MA550_HW1

Enbo Tian

MA 550 − Homework 1

Due Wed, 11:30am, September 16, 2020

Problem 1

Consider Slides #1, Page 28.

(a) Find the covariance matrices $Cov(\boldsymbol{e},\boldsymbol{Y})$ and $Cov(\boldsymbol{e},\hat{\boldsymbol{Y}})$.

$Cov(\boldsymbol{e},\boldsymbol{Y}) = Cov(\boldsymbol{Y} - \hat{\boldsymbol{Y}},\boldsymbol{Y}) = Cov(\boldsymbol{Y},\boldsymbol{Y}) - Cov(\hat{\boldsymbol{Y}},\boldsymbol{Y}) = Var(\boldsymbol{Y}) - Cov(\boldsymbol{HY},\boldsymbol{Y}) = Var(\boldsymbol{Y}) - \boldsymbol{H}Cov(\boldsymbol{Y},\boldsymbol{Y}) = Var(\boldsymbol{Y}) - \boldsymbol{H}Var(\boldsymbol{Y}) = (\boldsymbol{I} - \boldsymbol{H})Var(\boldsymbol{Y})$

$Cov(\boldsymbol{e},\hat{\boldsymbol{Y}}) = Cov(\boldsymbol{Y} - \hat{\boldsymbol{Y}},\hat{\boldsymbol{Y}}) = Cov(\boldsymbol{Y},\hat{\boldsymbol{Y}}) - Cov(\hat{\boldsymbol{Y}},\hat{\boldsymbol{Y}}) = \boldsymbol{H}Var(\boldsymbol{Y}) - Var(\boldsymbol{HY}) = \boldsymbol{H}Var(\boldsymbol{Y}) - \boldsymbol{H}^2 Var(\boldsymbol{Y}) = (\boldsymbol{H} - \boldsymbol{H}^2)Var(\boldsymbol{Y}) = \boldsymbol{H}(\boldsymbol{I} - \boldsymbol{H})Var(\boldsymbol{Y})$

(b) The estimator of σ^2 is $\hat{\sigma}^2 = (n - p)^{-1}\boldsymbol{e}'\boldsymbol{e}$. Show that $\mathrm{E}(\hat{\sigma}^2) = \sigma^2$ and $\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2}$ follows a Chi-square distribution with degrees of freedom $n - p$. *Hint: check Rencher & Schaalje (2008)[1] Page 107 Theorem 5.2a and Page118 Corollary 2.*

$\hat{\sigma}^2 = (n - p)^{-1}\boldsymbol{e}'\boldsymbol{e} = (n - p)^{-1}\boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}'\frac{(\boldsymbol{I} - \boldsymbol{H})}{n - p}\boldsymbol{Y}$.

Let$A = \frac{\boldsymbol{I} - \boldsymbol{H}}{n - p}$,

then $\hat{\sigma}^2 = \boldsymbol{Y}'\boldsymbol{AY}$, $\mathrm{E}(\hat{\sigma}^2) = \mathrm{E}(\boldsymbol{Y}'\boldsymbol{AY}) = \mu'\boldsymbol{A}\mu + tr(\boldsymbol{A}\Sigma)$.

As $\boldsymbol{A}\mu = 0$ and $\Sigma = \sigma^2\boldsymbol{I}$, $\boldsymbol{Y} \sim \boldsymbol{Np}(\mu, \sigma^2\boldsymbol{I})$ and $\hat{\sigma}^2 = \frac{\boldsymbol{Y}'\boldsymbol{AY}}{tr(\boldsymbol{A})} = \boldsymbol{Y}'\boldsymbol{AY}$.

So $tr(\boldsymbol{A}) = 1$, $\mathrm{E}(\hat{\sigma}^2) = \sigma^2 tr(\boldsymbol{A}) = \sigma^2$

Since $\boldsymbol{A} = \frac{\boldsymbol{I} - \boldsymbol{H}}{n - p}$, $\boldsymbol{A}(n - p) = \boldsymbol{I} - \boldsymbol{H}$ is idempotent with a rank r = 1.

