NJUAI 南京大学 人工智能学院
SCHOOL OF ARTIFICIAL INTELLIGENCE, NANJING UNIVERSITY
南京大學 NANJING UNIVERSITY 1902

06 时间序列的检验与拟合

时间序列的建模步骤

平稳条件下统计量的估计

- 均值（序列）

$$\hat{\mu} = \frac{1}{n}\sum_{t=1}^{n} Y_t$$

- 延迟k自协方差函数

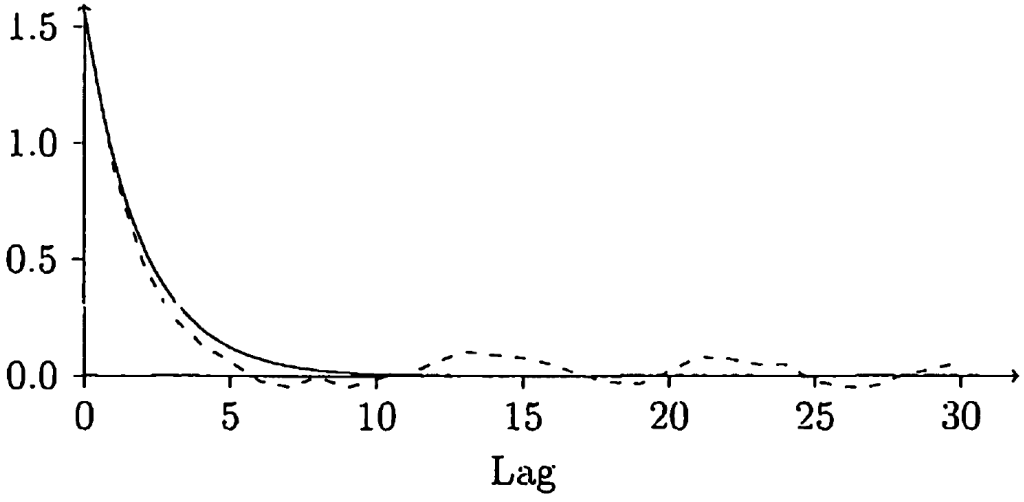
$$\hat{\gamma}(k) = \frac{1}{n}\sum_{t=1}^{n-k}(Y_t - \hat{\mu})(Y_{t+k} - \hat{\mu}), 0 < k < n$$

- 方差 $(k = 0)$

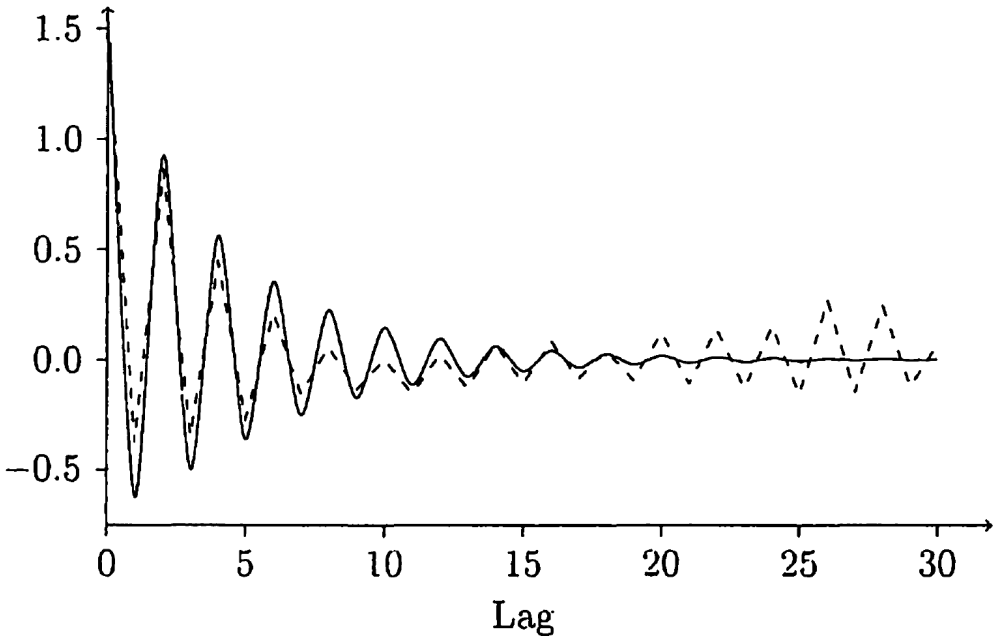
$$\hat{\gamma}(0) = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{\mu})^2$$

理论值 vs. 估计值

- ACF：
  - 理论值：实线
  - 估计值：虚线



$AR(1)$过程



$ARMA(2,2)$过程

概要

1. 模型检验定阶：
确定(p, d, q)

2. 模型优化：
如何学习模型参数

3. 模型检验：
建模是否充分

4. 模型拓展：
ARIMA的拓展

差分阶数d的识别

- 对于$ARMA(p,q)$模型，即$ARMA(p,0,q)$，$\phi(B)Y_t = \theta(B)\varepsilon_t$，其ACF满足

$$\phi(B)\rho_k = 0, \qquad k > q$$

- 若有

$$\phi(B) = \prod_{i=1}^{p} (1 - G_i B)$$

通过AR部分判断平稳性

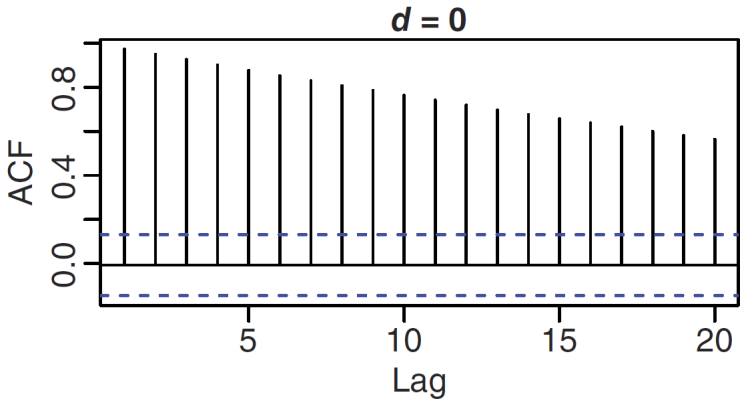
- 则ACF具有以下形式（假定没有重根）

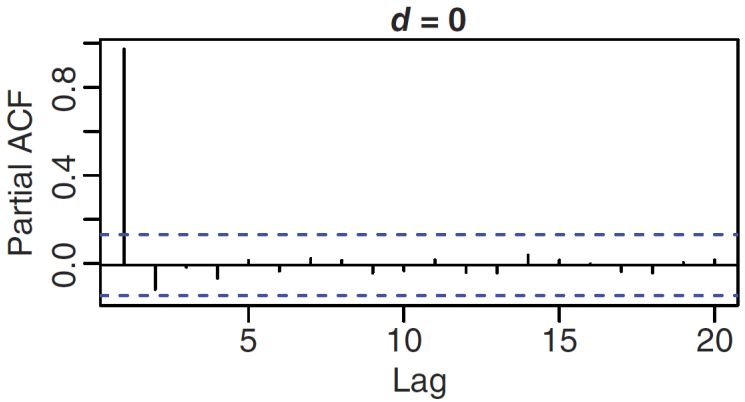
$$\rho_k = A_1 G_1^k + A_2 G_2^k + \cdots + A_p G_p^k, \qquad k > q - p$$

- 对于**平稳**的序列，$\phi(B)$的根在单位圆外，等价于要求$G_1, \ldots, G_p$在单位圆内
- 因此，对于平稳的序列，所有的根都不临近单位圆边界，ACF对于较大的k而言会指数衰减（"die out"）；因此，如果ACF估计值无法快速衰减，则说明应该进一步做差分

差分阶数d的识别

- 过差分（Overdifferencing）：由于过度差分引入不必要的相关性，且增加模型复杂度
- 对于随机游走序列Y_t，





- 进行一次差分，得到

$$w_t = (1 - B)Y_t = \varepsilon_t$$

为白噪声平稳序列，构建$ARIMA(0,1,0)$模型

- 进行第二次差分，有

$$(1 - B)w_t = (1 - B)\varepsilon_t$$

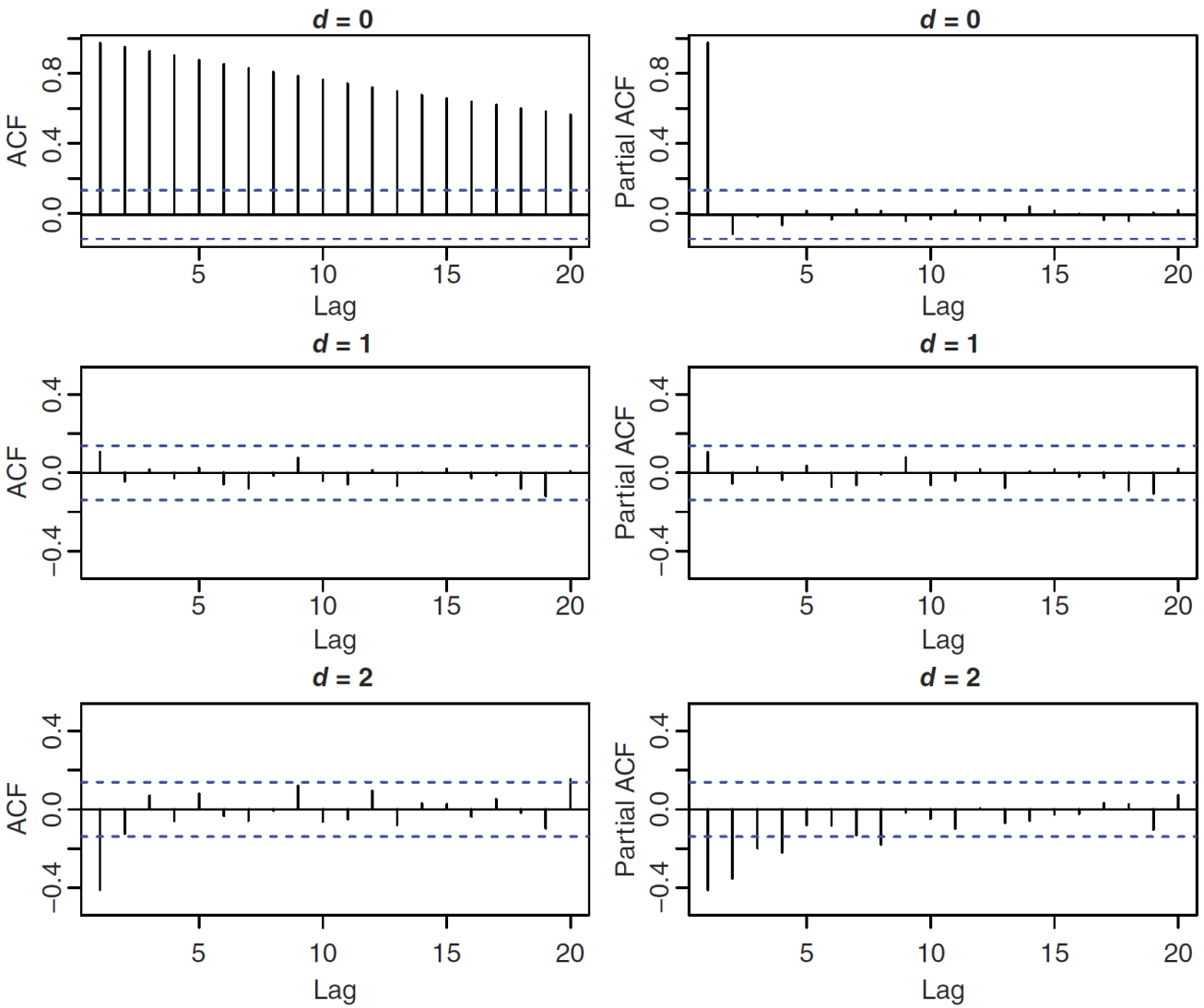
为关于w_t的$MA(1)$模型，参数$\theta = 1$（不可逆），此时有$Y_t = ARIMA(0,2,1)$

