时间序列分析第五章作业

Phlinsia

2024 年 4 月 22 日

Contents

0.1 练习 5.1

识别特定的 ARIMA 模型，即确定 p、d 和 q 的值以及参数 ϕ 和 θ 的取值。

（**a**）$Y_t = Y_{t-1} - 0.25Y_{t-2} + e_t - 0.1e_{t-1}$

此模型应该为 **ARMA(2,1)** 模型，其中 $\phi_1 = 1, \phi_2 = -0.25$。检查平稳性条件，按照不等式：

$$\begin{cases} \phi_1 + \phi_2 = 0.75 < 1 \\ \phi_2 - \phi_1 = -1.25 < 1 \\ |\phi_2| = 0.25 < 1 \end{cases}$$

由于所有条件均满足，该过程为平稳且可逆的 **ARMA(2,1)** 模型，其中 $\phi_1 = 1, \phi_2 = -0.25, \theta_1 = 0.1$。

（**b**）$Y_t = 2Y_{t-1} - Y_{t-2} + e_t$

初看似乎为一个 **AR(2)** 模型，但 $2 + (-1) = 1$，不严格小于 1。改写方程：

$$Y_t - Y_{t-1} = (Y_{t-1} - Y_{t-2}) + e_t$$

似乎是差分 $Y_t - Y_{t-1}$ 上的一个 **AR(1)** 模型，但 AR 系数将等于 1。然而，二阶差分：

$$Y_t - 2Y_{t-1} + Y_{t-2} = e_t$$

为白噪声，表明 $\{Y_t\}$ 为一个 **IMA(2,0)** 模型。

（**c**）$Y_t = 0.5Y_{t-1} - 0.5Y_{t-2} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$

对于 AR 部分，满足了平稳性检验不等式，确保了平稳性。将相同方程应用到模型的 MA 部分，发现 MA 部分可逆。因此，该模型为平稳且可逆的 **ARMA(2,2)** 模型，其中 $\phi_1 = 0.5, \phi_2 = -0.5, \theta_1 = 0.5, \theta_2 = -0.25$

0.2 练习 5.2

对于下列每个 ARIMA 模型，给出 $E(\nabla Y_t)$ 和 $Var(\nabla Y_t)$ 的值。

（**a**）$Y_t = 3 + Y_{t-1} + e_t - 0.75e_{t-1}$

$$\because \nabla Y_t = Y_t - Y_{t-1} = 3 + e_t - 0.75e_{t-1}$$
$$\therefore E(\nabla Y_t) = 3, Var(\nabla Y_t) = [1 + (0.75)^2]\sigma_e^2 = \left[\frac{25}{16}\right]\sigma_e^2$$

（**b**）$Y_t = 10 + 1.25Y_{t-1} - 0.25Y_{t-2} + e_t - 0.16e_{t-1}$

$$\nabla Y_t = Y_t - Y_{t-1} = 10 + 0.25(Y_{t-1} - Y_{t-2}) + e_t - 0.1e_{t-1}$$

所以该模型为平稳、可逆的 ARIMA(1,1,1) 模型，其中 $\phi = 0.25$，$\theta = 0.1$，$\theta_0 = 10$

$$E(\nabla Y_t) = \frac{\theta_0}{1-\phi} = \frac{10}{1-0.25} = \frac{10}{0.75} = \frac{40}{3}$$

$$Var(\nabla Y_t) = \frac{(1 - 2\phi\theta + \theta^2)}{1-\phi^2}\sigma_e^2 = \frac{(1 - 2(0.25)(0.1) + (0.1)^2)}{1-(0.25)^2}\sigma_e^2 = 1.024\sigma_e^2$$

（**c**）$Y_t = 5 + 2Y_{t-1} - 1.7Y_{t-2} + 0.7Y_{t-3} + e_t - 0.5e_{t-1} + 0.25e_{t-2}$

对 AR 特征多项式进行因式分解，得：

$$1 - 2x + 1.7x^2 - 0.7x^3 = (1-x)(1 - x + 0.7x^2)$$

需要进行一次差分才能得到一个平稳的 AR(2) 过程。模型可以重写为：

$$\nabla Y_t = 5 + \nabla Y_{t-1} - 0.7\nabla Y_{t-2} + e_t - 0.5e_{t-1} + 0.252e_{t-2}$$

所以模型是一个 ARIMA(2,1,2)，其中 $\theta_0 = 5, \theta_1 = 0.5, \theta_2 = -0.25, \phi_1 = 1, \phi_2 = -0.25, \phi_0 = 5$

0.3 练习 5.3

设 $\{Y_t\}$ 按以下方式生成：

$$Y_t = e_t + ce_{t-1} + ce_{t-2} + \cdots + ce_{t-n} + ce_0$$

又 $t > 0$

（**a**）$\{Y_t\}$ 的均值函数与协方差函数，它平稳吗？

Y_t 的均值和方差分别为：

$$E(Y_t) = 0$$
$$\mathrm{Var}(Y_t) = \mathrm{Var}(e_t + ce_{t-1} + ce_{t-2} + \cdots + ce_0) = (1 + tc^2)\sigma_e^2$$

均值函数与协方差函数均随 t 变化。

假设 $t < s$，则 Y_t 与 Y_s 之间的协方差为：

$$\begin{aligned}\mathrm{Cov}(Y_t, Y_s) &= \mathrm{Cov}(e_t + ce_{t-1} + ce_{t-2} + \cdots + ce_0, e_s + ce_{s-1} + \cdots + ce_t + ce_{t-1} + \cdots + ce_0) \\ &= (c + c^2 t)\sigma_e^2 = c(1+ct)\sigma_e^2\end{aligned}$$