Since $\boldsymbol{Y} \sim \boldsymbol{Np}(\mu, \sigma^2\boldsymbol{I})$, $\frac{\boldsymbol{e}'\boldsymbol{e}}{\sigma^2} = \frac{\boldsymbol{Y}'\boldsymbol{AY}}{\sigma^2}$ is$\chi^2(1, \frac{\mu'\boldsymbol{A}\mu}{2\sigma^2})$ with a degree freedom $n - p$.

(c) Show that $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y}$ and $\hat{\sigma}^2$ are independent. *Hint: check Rencher & Schaalje (2008) Page 119 Theorem 5.6a.*

Let $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{Y} = \frac{\boldsymbol{HY}}{\boldsymbol{X}} = \boldsymbol{BY}$, and let $\hat{\sigma}^2 = \boldsymbol{Y}'(\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y} = \boldsymbol{Y}'\boldsymbol{AY}$

Since $cov(\boldsymbol{BY}, \boldsymbol{AY}) = cov(\frac{\boldsymbol{HY}}{\boldsymbol{X}}, (\boldsymbol{I} - \boldsymbol{H})\boldsymbol{Y}) = \mathrm{E}((\frac{\boldsymbol{HY}}{\boldsymbol{X}})(\boldsymbol{Y} - \boldsymbol{HY})) - \mathrm{E}(\frac{\boldsymbol{HY}}{\boldsymbol{X}})\mathrm{E}(\boldsymbol{Y} - \boldsymbol{HY}) = \mathrm{E}(\frac{\boldsymbol{HY}^2}{\boldsymbol{X}} - \frac{(\boldsymbol{HY})^2}{\boldsymbol{X}}) - \mathrm{E}(\frac{\boldsymbol{HY}}{\boldsymbol{X}})\mathrm{E}(\boldsymbol{Y}) + \mathrm{E}(\frac{\boldsymbol{HY}}{\boldsymbol{X}})\mathrm{E}(\boldsymbol{HY}) = \mathrm{E}(\frac{\boldsymbol{HY}^2}{\boldsymbol{X}}) - \mathrm{E}(\frac{(\boldsymbol{HY})^2}{\boldsymbol{X}}) - \mathrm{E}(\frac{\boldsymbol{HY}^2}{\boldsymbol{X}}) + \mathrm{E}(\frac{(\boldsymbol{HY})^2}{\boldsymbol{X}}) = 0$

By Thm 5.6a, $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma}^2$ are independent.

---

[1]Rencher and Schaalje (2008): Linear Models in Statistics (2nd ed), Wiley.

Problem 2

Prove the following results in Slides #1:

(a) Page #26: $\frac{\partial Q}{\partial \boldsymbol{\beta}} = -2\boldsymbol{X'Y} + 2\boldsymbol{X'X}\boldsymbol{\beta}$;

Since $Q(\beta) = (\boldsymbol{Y} - \boldsymbol{X}\beta)'(\boldsymbol{Y} - \boldsymbol{X}\beta) = \boldsymbol{Y'Y} - \boldsymbol{Y'X}\beta - \boldsymbol{YX}\beta + \beta'\boldsymbol{X'X}\beta$,

$\frac{\partial Q}{\partial \boldsymbol{\beta}} = 0 - \boldsymbol{X'Y} - \boldsymbol{X'Y} + 2\boldsymbol{X'X}\beta = -2\boldsymbol{X'Y} + 2\boldsymbol{X'X}\beta$

(b) Page #27: $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$ and $\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$;

when $p = 2$, $\boldsymbol{Y} = \beta_0 + \boldsymbol{X}\beta_1$ AS we set $\frac{\partial S}{\partial \beta_0} = 0$, and $\frac{\partial S}{\partial \beta_1} = 0$,

