时间序列 HW02

18300290007 加兴华

2.6

a

$$\begin{gathered} Ex_t = \beta_0 + \beta_1 t \neq Const \\ \therefore x_t \text{ is nonstaitionary} \end{gathered}$$

b

$$\begin{aligned} Def: \quad & y_t = \Delta x_t = \beta_1 + w_t - w_{t-1} \\ then, \quad & \begin{cases} Ey_t = \beta_1 \\ \gamma(s,t) = Cov(y_s, y_t) = \begin{cases} 2\sigma_w^2, & |s-t| = 0 \\ \sigma_w^2, & |s-t| = 1 \\ 0, & o.w. \end{cases} \end{cases} \end{aligned}$$

y_t 的期望为常数，且自方差函数只与s和t的相对距离有关，因此y_t 是平稳的。

c

$$\begin{aligned} Def: \quad & z_t = \Delta x_t = \beta_1 + y_t - y_{t-1} \\ then\ , \quad & \begin{cases} Ey_t = \beta_1 + Ey_t - Ey_{t-1} = \beta_1 \\ \gamma(s,t) = Cov(z_s, z_t) = E(y_t - y_{t-1}, y_s - y_{s-1}) \end{cases} \\ Note: \quad & \gamma_y(h) = Cov(y_t, y_s) = E(y_t y_s) - \mu_y^2 \\ \Rightarrow \quad & \gamma(s,t) = \gamma_y(t-s) - \gamma_y(t-s-1) - \gamma_y(t-s+1) + \gamma_y(t-s) \\ \therefore \quad & |s-t| \mapsto \gamma(s,t) \end{aligned}$$

因此，由z_t 的期望为常数且自方差函数只与s和t的相对距离有关，知z_t 是平稳的。

2.7

由2.6（c）知，当y_t 平稳时，$\delta+y_t-y_{t-1}$ 也平稳。

因此只要证当$z_t=\delta+y_t-y_{t-1}$ 平稳时，z_t+w_t 也平稳即可。

$$Def:\quad a_t=z_t+w_t$$
$$then,\ \begin{cases}Ea_t=Ez_t+0=Ez_t=Const\\ \gamma(s,t)=E(z_t+w_t-\mu_z,z_s+w_s-\mu_z)=\gamma_z(t-s)+E(w_tw_s)\end{cases}$$

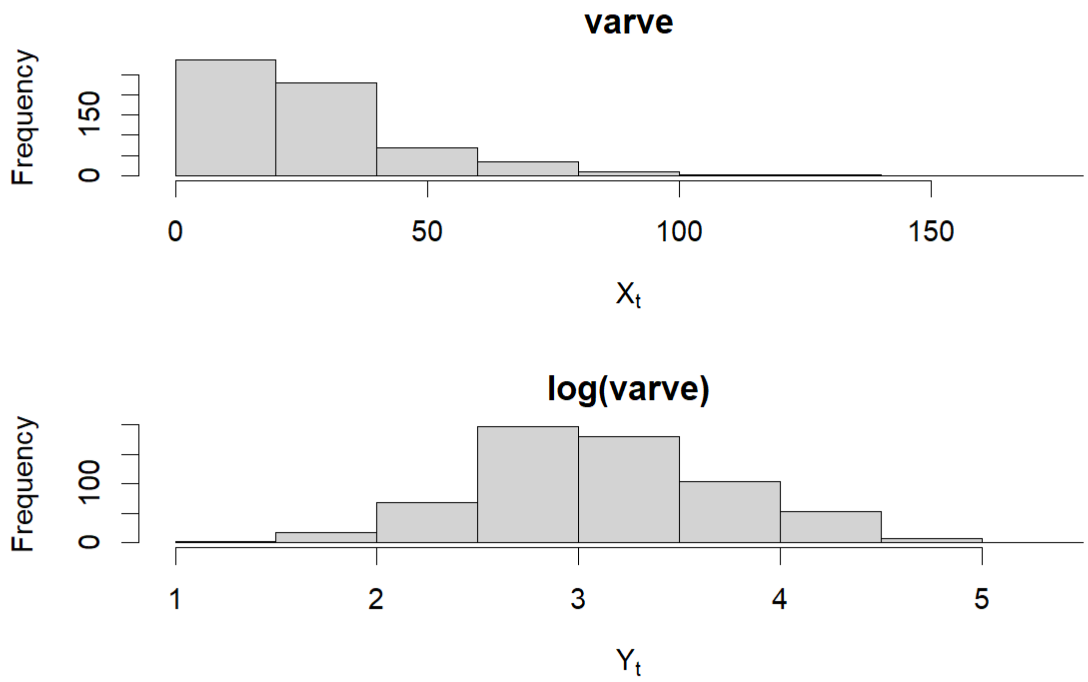
显然，a_t 的期望仍是常数，且自相关函数仍只与|t-s|有关，因此仍是平稳的。

