测

- 3. MA(q) 模型的预测

设  $\{x_t, t \in T\}$  是 MA(q) 模型, 则由 (3.35) 式和 (3.38) 式知, 在条件  $x_t, x_{t-1}, \dots$  下,  $x_{t+l}$  的预测值等于在条件  $\varepsilon_t, \varepsilon_{t-1}, \dots$  下,  $x_{t+l}$  的预测值. 而未来时刻的随机扰动  $\varepsilon_{t+1}, \varepsilon_{t+2}, \dots$  是不可观测的, 属于预测误差.

当预测步长  $l \geq 1$  小于等于 MA(q) 模型的阶数  $q$  时,  $x_{t+l}$  可分解为

$$x_{t+l} = \mu + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \dots - \theta_q \varepsilon_{t+l-q}$$

$$\begin{aligned} &= (\varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \dots - \theta_l \varepsilon_{t+1}) + (\mu - \theta_l \varepsilon_t - \dots - \theta_q \varepsilon_{t+l-q}) \\ &= e_t(l) + \hat{x}_{t+1}. \end{aligned}$$

当预测步长  $l \geq 1$  大于 MA(q) 模型的阶数  $q$  时,  $x_{t+l}$  可写为

$$\begin{aligned} x_{t+l} &= (\varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \dots - \theta_q \varepsilon_{t+l-q}) + \mu \\ &= e_t(l) + \hat{x}_{t+1}. \end{aligned}$$





# 自回归移动平均模型的预测

- 于是, MA(q) 模型  $l \geq 1$  的预测值为

$$\hat{x}_{t+l} = \begin{cases} \mu - \theta_l \varepsilon_t - \cdots - \theta_q \varepsilon_{t+l-q}, & l \leq q; \\ \mu, & l > q. \end{cases}$$

可见, MA(q) 模型只能预测  $q$  步之内的序列值.  $q$  步之外的序列值都是 0.

$$\text{预测方差为 } \text{Var}[e_t(l)] = \begin{cases} \sigma_\varepsilon^2 (1 + \theta_1^2 + \cdots + \theta_{l-1}^2), & l \leq q; \\ \sigma_\varepsilon^2 (1 + \theta_1^2 + \cdots + \theta_q^2), & l > q. \end{cases}$$

- 4. ARMA(p,q) 模型的预测

设  $\{x_t, t \in T\}$  是 ARMA(p,q) 模型, 则由 (3.38) 式得

# 自回归移动平均模型的预测

$$\begin{aligned}\hat{x}_{t+l} &= E(x_{t+l} | x_t, x_{t-1}, \dots) \\ &= E(\phi_1 x_{t+l-1} + \dots + \phi_p x_{t+l-p} + \varepsilon_{t+l} - \theta_1 \varepsilon_{t+l-1} - \dots - \theta_q \varepsilon_{t+l-q} | x_t, x_{t-1}, \dots) \\ &= \begin{cases} \sum_{k=1}^p \phi_k \hat{x}_{t+l-k} - \sum_{i=1}^q \theta_i \varepsilon_{t+l-i}, & l \leq q; \\ \sum_{k=1}^p \phi_k \hat{x}_{t+l-k}, & l > q. \end{cases}\end{aligned}$$

其中  $\hat{x}_{t+i} = \begin{cases} \hat{x}_{t+i}, & i \geq 1; \\ x_{t+i}, & i \leq 0. \end{cases}$

预测方差为

$$\text{Var}[e_t(l)] = E[e_t(l)]^2 = \sum_{i=0}^{l-1} G_i^2 \sigma_\varepsilon^2, \quad l \geq 1.$$



# 自回归移动平均模型的预测

**例 3.11** 假定 Acme 公司的年销售额 (单位: 百万美元) 符合 AR(2) 模型:

$$x_t = 5 + 0.6x_{t-1} + 0.3x_{t-2} + \varepsilon_t, \varepsilon_t \sim N(0, 2).$$

已知 2005 年、2006 年和 2007 年的销售额分别是 900 万、1100 万和 1000 万美元, 那么

- (1) 预测 2008 年和 2009 年的销售额;
- (2) 确定 2008 年和 2009 年的销售额的 95% 的置信区间.