差分阶数d的识别

- 对于随机游走序列Y_t

$ARIMA(0,1,0)$

$ARIMA(0,2,1)$



ARMA模型定阶

- 利用ACF、PACF估计值，确定差分之后的序列$w_t \sim ARMA(p,q)$的参数

	Autoregressive Process	Moving Average Processes	Mixed Processes
Model in terms of previous $\tilde{z}'s$	$\phi(B)\tilde{z}_t = a_t$	$\theta^{-1}(B)\tilde{z}_t = a_t$	$\theta^{-1}(B)\phi(B)\tilde{z}_t = a_t$
Model in terms of previous a's	$\tilde{z}_t = \phi^{-1}(B)a_t$	$\tilde{z}_t = \theta(B)a_t$	$\tilde{z}_t = \phi^{-1}(B)\theta(B)a_t$
π weights	Finite series	Infinite series	Infinite series
ψ weights	Infinite series	Finite series	Infinite series
Stationarity condition	Roots of $\phi(B) = 0$ lie outside the unit circle	Always stationary	Roots of $\phi(B) = 0$ lie outside the unit circle
Invertibility condition	Always invertible	Roots of $\theta(B) = 0$ lie outside the unit circle	Roots of $\theta(B) = 0$ lie outside the unit circle
Autocorrelation function	Infinite (damped exponentials and/or damped sine waves)	Finite	Infinite (damped exponentials and/or damped sine waves after first $q - p$ lags)
	Tails off	Cuts off after lag q	Tails off
Partial autocorrelation function	Finite	Infinite (dominated by damped exponentials and/or damped sine waves)	Infinite (dominated by damped exponentials and/or damped sine waves after first $p - q$ lags)
	Cuts off after lag p	Tails off	Tails off

ARMA模型定阶

- 利用ACF、PACF估计值，确定差分之后的序列$w_t \sim ARMA(p,q)$的参数

	Order				
	$(1, d, 0)$	$(0, d, 1)$	$(2, d, 0)$	$(0, d, 2)$	$(1, d, 1)$
Behavior of ρ_k	Decays exponentially	Only ρ_1 nonzero	Mixture of exponentials or damped sine wave	Only ρ_1 and ρ_2 nonzero	Decays exponentially from first lag
Behavior of ϕ_{kk}	Only ϕ_{11} nonzero	Exponential dominates decay	Only ϕ_{11} and ϕ_{22} nonzero	Dominated by mixture of exponential or damped sine wave	Dominated by exponential decay from first lag
Preliminary estimates from	$\phi_1 = \rho_1$	$\rho_1 = \dfrac{-\theta_1}{1+\theta_1^2}$	$\phi_1 = \dfrac{\rho_1(1-\rho_2)}{1-\rho_1^2}$	$\rho_1 = \dfrac{-\theta_1(1-\theta_2)}{1+\theta_1^2+\theta_2^2}$	$\rho_1 = \dfrac{(1-\theta_1\phi_1)(\phi_1-\theta_1)}{1+\theta_1^2-2\phi_1\theta_1}$
			$\phi_2 = \dfrac{\rho_2-\rho_1^2}{1-\rho_1^2}$	$\rho_2 = \dfrac{-\theta_2}{1+\theta_1^2+\theta_2^2}$	$\rho_2 = \phi_1\rho_1$
Admissible region	$-1 < \phi_1 < 1$	$-1 < \theta_1 < 1$	$-1 < \phi_2 < 1$ $\phi_2 + \phi_1 < 1$ $\phi_2 - \phi_1 < 1$	$-1 < \theta_2 < 1$ $\theta_2 + \theta_1 < 1$ $\theta_2 - \theta_1 < 1$	$-1 < \phi_1 < 1$ $-1 < \theta_1 < 1$

ARMA模型定阶

- 假定N是平稳时间序列观测样本数目
- 对于$AR(p)$模型，有

$$\left.\begin{aligned} \mathrm{E}[\hat{\phi}_{kk}] &\simeq 0 \\ \mathrm{Var}[\hat{\phi}_{kk}] &\simeq \frac{1}{N} \end{aligned}\right\}, \qquad k = p+1, p+2, \ldots$$

- 对于$MA(q)$模型，有

$$\left.\begin{aligned} \mathrm{E}[\hat{\rho}(k)] &\simeq 0 \\ \mathrm{Var}[\hat{\rho}(k)] &\simeq \frac{1}{N}[1 + 2(\hat{\rho}^2(1) + \cdots + \hat{\rho}^2(q))] \end{aligned}\right\}, \qquad k = q+1, q+2, \ldots$$

- $\hat{\rho}(k), (k > q)$和$\hat{\phi}_{kk}(k > p)$近似服从于高斯分布

标准差和显著性

- 需要确定ACF或PACF的估计值是否足够大
- ACF：假设当lag较大时为$MA(q)$过程，根据Bartlett定理，有

$$\hat{\sigma}[\rho_k] \simeq \frac{1}{\sqrt{n}}\left[1 + 2\left(\rho_1^2 + \rho_2^2 + \cdots + \rho_q^2\right)\right]^{\frac{1}{2}}, \qquad k > q$$

- PACF: 假设服从$AR(p)$过程，则$p+1$阶PACF估计值的标准差为

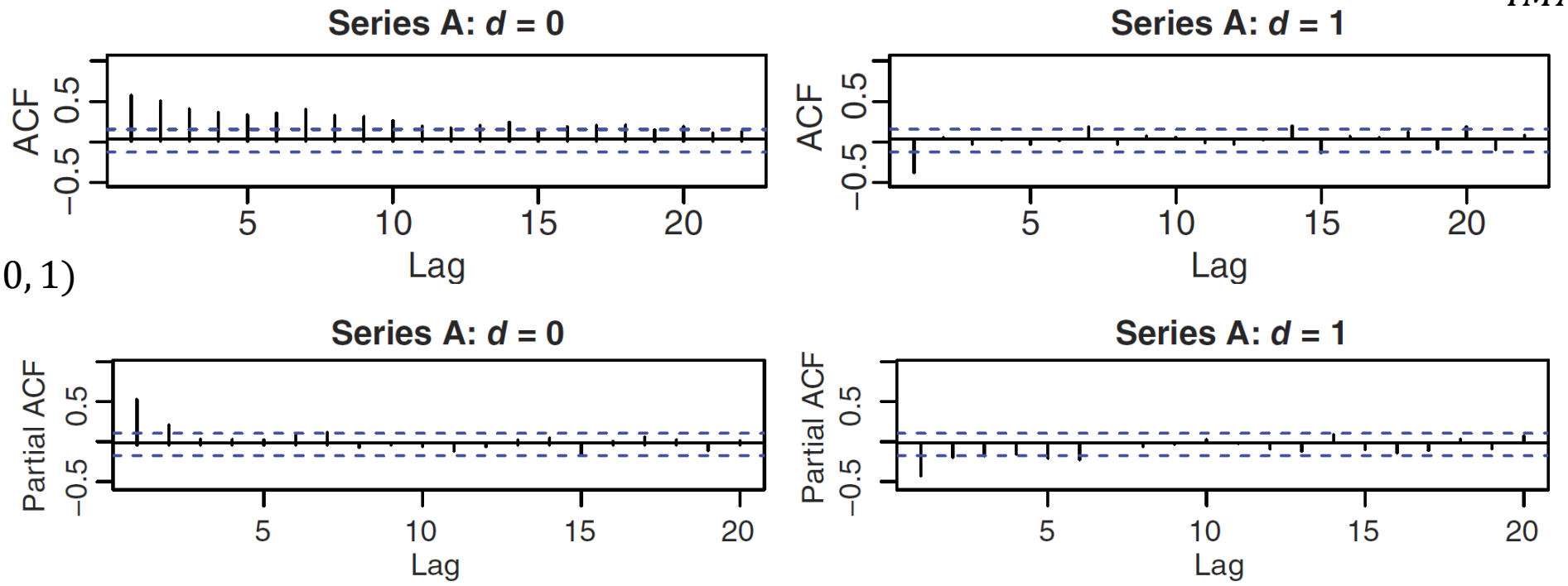
$$\hat{\sigma}\left[\hat{\phi}_{kk}\right] \simeq \frac{1}{\sqrt{n}}, \qquad k > p$$