2.8

a

```
library(astsa)
print(paste('varve的前半段样本方差为：',var(varve[1:317])))
print(paste('varve的后半段样本方差为：',var(varve[318:634])))
print(paste('log(varve)的前半段样本方差为：',var(log(varve[1:317]))))
print(paste('log(varve)的后半段样本方差为：',var(log(varve[318:634]))))
x_t=varve
y_t=log(varve)
par(mfrow=c(2,1))
xt=expression(X[t])
yt=expression(Y[t])
hist(x_t,main='varve',xlab =xt)
hist(y_t,main='log(varve)',xlab=yt)
```

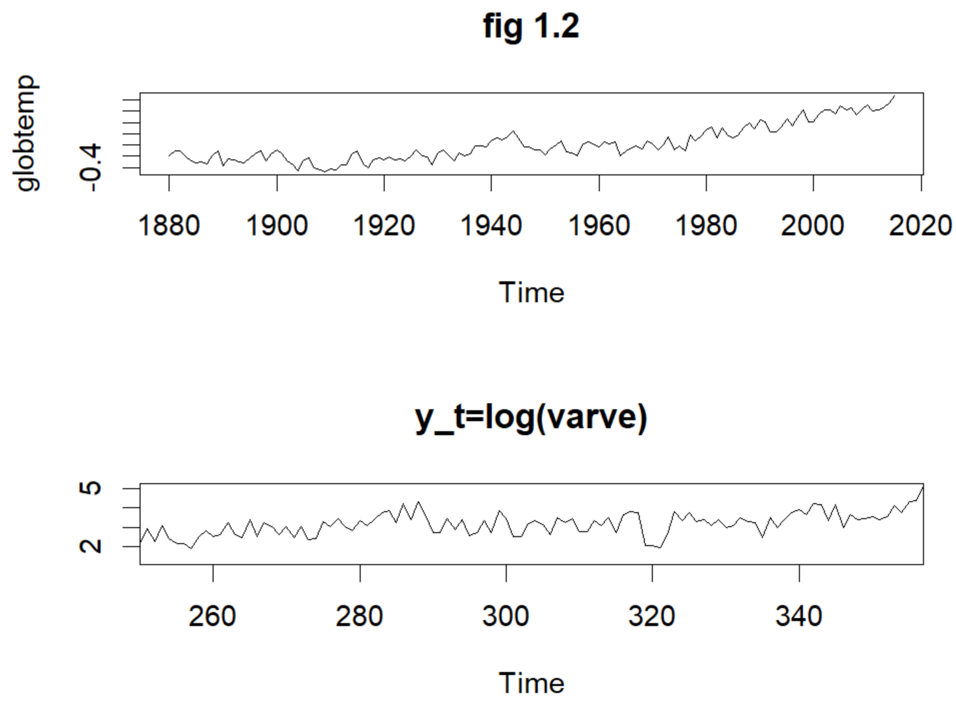
[1] "varve的前半段样本方差为： 133.457415667053"
[1] "varve的后半段样本方差为： 594.490438823224"
[1] "log(varve)的前半段样本方差为： 0.270721652653357"
[1] "log(varve)的后半段样本方差为： 0.451371011716303"



可见在对数化后时间序列的近似正态性确实有所改善。

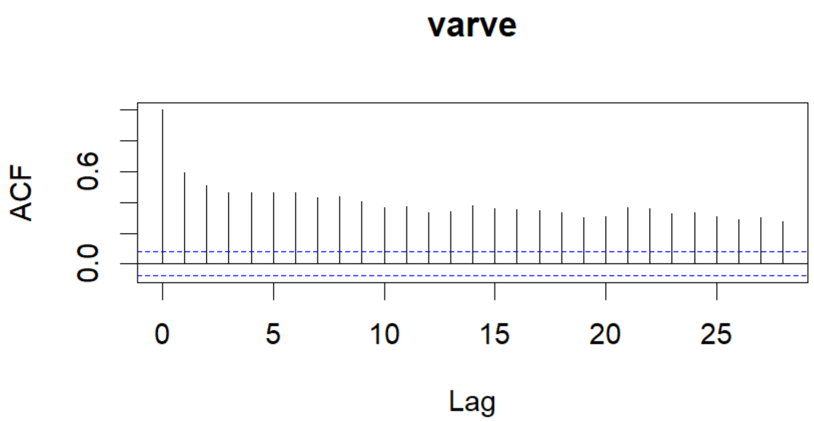
b

```
library(astsa)
par(mfrow=c(2,1))
plot(globtemp, main='fig 1.2')
summary(globtemp)
plot.ts(log(varve), main="y_t=log(varve)", ylab="",xlim=c(254,353) )
```