**解:** (1) 计算预测值:

2008 年的销售额:  $\hat{x}_{2008} = 5 + 0.6x_{2007} + 0.3x_{2006} = 935;$

2009 年的销售额:  $\hat{x}_{2009} = 5 + 0.6\hat{x}_{2008} + 0.3x_{2007} = 866.$



# 自回归移动平均模型的预测

解: (1) 计算预测值:

$$2008 \text{ 年的销售额: } \hat{x}_{2008} = 5 + 0.6x_{2007} + 0.3x_{2006} = 935;$$

$$2009 \text{ 年的销售额: } \hat{x}_{2009} = 5 + 0.6\hat{x}_{2008} + 0.3x_{2007} = 866.$$

(2) 确定置信区间:

根据 Green 函数的递推计算得

$$G_0 = 1; G_1 = \phi_1 G_0 = 0.6; G_2 = \phi_1 G_1 + \phi_2 G_0 = 0.66.$$

由预测方差公式得

$$\text{Var}[e_{2007}(1)] = G_0^2 \sigma_\varepsilon^2 = 2;$$

$$\text{Var}[e_{2007}(2)] = (G_0^2 + G_1^2) \sigma_\varepsilon^2 = 2.72.$$

于是得 2008 年销售额的 95% 的置信区间为



# 自回归移动平均模型的预测

于是得 2008 年销售额的 95% 的置信区间为

$$(935 - 1.96\sqrt{2}, 935 + 1.96\sqrt{2}) = (932.23, 937.77);$$

2009 年销售额的 95% 的置信区间为

$$(866 - 1.96\sqrt{2.72}, 866 + 1.96\sqrt{2.72}) = (862.77, 869.23).$$

**例 3.12** 假定 Deere 公司生产的某零件月不合格率符合 ARMA(1,1) 模型:

$$x_t = 0.5x_{t-1} + \varepsilon_t - 0.25\varepsilon_{t-1}, \varepsilon_t \sim N(0, 0.2).$$

已知一月份的不合格率为 6%, 误差为 0.015, 请预测未来 3 个月不合格率的 95% 的置信区间.

**解:** 未来 3 个月不合格率的预测值:



# 自回归移动平均模型的预测

解：未来 3 个月不合格率的预测值：

$$2 \text{ 月份不合格率: } \hat{x}_2 = 0.5x_1 - 0.25\varepsilon_1 = 0.02625;$$

$$3 \text{ 月份不合格率: } \hat{x}_3 = 0.5\hat{x}_2 = 0.013125;$$

$$4 \text{ 月份不合格率: } \hat{x}_4 = 0.5\hat{x}_3 = 0.0065625.$$

根据 Green 函数的递推计算得

$$G_0 = 1; G_1 = \phi_1 G_0 - \theta_1 = 0.25; G_2 = \phi_1 G_1 = 0.125.$$

由预测方差公式得

$$\text{Var}[e_1(1)] = G_0^2 \sigma_\varepsilon^2 = 0.2;$$

$$\text{Var}[e_1(2)] = (G_0^2 + G_1^2) \sigma_\varepsilon^2 = 0.2125;$$

$$\text{Var}[e_1(3)] = (G_0^2 + G_1^2 + G_2^2) \sigma_\varepsilon^2 = 0.215625.$$



# 自回归移动平均模型的预测

于是得 2 月份不合格率的 95% 的置信区间为

$$(0.02625 - 1.96\sqrt{0.2}, 0.02625 + 1.96\sqrt{0.2}) = (-0.8503, 0.9028);$$

3 月份不合格率的 95% 的置信区间为

$$(0.013125 - 1.96\sqrt{0.2125}, 0.013125 + 1.96\sqrt{0.2125}) = (-0.8904, 0.9166);$$

4 月份不合格率的 95% 的置信区间为

$$(0.0065625 - 1.96\sqrt{0.215625}, 0.0065625 + 1.96\sqrt{0.215625}) = (-0.9034, 0.9177).$$

- 在 R 中, 我们可以通过调用函数 `forecast()` 来进行预测. 在应用该命令之前, 首先应该下载安装 `forecast` 程序包, 并加载. `forecast()` 函数的命令格式如下:

`forecast(object, h=, level= )`



# 自回归移动平均模型的预测

`forecast(object, h=, level= )`

该函数的参数说明:

- object: 拟合信息文件名.
- h: 预测期数.
- level: 置信区间的置信水平. 不特殊指定的话, 系统会自动给出置信水平分别为 80% 和 95% 的双层置信区间.