模型识别：序列A

- **SERIES A** Chemical process concentration readings: every 2 hours

Series A "Uncontrolled'' concentration, two hourly readings: chemical process

$IMA(0,1,1)$

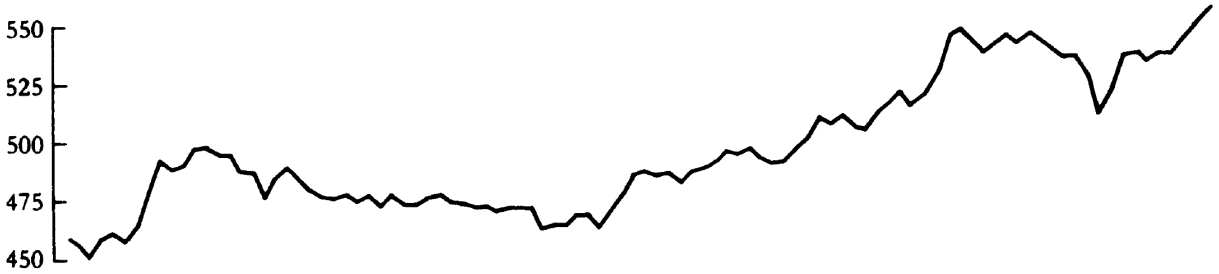


$ARMA(1,0,1)$

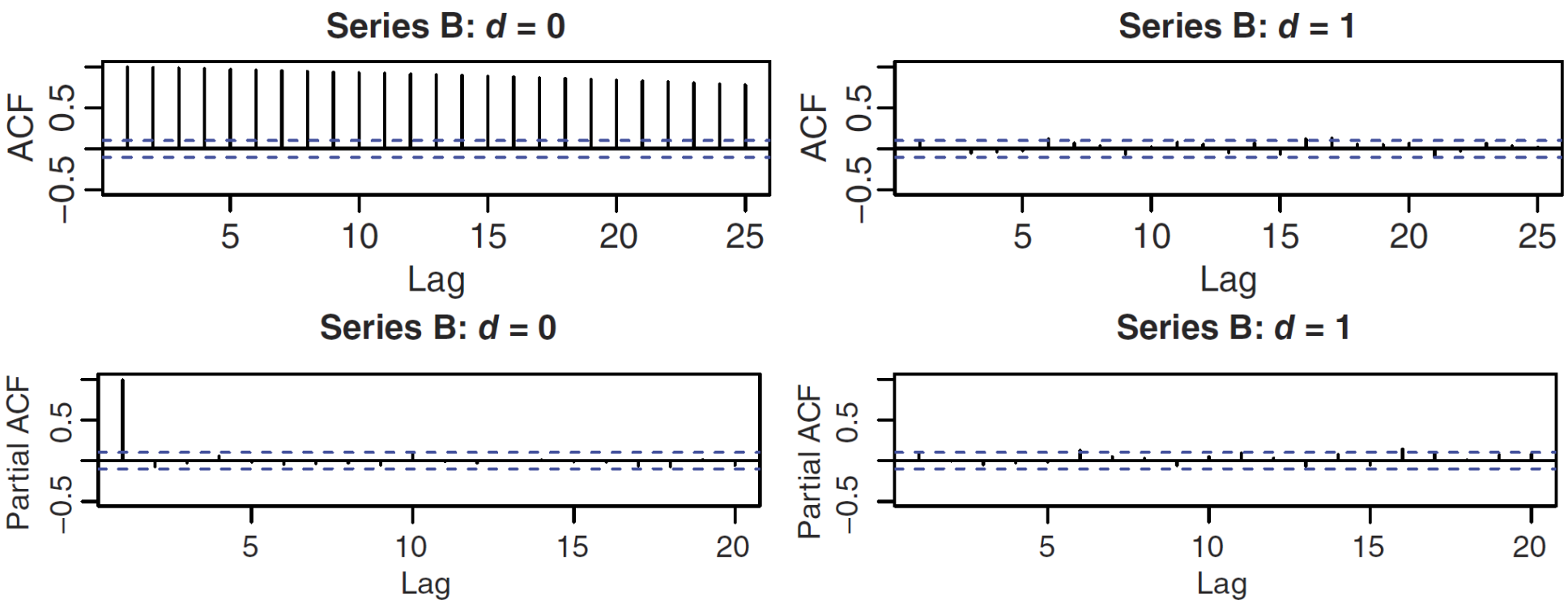
模型识别：序列B

- **SERIES B** IBM common stock closing prices: daily, May 17, 1961--November 2, 1962