254-353年varve曲线的趋势是缓慢上升，和Fig1.2行为相同，但如果对相似行为的判断更严格一些，进一步考虑‘是否穿过零点’和‘波动性’，则不存在这样的100年。

c

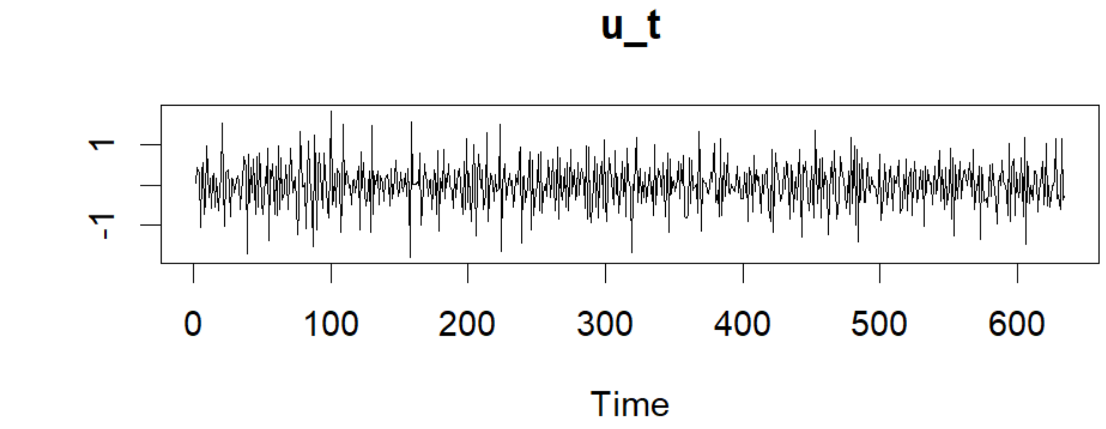


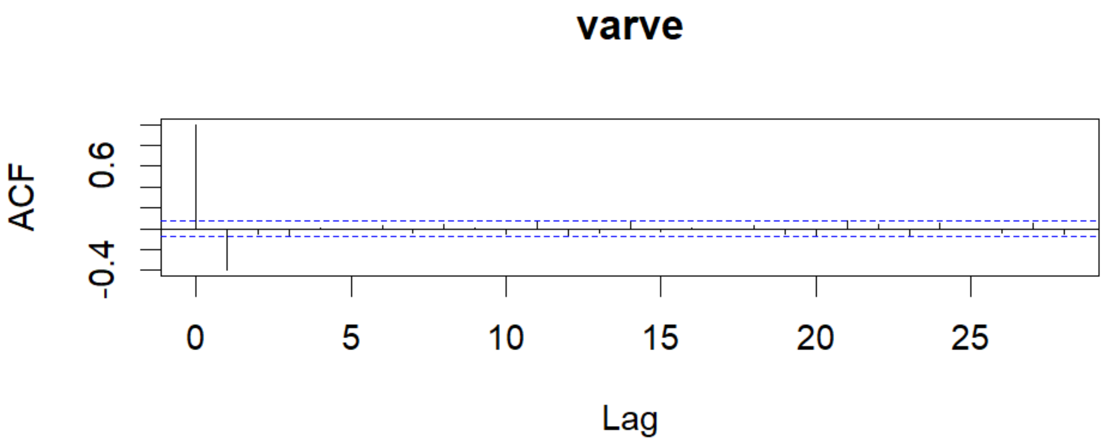
可以看到y_t 的样本ACF随着lag的增大而缓慢下降，但始终保持正相关，说明不平稳。

d

```
library(astsa)
u = diff(log(varve)) # approximate returns
summary(u)
par(mfrow=c(2,1))
plot(u, main="u_t",ylab="" )
acf(u,type = c("correlation"),main='varve')
```

```
 Min.       1st Qu.     Median     Mean     3rd Qu.      Max.
 -1.785514 -0.381677   0.005811 -0.001125   0.376144    1.821727
```





从时间曲线上看，曲线的均值保持为0且波动程度不受绝对时间影响；从ACF图上看，随着滞后阶数提高值快速趋于0，说明对数差分后的时间序列是平稳的。

$$u_t = y_t - y_{t-1} = log(\frac{x_t}{x_{t-1}}) = log(1 + \frac{\Delta x_t}{x_{t-1}}) \approx \frac{\Delta x_t}{x_{t-1}}, \quad (\frac{\Delta x_t}{x_{t-1}} \text{ is small})$$

所以u_t是x_t的逐点变化率的近似序列，一些附加性的的不平稳在分子中被消去.

e

$$\begin{aligned} Eu_t &= u, \\ \gamma_u(s,t) &= E(u_t - u, u_s - u) \\ &= E(w_t - \theta w_{t-1}, w_s - \theta w_{s-1}) \\ &= \sigma_w^2 * [I(t=s) + \theta^2 I(t=s) - \theta I(t=s-1) - \theta I(t=s+1)] \\ &= \begin{cases} \sigma_w^2\left(1+\theta^2\right) & \text{if } t=s \\ \theta\sigma_w^2 & \text{if } t-s=\pm 1 \\ 0 & \text{if } |t-s|>1 \end{cases} \end{aligned}$$

