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

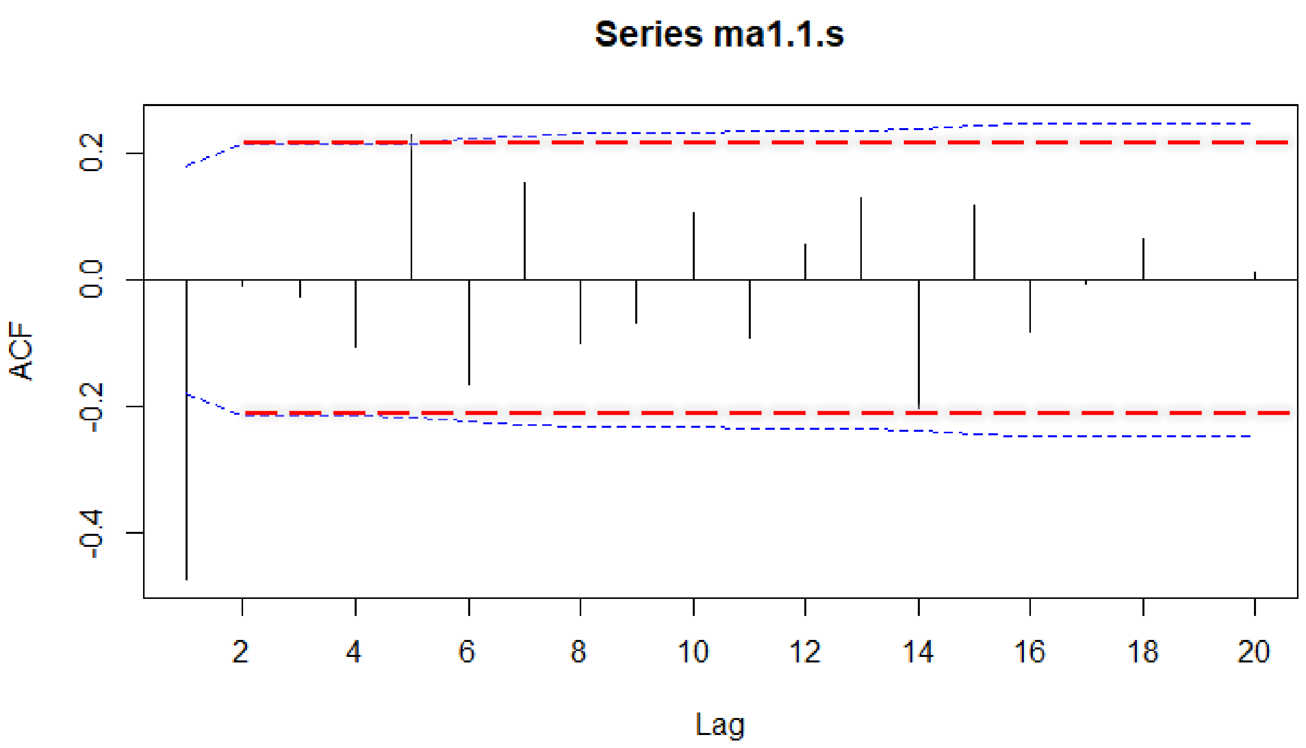
MA模型识别

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

- 对于白噪声(MA(0))， 当$k > 0$时，$Var(r_k) \approx \frac{1}{n}$
- 对于MA(q)，当$k > q$时， $Var(r_k) \approx \frac{1}{n}\left[1 + 2\sum_{j=1}^{q}\rho_j^2\right]$
- r_k在正负2倍的标准差$\pm 2\hat{s}_q$之间的概率约为0.95，其中

  $\hat{s}_q = \sqrt{\frac{1}{n}\left[1 + 2\sum_{j=1}^{q}r_j^2\right]}$
- 先判断序列是否为白噪声，再依次判断是否为MA(1), MA(2), ……

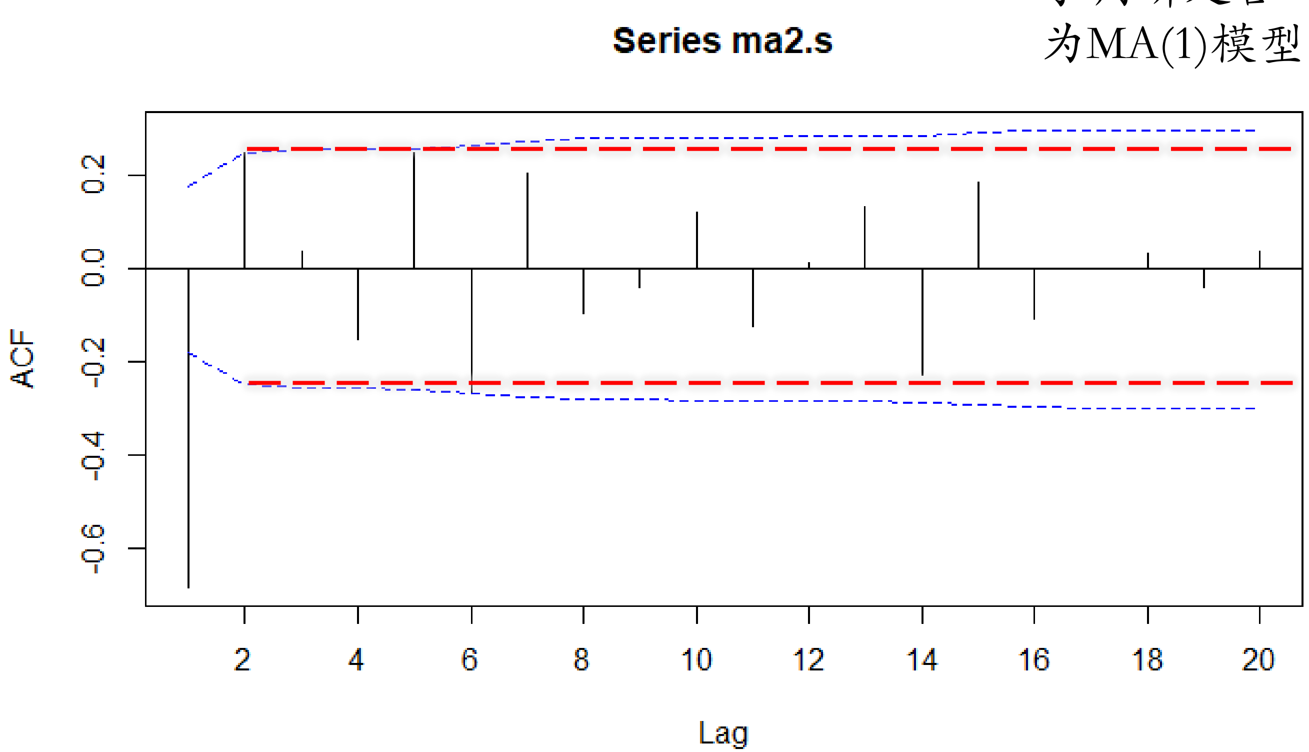
MA(1)模型识别范例




```
> acf(ma1.1.s,ci.type='ma',xaxp=c(0,20,10))
```