random walk model $(1 - B)Y_t = \varepsilon_t$



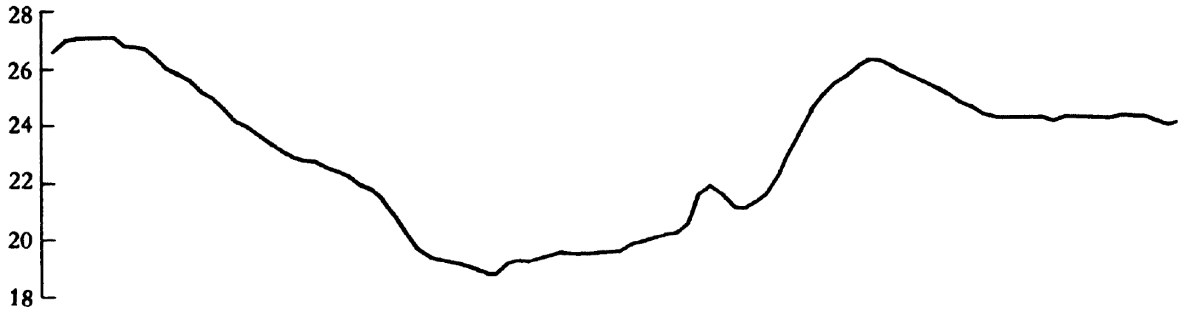
Series B Daily IBM stock prices



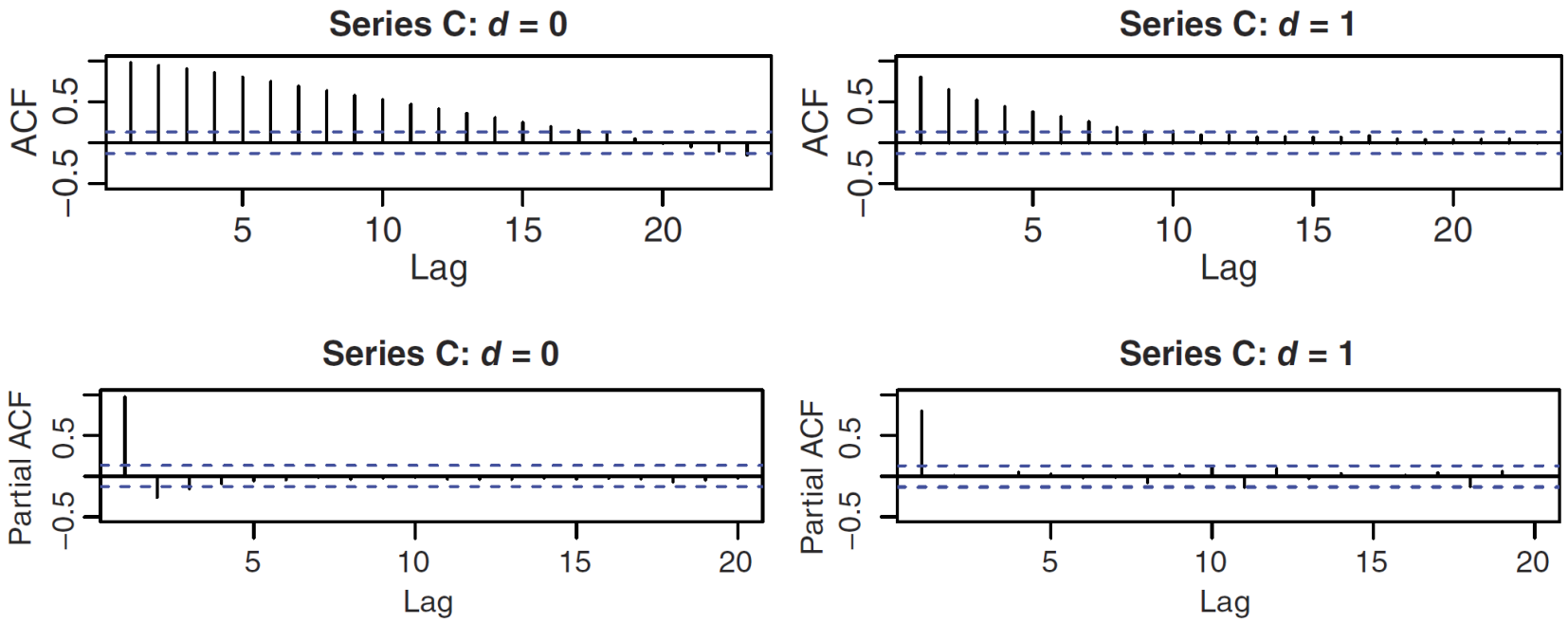
模型识别：序列C

- **SERIES C** Chemical process temperature readings: every minute

$$(1-0.8B)(1-B)Y_t = \varepsilon_t$$
$$(1-B)^2 Y_t = \varepsilon_t$$



Series C "Uncontrolled'' temperature, readings every minute : chemical process



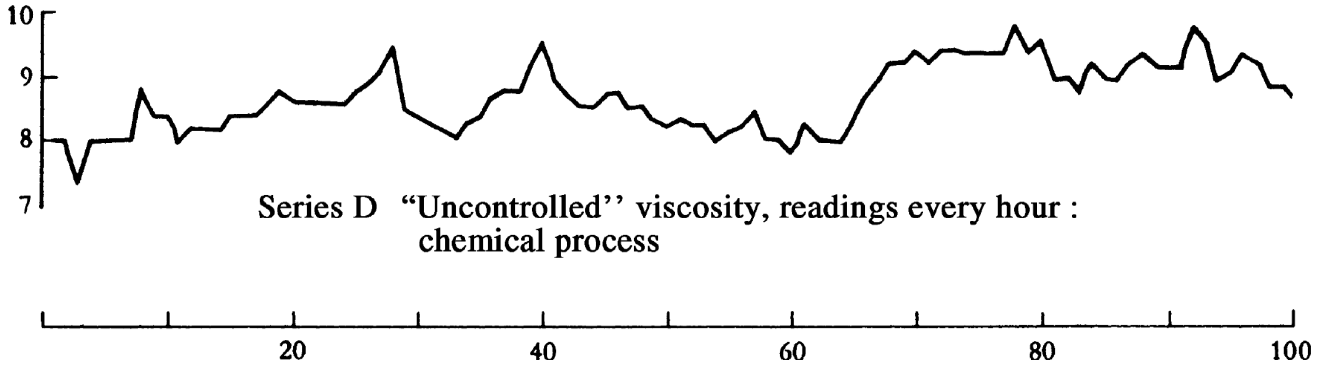
$ARMA(1,1,0)$

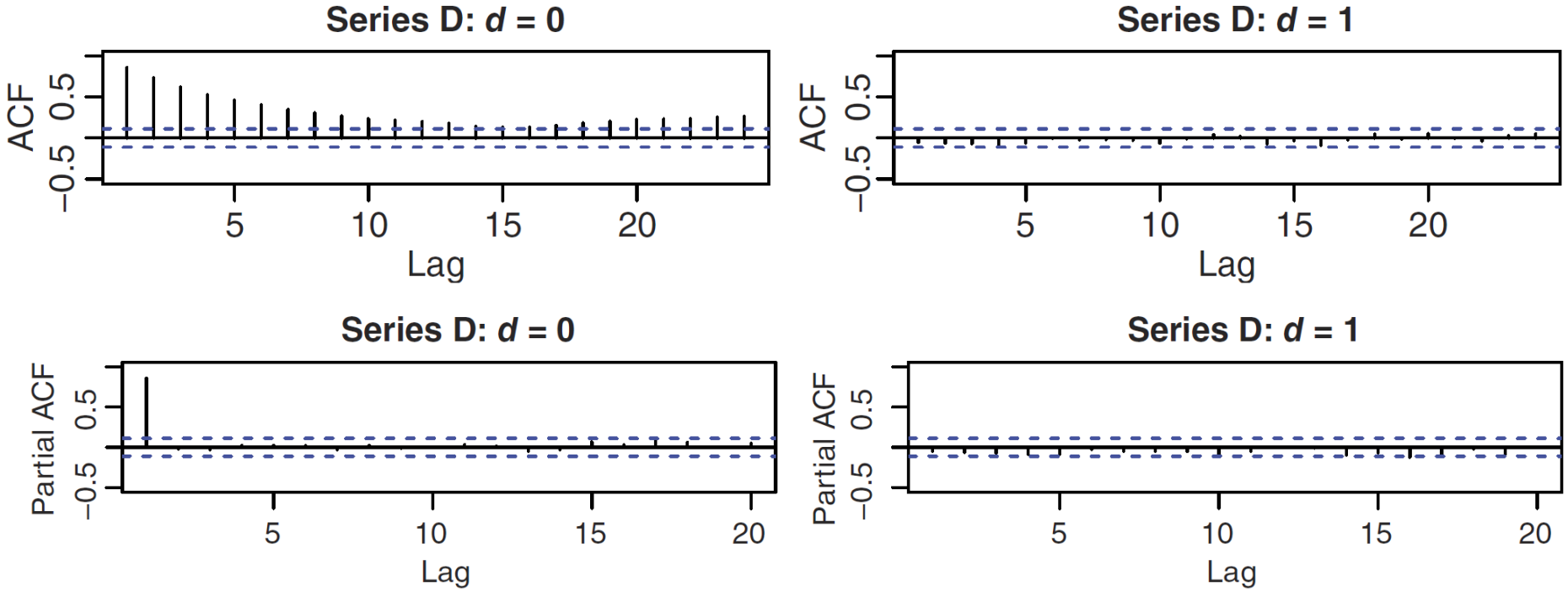
$IMA(0,2,0)$

模型识别：序列D

- **SERIES D** Chemical process viscosity readings: every hour

$$(1 - \phi B)Y_t = \varepsilon_t$$
$$(1 - B)Y_t = \varepsilon_t$$

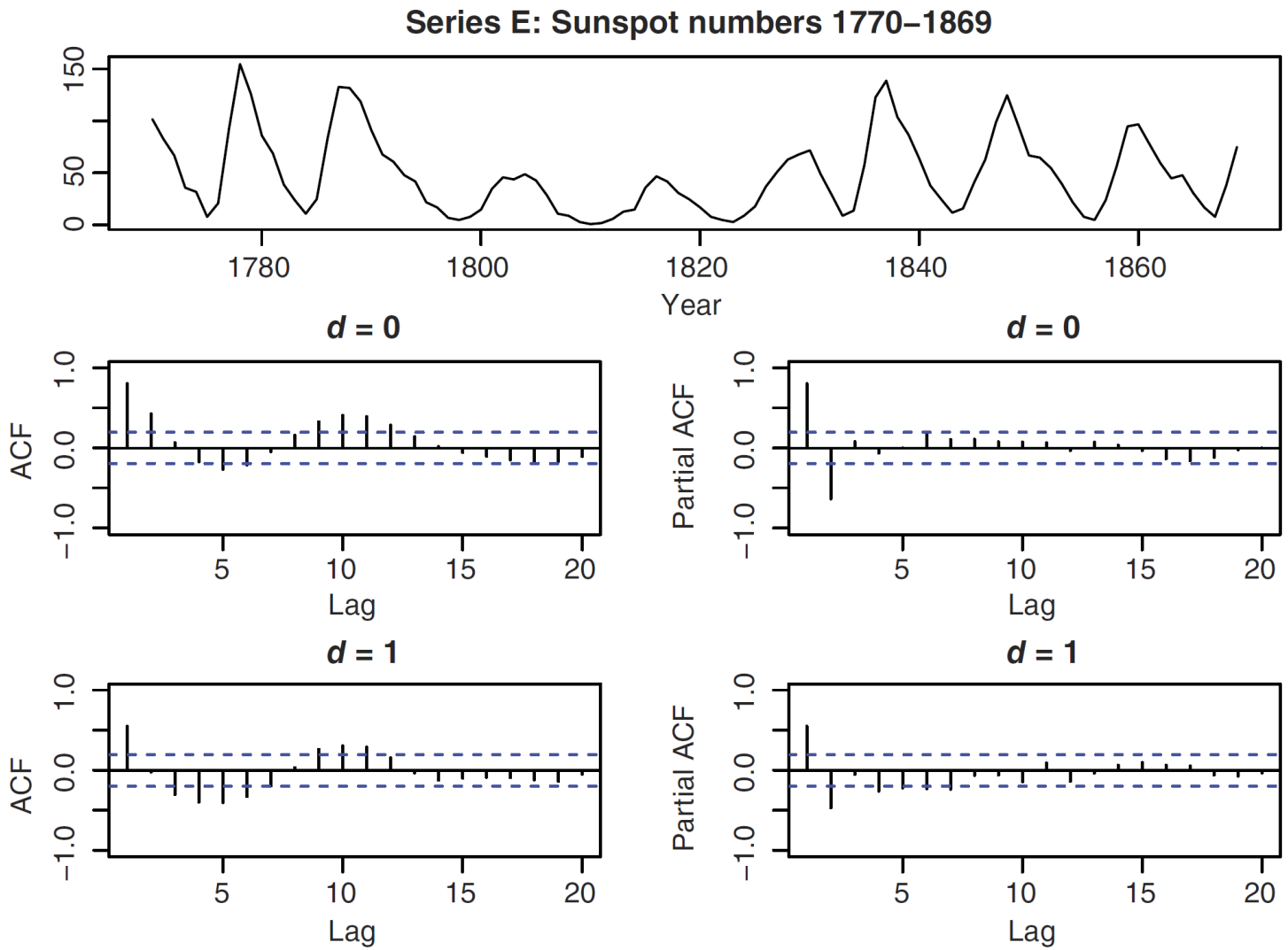




模型识别：序列E

- **SERIES E** Wolfer sunspot numbers: yearly

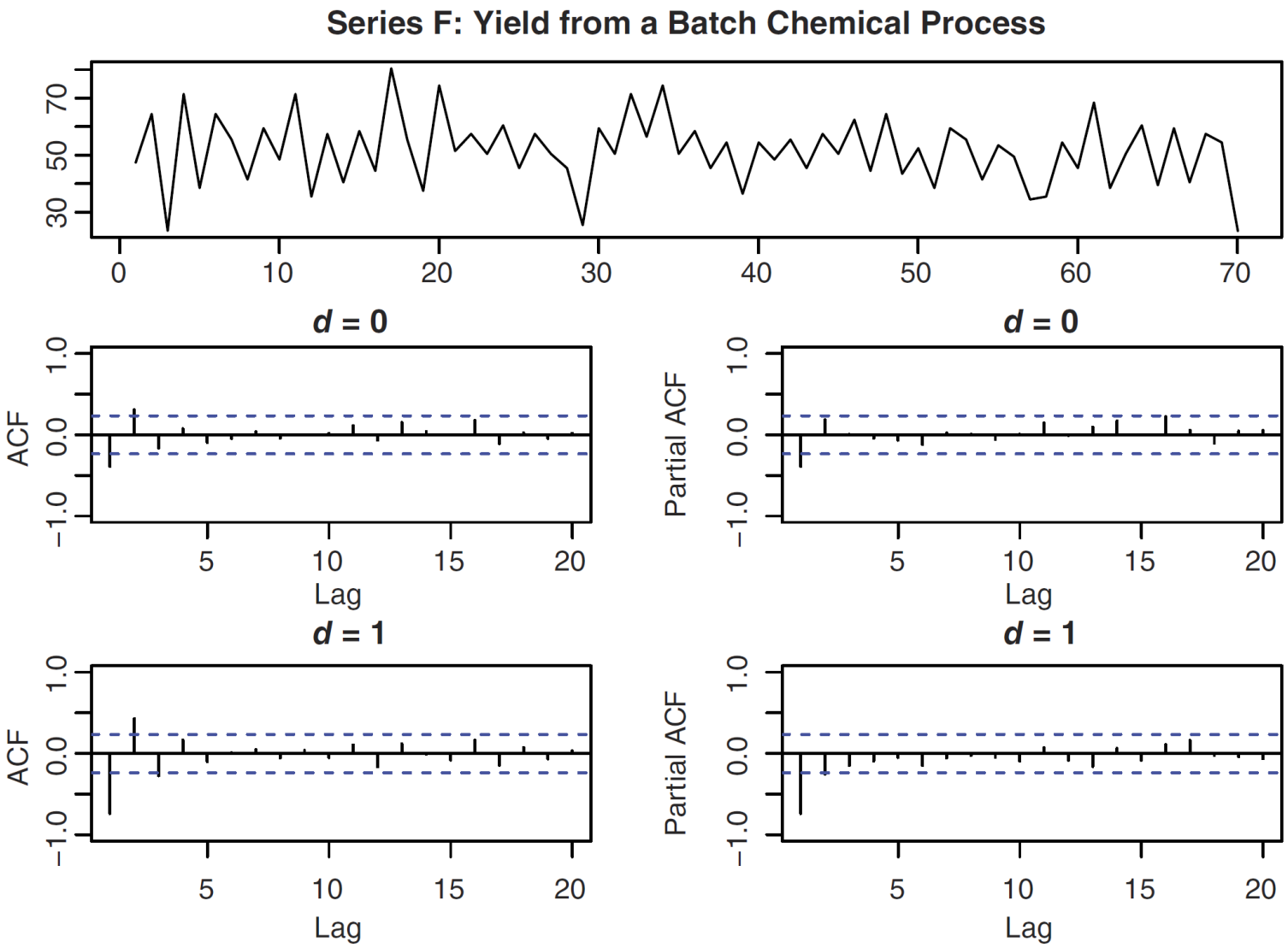
$AR(2)$ process



模型识别：序列F

- **SERIES F** Yields from a batch chemical process: consecutive

$AR(1)$ process



概要

1. 模型检验定阶：
确定(p, d, q)

2. 模型优化：
如何学习模型参数

3. 模型检验：
建模是否充分

4. 模型拓展：
ARIMA的拓展

ARMA模型参数估计

- 具有如下结构的模型称为自回归移动平均模型，简记为$ARMA(p,q)$

$$\begin{cases} Y_t = \varphi_0 + \varphi_1 Y_{t-1} + \cdots + \varphi_p Y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \cdots - \theta_q \varepsilon_{t-q} \\ \varphi_p \neq 0, \ \ \theta_q \neq 0 \\ E(\varepsilon_t) = 0, \ \ Var(\varepsilon_t) = \sigma_\varepsilon^2, E(\varepsilon_t \varepsilon_s) = 0, s \neq t \\ E(Y_s \varepsilon_t) = 0, \forall s < t \end{cases}$$

- 待估参数
  - $p+q+2$个未知参数

$$\varphi_1, \cdots, \varphi_p, \theta_1, \cdots, \theta_q, \mu, \sigma_\varepsilon^2$$

