时间序列分析（初级）

非平稳时间序列分析 (II)

陈堃翰

安徽大学 大数据与统计学院

本章结构

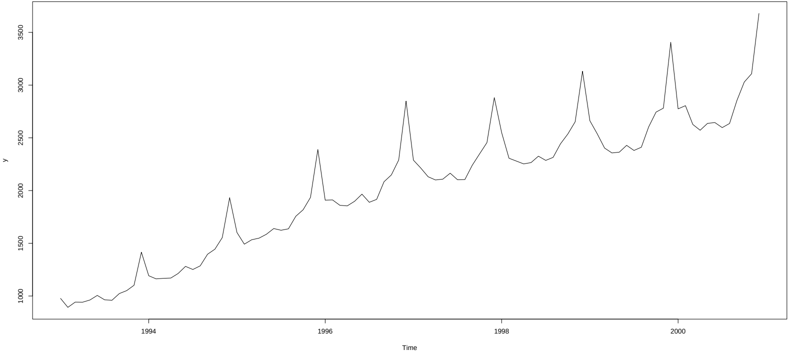
- 因素分解
  - 趋势效应的提取
  - 季节效应的提取
- 指数平滑预测模型
- ARIMA 加法(乘法)模型

因素分解

因素分解理论

- 因素分解：$x_t = f\left(T_t, C_t, S_t, I_t\right)$
- 常用的因素分解模型
  - 加法模型：$x_t = T_t + S_t + D_t + I_t$
  - 乘法模型：$x_t = T_t \times S_t \times D_t \times I_t$
  - 伪加法模型：$x_t = T_t \times (S_t + D_t + I_t)$
  - 对数加法模型：$\ln x_t = \ln T_t + \ln S_t + \ln D_t + \ln I_t$
- 如何对 T_t, S_t 建模？

例:中国社会消费品零售总额

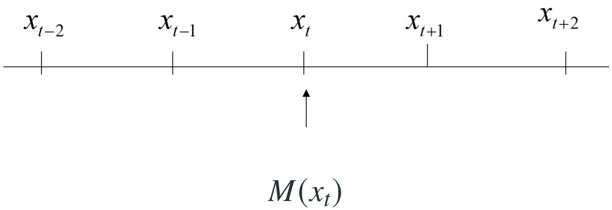


趋势效应的提取:移动平均法

- 移动平均法

$$M(x_t) = \sum_{i=-k}^{f} \theta_i x_{t-i}, \ \forall \ k, f \in \mathbf{N}^+ \tag{1}$$

- 当 $k = 0$ 且 $f \neq 0$ 时,移动平均
- 当 $k = f \neq 0$ 时,中心移动平均



一元一次线性趋势的提取：中心移动平均法

- 对于一元一次线性趋势 $x_t = a + bt + \varepsilon_t$， $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$

$$
\begin{aligned}
M(x_t) &= \sum_{i=-k}^{k} \theta_i x_{t-i} \\
&= \sum_{i=-k}^{k} \theta_i \left[a + b(t-i) + \varepsilon_{t-i}\right] \\
&= a \sum_{i=-k}^{k} \theta_i + bt \underbrace{\sum_{i=-k}^{k} \theta_i}_{=1} - b \underbrace{\sum_{i=-k}^{k} i\theta_i}_{=0} + \sum_{i=-k}^{k} \theta_i \varepsilon_{t-i}
\end{aligned}
$$

- $\mathrm{Var}\left[M(x_t)\right] = \sum_{i=-k}^{k} \theta_i^2 \sigma^2$
- $\mathbb{E}\left[M(x_t)\right] = \mathbb{E}(x_t)$，且当 $\theta_i = \frac{1}{2k+1}$ 时，$\mathrm{Var}\left[M(x_t)\right]$ 最小
- 当 $\theta_i = \frac{1}{2k+1}$ 时，称为简单中心移动平均法