MA(2)模型识别范例

该红线是用
于判断是否
为MA(1)模型




```
> acf(ma2.s,ci.type='ma',xaxp=c(0,20,10))
```

偏自相关函数（PACF）

- 对于任意平稳序列，已知$\rho_1, \rho_2, \dots$，对于给定的k，考虑方程组

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

- 我们把 $\{\phi_{kk}\}$称为 偏自相关函数（PACF）序列，有如下重要结论：
- 对于AR(p)过程，当$k \geq p$时，

$$\phi_{kj} = \begin{cases} \phi_j & j = 1, \dots, p \\ 0 & j = p+1, \dots, k \end{cases}$$

- 特别地，$\phi_{pp} = \phi_p \neq 0, \phi_{kk} = 0, k > p$, 即AR(p)过程的偏自相关函数序列$p$阶截尾。

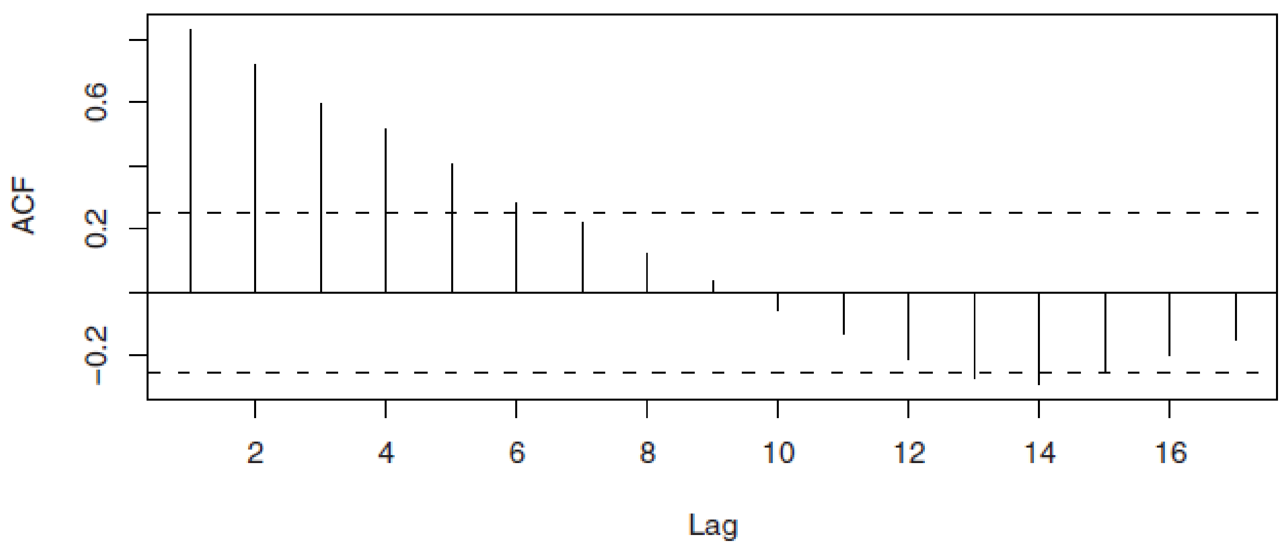
样本偏自相关函数

- 前面所定义的偏自相关函数都是基于序列的性质，由其内在的自相关函数计算得出（与自相关函数类似，是模型本身的参数）；当我们有具体样本数值时，可以用样本自相关函数r_k代替ρ_k来计算，例如求解方程组、递推式等，记为$\hat{\phi}_{kk}$.
- 由于AR(p)过程的PACF序列p阶截尾，当$k > p$时，$\phi_{kk} = 0$, 所以$\hat{\phi}_{kk}$应该也接近0，实际上，Quenoulle (1949) 证明了对于AR(p)过程，当$k > p$时，$\hat{\phi}_{kk}$近似服从均值为0，方差为$1/n$的正态分布，所以其在正负两倍标准差$\pm 2/\sqrt{n}$之间的概率约为0.95.

$$P\left(-\frac{2}{\sqrt{n}} \le \hat{\phi}_{kk} \le \frac{2}{\sqrt{n}}\right) \approx 0.95$$