令h=t-s,得证。

Example 1.26即$\theta = 0.7$ 时的情况。

f

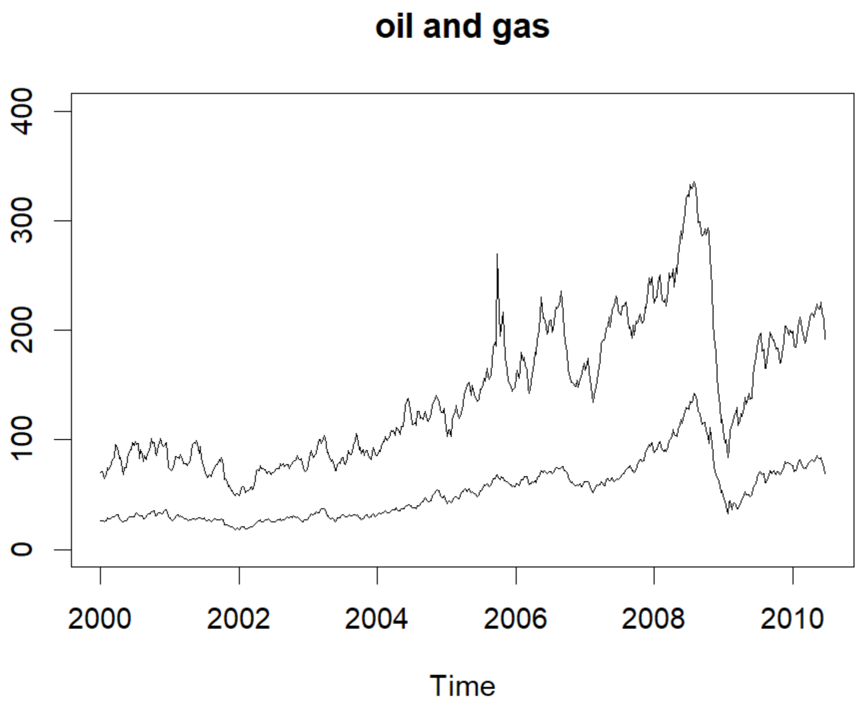
$$
\begin{aligned}
\rho_u(1) = \frac{\gamma_u(1)}{\gamma_u(0)} &= \frac{\theta}{1+\theta^2} \\
\Rightarrow \qquad \hat{\rho}_u(1) &= \frac{\hat{\theta}}{1+\hat{\theta}^2} \\
\Rightarrow \qquad 1+\hat{\theta}^2 - \frac{1}{\hat{\rho}_u(1)}\hat{\theta} &= 0 \\
\Rightarrow \qquad \hat{\theta} &= \frac{\frac{1}{\hat{\rho}_u(1)} \pm \sqrt{\frac{1}{\hat{\rho}_u(1)}^2 - 4}}{2} \\
&= \frac{1 \pm \sqrt{1-4\hat{\rho}_u(1)^2}}{2\hat{\rho}_u(1)}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\gamma}_u(0) &= \hat{\sigma}_w^2(1+\hat{\theta}^2) \\
\Rightarrow \qquad \hat{\sigma}_w^2 &= \frac{\hat{\gamma}_u(0)}{1+\hat{\theta}^2} \\
&= \frac{2\hat{\gamma}_u(0)\hat{\rho}_u(1)^2}{1 \pm \sqrt{1-4\hat{\rho}_u(1)^2}}
\end{aligned}
$$

2.10

a

```
library(astsa)
plot(gas, main="oil and gas", ylab="",ylim=c(0,400))
lines(oil)
```