**例 3.13** 使用在例 3.4 中所建立的模型, 预测 2017 年 7 月至 2017 年 12 月青海省居民的消费价格指数.

**解:** 利用例 3.4 所建立的 AR(2) 模型, 进行未来 6 期的预测. 预测结果中自动给出了置信水平分别为 80% 和 95% 的置信区间. 具体命令及运行结果如下:





# 自回归移动平均模型的预测

```
> QHCPI.fore <- forecast(QHCPI.fix,h=6)
```

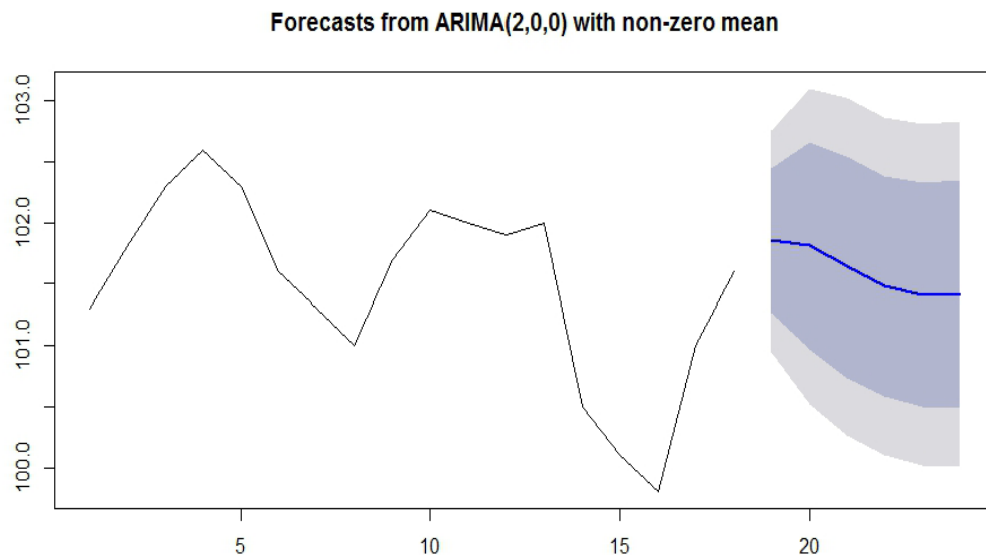
```
> QHCPI.fore
```

|    | Point Forecast | Lo 80    | Hi 80    | Lo 95    | Hi 95    |
|----|----------------|----------|----------|----------|----------|
| 19 | 101.8557       | 101.2653 | 102.4462 | 100.9527 | 102.7587 |
| 20 | 101.8134       | 100.9687 | 102.6582 | 100.5214 | 103.1055 |
| 21 | 101.6406       | 100.7375 | 102.5437 | 100.2595 | 103.0217 |
| 22 | 101.4851       | 100.5818 | 102.3884 | 100.1036 | 102.8666 |
| 23 | 101.4136       | 100.4994 | 102.3278 | 100.0155 | 102.8118 |
| 24 | 101.4192       | 100.4921 | 102.3464 | 100.0012 | 102.8372 |



# 自回归移动平均模型的预测

利用预测结果,可以直接绘制预测图进行观察.