- 常用估计方法
  - 矩估计 （Moment Estimation）
  - 极大似然估计
  - 最小二乘估计

最小二乘法

- 1步预测

$$\hat{y}_{t+1|t} = -\phi_1 y_t - \phi_2 y_{t-1}$$

- 观测值：$y_1, y_2, \dots, y_N$
- 预测误差：

$$e_{t+1|t} = y_{t+1} - \hat{y}_{t+1|t} = y_{t+1} - (-\phi_1 y_t - \phi_2 y_{t-1})$$

$$\begin{aligned}
e_{3|2} &= y_3 + \phi_1 y_2 + \phi_2 y_1 \\
e_{4|3} &= y_4 + \phi_1 y_3 + \phi_2 y_2 \\
e_{5|4} &= y_5 + \phi_1 y_4 + \phi_2 y_3 \\
&\vdots \\
e_{N|N-1} &= y_N + \phi_1 y_{N-1} + \phi_2 y_{N-2}
\end{aligned}$$

$$\begin{bmatrix} y_3 \\ \vdots \\ y_N \end{bmatrix} = \begin{bmatrix} -y_2 & -y_1 \\ \vdots & \vdots \\ -y_{N-1} & -y_{N-2} \end{bmatrix} \begin{bmatrix} \phi_1 \\ \phi_2 \end{bmatrix} + \begin{bmatrix} e_{3|2} \\ \vdots \\ e_{N|N-1} \end{bmatrix}$$

$$Y = X\theta + \varepsilon$$

最小二乘方法求解

- 最小化1步预测误差的平方和$\varepsilon^{\mathrm{T}}\varepsilon$，得到最优解

$$\widehat{\boldsymbol{\theta}} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y}$$

$$\boldsymbol{X} = \begin{bmatrix} -y_2 & -y_1 \\ \vdots & \vdots \\ -y_N - 1 & -y_{N-2} \end{bmatrix}, \quad 且\boldsymbol{Y} = \begin{bmatrix} y_3 \\ \vdots \\ y_N \end{bmatrix}$$

- 推广，考虑一般性的$AR(p)$模型

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t, \qquad \boldsymbol{t = p+1, p+2, \ldots}$$

$\{\varepsilon_t\}$为白噪声，观察值为$Y_1, \ldots, Y_N$

$$\boldsymbol{X}_t^T = (-Y_{t-1}, \ldots, -Y_{t-p}), \quad t = p+1, p+2, \ldots$$
$$\theta^{T'} = (\phi_1, \ldots, \phi_p)$$

$$\Rightarrow \quad Y_t = \boldsymbol{X}_t^T\theta + \varepsilon_t, \quad t = p+1, p+2, \ldots$$

$$\mathcal{Y}^T = (Y_{p+1}, \ldots, Y_N)$$
$$\boldsymbol{X}^T = (\boldsymbol{X}_{p+1}, \ldots, \boldsymbol{X}_N)$$
$$\varepsilon^T = (\varepsilon_{p+1}, \ldots, \varepsilon_N)$$

$$\Rightarrow \quad Y = X\theta + \varepsilon$$

最小二乘估计

- $AR(p)$模型

$$Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} = \varepsilon_t,$$
$$t = p+1, p+2, \ldots$$

$\{\varepsilon_t\}$为白噪声，观察值为$Y_1, \ldots, Y_N$

- 令

$$\mathcal{Y}^T = \left(Y_{p+1}, \ldots, Y_N\right)$$
$$\boldsymbol{X}^T = \left(\boldsymbol{X}_{p+1}, \ldots, \boldsymbol{X}_N\right)$$
$$\varepsilon^T = \left(\varepsilon_{p+1}, \ldots, \varepsilon_N\right)$$

- 则有

$$Y = X\theta + \varepsilon$$

- 最小二乘估计（LS Estimator）最小化

$$S(\boldsymbol{\theta}) = (\boldsymbol{Y} - \boldsymbol{X\theta})^T (\boldsymbol{Y} - \boldsymbol{X\theta})$$

- 最优解

$$(\boldsymbol{X}^T \boldsymbol{X})\hat{\theta} = \boldsymbol{X}^T \boldsymbol{Y}$$

- 噪声方差估计（ $\hat{\sigma}_\varepsilon^2 = V[\varepsilon_t]$（

$$\hat{\sigma}_\varepsilon^2 = \frac{S(\hat{\theta})}{N-p} = \frac{\sum_{t=p+1}^{N} \varepsilon_t^2(\hat{\theta})}{N-p}$$

针对$AR(2)$的最小二乘估计

- 观察值为$Y_1,\dots,Y_N$，针对$AR(2)$模型

$$Y_t = -\phi_1 Y_{t-1} - \phi_2 Y_{t-2} + \varepsilon_t$$

- 令

$$\boldsymbol{Y}^T = (Y_3, Y_4, \dots, Y_N)$$

$$\boldsymbol{X} = \begin{pmatrix} \boldsymbol{X}_3^T \\ \vdots \\ \boldsymbol{X}_N^T \end{pmatrix} = \begin{pmatrix} -Y_2 & -Y_1 \\ \vdots & \vdots \\ -Y_{N-1} & -Y_{N-2} \end{pmatrix}$$

- 则

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{pmatrix} \sum_{t=2}^{N-1} Y_t^2 & \sum_{t=2}^{N-1} Y_t Y_{t-1} \\ \sum_{t=2}^{N-1} Y_t Y_{t-1} & \sum_{t=1}^{N-2} Y_t^2 \end{pmatrix}, \boldsymbol{X}^T\boldsymbol{Y} = \begin{pmatrix} -\sum_{t=2}^{N-1} Y_t Y_{t+1} \\ -\sum_{t=1}^{N-2} Y_t Y_{t+2} \end{pmatrix}$$

- 优化变量为$\theta = (\theta_1, \theta_2)^T$
- 求解

$$\begin{pmatrix} \sum_{t=2}^{N-1} Y_t^2 & \sum_{t=1}^{N-2} Y_{t+1} Y_t \\ \sum_{t=1}^{N-2} Y_{t+1} Y_t & \sum_{t=1}^{N-2} Y_t^2 \end{pmatrix} \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} = \begin{pmatrix} -\sum_{t=2}^{N-1} Y_t Y_{t+1} \\ -\sum_{t=1}^{N-2} Y_t Y_{t+2} \end{pmatrix}$$

- 得到

$$\hat{\theta} = \left(\hat{\phi}_1, \hat{\phi}_2\right)^T = (X^T X)^{-1} X^T Y$$

- 注意最优解不是关于Y的线性函数
- 当 ε_t为 白噪声时，$\hat{\theta}$ 是一致的（Consistent），且$\mathrm{Var}[\hat{\theta}] \approx \sigma_\varepsilon^2 (X^T X)^{-1}$

极大似然估计

- $ARMA(p,q)$过程

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

- 记

$$\boldsymbol{\theta}^T = \left(\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q\right)$$
$$\mathbf{Y}_t^T = (Y_t, Y_{t-1}, \dots, Y_1)$$

- 似然函数是给定$\boldsymbol{\theta}$和σ_ε^2的样本联合分布

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = f(\mathbf{Y}_N \mid \boldsymbol{\theta}, \sigma_\varepsilon^2)$$

优化$\boldsymbol{\theta}$和σ_ε^2使联合分布最大

极大似然估计

- 随机变量$Y_N|\boldsymbol{Y}_{N-1}$只包含ε_N这一随机成分
- ε_N在时刻N为白噪声，因此$Y_N|\boldsymbol{Y}_{N-1}$与$\boldsymbol{Y}_{N-1}$独立

$$f(\mathbf{Y}_N \mid \boldsymbol{\theta}, \sigma_\varepsilon^2) = f(Y_N \mid \mathbf{Y}_{N-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2) f(\mathbf{Y}_{N-1} \mid \boldsymbol{\theta}, \sigma_\varepsilon^2)$$

- 因此

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = \left(\prod_{t=p+1}^{N} f(Y_t \mid \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2) \right) f\left(\mathbf{Y}_p \mid \boldsymbol{\theta}, \sigma_\varepsilon^2\right)$$

- 使用**条件**似然函数

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = \prod_{t=p+1}^{N} f(Y_t \mid \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2)$$

当观察值较多时和精确极大似然得到同样的解

估计条件似然

- 给定$\boldsymbol{\theta}$和σ_ε^2确定条件似然函数
- $Y_t|\boldsymbol{Y}_{t-1}$的均值：1步预测$\hat{Y}_{t|t-1}$

$$\hat{Y}_{t|t-1}(\theta) = \mathrm{E}[Y_t \mid Y_{t-1}, \theta] = -\sum_{i=1}^{p} \phi_i Y_{t-i} + \sum_{i=1}^{q} \theta_i \varepsilon_{t-i}(\theta)$$

- 预测误差：$\varepsilon_t(\theta) = Y_t - \hat{Y}_{t|t-1}(\theta)$方差为$\sigma_\varepsilon^2$
- 假设高斯过程

$$f(Y_t \mid \mathbf{Y}_{t-1}, \boldsymbol{\theta}, \sigma_\varepsilon^2) = \frac{1}{\sigma_\varepsilon\sqrt{2\pi}} \mathrm{e}^{-\left(Y_t - \hat{Y}_{t|t-1}(\boldsymbol{\theta})\right)^2/2\sigma_\varepsilon^2}$$

- 因此**Conditional** Likelihood:

$$L(\mathbf{Y}_N; \boldsymbol{\theta}, \sigma_\varepsilon^2) = (\sigma_\varepsilon^2 2\pi)^{-\frac{N-p}{2}} \exp\left(-\frac{1}{2\sigma_\varepsilon^2} \sum_{t=p+1}^{N} \varepsilon_t^2(\boldsymbol{\theta})\right)$$

模型选择准则

- AIC, Akaike Information Criterion [Akaike, 1974]
  - 似然函数值越大越好
  - 未知参数的个数越少越好

$$\mathrm{AIC}_{p,q} = \frac{-2\ln L + 2r}{n} \approx \ln(\hat{\sigma}_a^2) + r\frac{2}{n} + \text{ constant}$$

  - n，样本数量
  - $\hat{\sigma}_a^2$，极大似然σ_a^2的估计值
  - $r = p + q + 1$是待估计的参数

- ARIMA模型的AIC修正
- corrected AIC， AICc

AICc

$$= \mathrm{AIC} + \frac{2r(r+2)}{n-(r+1)}$$

$$= \mathrm{AIC} + \frac{2(p+q+1)(p+q+2)}{n-p-q-2}$$

模型选择准则

- BIC, Bayesian Information Criteria [Schwarz, 1978]
- Schwarz's Bayesian Criterion, (SBC)
- n，样本数量
- $\hat{\sigma}_a^2$，极大似然σ_a^2的估计值
- $r = p + q + 1$是待估计的参数

$$\begin{aligned}\mathrm{BIC} &= \ln(\hat{\sigma}_a^2) + r\frac{\ln(n)}{n} \\ &= \mathrm{AIC} + r\frac{\ln(n) - 2}{n}\end{aligned}$$

- AIC准则的缺陷
  - 在样本容量趋于无穷大时，由AIC准则选择的模型不收敛于真实模型，它通常比真实模型所含的未知参数个数要多

- AIC, BIC只能用于选择p, q无法确定d **(差分将改变对数似然，使得不同d下的对数似然不可比)**

ARIMA模型的过程

- Hyndman-Khandakar algorithm [Hyndman andKhandakar, 2008]

1.The number of differences $0 \leq d \leq 2$ is determined using repeated KPSS tests.