它们和Fig.1.10 对应的时间序列最像，我相信它们不平稳，因此读图发现两个序列的均值线不可能水平。

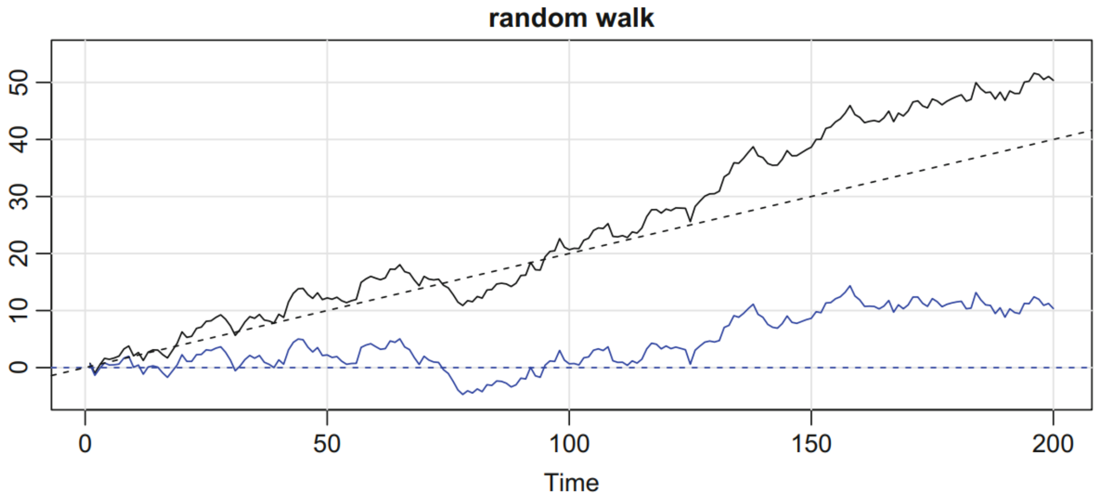


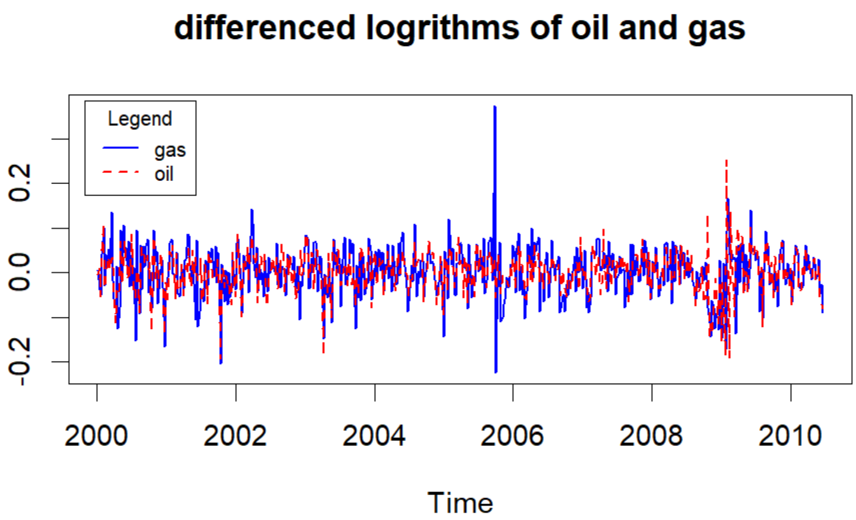
Fig. 1.10. Random walk, $\sigma_w = 1$, with drift $\delta = .2$ (*upper jagged line*), without drift, $\delta = 0$ (*lower jagged line*), and *straight (dashed) lines* with slope δ

b

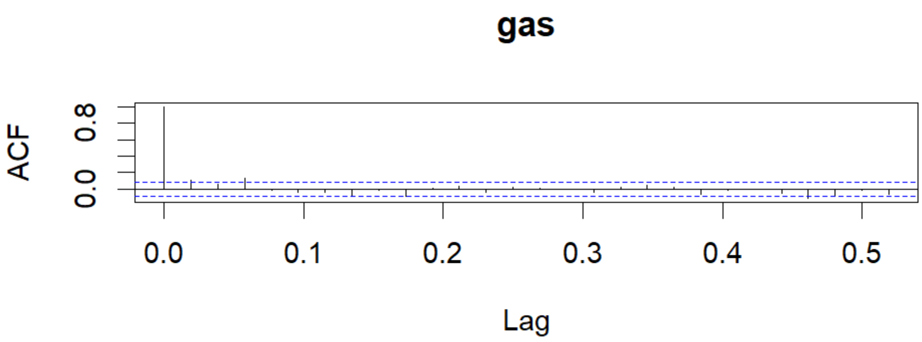
与2.8（d）同理，当 $\frac{\Delta x_t}{x_{t-1}}$ 接近0时，$y_t = diff(log(x_t)) \approx \frac{\Delta x_t}{x_{t-1}}$，即百分比变化。

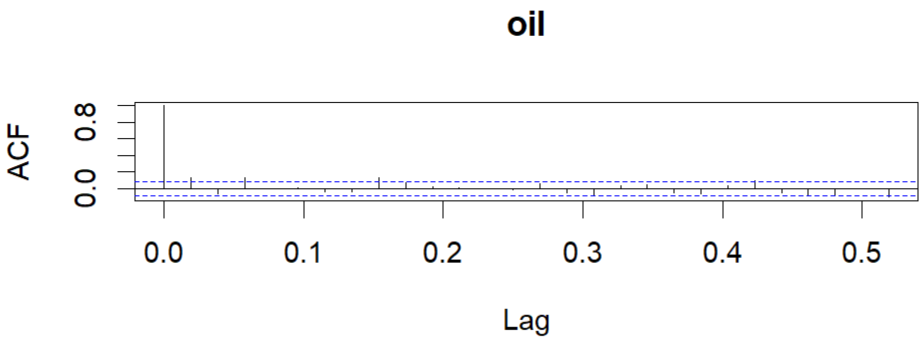
c

```
library(astsa)
g1=diff(log(gas))
o1=diff(log(oil))
plot.ts(g1, main="differenced logrithms of oil and gas",
ylab="",lty=1,col='blue',lwd=1.5)
lines(o1,lty=2,col='red',lwd=1.5)
legend('topleft', inset=.02, title='Legend',bty='o',
legend=c("gas","oil"),lty=c(1, 2),cex=0.7,col=c('blue','red'),lwd=c(1.5,1.5))
```




```
library(astsa)
g1=diff(log(gas))
o1=diff(log(oil))
par(mfrow=c(2,1))
acf(g1,type = c("correlation"),main='gas')
acf(o1,type = c("correlation"),main='oil')
```