（**b**）$\{\nabla Y_t\}$ 的均值函数与协方差函数

差分过程 $\{\nabla Y_t\}$ 的均值和协方差函数：

$$\nabla Y_t = (e_t + ce_{t-1} + ce_{t-2} + \cdots + ce_0) - (ce_t + ce_{t-1} + \cdots + ce_0) = e_t - (1-c)e_{t-1}$$

c 取何值都平稳。

（**c**）将 $\{Y_t\}$ 识别为特定 **ARIMA** 过程

过程 $\{\nabla Y_t\}$ 为 MA(1) 过程，$\therefore\{Y_t\}$ 为 IMA(1,1) 或 ARIMA(0,1,1) 其中 $\theta = 1 - c$。若 $|c| < 1$，则 $\{\nabla Y_t\}$ 过程可逆。

0.4 练习 5.4

设 $Y_t = A + Bt + X_t$，$\{X_t\}$ 随机游走。考虑 A 和 B 为常数时的情况。

（**a**）$\{Y_t\}$ 的平稳性

$E(Y_t) = A + Bt$ 随 t 变化，过程 $\{Y_t\}$ 非平稳。

（**b**）$\{\nabla Y_t\}$ 的平稳性

$$\begin{gathered}\nabla Y_t = Y_t - Y_{t-1} = (A + Bt + X_t) - (A + B(t-1) + X_{t-1}) = B + X_t - X_{t-1} = B + \nabla X_t \\ \therefore E(\nabla Y_t) = E(B) = B \\ \because \nabla X_t \text{是白噪声，而} B \text{为常数} \\ \mathrm{Cov}(\nabla Y_t, \nabla Y_{t-k}) = \mathrm{Cov}(B + \nabla X_t, B + \nabla X_{t-k}) = 0 \quad (k > 0)\end{gathered}$$

因此，$\{\nabla Y_t\}$ 平稳。

现在假设 A 和 B 是独立于随机游走的 $\{X_t\}$ 的随机变量。

（**c**）$\{Y_t\}$ 的平稳性

在此情况下，$E(Y_t) = E(A) + E(B)t$。由于期望随 t 变化，过程 $\{Y_t\}$ 非平稳。

（**d**）$\{\nabla Y_t\}$ 的平稳性

当 A 和 B 为随机变量时，我们仍有 $\nabla Y_t = B + \nabla X_t$。

\$ E(Y_t) = E(B) t\$ 上恒定。

$$\mathrm{Cov}(\nabla Y_t, \nabla Y_{t-k}) = \mathrm{Cov}(B + \nabla X_t, B + \nabla X_{t-k}) = \mathrm{Var}(B) \quad k\text{任取}$$

协方差函数仅依赖于滞后 k（不依赖 t），即使 A 和 B 为随机变量，$\{\nabla Y_t\}$ 仍保持平稳。

0.5 练习 5.7

考虑两个模型：

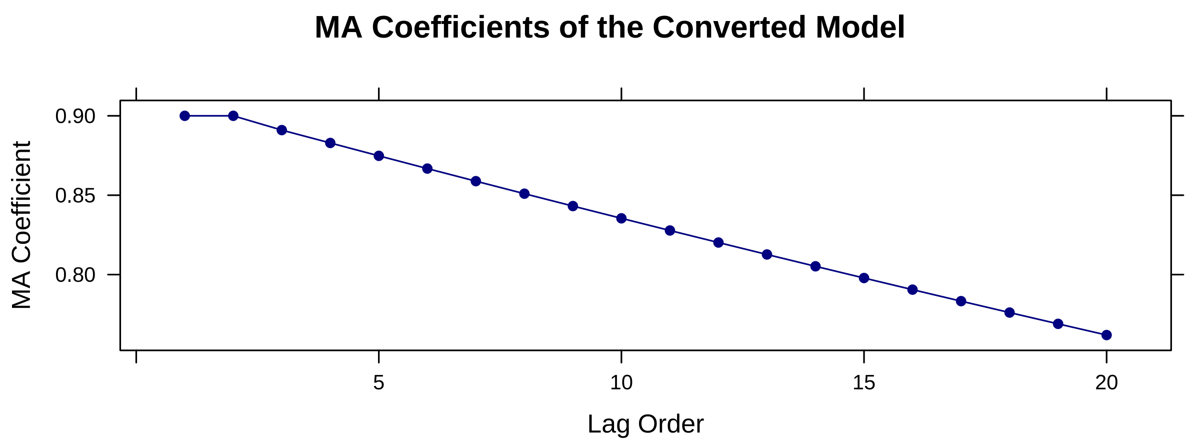
$$A: Y_t = 0.9Y_{t-1} + 0.09Y_{t-2} + e_t$$

由于

$$\begin{cases} \phi_1 + \phi_2 < 1 \\ \phi_2 - \phi_1 < 1 \\ |\phi_2| < 1 \end{cases}$$

故该过程为平稳 AR(2) 过程，其中 $\phi_1 = 0.9, \phi_2 = 0.09$ 。

```
# 使用 xyplot 绘制图形
xyplot(Coefficient ~ Lag, data = data.frame(Lag = 1:20,
      Coefficient = ARMAtoMA(ar=c(0.9,0.09),lag.max=20)),
       type = "o",       # 画出点并连线
       pch = 16,         # 点的样式
       col = "navy",     # 线和点的颜色
       xlab = "Lag Order", ylab = "MA Coefficient",
       main = "MA Coefficients of the Converted Model")
```



$$B: Y_t = Y_{t-1} + e_t - 0.1e_{t-1}$$

由于

$$Y_t - Y_{t-1} = e_t - 0.1e_{t-1}$$

故该过程为 IMA(1,1) 过程，其中 $\theta = 0.1$

$$\because Y_t \ = \ e_t + (1-\theta)e_{t-1} + (1-\theta)e_{t-2} + \cdots + (1-\theta)e_{-m} - \theta e_{-m-1}$$

$IMA(1,1)$ 模型的 ψ 权重为 $1, 1-0.1 = 0.9, 1-0.1 = 0.9, 1-0.1 = 0.9, \ldots$

因此，滞后阶数很大时，两个模型的 ψ 权重非常相似。

$$\because Y_t = \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \pi_3 Y_{t-3} + \cdots + e_t$$