一元二次趋势的提取:简单中心移动平均法

- 对于一元二次趋势函数 $x_t = a+bt+ct^2+\varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}\left(0, \sigma^2\right)$

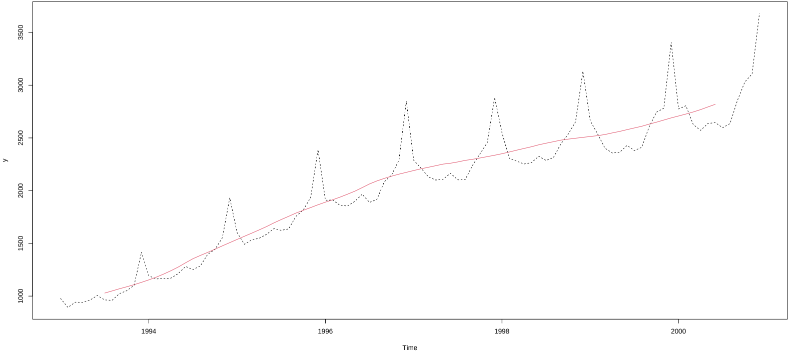
$$\begin{aligned} M\left(x_t\right) &= \frac{1}{2k+1}\sum_{i=-k}^{k} x_{t-i} \\ &= \frac{1}{2k+1}\sum_{i=-k}^{k}\left[a+b(t-i)+c(t-i)^2+\varepsilon_{t-i}\right] \\ &= a+bt+ct^2+c\frac{k(k+1)}{3}+\frac{1}{2k+1}\sum_{i=-k}^{k}\varepsilon_{t-i} \end{aligned}$$

- $\mathbb{E}\left[M(x_t)\right] \neq \mathbb{E}(x_t)$, 且

$$\mathbb{E}\left[M(x_t)\right]-\mathbb{E}\left(x_t\right)=\frac{k(k+1)}{3}$$

中国社会消费品零售总额趋势效应提取:移动平均法

- $M_{2\times12}(x_t)$:红色实线



季节效应的提取：加法（乘法）模型中季节指数的构造

加法模型：

- 去除趋势效应

$$y_t = x_t - T_t = S_t + I_t$$

- 季节效应

$$y_{(i-1)m+j} = \bar{y} + S_j + I_{(i-1)m+j}$$

- 计算总均值

$$\bar{y} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m} y_{m(i-1)+j}}{km}$$

- 计算每季度均值

$$\bar{y}_j = \frac{1}{k} \sum_{i=1}^{k} y_{(i-1)m+j}$$

- 季节指数

$$S_j = \bar{y}_j - \bar{y}$$

乘法模型：

- 去除趋势效应

$$y_t = \frac{x_t}{T_t} = S_t \times I_t$$

- 季节效应

$$y_{(i-1)m+j} = \bar{y} \times S_j \times I_{(i-1)m+j}$$

- 计算总均值

$$\bar{y} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{m} y_{m(i-1)+j}}{km}$$