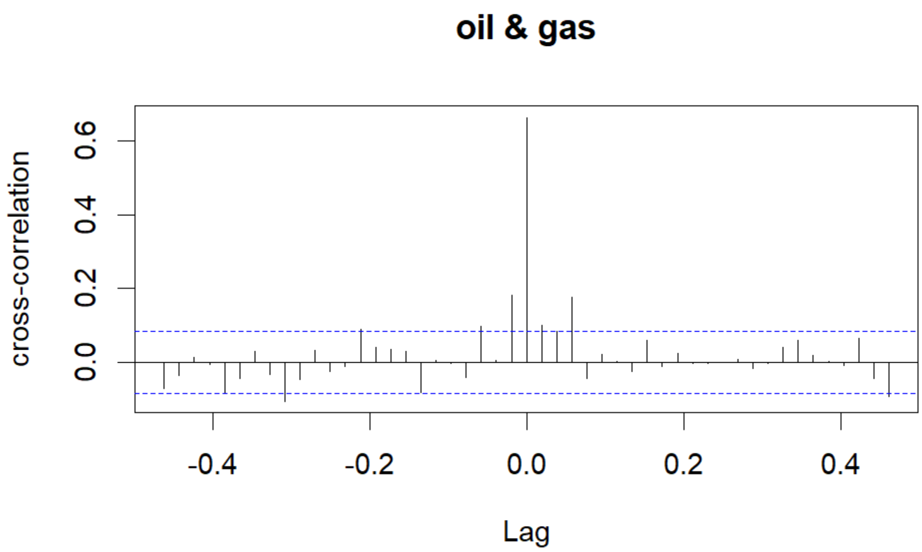




可以看到时间曲线上两者均值都是稳定在0点的；ACF图中两者也都快速衰减到0，因此基本是稳定的。

d

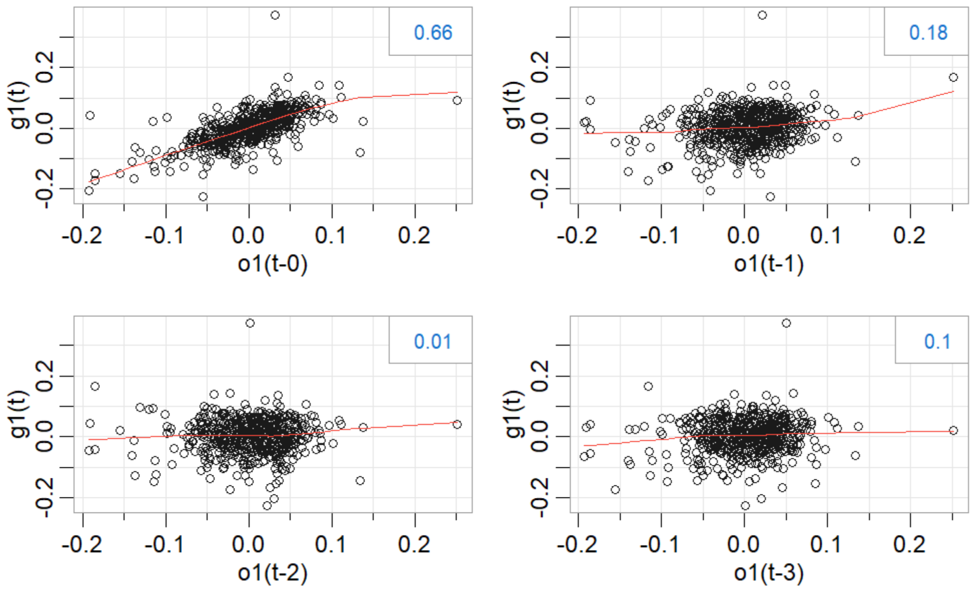
```
library(astsa)
g1=diff(log(gas))
o1=diff(log(oil))
ccf(x=o1,y=g1,main="oil & gas",ylab = "cross-correlation")
```



R语言函数编码为$ccf(n) = Cov(x_{t+n}, y_t)$,正半轴部分为gas lead oil，负半轴部分为oil lead gas，结合题干可以做以下解读：gas对oil变化十分敏感，且响应后在将来又偶尔会给oil一些反馈。

e

```
library(astsa)
g1=diff(log(gas))
o1=diff(log(oil))
lag2.plot(o1,g1, 3)
```