AR(1)模型识别范例

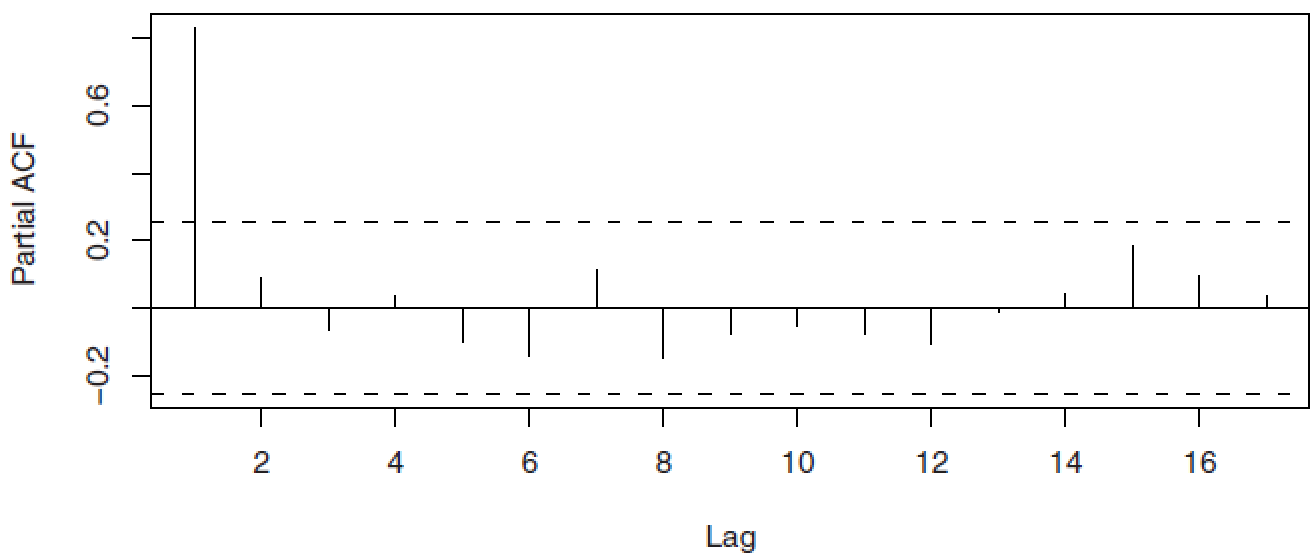
Exhibit 6.10 Sample ACF for an AR(1) Process with $\phi = 0.9$




```
> data(ar1.s); acf(ar1.s,xaxp=c(0,20,10))
```

AR(1)模型识别范例

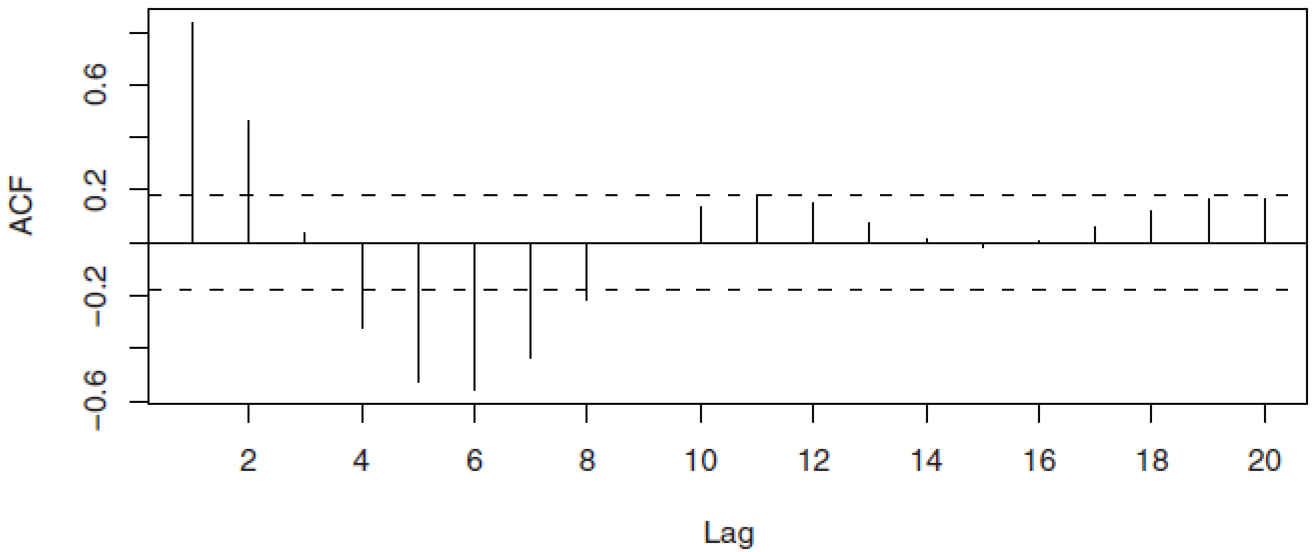
Exhibit 6.11 Sample Partial ACF for an AR(1) Process with $\phi = 0.9$




```
> pacf(ar1.s,xaxp=c(0,20,10))
```