$\Sigma_i \boldsymbol{Y}_i = n\hat{\beta}_0 + \hat{\beta}_1 \Sigma_i \boldsymbol{X}_i$ and $\Sigma_i \boldsymbol{X}_i \boldsymbol{Y}_i = \hat{\beta}_0 \Sigma_i \boldsymbol{X}_i + \hat{\beta}_1 \Sigma_i \boldsymbol{X}_i^2$

then $\hat{\beta}_0 = \frac{\Sigma \boldsymbol{X}_1^2 \Sigma \boldsymbol{Y}_i - \Sigma \boldsymbol{X}_i \Sigma \boldsymbol{X}_i \boldsymbol{Y}_i}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2}$ and $\hat{\beta}_1 = \frac{n\Sigma \boldsymbol{X}_i \boldsymbol{Y}_i - \Sigma \boldsymbol{X}_i \Sigma \boldsymbol{Y}_i}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2} = \hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}$

then $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

(c) Page #29: $E(\hat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$ and $Var(\hat{\boldsymbol{\beta}}) = \sigma^2 (\boldsymbol{X'X})^{-1}$;

Since $\boldsymbol{Y}_i = \beta_0 + \beta_1 \boldsymbol{X}_i + e_i$ and $E[e_i] = 0$

We can get $E[\hat{\beta}_0] = \frac{\Sigma \boldsymbol{X}_i^2 \Sigma E[\boldsymbol{Y}_i] - \Sigma \boldsymbol{X}_i \Sigma \boldsymbol{X}_i \Sigma E[\boldsymbol{Y}_i]}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2} = \frac{\Sigma \boldsymbol{X}_i^2 (n\beta_0 + \beta_1 \Sigma \boldsymbol{X_i}) - \Sigma \boldsymbol{X}_i (\beta_0 \Sigma \boldsymbol{X}_i + \beta_1 \Sigma \boldsymbol{X}_i^2)}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2} = \beta_0$

Similarly $E[\hat{\beta}_1] = \beta_1$

By the way, $E[\hat{\beta}] = \beta$

Since $Var[\boldsymbol{Y}_i] = \sigma^2$ and $Cov[\boldsymbol{Y}_i, \boldsymbol{Y}_j] = 0$ $Var[\hat{\beta}_0] = \frac{\sigma^2 \Sigma_i \boldsymbol{X}_i^2}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2}$

$Var[\hat{\beta}_1] = \frac{n\sigma^2}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2}$

$Cov[\hat{\beta}_0, \hat{\beta}_1] = \frac{-\sigma^2 \Sigma \boldsymbol{X}_i^2}{n\Sigma \boldsymbol{X}_i^2 - (\Sigma \boldsymbol{X}_i)^2}$ and from (b) we can write $\hat{\beta}_1 = \frac{\Sigma_i (x_i - \bar{x})(y_i - \bar{y})}{\Sigma_i (x_i - \bar{x})^2} = \frac{\Sigma_i (x_i - \bar{x})(y_i)}{\Sigma_i (x_i - \bar{x})^2}$ then $Var[\hat{\beta}_1] = \frac{\sigma^2}{\Sigma_i (x_i - \bar{x})^2}$ Similarly $Var[\hat{\beta}] = \sigma^2 (\boldsymbol{X'X})^{-1}$

(d) Page #30: $\frac{\hat{\beta}_k - \beta_k}{S\{\hat{\beta}_k\}} \sim T_{n-p}$. *Hint: consider Problem 1 (b) and (c).*

Since $\hat{\beta} - \beta \sim N(0, \frac{\sigma^2}{\sum_{i=1}^n (X_i - \bar{X}^2)})$

then the sampling distribution is $\hat{\beta}_k - \beta_k \sim N(0, \sigma^2 (X'X)_{kk}^{-1})$ Since from the varience of (C),the eatimates $S_{\hat{\beta}_k}$ is RV's So the hypothesis test is normally distributed. then it can be shown as $\frac{\hat{\beta}_k - \beta_k}{S\{\hat{\beta}_k\}} \sim T_{n-p}$