2.The values of p and q are then chosen by minimising the AICc after differencing the data d times.

2.1 Four **initial** models are fitted:

$ARIMA(0,d,0)$, $ARIMA(2,d,2)$, $ARIMA(1,d,0)$, $ARIMA(0,d,1)$.

A constant is included unless $d = 2$. If $d \leq 1$, an additional model is also fitted: $ARIMA(0,d,0)$ without a constant.

2.2 The **best** model (with the smallest AICc value) fitted in step (a) is set to be the "current model".

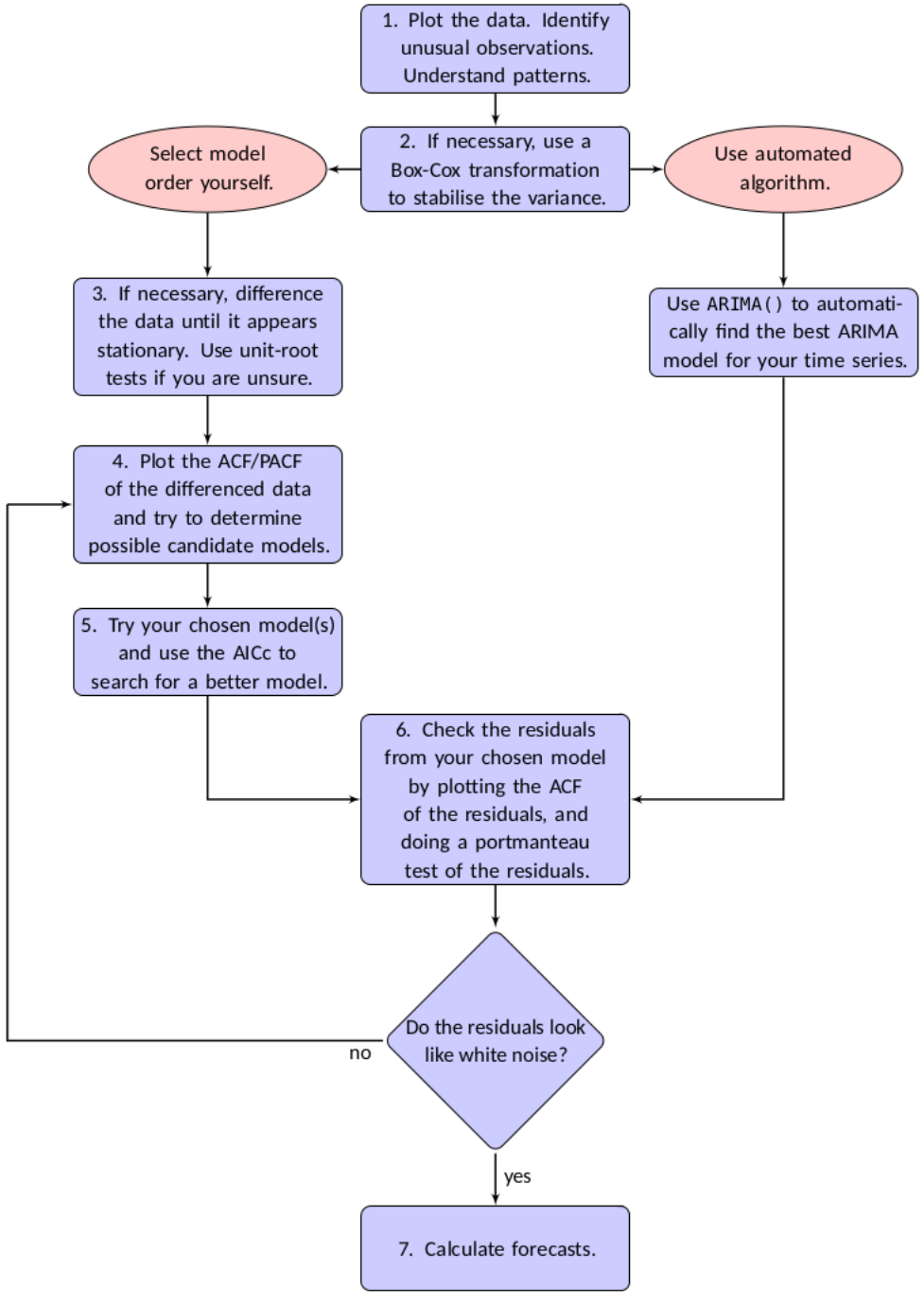
2.3 **Variations** on the current model are considered:

2.3.1 vary p and/or q from the current model by ± 1;

2.3.2 include/exclude c from the current model.

The **best** model considered so far (either the current model or one of these variations) becomes the new current model.

2.4 Repeat Step 2(c) until no lower AICc can be found.



概要

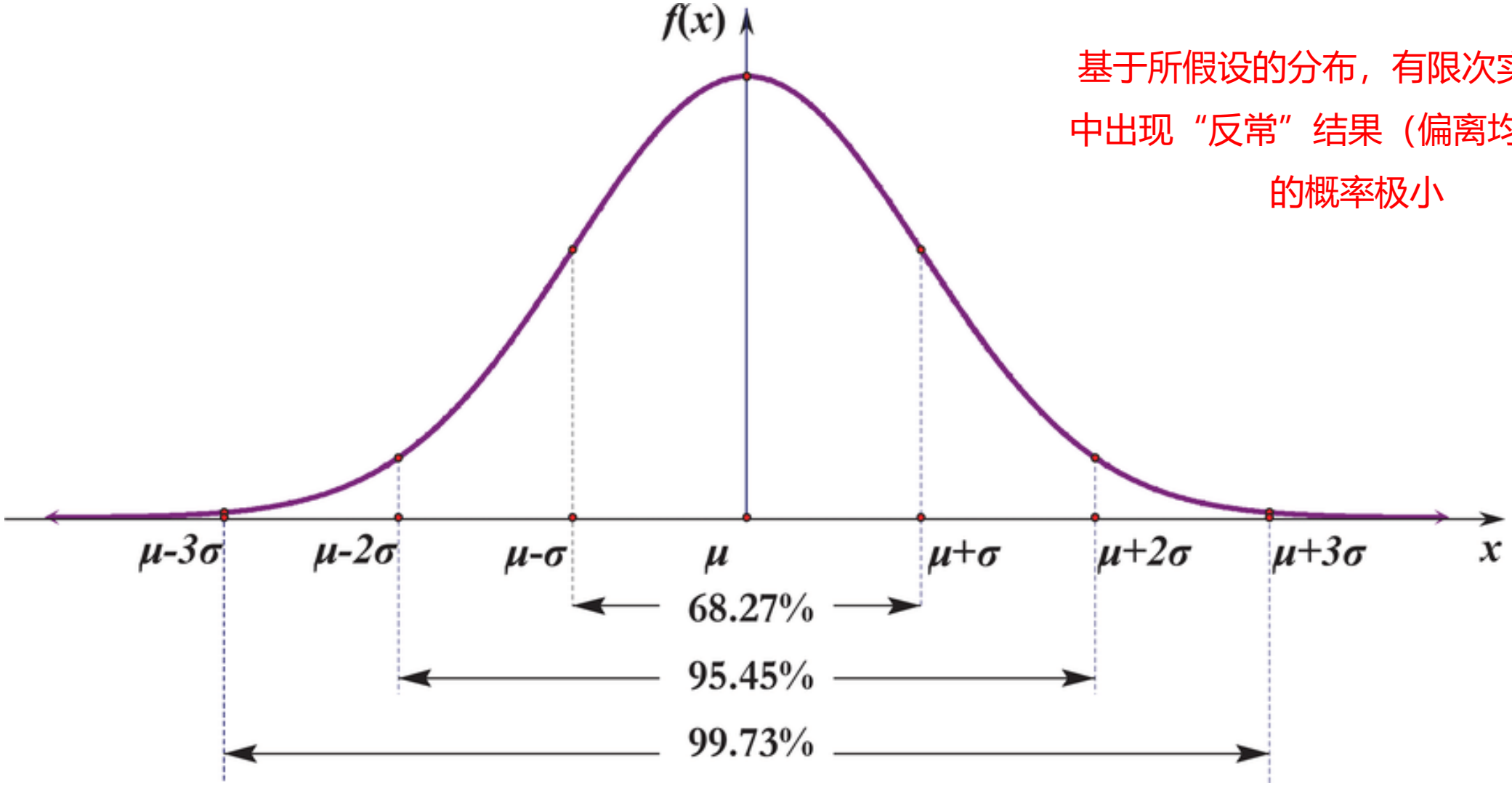
1. 模型检验定阶：
确定(p, d, q)