AR(2)模型识别范例

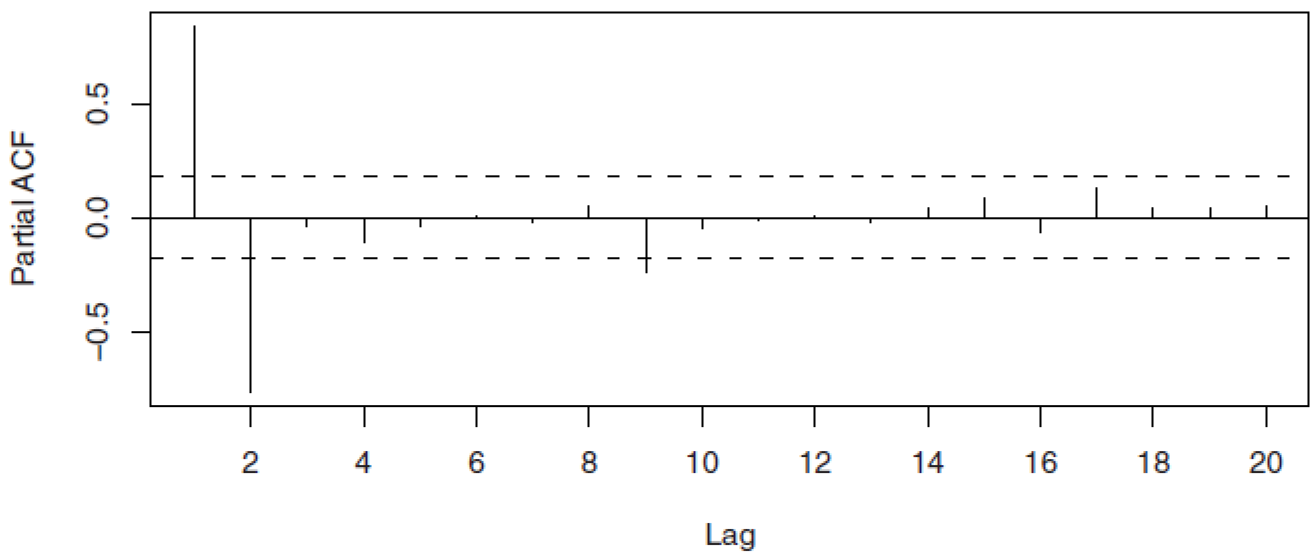
Exhibit 6.12 Sample ACF for an AR(2) Process with $\phi_1 = 1.5$ and $\phi_2 = -0.75$




```
> acf(ar2.s,xaxp=c(0,20,10))
```

AR(2)模型识别范例

Exhibit 6.13 Sample PACF for an AR(2) Process with ϕ_1 = 1.5 and ϕ_2 = −0.75




```
> pacf(ar2.s,xaxp=c(0,20,10))
```

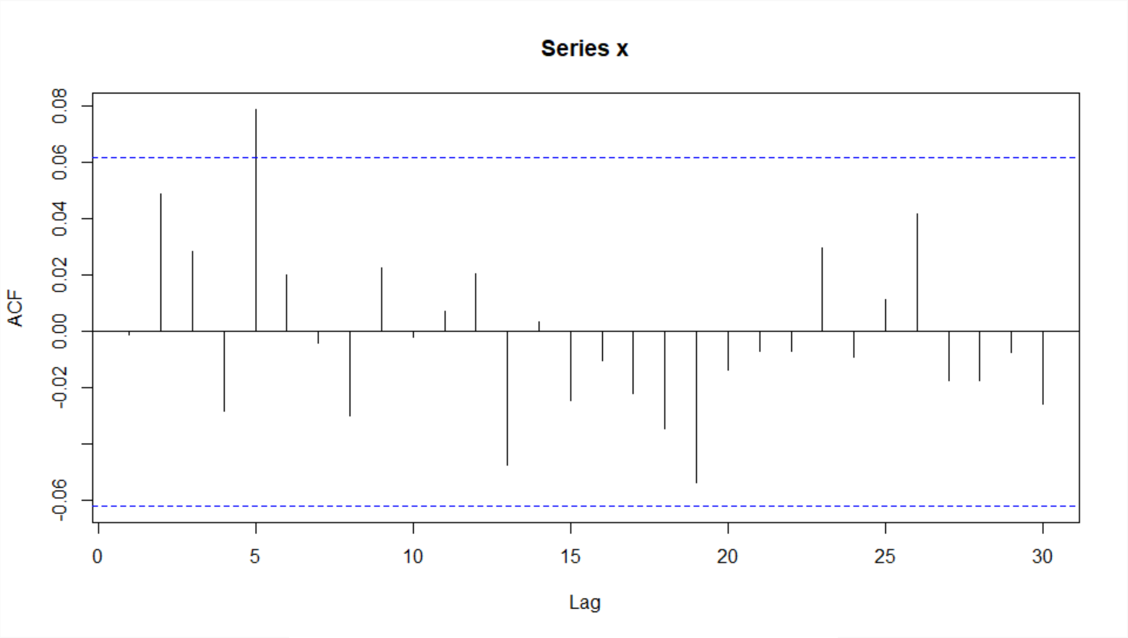
ARMA模型识别和EACF

- EACF：滤出AR部分，剩下MA部分，余下部分ACF截尾。
- 经过$j+1$次回归，得到了$Y_{t-1},\ldots,Y_{t-k}$的系数$\tilde{\phi}_1,\ldots\tilde{\phi}_k$，令$W_{t,k,j}=Y_t-\tilde{\phi}_1 Y_{t-1}-\cdots-\tilde{\phi}_k Y_{t-k}$
- 如果真实模型为ARMA(p, q), 对于$k=p,j\geq q$，$\{W_{t,k,j}\}$近似是MA(q)模型，所以其ACF序列q阶截尾。
- 若$\{W_{t,k,j}\}$的$j+1$阶样本ACF显著不为0，则EACF表的第k行第j列标为“x”，否则标为“o”
- 当$k>p$时，$\{W_{t,k,j}\}$的MA阶数同步增加。
- 问答：EACF图的第k行代表什么？怎么解释?