对于 $k = 1, 2, \ldots$，有 $\pi_k = (1-\theta)\theta^{k-1}$。则 $\pi_1 = (1-0.1) = 0.9, \pi_2 = (1-0.1)(0.1) = 0.09$, $\pi_3 = (1-0.1)(0.1)^2 = 0.009$, 依此类推。

两个模型的前两个 π 权重完全相同，其余的 π 权重几乎一样。实践层面这两个模型几乎无法区分。

0.6 5.11.Winnebago

a

图 1 中的趋势看似接近指数增长，同时呈现出季节性模式，即销售在年尾有所下滑，春季月份则出现显著增长。

```
data(winnebago)
winnebago <- as.xts(winnebago)
xyplot(winnebago, ylab = " 销量")
```

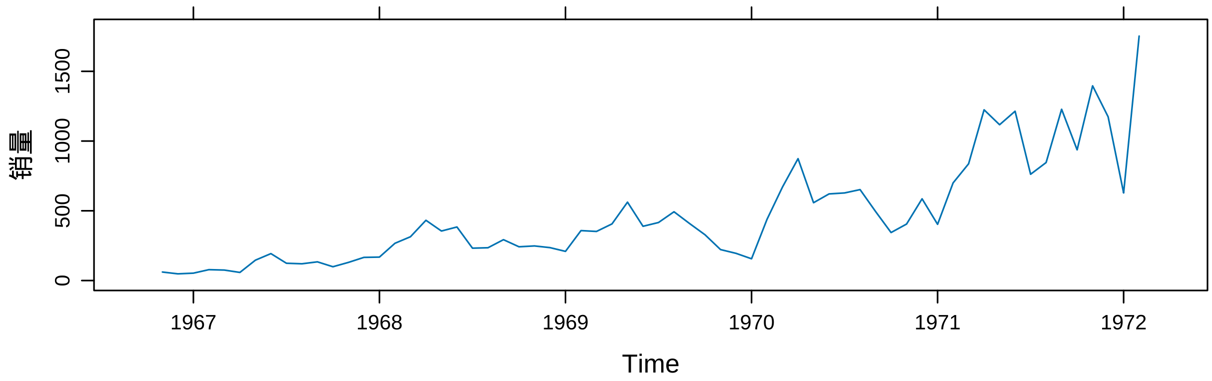


Figure 1: 休闲车月度销量.

b

我们对销量取对数后，发现趋势似乎变为随时间线性变化了 2。

```
winnebago_log <- log(winnebago)
xyplot(winnebago_log, ylab = expression(log(销量)))
```

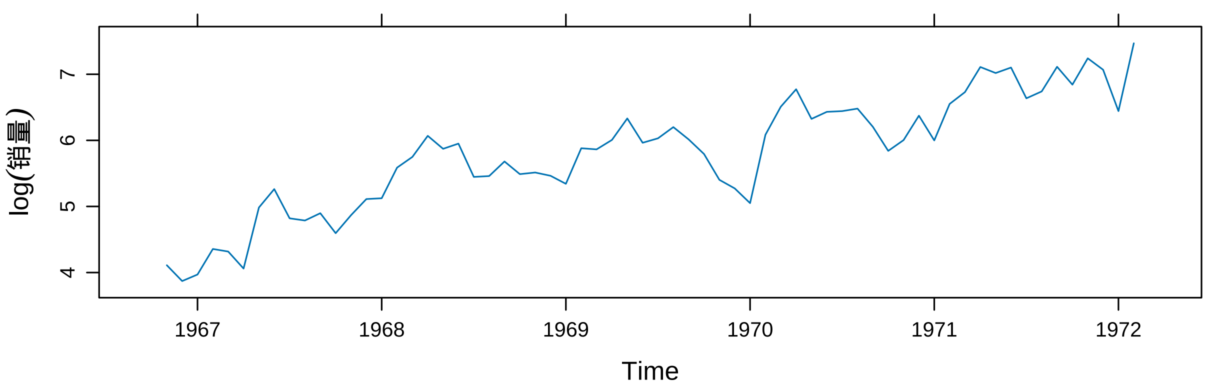


Figure 2: 对数形式的月度销量.

c

我们在图 3 中展示了结果。两种模式相似，但当销售额较大时，分数相对变化的幅度似乎更大。

```r
winnebago_frac <- diff(winnebago) / lag(winnebago, 1)
winnebago_logdiff <- diff(log(winnebago))
winnebago_comp <- merge.xts(frac = winnebago_frac, logdiff = winnebago_logdiff)
xyplot(winnebago_comp, screens = c(1, 1), auto.key = TRUE,
       ylab = " 销量",
       col = c("darkorange2", "navy"))
```

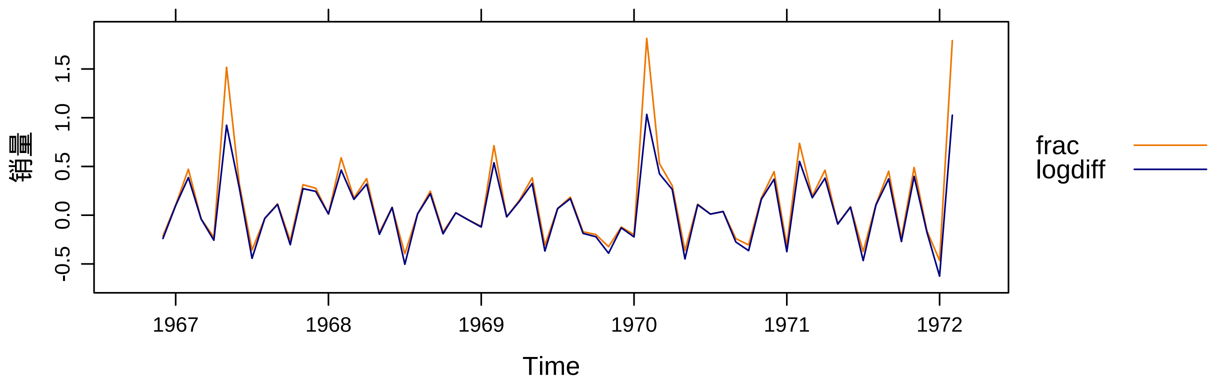


Figure 3: 对数差分与分数相对变化的比较.