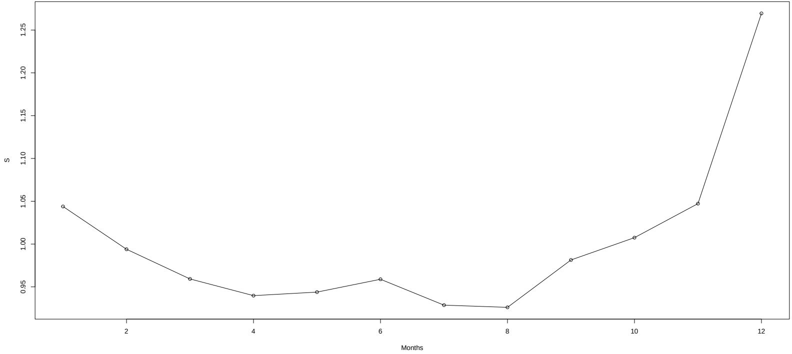
- 计算每季度均值

$$\bar{y}_j = \frac{1}{k} \sum_{i=1}^{k} y_{(i-1)m+j}$$

- 季节指数

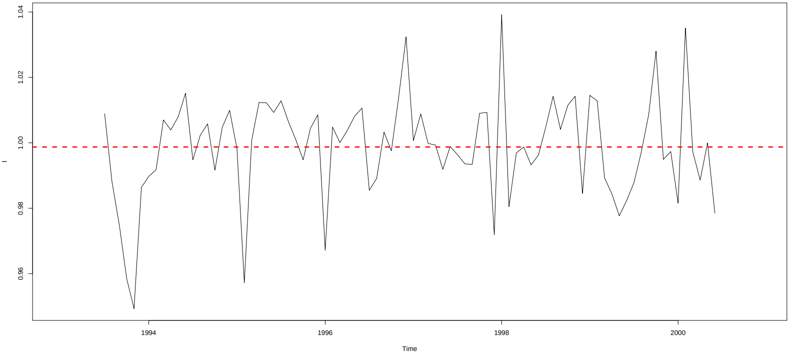
$$S_j = \frac{\bar{y}_j}{\bar{y}}$$

中国社会消费品零售总额季节效应提取:构造季节指数



中国社会消费品零售总额:剩余随机效应, I_t

- $I_t = \frac{x_t}{T_t S_t}$



指数平滑预测模型

简单指数平滑

- 几何级数（geometric sequence）和指数函数（expoential function）
- 指数平滑（平滑和预测）

$$\begin{aligned}\widehat{x}_{t+1} &= \alpha x_t + \alpha(1-\alpha)x_{t-1} + \alpha(1-\alpha)^2 x_{t-2} + \cdots \\ &= \alpha x_t + (1-\alpha)\left[\alpha x_{t-1} + \alpha(1-\alpha)x_{t-2}\right] \\ &= \alpha x_t + (1-\alpha)\widehat{x}_t \end{aligned} \tag{1}$$

- 简单指数平滑预测

$$\widehat{x}_{t+k} = \widehat{x}_{t+1} = \alpha x_t + (1-\alpha)\widehat{x}_t, \ \forall\, k \geqslant 1$$

Holt 两参数指数平滑

- 线性趋势序列 $x_t = a_0 + bt + \varepsilon_t$
- 递归表达

$$x_t = \underbrace{a_0 + b\,(t-1)}_{:=a_{t-1}} + \underbrace{b + \varepsilon_t}_{:=b_{t-1}}; \quad x_{t+1} = \underbrace{a_0 + bt}_{:=a_t} + \underbrace{b + \varepsilon_{t+1}}_{:=b_t}$$

如何得到 $\widehat{a}_t$ 和 $\widehat{b}_t$?

- Holt 两参数指数平滑
  - 初始化

$$\widehat{a}_0 = x_0$$

$$\widehat{b}_0 = x_1 - x_0 \quad \text{或者 } \{\nabla x_t\} \text{ 的均值}$$

  - 指数平滑 $\widehat{a}_t, \widehat{b}_t$

$$\widehat{a}_t = \alpha x_t + (1-\alpha)\left(\widehat{a}_{t-1} + \widehat{b}_{t-1}\right)$$

$$\widehat{b}_t = \beta\left(\widehat{a}_t - \widehat{a}_{t-1}\right) + (1-\beta)\widehat{b}_{t-1}$$

- Holt 两参数指数平滑预测

$$\widehat{x}_{t+k} = \widehat{a}_t + \widehat{b}_t k, \quad \forall k \geqslant 1$$

Holt-Winters 三参数指数平滑:季节加法模型

- 考虑季节加法模型 $x_t = a_0 + bt + c_t + \varepsilon_t$
- 递归表达

$$x_t = \underbrace{a_0 + b(t-1)}_{:=a_{t-1}} + \underbrace{b + \varepsilon_t}_{:=b_{t-1}} + \underbrace{c_t}_{:=S_j+e_t}$$

$$x_{t+1} = \underbrace{a_0 + bt}_{:=a_t} + \underbrace{b + \varepsilon_{t+1}}_{:=b_t} + \underbrace{c_{t+1}}_{:=S_j+e_{t+1}}$$

其中,$S_1, S_2, \ldots, S_m$ 为 m 期季节指数。

- $\widehat{a}_t, \widehat{b}_t, \widehat{c}_t$?