例一：白噪声

- 对于白噪声序列$\{e_t\}$，假设它的AR部分为0阶，则无论做几次回归都有$W_{t,0,j} = e_t$， ACF为0阶截尾。
- 如果假设它的AR部分为1阶，则无论做几次回归，$\{W_{t,1,j}\}$都满足MA(1)模型，ACF为1阶截尾。
- 如果假设它的AR部分为k阶，则无论做几次回归，$W_{t,k,j} = e_t - \tilde{\phi}_1 e_{t-1} - \cdots - \tilde{\phi}_k e_{t-k}$都满足MA(k)模型，ACF为k阶截尾。
- 所以白噪声的EACF表的第k行从第k列开始大概率为“o”，但是也有意外情况，如果该白噪声的样本ACF在第q阶显著不为0，则EACF表的第q-1列会出现一定的“x”,这是EACF表中常见的一种现象。

例一：白噪声




```
> eacf(x)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 o o o o x o o o o o o  o  o  o
1 o o o o x o o o o o o  o  o  o
2 x x o o x o o o o o o  o  o  o
3 x x x o x o o o o o o  o  o  o
4 x x x x x o o o o o o  o  o  o
5 x x x x x o o o o o o  o  o  o
6 x x x x x o o o o o o  o  o  o
7 x x x x o o o o o o o  o  o  o
```

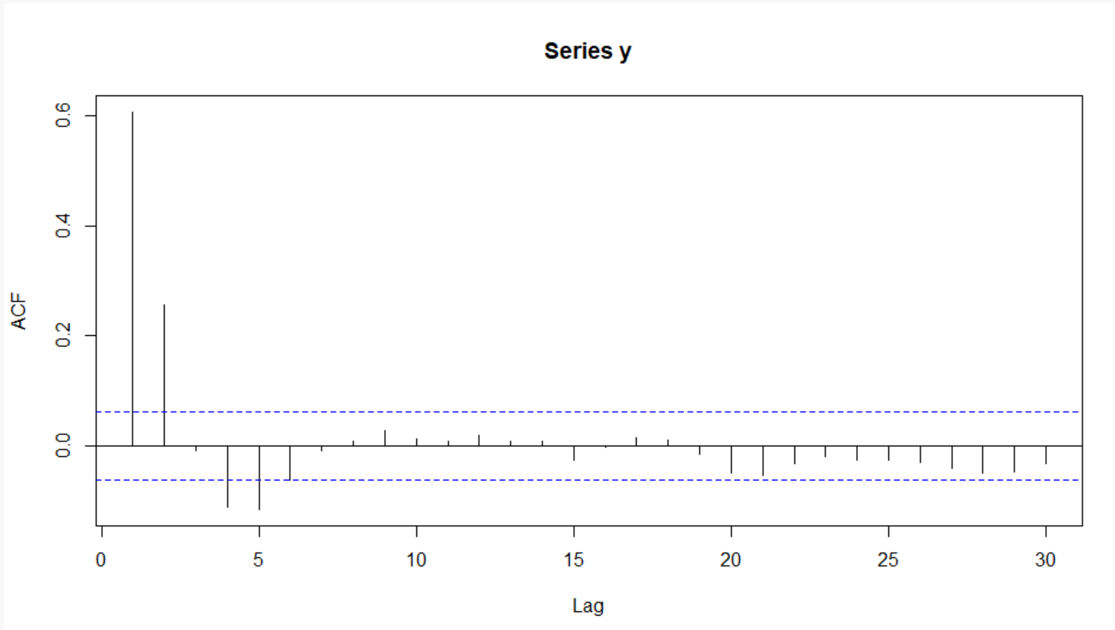
例二：ARMA(2, 2)

- 对于ARMA(2, 2)序列$\{Y_t\}$，当$k = 0, 1$时，做回归无法有效去除AR部分，所以ACF无明显截尾。
- EACF的第2行有什么特点?

- EACF的第k行有什么特点?

- EACF表中的“o”大致呈现出以第2行第2列为“顶点”的三角模式。

例二：ARMA(2, 2)