可以看到图中总是有一些离群点，只有lag=0时线性相关性强，lag=1时有弱相关性，其余时候几乎没有。

f

i

```
poil = diff(log(oil))
pgas = diff(log(gas))
indi = ifelse(poil < 0, 0, 1)
mess = ts.intersect(pgas, poil, poilL = lag(poil,-1), indi)
summary(fit <- lm(pgas~ poil + poilL + indi, data=mess))
```

```
lm(formula = pgas ~ poil + poilL + indi, data = mess)

Residuals:
      Min        1Q    Median        3Q       Max
-0.18451 -0.02161 -0.00038   0.02176   0.34342

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.006445   0.003464  -1.860  0.06338 .
poil         0.683127   0.058369  11.704  < 2e-16 ***
poilL        0.111927   0.038554   2.903  0.00385 **
indi         0.012368   0.005516   2.242  0.02534 *
```

```
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.04169 on 539 degrees of freedom
Multiple R-squared:  0.4563,	Adjusted R-squared:  0.4532 
F-statistic: 150.8 on 3 and 539 DF,  p-value: < 2.2e-16
```

ii

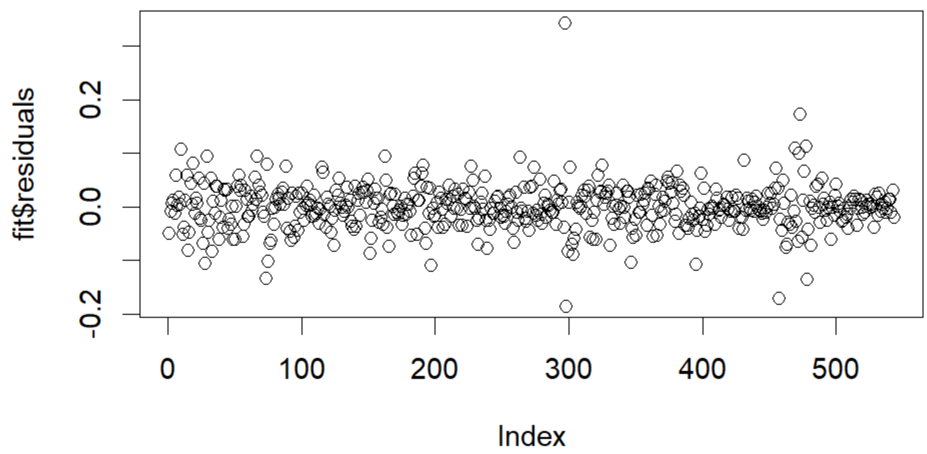
油价负增长时$I_t = 0, \quad G_t = -0.006445 + 0.683127 O_t + 0.111927 O_{t-1}$