Holt-Winters 三参数指数平滑:季节加法模型

- Holt-Winters 三参数指数平滑
  - 初始化

$$
\begin{aligned}
&\widehat{a}_0 = x_0 \\
&\widehat{b}_0 = \frac{1}{m}\left(\frac{x_{m+1}-x_1}{m} + \frac{x_{m+2}-x_2}{m} + \cdots + \frac{x_{m+m}-x_m}{m}\right) \\
&\widehat{c}_j = \frac{1}{k}\sum_{i=1}^{k}\left[x_{m(i-1)+j} - A_i\right] \qquad \text{for } j = 1, 2, \ldots, m
\end{aligned}
$$

其中, $A_i = \frac{\sum_{j=1}^{m} x_{m(i-1)+j}}{m} \qquad \text{for } i = 1, 2, \ldots, k$

- 指数平滑 $\widehat{a}_t, \widehat{b}_t, \widehat{c}_t$

$$\begin{aligned}
\widehat{a}_t &= \alpha\left[x_t - \widehat{c}_{t-m}\right] + (1-\alpha)\left(\widehat{a}_{t-1} + \widehat{b}_{t-1}\right) \\
\widehat{b}_t &= \beta\left[\widehat{a}_t - \widehat{a}_{t-1}\right] + (1-\beta)\,\widehat{b}_{t-1} \\
\widehat{c}_t &= \gamma^*\left[x_t - \widehat{a}_{t-1} - \widehat{b}_{t-1}\right] + (1-\gamma^*)\,\widehat{c}_{t-m} \\
&= \frac{\gamma^*}{1-\alpha}\left[x_t - \widehat{a}_t\right] + \left(1 - \frac{\gamma^*}{1-\alpha}\right)\widehat{c}_{t-m}
\end{aligned}$$

- Holt-Winters 三参数指数平滑预测

$$\widehat{x}_{t+k} = \widehat{a}_t + \widehat{b}_t k + \widehat{c}_{t+k}, \quad \forall k \geqslant 1,$$

$$\text{且 } \widehat{c}_{t+k} = \widehat{c}_{\mathrm{mod}(t+k,m)} \left(\text{或 } \widehat{c}_{t+k} = \widehat{S}_{\mathrm{mod}(t+k,m)}\right)$$

Holt-Winters 三参数指数平滑:季节乘法模型

- 考虑季节乘法模型 $x_t = (a_0 + bt + \varepsilon_t)\, c_t$
- 递归表达

$$x_t = \Bigg[\underbrace{a_0 + b(t-1)}_{:=a_{t-1}} + \underbrace{b + \varepsilon_t}_{:=b_{t-1}}\Bigg]\underbrace{c_t}_{:=S_j+e_t}$$

Holt-Winters 三参数指数平滑:季节乘法模型

- Holt-Winters 三参数指数平滑
  - 初始化

$$
\begin{aligned}
\widehat{a}_0 &= x_0 \\
\widehat{b}_0 &= \frac{1}{m}\left(\frac{x_{m+1}-x_1}{m}+\frac{x_{m+2}-x_2}{m}+\cdots+\frac{x_{m+m}-x_m}{m}\right) \\
\widehat{c}_j &= \left[\frac{1}{k}\sum_{i=1}^{k} x_{m(i-1)+j}\right] \Big/ \bar{A} \quad \text{for } j = 1, 2, \ldots, m
\end{aligned}
$$