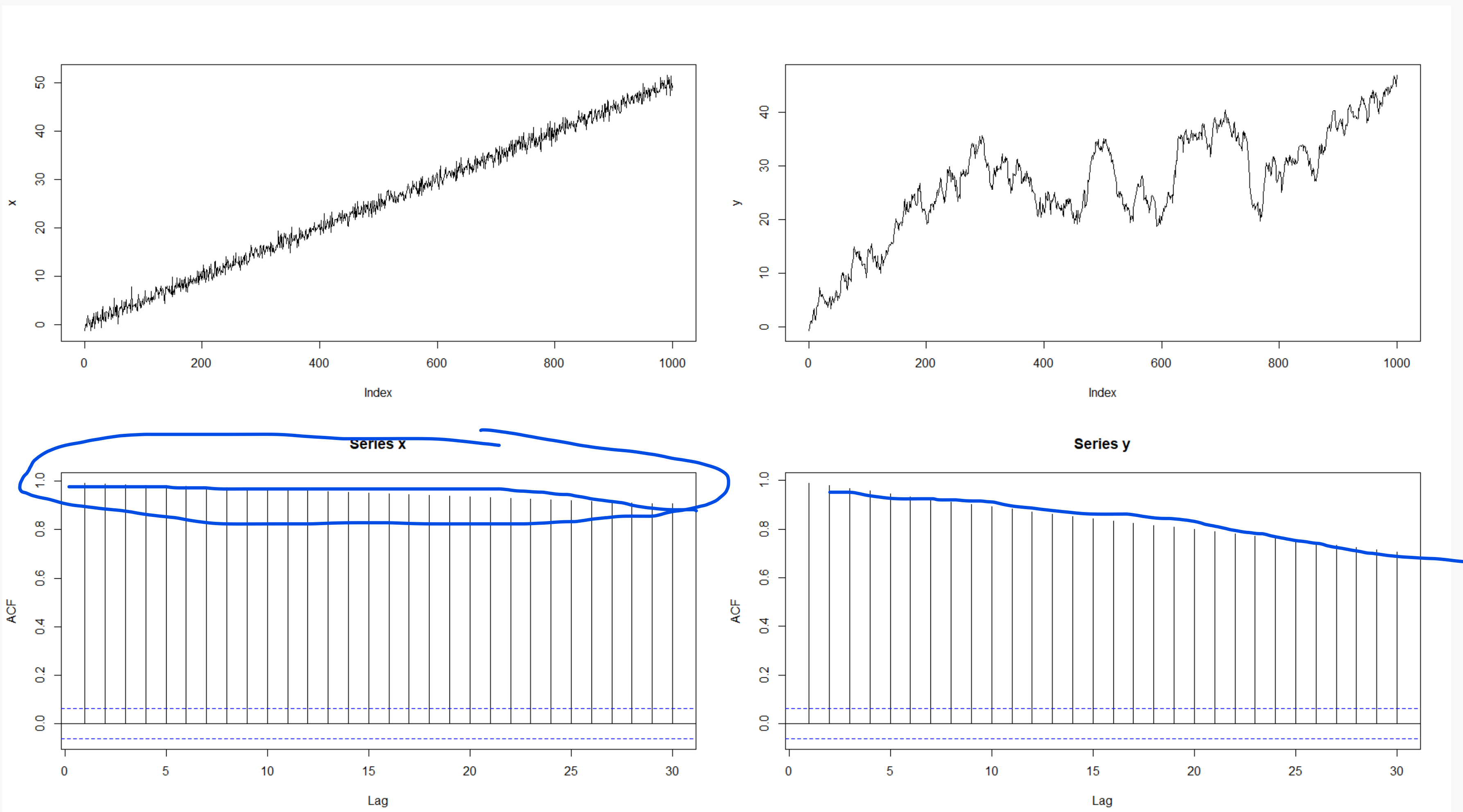

```
> eacf(y)
AR/MA
  0 1 2 3 4 5 6 7 8 9 10 11 12 13
0 x x o x x o o o o o o  o  o  o
1 x x o x x o o o o o o  o  o  o
2 x x o o o o o o o o o  o  o  o
3 o o x o o o o o o o o  o  o  o
4 x o x o o o o o o o o  o  o  o
5 x x o o o o o o o o o  o  o  o
6 x x x o o o o o o o o  o  o  o
7 x x x x o o o o o o o  o  o  o
```

非平稳模型识别

- 当样本ACF呈现出缓慢下降的趋势时，我们可以尝试对原序列进行一次差分，对差分后的序列观察ACF、PACF、EACF等，尝试建立平稳模型。
- 如果一次差分后的样本ACF仍然缓慢下降，可以考虑再次差分，但为了避免过度差分，建议仔细查看每次差分后的序列本身及其自相关特性：模型尽量简洁，但也不能草率。
- 可以结合数据实际背景做变换，如对数差分变换，Box-Cox变换等。

例：确定趋势VS随机趋势

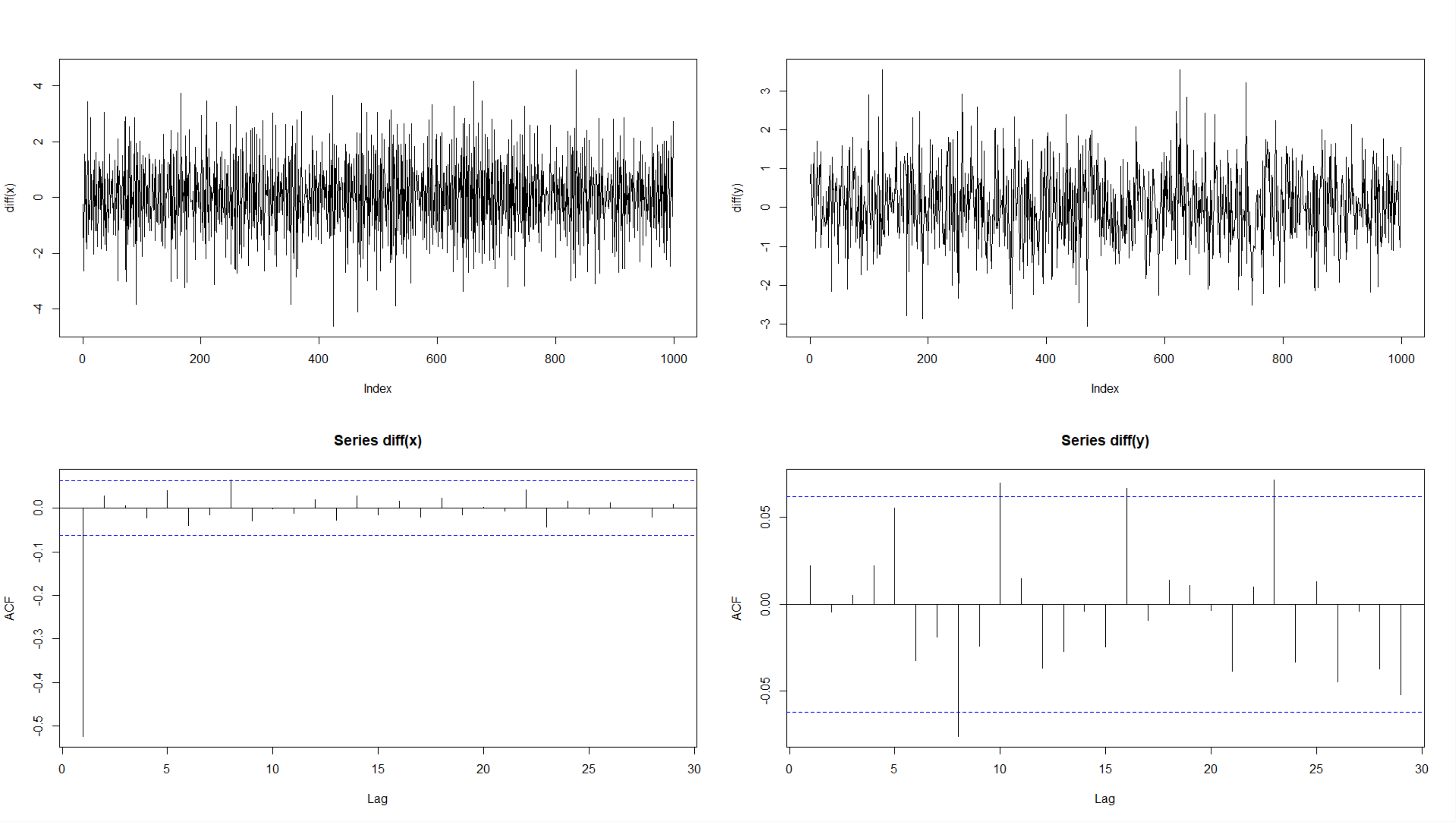
$X_t = 0.05t + e_t$ $\qquad\qquad$ $Y_t = Y_{t-1} + e_t$