0.7 5.12.Standard & Poor

a

时间序列中存在指数增长趋势 (图 4)，不过该趋势似乎在 1970 年左右开始趋于平稳。

```r
data(SP)
SP <- as.xts(SP)
xyplot(SP, grid = TRUE)
```

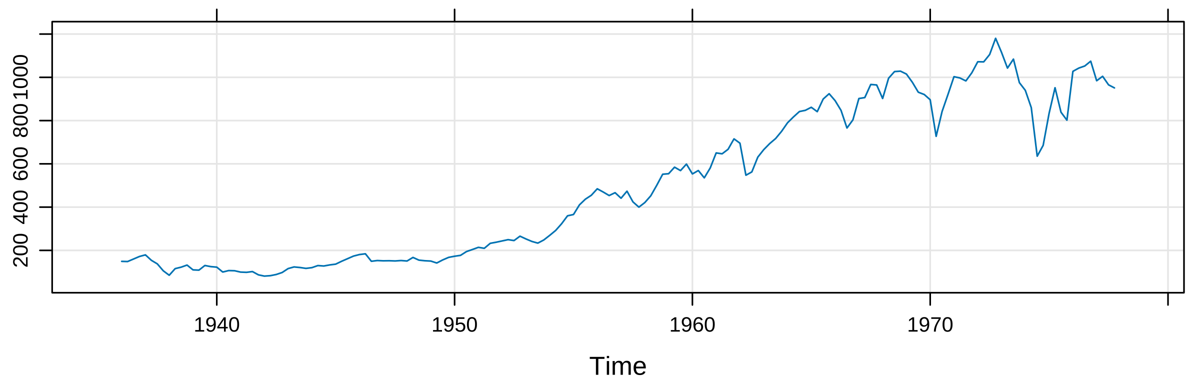


Figure 4: 标准普尔指数季度值.

b

我们在图 5 中对标准普尔指数的时间序列取对数。序列变得更加线性，但仍然存在指数模式。

```
sp_log <- log(SP)
xyplot(sp_log, ylab = expression(log(值)), grid = TRUE)
```

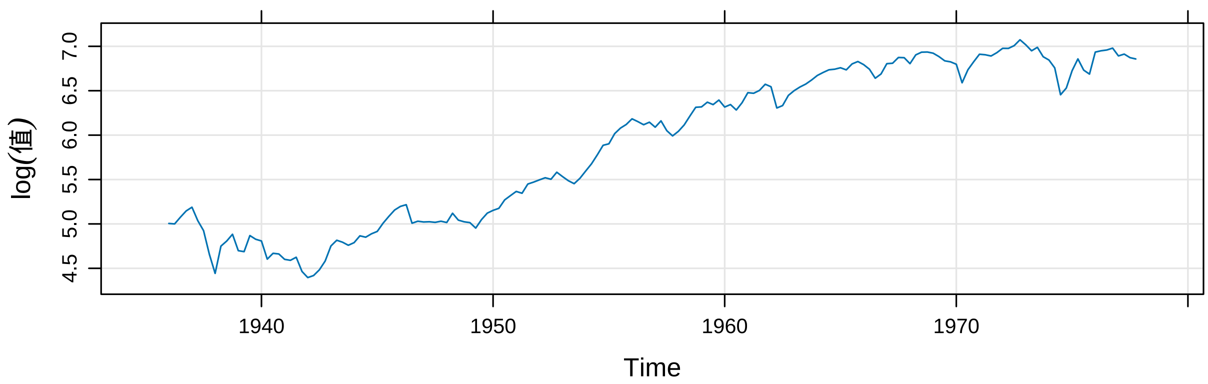


Figure 5: 标准普尔指数的对数形式.

c

接着，我们计算分数相对变化和自然对数差分 (图 6)。我们看到两序列之间差异很小，仅在较高的数值上略有不同。

```
SP_frac <- diff(SP) / lag(SP, 1)
SP_logdiff <- diff(log(SP))
SP_comp <- merge.xts(frac = SP_frac, logdiff = SP_logdiff)
xyplot(SP_comp, screens = c(1, 1), auto.key = TRUE,
       ylab = " 销量",
       col = c("darkorange2", "navy"))
```

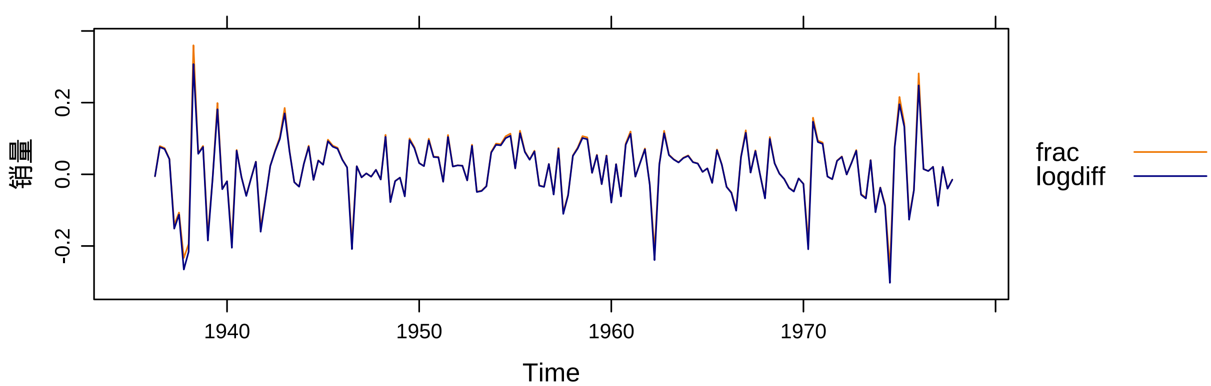


Figure 6: 对数差分与分数相对差异的比较.