2. 模型优化：
如何学习模型参数

3. 模型检验：
建模是否充分

4. 模型拓展：
ARIMA的拓展

假设检验



基于所假设的分布，有限次实验结果中出现“反常”结果（偏离均值过大）的概率极小

什么是好的预测

1.预测残差具有如下性质：

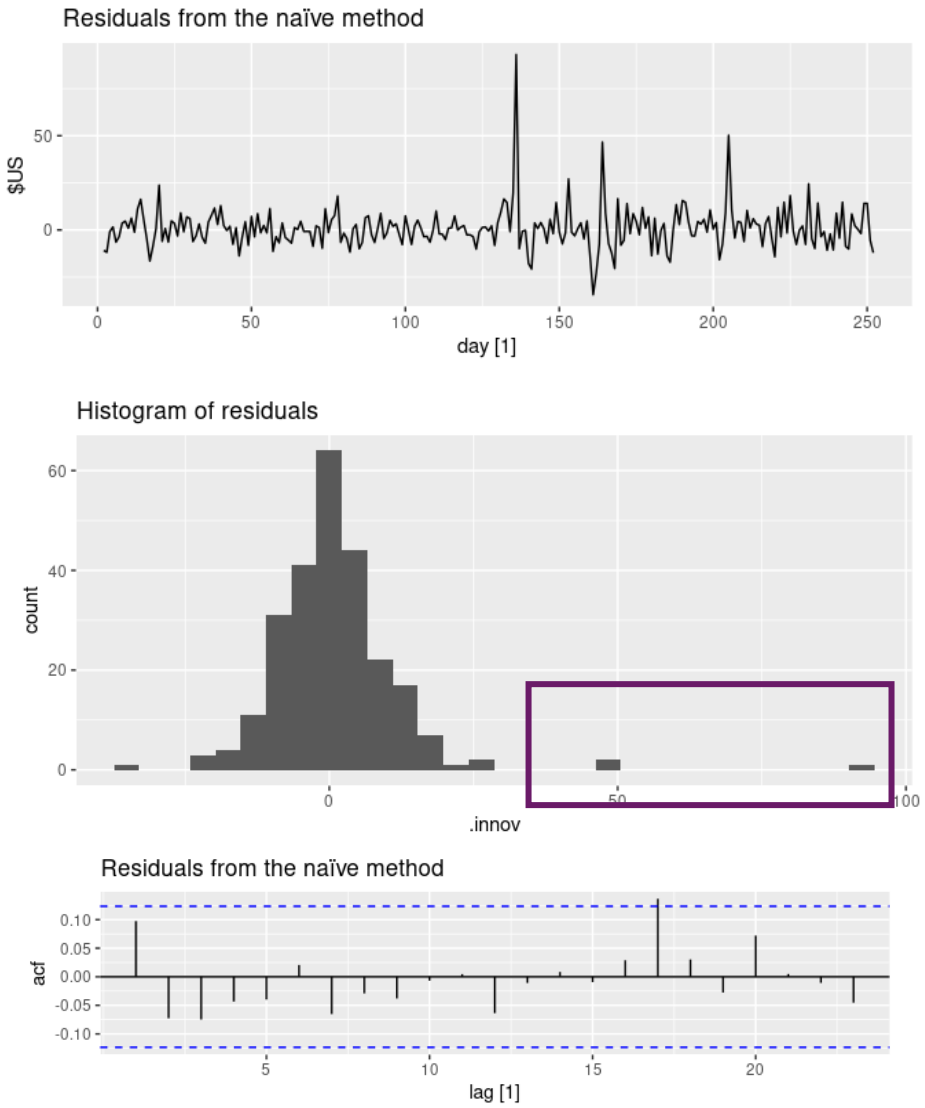
1.互不相关；否则可以使用其相关性进一步预测；

2.零均值；否则预测有偏

以下性质有助于计算预测区间

1.具有相同的方差（homoscedasticity）

2.为正态分布 → 使用Box-Cox变换



常用的假设检验

- **纯随机**检验（相当于一种**模型显著性**检验）
- 分别对r_k的取值进行检验
  - 对一组$\{r_k\}$进行检验，前l个ACF取值是否和白噪声有过大的差异（Portmanteau tests for autocorrelation ）
    - Box-Pierce test（1970）

$$Q = n\sum_{k=1}^{\ell} r_k^2$$

    - Ljung-Box test （1978）

$$Q^* = n(n+2)\sum_{k=1}^{\ell}(n-k)^{-1}r_k^2.$$

    - Q越大，越不可能是白噪声

- **平稳性**检验
  - 平稳序列的所有特征根都应该在单位圆内，单位根检验（unit root test）
  - DF检验（Dickey and Fuller），ADF检验
  - Kwiatkowski-Phillips-Schmidt-Shin (KPSS) 检验（Kwiatkowski et al., 1992）：备择假设为序列稳定

概要

1. 模型检验定阶：
确定(p, d, q)

2. 模型优化：
如何学习模型参数

3. 模型检验：
建模是否充分

4. 模型拓展：
ARIMA的拓展

动态回归模型

- Dynamic regression models
- Xreg："external regressors"
- 结合线性回归和ARIMA模型：

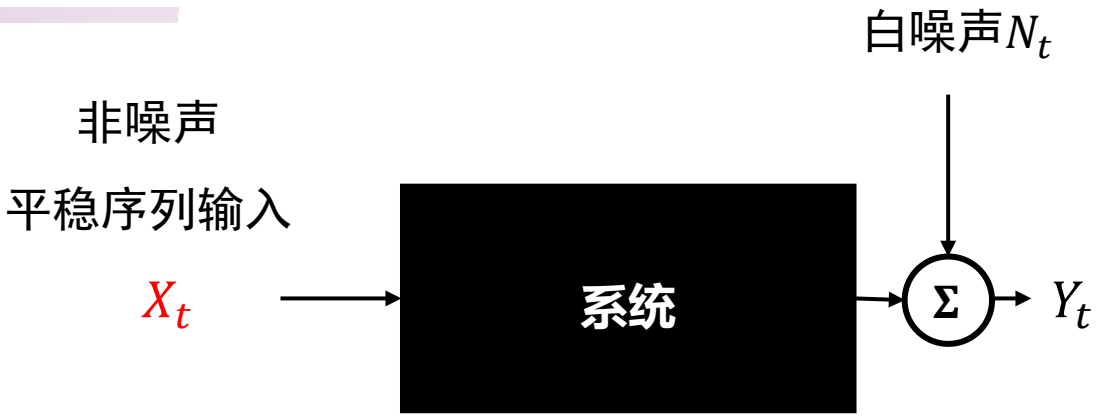
$$y_t = \beta_0 + \beta_1 x_{1,t} + \cdots + \beta_k x_{k,t} + \eta_t,$$
$$(1-\phi_1 B)(1-B)\eta_t = (1+\theta_1 B)\varepsilon_t$$

  - η_t：回归模型的误差
  - ε_t：ARIMA模型误差，白噪声
- 难点：
  - η_t不是白噪声，最小二乘估计相关的统计性质失效（AIC，假设检验等）
  - 若y_t以及$(x_{1,t},\ldots,x_{k,t})$不平稳，则无法得到一致的估计
- 对y_t进行差分，使用ARMA建模误差

$$y'_t = \beta_1 x'_{1,t} + \cdots + \beta_k x'_{k,t} + \eta'_t,$$
$$(1-\phi_1 B)\eta'_t = (1+\theta_1 B)\varepsilon_t,$$

- $y'_t = y_t - y_{t-1}$,
- $x'_{t,i} = x_{t,i} - x_{t-1,i}$
- $\eta'_t = \eta_t - \eta_{t-1}$

线性系统角度进行扩展



使用卷积刻画输入输出之间的关系：

$$Y = X * h$$

$$Y_t = \sum_{k=-\infty}^{\infty} h_k X_{t-k}$$

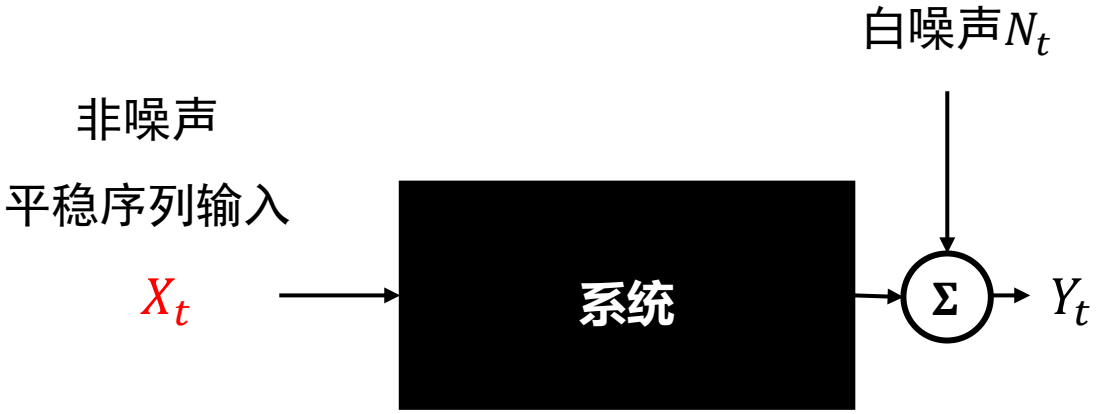
假设X和Y之间只有**单向（No feedback）**影响

- 输出包含**观测噪声**N_t

$$Y_t = \sum_{i=-\infty}^{\infty} h_i X_{t-i} + N_t$$

利用输入、输出的互相关系数$\hat{\gamma}_{XY}(k)$估计系统的性质，若输入为**白噪声**，可通过$\gamma_{XY}(k) = h_k \sigma_\varepsilon^2$估计该系统的**单位脉冲响应**

输入输出模型



- 输出就有观测噪声N_t

$$Y_t = \sum_{i=-\infty}^{\infty} h_i X_{t-i} + N_t$$

参数过多，需要进行**简化**

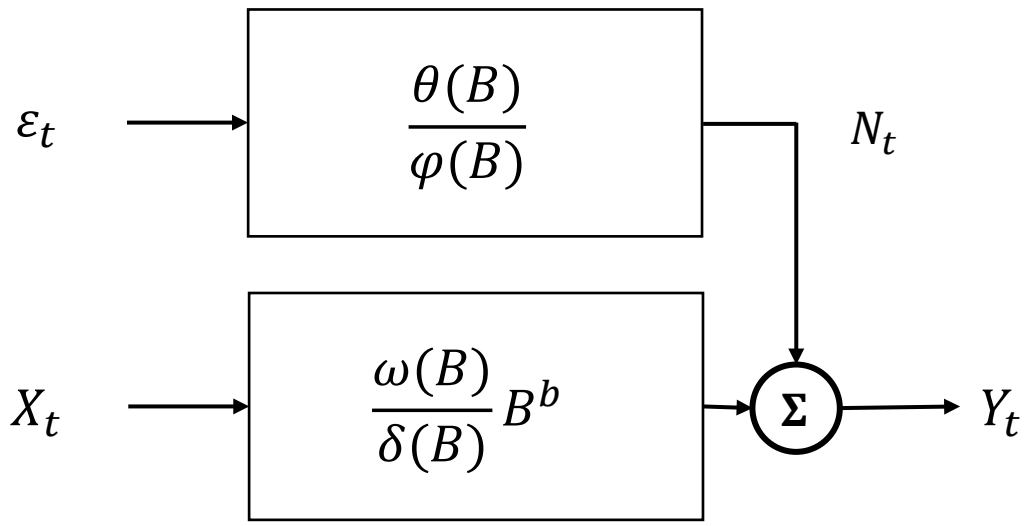
- 类比ARMA模型进行"因果形式"修改，有

$$Y_t + \delta_1 Y_{t-1} + \cdots + \delta_r Y_{t-r} = \omega_0 X_{t-b} + \omega_1 X_{t-b-1} + \cdots + \omega_s X_{t-b-s} + N_t'$$

b是时间延迟

- 输入输出模型（Input-Output Models）可以变化为不同形式的模型
  - 传递函数模型（Transfer Function Models）
  - 差分方程模型 （Difference Equation Models）
  - 输出误差模型（Output Error Models）

传递函数模型



- 假设N_t可使用ARMA模型刻画

$$N_t + \varphi_1 N_{t-1} + \cdots + \varphi_p N_{t-p}$$
$$= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