油价非负增长时$I_t = 1, \quad G_t = 0.005923 + 0.683127 O_t + 0.111927 O_{t-1}$

结合表达式和所有因素均显著来看，题干中的假设合理

iii

```
plot(fit$residuals)
```

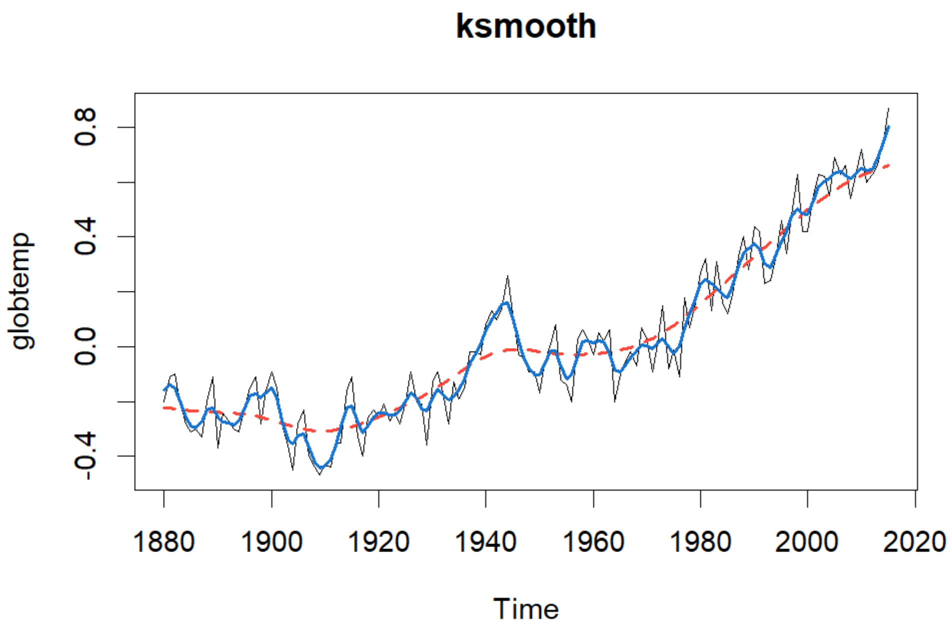


残差的均值保持为零且除了极端点外波动性还是比较随机的，说明线性模型拟合后的残差确实只剩了某种简单噪声。

2.11

kernel smoothing

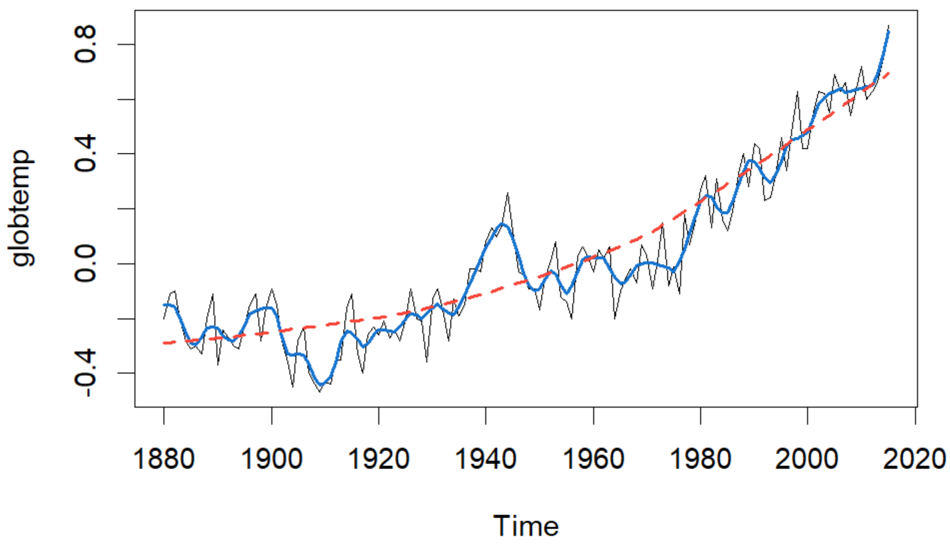
```
library(astsa)
plot.ts(globtemp,main='ksmooth')
lines(ksmooth(time(globtemp), globtemp, "normal", bandwidth=20), lwd=2,
col=2,lty=2)
lines(ksmooth(time(globtemp), globtemp, "normal", bandwidth=3), lwd=2, col=4)
```



kernel smooth 中随着参数bandwidth 的增大曲线变得越来越光滑,越反应整体趋势

lowess

```
library(astsa)
plot(globtemp)
lines(lowess(globtemp, f=.05), lwd=2, col=4) # El Nino cycle
lines(lowess(globtemp), lty=2, lwd=2, col=2) # trend (with default span)
```



lowess 中随着参数f 的增大曲线变得越来越光滑,越反应整体趋势

附加题

a

第一个等号：

$$\int_a^b g''(x)h''(x)dx = \int_a^b g''(x)dh'(x) = [g''(x)h'(x)]\Big|_{x=a}^{b} - \int_a^b h'(x)dg''(x)$$

自然样条 $g''(a)=g''(b)=0$.

$$= 0 - \int_a^b h'(x)g'''(x)dx$$

$$= -\int_a^b g'''(x)dh(x) \quad \because g'''(x) \equiv g \quad \forall x \in [x_j^+, x_{j+1})$$

$$= -\sum_{j=1}^{N-1} g'''(x_j^+)\cdot\Delta h(x_j) \quad \#.$$

第二个等号：

$$\int_a^b h'(x)dg'''(x) = \int_a^b g''''(x)h'(x)dx$$

$$= [h(x)\cdot g''''(x)]\Big|_{x=a}^{b} - \int_a^b h(x)\underbrace{dg''''(x)}_{\text{四阶为零}}$$

$$= [h(x)\cdot \underbrace{g'''(x)}_{\text{在端点为零}}]\Big|_{x=a}^{b}$$

$$= 0.$$

b

$$\int_a^b \tilde{g}''(t)^2dt = \int_a^b (g''(t)+h''(t))^2dt$$

$$= \int_a^b [g''(t)^2 + h''(t)^2]dt + 2\int_a^b g''(t)h''(t)dt$$

$$= \int_a^b g''(t)^2dt + \int_a^b h''(t)^2dt$$

$$\geq \int_a^b g''(t)^2dt + (b-a)\cdot\min_{[a,b]} h''(t)^2$$

$$\geq \int_a^b g''(t)^2dt.$$

当 $h''(t)\equiv 0$ 时取等 $\Leftrightarrow$ 当 $h'(t)\equiv C$ 时取等

又很明显 $h(t)$ 在样本点处为0. 从而 $\Leftrightarrow h(t)\equiv 0$ 时取等

c

假设 $\arg\min\left\{\sum (y_i - f(x_i))^2 + \lambda \int_a^b f''(t)^2 dt\right\} = \tilde{g}(t)$

而三次样条函数为 $g(t)$.

$$\begin{cases} \sum (y_i - \tilde{g}(t))^2 \geq 0 = \sum (y_i - g(t))^2 \\ \int_a^b \tilde{g}''(t)^2 dt \geq \int_a^b g(t)^2 dt \end{cases}$$

$\Rightarrow$ $\tilde{g}(t)$ 并非 minimizer. 矛盾.

$\therefore$ minimizer 必为过样本点的三次样条函数. 得证.