其中, $\bar{A} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{m} x_{m(i-1)+j}}{km} \quad \text{for } i = 1, 2, \ldots, k$

- 指数平滑 $\widehat{a}_t, \widehat{b}_t, \widehat{c}_t$

$$
\begin{aligned}
\widehat{a}_t &= \alpha\left[x_t/\widehat{c}_{t-m}\right] + (1-\alpha)\left(\widehat{a}_{t-1} + \widehat{b}_{t-1}\right) \\
\widehat{b}_t &= \beta\left[\widehat{a}_t - \widehat{a}_{t-1}\right] + (1-\beta)\,\widehat{b}_{t-1} \\
\widehat{c}_t &= \gamma^*\left[x_t \Big/ \left(\widehat{a}_{t-1} + \widehat{b}_{t-1}\right)\right] + (1-\gamma^*)\,\widehat{c}_{t-m}
\end{aligned}
$$

- Holt-Winters 三参数指数平滑预测

$$
\widehat{x}_{t+k} = \left[\widehat{a}_t + \widehat{b}_t k\right]\widehat{c}_{t+k}, \quad \forall k \geqslant 1
$$

$$
\text{且 } \widehat{c}_{t+k} = \widehat{c}_{\mathrm{mod}(t+k,m)} \left(\text{或 } \widehat{c}_{t+k} = \widehat{S}_{\mathrm{mod}(t+k,m)}\right)
$$

ARIMA 加法（乘法）模型

ARIMA 加法模型

- 季节效应、趋势效应、随机效应线性叠加

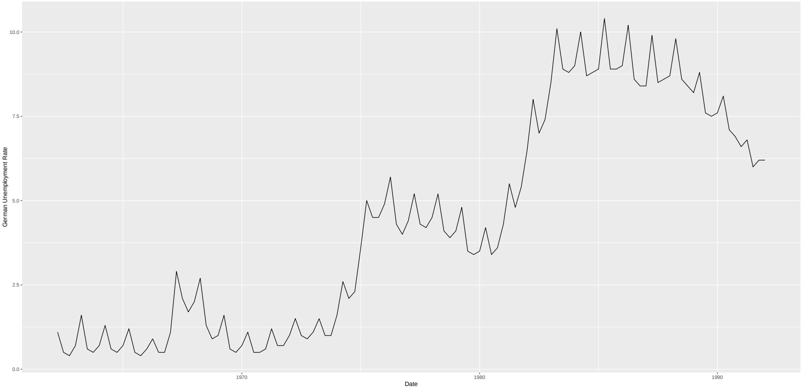
$$x_t = S_t + T_t + I_t$$

- 原始时间序列数据 x_t 在经过趋势差分、季节差分之后，可转换为平稳的基于 ARMA 模型进行分析的平稳序列

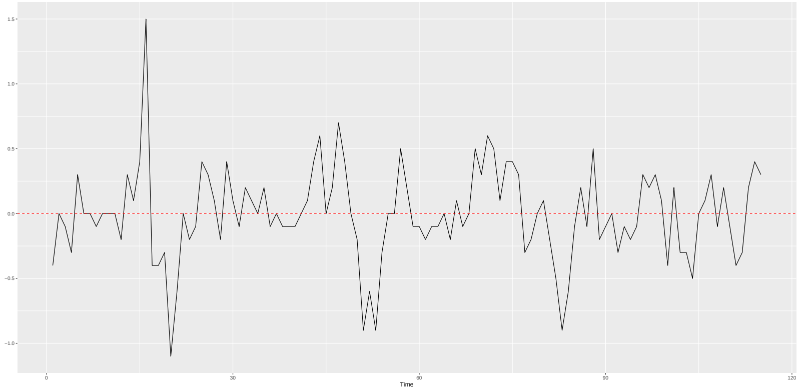
$$\nabla_S \nabla^d x_t = \frac{\Theta(B)}{\Phi(B)} \varepsilon_t$$