0.8 5.13.Air passengers

a

我们在图 7 中绘制了月度国际航空乘客人数，注意到存在明显的季节性趋势，可能还存在轻微的指数增长。方差似乎也随着增大。

```
data(airpass)
airpass <- as.xts(airpass)
xyplot(airpass, ylab = " 乘客数")
```

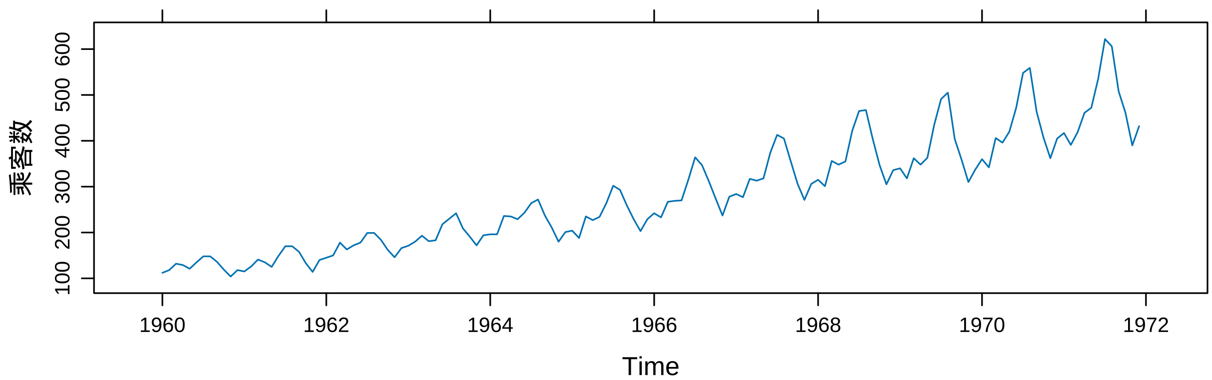


Figure 7: 月度航空乘客总数.

b

如同前面的练习，我们对因变量（本例中为乘客总数）取自然对数，发现已将趋势线性化（或者可能是指数递减?），同时也稳定了序列的方差。

```
airpass_log <- log(airpass)
xyplot(airpass_log, ylab = expression(log(航空乘客数)))
```

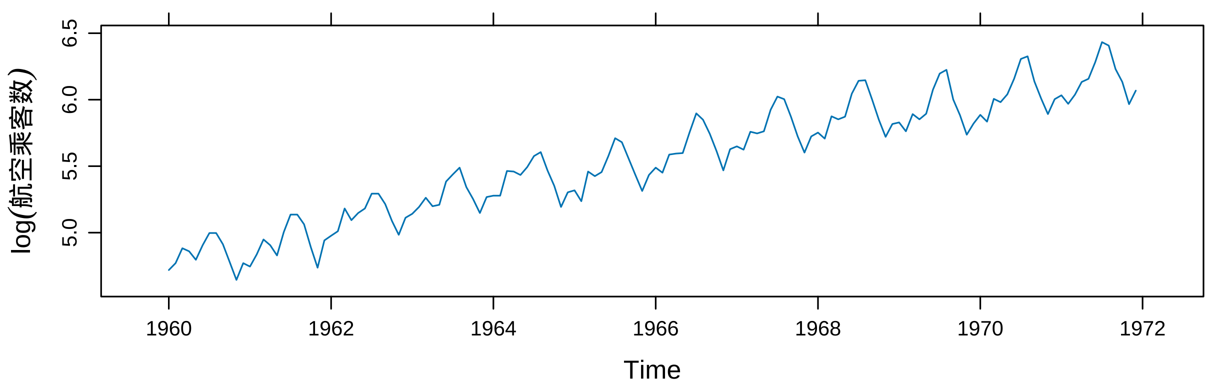


Figure 8: 对数形式的月度航空乘客总数.

c

接下来，我们计算分数相对变化和自然对数差分 (图 9)。我们看到两序列之间差异很小，仅在较高的数值上略有不同。

```
airpass_frac <- diff(airpass) / lag(airpass, 1)
airpass_logdiff <- diff(log(airpass))
airpass_comp <- merge.xts(frac = airpass_frac, logdiff = airpass_logdiff)
xyplot(airpass_comp, screens = c(1, 1), auto.key = TRUE,
       ylab = " 销量",
       col = c("darkorange2", "navy"))
```

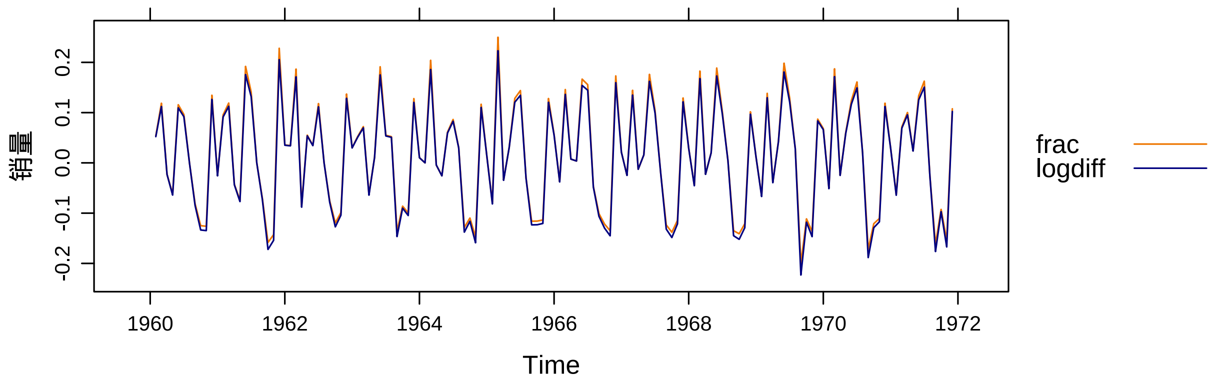


Figure 9: 对数差分与分数相对差异的比较.