例：确定趋势VS随机趋势

$\nabla X_t = 0.05 + e_t - e_{t-1}$

$\nabla Y_t = e_t$



ARMA模型参数估计

- 矩估计
- 最小二乘估计
- 最大似然估计

待估参数

- 估计对象为差分后的平稳模型：
- $Y_t - \mu = \phi_1(Y_{t-1} - \mu) + \cdots + \phi_p\left(Y_{t-p} - \mu\right) + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$

- 待估参数为：
- 均值 μ
- AR系数$\phi_1, \phi_2, \cdots, \phi_p$，MA系数$\theta_1, \theta_2, \cdots, \theta_q$
- 白噪声的方差σ_e^2
- 一共$p + q + 2$个参数

矩估计

- 什么是矩?
  - *均值，方差，协方差，相关系数，高阶矩等*

- 矩估计，就是用样本矩估计理论矩，然后用相应的对应关系（函数、方程）等估计参数。

- 用样本均值估计理论均值，$\hat{\mu} = \bar{Y}$，用样本方差估计理论方差，$\hat{\gamma}_0 = S^2$

Yule-Walker估计

假设拟合模型为AR(p)，用r_k替换ρ_k

$$
\begin{aligned}
r_1 &= \phi_1 + \phi_2 r_1 + \cdots + \phi_p r_{p-1} \\
r_2 &= \phi_1 r_1 + \phi_2 + \cdots + \phi_p r_{p-2} \\
&\vdots \\
r_p &= \phi_1 r_{p-1} + \phi_2 r_{p-2} + \cdots + \phi_p
\end{aligned}
$$

通过解方程，得到估计$\hat{\phi}_1, \hat{\phi}_2, \cdots, \hat{\phi}_p$

MA矩估计

- MA(1)，$\rho_1 = -\frac{\theta}{1+\theta^2}$，替换后得$r_1 = -\frac{\theta}{1+\theta^2}$，解得
- $\hat{\theta} = -\frac{1}{2r_1} \pm \sqrt{\frac{1}{4r_1^2} - 1}$, 只有一个满足可逆条件$|\hat{\theta}| < 1$
- $\rho_k = \begin{cases} 1, & k = 0 \\ \frac{-\theta_k + \sum_{i=1}^{q-k} \theta_i \theta_{k+i}}{1+\theta_1^2+\cdots+\theta_q^2}, & 1 \le k \le q \\ 0, & k > q \end{cases}$
- 方程组可能没有实数解！且估计误差较大！

ARMA矩估计

- $\rho_k = \frac{(1-\theta\phi)(\phi-\theta)}{1-2\theta\phi+\theta^2}\phi^{k-1}, k \geq 1$

- 注意到$\frac{\rho_2}{\rho_1} = \phi$, 因此$\hat{\phi} = \frac{r_2}{r_1}$，也可以是任意的 $\hat{\phi} = \frac{r_{k+1}}{r_k}$, $k \geq 1$

- 再用$r_1 = \frac{(1-\theta\hat{\phi})(\hat{\phi}-\theta)}{1-2\theta\hat{\phi}+\theta^2}$解得$\theta$

噪声方差估计

- 同理，只要有公式，就能估计
- 如MA(q), $\gamma_0 = \left(1 + \theta_1^2 + \cdots + \theta_q^2\right)\sigma_e^2$, 故

$$\hat{\sigma}_e^2 = \frac{S^2}{(1 + \hat{\theta}_1^2 + \cdots + \hat{\theta}_q^2)}$$

- 如AR(p), $\gamma_0 = \frac{\sigma_e^2}{1-\phi_1\rho_1-\phi_2\rho_2-\cdots-\phi_p\rho_p}$, 故

$$\hat{\sigma}_e^2 = S^2(1 - \hat{\phi}_1 r_1 - \hat{\phi}_2 r_2 - \cdots - \hat{\phi}_p r_p)$$

- 如ARMA(1, 1), $\gamma_0 = \frac{1-2\theta\phi+\theta^2}{1-\phi^2}\sigma_e^2$, 故$\hat{\sigma}_e^2 = \frac{1-\hat{\phi}^2}{1-2\hat{\theta}\hat{\phi}+\hat{\theta}^2}S^2$