- 记为 ARIMA$(p,(d,S),q)$ 或者 ARIMA$(p,d,q) \times (0,1,0)_S$

例:1962–1991 年德国工人季度失业率



1962–1991 年德国工人季度失业率：1 阶 4 步差分（$\nabla_4\nabla^1$）



1962–1991 年德国工人季度失业率: 1 阶 4 步差分序列平稳性检验

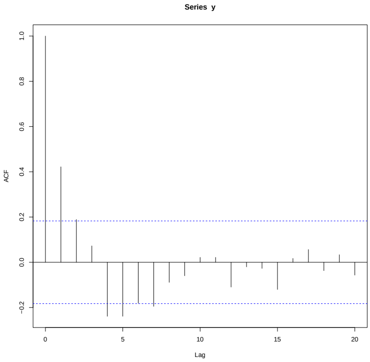
```
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag   ADF p.value
[1,]   0 -6.77    0.01
[2,]   1 -5.51    0.01
[3,]   2 -4.89    0.01
[4,]   3 -6.26    0.01
[5,]   4 -5.51    0.01
Type 2: with drift no trend
     lag   ADF p.value
[1,]   0 -6.74    0.01
[2,]   1 -5.48    0.01
[3,]   2 -4.86    0.01
[4,]   3 -6.23    0.01
[5,]   4 -5.49    0.01
Type 3: with drift and trend
     lag   ADF p.value
[1,]   0 -6.71    0.01
[2,]   1 -5.45    0.01
[3,]   2 -4.84    0.01
[4,]   3 -6.22    0.01
[5,]   4 -5.46    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

1962–1991 年德国工人季度失业率: 1 阶 4 步差分序列随机性检验

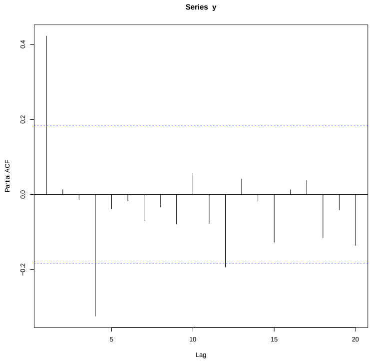
```
        Box-Ljung test

data:  y
X-squared = 43.837, df = 6, p-value = 7.964e-08


        Box-Ljung test

data:  y
X-squared = 51.708, df = 12, p-value = 6.982e-07


        Box-Ljung test

data:  y
X-squared = 54.476, df = 18, p-value = 1.547e-05
```

1962–1991 年德国工人季度失业率:1 阶 4 步差分序列自相关图和偏自相关图

- ACF



- PACF



1962–1991 年德国工人季度失业率: ARIMA 加法模型分析

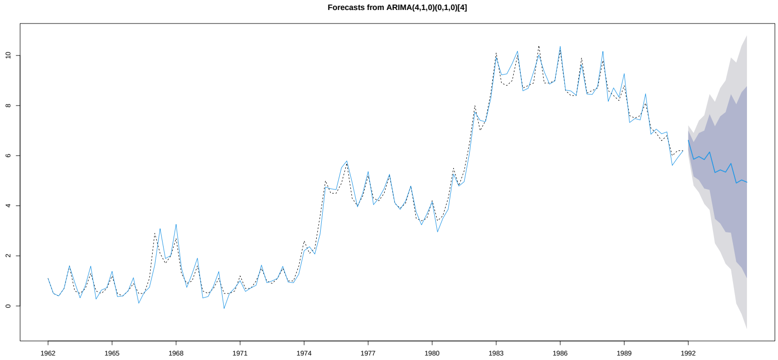
- stats::arima

```
Call:
arima(x = x, order = c(4, 1, 0), seasonal = list(order = c(0, 1, 0), period = 4),
    transform.pars = F, fixed = c(NA, 0, 0, NA))