0.9 5.15 强生公司季度数据

a

```
data(JJ); xyplot(JJ,type='l',ylab=" 收入",col="navy")
```

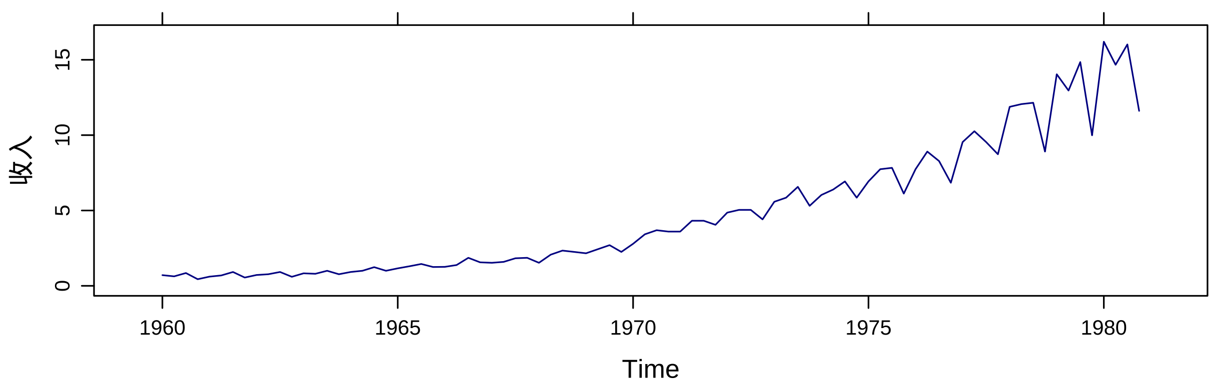
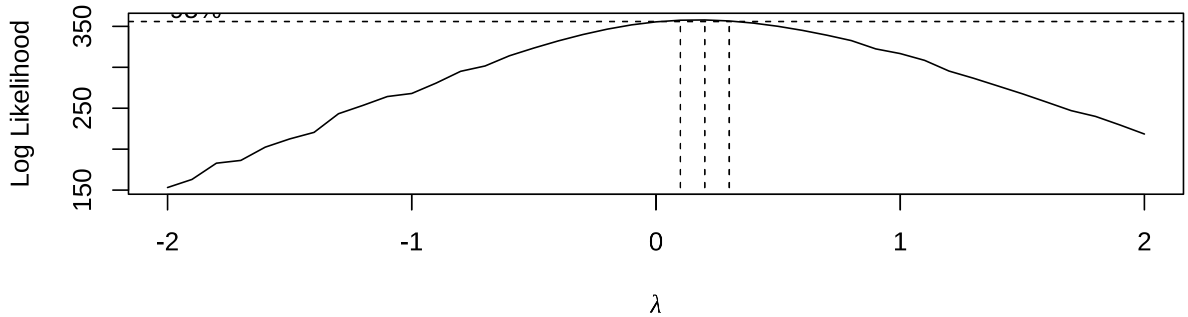


Figure 10: 强生公司季度数据.

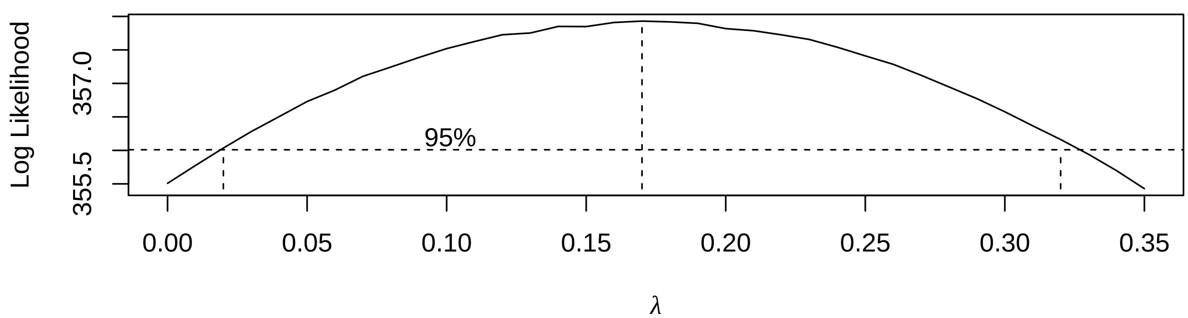
可见在 1960-1975 年每股收益呈现小幅度波动上升，从 75 年开始每股收益激增，且更不稳定。

b

```
BC <- BoxCox.ar(JJ)
```



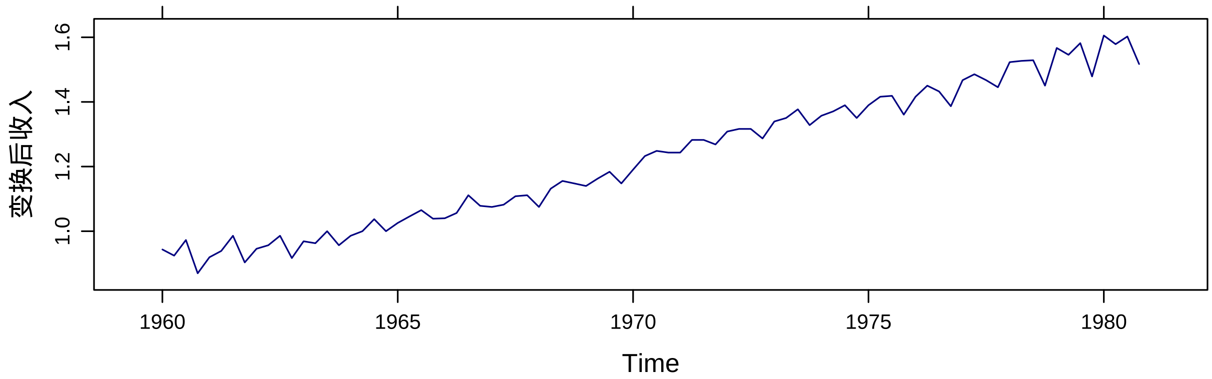

```
BC <- BoxCox.ar(JJ, lambda = seq(0.0, 0.35, 0.01))
```