矩估计特点

- 通常要估计多少个参数，就需要多少个样本矩
- 不唯一
- 优点：思想简单，易计算，不需要假设总体分布
- 缺点：只用到几个样本矩信息，信息损失，估计误差大，特别对于含有MA项的模型。

最小二乘估计

- 使得残差平方和$\sum_{t=p+1}^{n} e_t^2$最小的那组参数即为最小二乘估计。
- 带均值AR(1)模型，$Y_t - \mu = \phi(Y_{t-1} - \mu) + e_t$，残差平方和为

$$S_c(\phi, \mu) = \sum_{t=2}^{n} [Y_t - \mu - \phi(Y_{t-1} - \mu)]^2$$

- 该平方和称为条件平方和函数，最小化该函数将得到最小二乘估计。
- 通常，可以通过求偏导来完成。
- 可以发现，当n较大时，估计值非常接近Yule-Walker矩估计。

MA(1)模型的最小二乘估计

- 考虑零均值模型 $Y_t = e_t - \theta e_{t-1}$，可变形为$e_t = Y_t + \theta e_{t-1}$,如果令$e_0 = 0$，则$e_1 = Y_1$，$e_2 = Y_2 + \theta e_1$，$\cdots$，$e_n = Y_n + \theta e_{n-1}$，需要最小化的函数为$S_c(\theta) = \sum_{t=1}^{n} e_t^2$.
- 当$|\theta| < 1$时，模型可逆，e_0的选取对估计的影响微乎其微。
- 当给定θ的值时，$S_c(\theta)$可直接算出来，但是难以求极值，可以通过网格遍历区间$(-1,1)$来求解最优值。

MA(q)模型的最小二乘估计

- $Y_t = e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$可以写成

$$e_t = Y_t + \theta_1 e_{t-1} + \cdots + \theta_q e_{t-q}$$

- 令$e_0 = e_{-1} = \cdots = e_{1-q} = 0$，则$e_1 = Y_1$，递推求出$e_t$的值，再最小化$S_c(\theta_1, \ldots, \theta_q) = \sum_{t=1}^{n} e_t^2$.

- 只能用数值优化算法求解，一般统计软件都有内置。

ARMA(p, q)模型的最小二乘估计

- $Y_t = \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} + e_t - \theta_1 e_{t-1} - \cdots - \theta_q e_{t-q}$可以写成

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

- 令$e_p = e_{p-1} = \cdots = e_{p+1-q} = 0$，则$e_{p+1} = Y_{p+1} - \phi_1 Y_p - \cdots - \phi_p Y_1$，递推求出$e_t$的值，再最小化$S_c(\theta_1, \ldots, \theta_q) = \sum_{t=p+1}^{n} e_t^2$.

最大似然估计

- 使得似然函数（等于概率密度函数）最大的参数值就是最大似然估计。

$$L(\boldsymbol{\theta}|Y_1, \dots, Y_n) = f(Y_1, \dots, Y_n|\boldsymbol{\theta})$$

$$\widehat{\boldsymbol{\theta}} = arg\max_{\boldsymbol{\theta}} L(\boldsymbol{\theta}|Y_1, \dots, Y_n)$$

AR(1)模型的最大似然估计

- 假设白噪声为正态分布$N(0,\sigma_e^2)$，其密度函数为

$$(2\pi\sigma_e^2)^{-1/2}\exp\left(-\frac{e_t^2}{2\sigma_e^2}\right)$$

- 则$e_2,\dots,e_n$的联合概率密度为

$$(2\pi\sigma_e^2)^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}e_t^2\right)$$

- 代入$e_t=Y_t-\mu-\phi(Y_{t-1}-\mu)$可得

$$(2\pi\sigma_e^2)^{-(n-1)/2}\exp\left(-\frac{1}{2\sigma_e^2}\sum_{t=2}^{n}[Y_t-\mu-\phi(Y_{t-1}-\mu)]^2\right)$$

- 问答：这是什么的密度函数?

AR(1)模型的似然函数

- 乘上初始值$Y_1 \sim N(\mu, \frac{\sigma_e^2}{1-\phi^2})$的概率密度，得到$Y_1, \dots, Y_n$的联合概率密度，亦即我们要的似然函数

$$L(\phi, \mu, \sigma_e^2 | Y_1, \dots, Y_n) = (2\pi\sigma_e^2)^{-\frac{n}{2}}(1-\phi^2)^{1/2}\exp\left(-\frac{1}{2\sigma_e^2}S(\phi, \mu)\right)$$

$$S(\phi, \mu) = (1-\phi^2)(Y_1 - \mu) + \sum_{t=2}^{n}[Y_t - \mu - \phi(Y_{t-1} - \mu)]^2$$

对数似然函数为$\log L(\phi, \mu, \sigma_e^2 | Y_1, \dots, Y_n) =$

$$l(\phi, \mu, \sigma_e^2 | Y_1, \dots, Y_n) = -\frac{n}{2}\log(2\pi\sigma_e^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_e^2}S(\phi, \mu)$$

无条件平方和与简化估计

$$l(\phi,\mu,\sigma_e^2|Y_1,\dots,Y_n) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma_e^2) + \frac{1}{2}\log(1-\phi^2) - \frac{1}{2\sigma_e^2}S(\phi,\mu)$$

- $S(\phi,\mu) = S_c(\phi,\mu) + (1-\phi^2)(Y_1-\mu)$，称为无条件平方和函数，简化起见，我们可以最小化函数$S(\phi,\mu)$，得到$\hat{\phi},\hat{\mu}$，而对σ_e^2求偏导可以发现：

$$\hat{\sigma}_e^2 = \frac{S(\hat{\phi},\hat{\mu})}{n}$$