Coefficients:
         ar1  ar2  ar3      ar4
      0.4449    0    0  -0.2720
s.e.  0.0807    0    0   0.0804

sigma^2 estimated as 0.09266:  log likelihood = -26.7,  aic = 59.39
```

- 模型

$$(1-B)\left(1-B^4\right)x_t = \frac{1}{1-0.4449B+0.272B^4}\varepsilon_t, \ \mathrm{Var}(\varepsilon_t) = 0.09266$$

1962–1991 年德国工人季度失业率: ARIMA 加法模型拟合、预测



- 实线表示失业率拟合值和预测值
- 虚线表示失业率观测值
- 深色阴影部分表示 80% 置信区间,浅色阴影部分表示 95% 置信区间

ARIMA 乘法模型

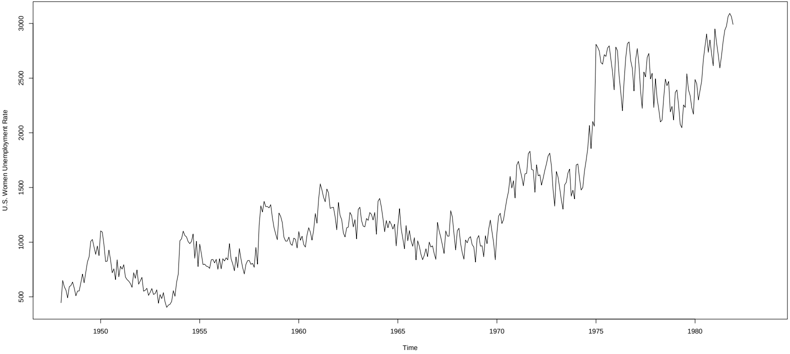
- 短期相关性用 ARMA(p, q) 提取
- 季节相关性用 S（周期长度）步长的 ARMA(P, Q) 提取
- 短期相关性和季节效应具有乘积关系

$$\nabla^d \nabla_S^D x_t = \frac{\Theta(B)}{\Phi(B)} \frac{\Theta_S(B)}{\Phi_S(B)} \varepsilon_t$$

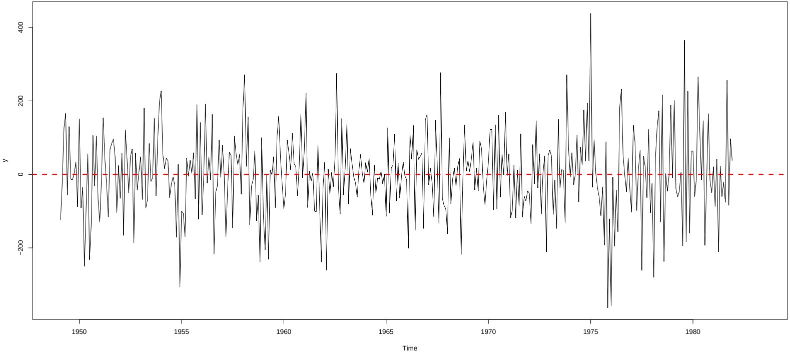
或者，

$$\Phi(B)\Phi_S(B)\nabla^d \nabla_S^D x_t = \Theta(B)\Theta_S(B)\varepsilon_t$$

例: 美国女性(20岁以上)月度失业率



美国女性(20岁以上)月度失业率: 1 阶 12 步差分 ($\nabla^1\nabla^1_{12}$)



美国女性(20岁以上)月度失业率: 1 阶 12 步差分序列平稳性检验

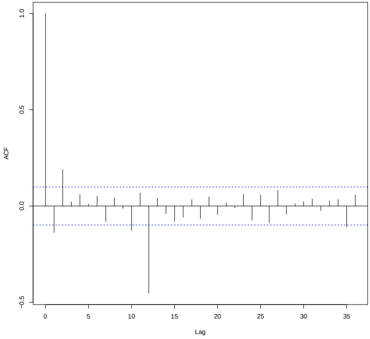
```
Augmented Dickey-Fuller Test
alternative: stationary