- 定义

$$\delta(\mathrm{B}) = 1 + \delta_1 \mathrm{B} + \cdots + \delta_r \mathrm{B}^r$$
$$\omega(\mathrm{B}) = \omega_0 + \omega_1 \mathrm{B} + \cdots + \omega_s \mathrm{B}^s$$
$$\varphi(\mathrm{B}) = 1 + \varphi_1 \mathrm{B} + \cdots + \varphi_p \mathrm{B}^p$$
$$\theta(\mathrm{B}) = 1 + \theta_1 \mathrm{B} + \cdots + \theta_q \mathrm{B}^q$$

则

$$Y_t + \delta_1 Y_{t-1} + \cdots + \delta_r Y_{t-r} = \omega_0 X_{t-b} + \omega_1 X_{t-b-1} + \cdots + \omega_s X_{t-b-s} + N_t'$$

可写为

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\varphi(B)} \varepsilon_t$$

传递函数成分

也称为Box-Jenkins模型

差分方程模型ARMAX

- 已知

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\varphi(B)} \varepsilon_t$$

其中

$$\delta(\mathrm{B}) = 1 + \delta_1 \mathrm{B} + \cdots + \delta_r \mathrm{B}^r$$
$$\omega(\mathrm{B}) = \omega_0 + \omega_1 \mathrm{B} + \cdots + \omega_s \mathrm{B}^s$$
$$\varphi(\mathrm{B}) = 1 + \varphi_1 \mathrm{B} + \cdots + \varphi_p \mathrm{B}^p$$
$$\theta(\mathrm{B}) = 1 + \theta_1 \mathrm{B} + \cdots + \theta_q \mathrm{B}^q$$

- 进一步简化

$$\delta(\mathrm{B}) = \varphi(B)$$

则

$$\varphi(\mathrm{B}) Y_t = \omega(\mathrm{B}) X_{t-b} + \theta(\mathrm{B}) \varepsilon_t$$

- ARMAX （AutoRegressive-Moving Average with eXogenous input ）模型
- 也被称为CARMA（Controlled ARMA）模型
- ARMAX模型的一些特殊形式
  - ARX模型：$\theta(B) = 1$

$$\varphi(\mathrm{B}) Y_t = \omega(\mathrm{B}) X_{t-b} + \varepsilon_t$$

  - FIR (Finite Impulse Response) : $\theta(B) = \varphi(B) = 1$

$$Y_t = \omega(\mathrm{B}) X_{t-b} + \varepsilon_t$$

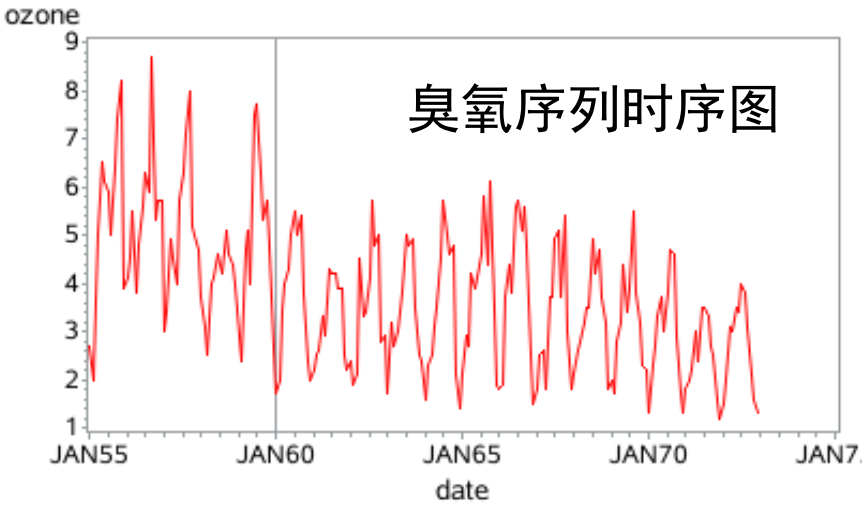
干预模型

- 稳定的时间序列可能由于某些事件的发生变得不稳定
- 干预模型（Intervention Model）描述由于外界因素导致的输出的变化，将输入建模为定性变量（qualitative variable）。引入干预函数

$$I_t = \begin{cases} 1 & t = t_0 \\ 0 & t \neq t_0 \end{cases}$$

则干预模型为

$$Y_t = \frac{\omega(B)}{\delta(B)} I_t + \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

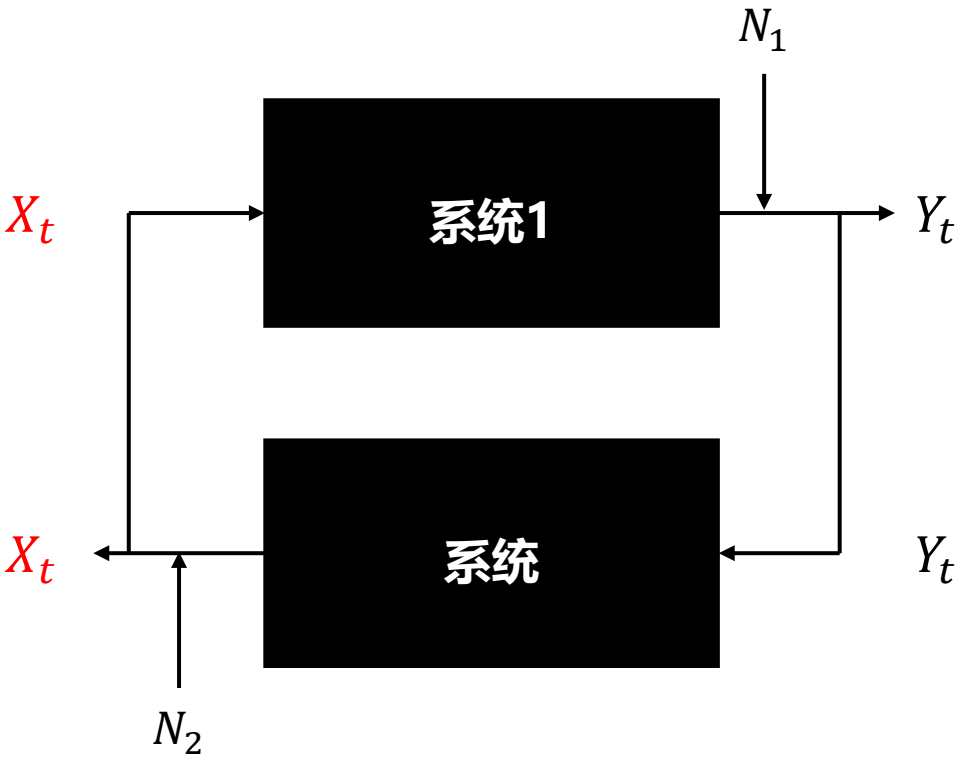


- Box和Tiao把干预变量（63号法令的颁布和执行）以虚拟变量的方式进行处理

$$X_{1t} = \begin{cases} 0, t < 1960\text{年}1\text{月} \\ 1, t \geq 1960\text{年}1\text{月} \end{cases}$$

类似地，也定义虚拟变量描述季节的影响

基于系统构建多元时间序列



$$Y_t = h_1(B)X_t + N_{1,t}$$
$$X_t = h_2(B)Y_t + N_{2,t}$$

- 矩阵形式表示

$$\begin{bmatrix} \delta(B)\varphi_\varepsilon(B) & -\varphi_\varepsilon(B)\omega(B)B^b \\ 0 & \varphi_\eta(B) \end{bmatrix} \begin{bmatrix} Y_t \\ X_t \end{bmatrix} = \begin{bmatrix} \delta(B)\theta_\varepsilon(B) & 0 \\ 0 & \theta_\eta(B) \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ \eta_t \end{bmatrix}$$

- 对于多元ARMA模型，将0替换为关于B的多项式，使ε_t和η_t之间相关性非零，并扩展到不同维度
- 多元ARMA模型

$$\phi(B)(\boldsymbol{Y}_t - \boldsymbol{c}) = \boldsymbol{\theta}(B)\boldsymbol{\epsilon}_t$$

- $\boldsymbol{\epsilon}_t$的方差矩阵为$\boldsymbol{\Sigma}$
- $\phi(B)$和$\boldsymbol{\theta}(B)$的元素是关于B的多项式

ARCH

- $ARMA(p,q)$模型：$\phi(\mathbf{B})Y_t = \theta_0 + \theta(\mathbf{B})\eta_t \Rightarrow$

$$Y_t = \mathbb{E}[Y_t|F_{t-1}] + \eta_t\text{，且}V[\eta_t] = \sigma_\varepsilon^2$$

- AutoRegressive Conditional Heteroscedastic (ARCH) [Engle, 1982]
  - 对（条件）**方差**进行假设$\sigma_\varepsilon^2 = V[\eta_t|F_{t-1}]$

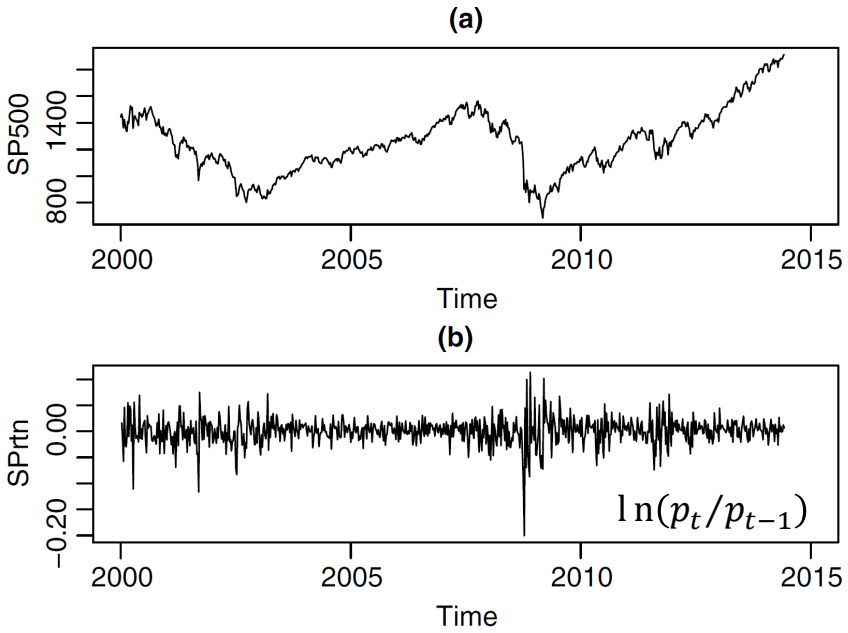
$$\eta_t = \sigma_t \varepsilon_t$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \eta_{t-1}^2 + \cdots + \alpha_s \eta_{t-s}^2$$

其中， ε_t为均值为0，方差为1的i.i.d.变量；

$$\alpha_0, \alpha_s > 0, \alpha_i \geq 0 \ (1 \leq i \leq s-1)$$

$$\sum_{i=1}^{s} \alpha_i < 1$$



- 扩展：Generalized Conditional Heteroscedastic (GARCH) [Bollerslev, 1986]