第一张图显示了初始默认的 Box-Cox 分析。第二张图则显示了更多的细节，因为 $\lambda \in (0, 0.35)$，所以 λ 的最大似然估计为 0.17，其 95% 置信区间为 0.02 到 0.32。后续分析，我们使用 $\lambda = 0.17$。

c

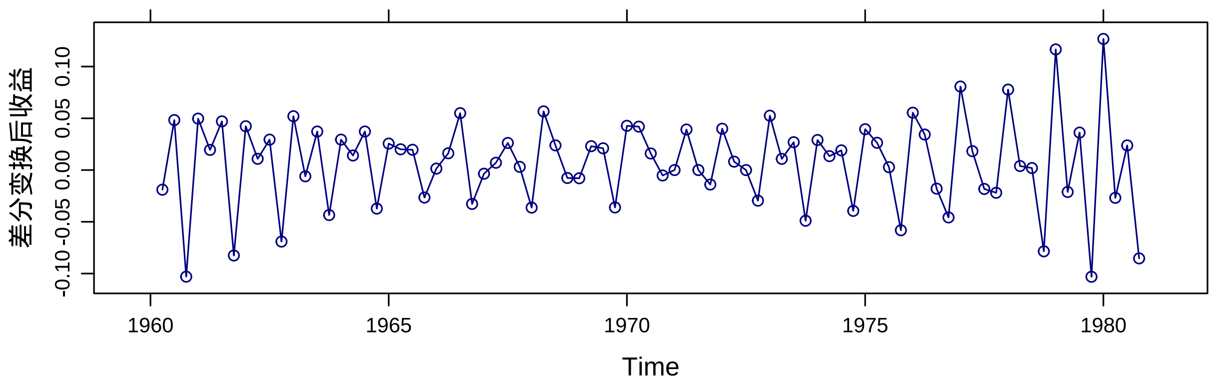
```
xyplot((JJ)^0.17,type='l',ylab='变换后收入',col="navy")
```



方差已稳定，但在建立平稳模型前，必须对明显的趋势进行处理。

d

```
xyplot(diff((JJ)^0.17),type='o',ylab='差分变换后收益',col="navy")
```

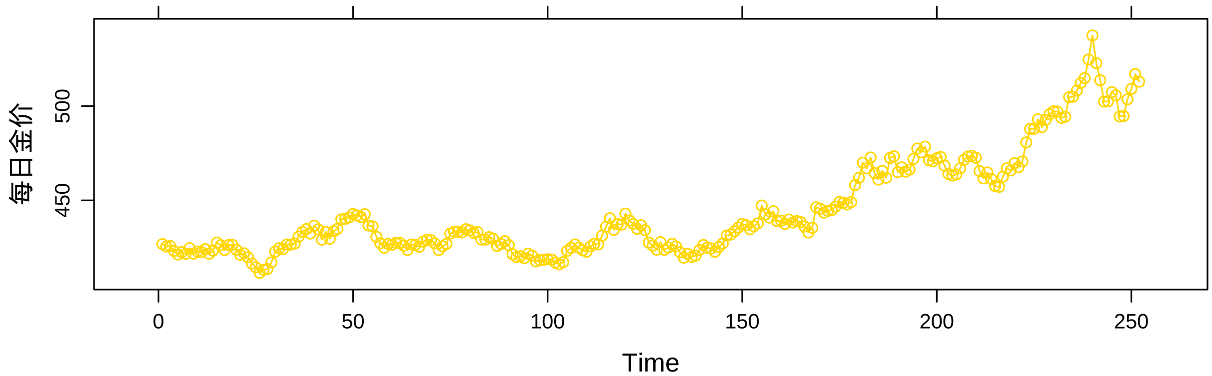


现在增长趋势已经没了，但可能还有波动的季节效应需要处理。

0.10 5.16

a

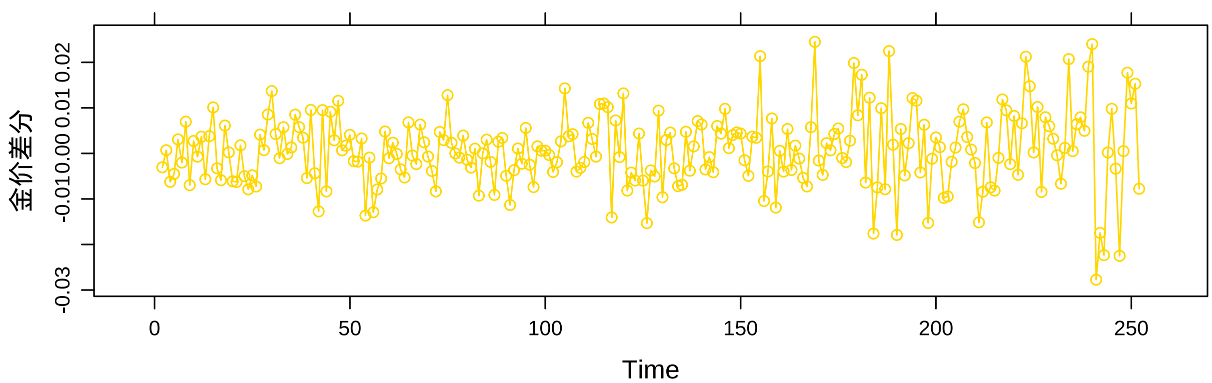
```
data(gold); xyplot(gold,type='o',ylab='每日金价',col="gold")
```