Type 1: no drift no trend
     lag    ADF p.value
[1,]   0 -22.81    0.01
[2,]   1 -12.52    0.01
[3,]   2  -9.79    0.01
Type 2: with drift no trend
     lag    ADF p.value
[1,]   0 -22.79    0.01
[2,]   1 -12.51    0.01
[3,]   2  -9.77    0.01
Type 3: with drift and trend
     lag    ADF p.value
[1,]   0 -22.76    0.01
[2,]   1 -12.49    0.01
[3,]   2  -9.77    0.01
----
Note: in fact, p.value = 0.01 means p.value <= 0.01
```

美国女性(20岁以上)月度失业率: 1 阶 12 步差分序列随机性检验

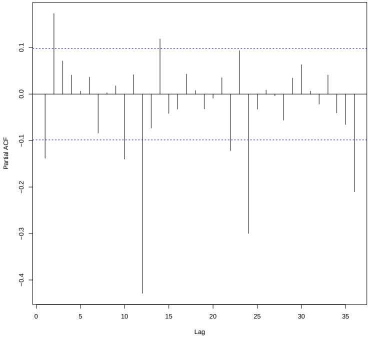
```
        Box-Ljung test

data:  y
X-squared = 24.692, df = 6, p-value = 0.0003894


        Box-Ljung test

data:  y
X-squared = 121.08, df = 12, p-value < 2.2e-16


        Box-Ljung test

data:  y
X-squared = 128.87, df = 18, p-value < 2.2e-16
```

美国女性(20岁以上)月度失业率: 1 阶 12 步差分序列自相关图和偏自相关图

- ACF



- PACF



美国女性(20岁以上)月度失业率: ARIMA 乘法模型分析

- `stats::arima`

```
Call:
arima(x = x, order = c(1, 1, 1), seasonal = list(order = c(0, 1, 1), period = 12))

Coefficients:
          ar1     ma1     sma1
      -0.7265  0.6030  -0.7918
s.e.   0.1511  0.1742   0.0337

sigma^2 estimated as 7444:  log likelihood = -2327.14,  aic = 4662.28
```

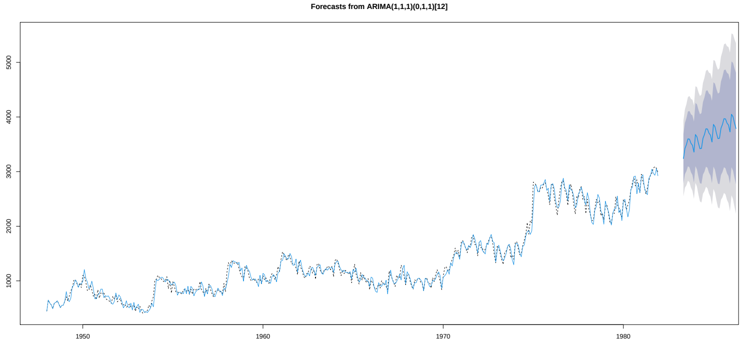
- 模型

$$\nabla\nabla_{12}x_t = \frac{1+0.6030B}{1+0.7265B}\left(1-0.7918B^{12}\right)\varepsilon_t$$

- 注意,在 R 的 `stats::arima` 函数中,ARMA(p, q) 模型表示为

$$x_t - \phi_1 x_{t-1} - \phi_2 x_{t-2} - \cdots - \phi_p x_{t-p} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

美国女性(20岁以上)月度失业率: ARIMA 乘法模型拟合、预测



- 实线表示失业率拟合值和预测值
- 虚线表示失业率观测值
- 深色阴影部分表示 80% 置信区间,浅色阴影部分表示 95% 置信区间