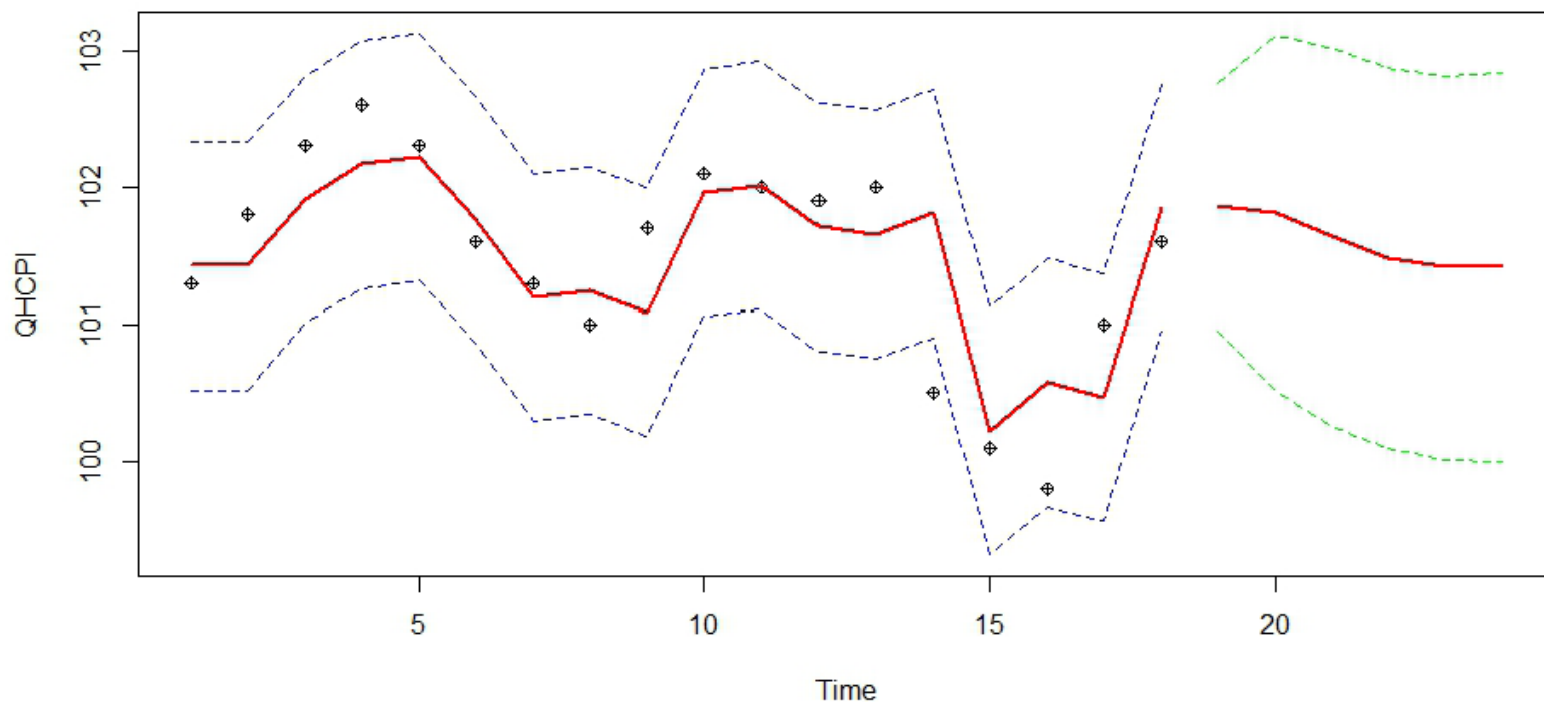
也可以绘制个性化的预测图. 具体命令如下:

# 自回归移动平均模型的预测

```
> Q1 <- QHCPI.fore$fitted-1.96*sqrt(QHCPI.fix$sigma2)
> Q2 <- QHCPI.fore$fitted+1.96*sqrt(QHCPI.fix$sigma2)
> B1 <- ts(QHCPI.fore$lower[,2],start=19)
> B2 <- ts(QHCPI.fore$upper[,2],start=19)
> J1 <- min(QHCPI,Q1,B1)
> J2 <- max(QHCPI,Q2,B2)
> plot(QHCPI,type="p",pch=10,xlim=c(1,24),ylim=c(J1,J2))
> lines(QHCPI.fore$fitted,col=2,lwd=2)
> lines(QHCPI.fore$mean,col=2,lwd=2)
> lines(Q1,col="blue",lty=2)
> lines(Q2,col="blue",lty=2)
> lines(B1,col="green",lty=2)
> lines(B2,col="green",lty=2)
```



# 自回归移动平均模型的预测



散点图为观察值序列, 实线是拟合值, 虚线是 95% 的置信线.