可见 05 年上半年金价稳定，仅有小幅度的波动，而年末交易日价格激增。

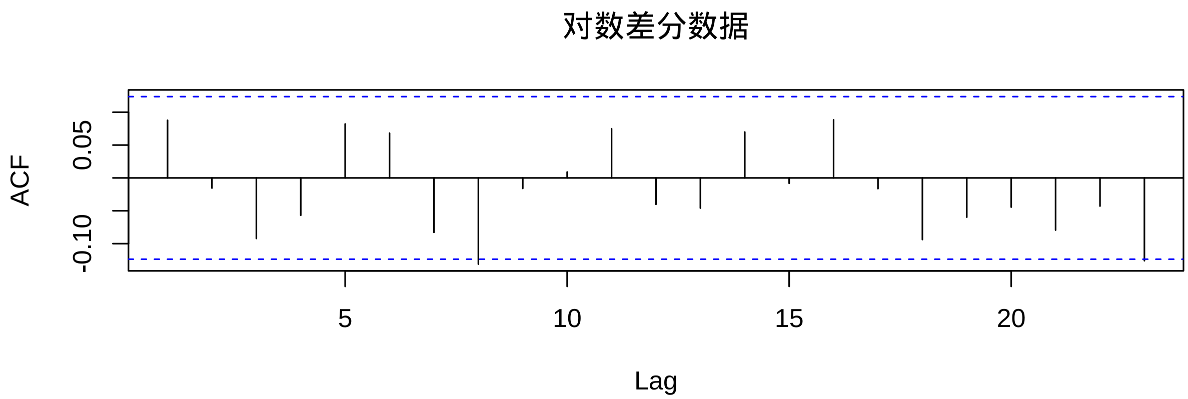
b

```
xyplot(diff(log(gold)),type='o',ylab='金价差分',col="gold")
```



c

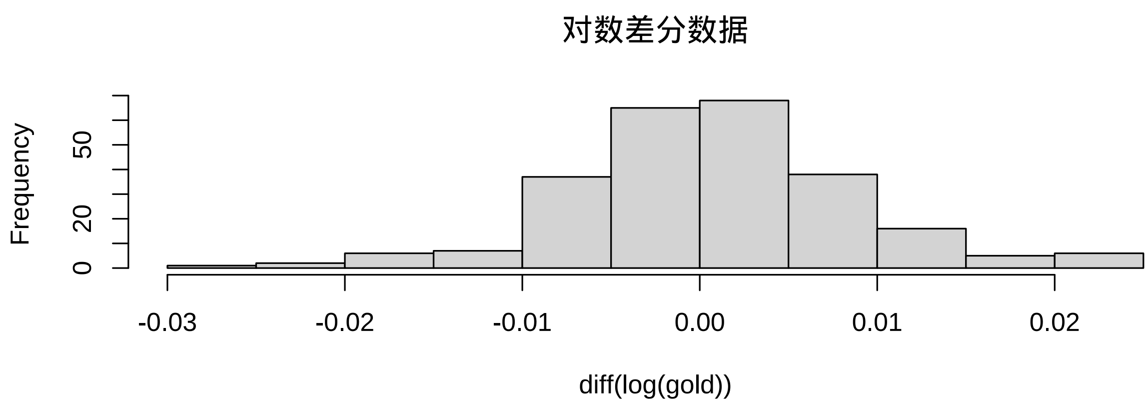
```
acf(diff(log(gold)),main='对数差分数据')
```



自相关系数均接近 0 且小范围波动。所以数据独立，似乎遵循随机游动模型。

d

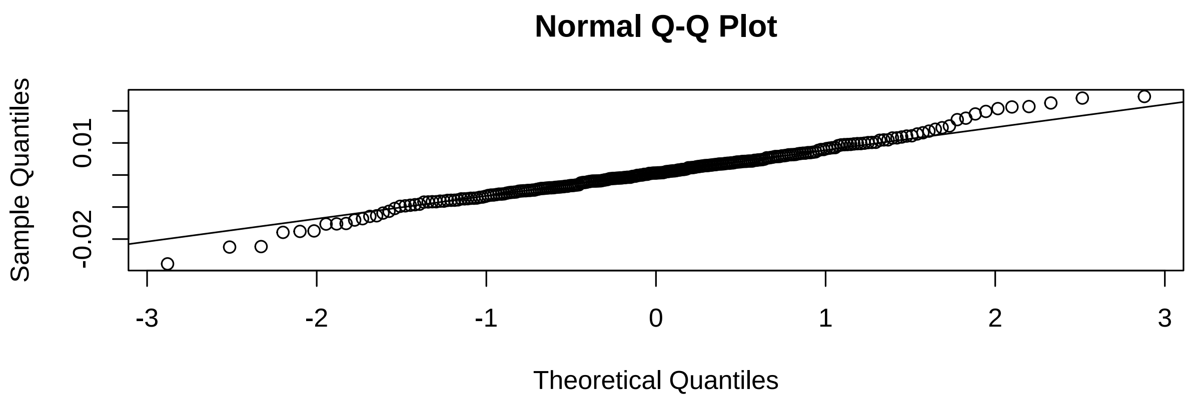
```
hist(diff(log(gold)),main='对数差分数据')
```



从直方图可以看出，数据经过对数变换和差分之后近乎为白噪声序列，因为差分数据的数值集中在 0 附近波动，说明数据独立，互不影响。

e

```
qqnorm(diff(log(gold))); qqline(diff(log(gold)));
```




```r
shapiro.test(diff(log(gold)))
```

```
##
##  Shapiro-Wilk normality test
##
## data:  diff(log(gold))
## W = 0.98607, p-value = 0.01519
```

Q-Q 图表明该分布偏离正态性。特别是末尾比正态分布要高。

Shapiro-Wilk 检验的 p 值为 0.015，进一步验证了正态性的偏离。