Problem 3

Consider the model $Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$ where $\varepsilon_i \sim \text{ind} N(0, \sigma_i^2)$ for observations $i = 1, \ldots, n$. Suppose $X_i > 0$ and $\sigma_i^2 = \sigma^2 X_i$. Let $\tilde{Q}(\boldsymbol{\beta}) = \sum_{i=1}^n \sigma_i^{-2}\{Y_i - (\beta_0 + \beta_1 X_i)\}^2$ denote the weighted error sum of squares. Define

$\boldsymbol{Y} = (Y_1, \ldots, Y_n)'$: $n \times 1$ vector of response variables

$\boldsymbol{X}$: $n \times 2$ design matrix with 1's in the first column and $(X_1, \ldots, X_n)'$ in the second column.

$\boldsymbol{\beta} = (\beta_0, \beta_1)'$: 2×1 vector of regression coefficients

$\boldsymbol{V} = \text{diag}\{X_1, \ldots, X_n\}$: $n \times n$ diagonal matrix with $X_1, \ldots, X_n$ along the diagonal

(a) Derive the distribution of $\boldsymbol{Y}$ including the analytical forms of the mean vector $\boldsymbol{\mu}$ and variance-covariance matrix $\boldsymbol{\Sigma}$.

$$E[\boldsymbol{Y}_i] = E(\beta_0 + \beta_1\boldsymbol{X}_i + \varepsilon_i) = E(\beta_0) + \beta_1 E(\boldsymbol{X}_i) + E(\varepsilon_i) = \beta_0 + \beta_1\frac{\sigma_i^2}{\sigma^2}$$

$$\mu = \begin{bmatrix} \beta 0 + \beta 1\frac{\sigma_1^2}{\sigma^2} \\ \beta 0 + \beta 1\frac{\sigma_2^2}{\sigma^2} \\ \vdots \\ \beta 0 + \beta 1\frac{\sigma_n^2}{\sigma^2} \end{bmatrix}$$

$Var(\boldsymbol{Y}_i) = Var(\beta_0 + \beta_1\boldsymbol{X}_i + \varepsilon_i) = \sigma_i^2$

$Cov(\boldsymbol{Y}_i, \boldsymbol{Y}_j) = E[(\boldsymbol{Y}_i - E(\boldsymbol{Y}_i))(\boldsymbol{Y}_j - E(\boldsymbol{Y}_j))] = E[(\beta_0 + \beta_1\boldsymbol{X}_i + \varepsilon_i - \beta_0 - \beta_1\frac{\sigma_i^2}{\sigma^2})(\beta_0 + \beta_1\boldsymbol{X}_j + \varepsilon_j - \beta_0 - \beta_1\frac{\sigma_j^2}{\sigma^2})] = 0$

Therefore

$$Cov(\boldsymbol{X}, \boldsymbol{Y}) = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

(b) Write the weighted error sum of squares $\tilde{Q}(\boldsymbol{\beta})$ in matrix terms.

$$\tilde{Q}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \sigma_i^{-2}(\boldsymbol{Y}_i - \beta_0 - \beta_1\boldsymbol{X}_i)^2 = (\frac{1}{\sigma_1}(\boldsymbol{Y}_1 - \beta_0 - \beta_1\boldsymbol{X}_1), \cdots, \frac{1}{\sigma_n}(\boldsymbol{Y}_n - \beta_0 - \beta_1\boldsymbol{X}_n)) \begin{pmatrix} \frac{1}{\sigma_1}(\boldsymbol{Y}_1 - \beta_0 - \beta_1\boldsymbol{X}_1) \\ \vdots \\ \frac{1}{\sigma_n}(\boldsymbol{Y}_n - \beta_0 - \beta_1\boldsymbol{X}_n) \end{pmatrix}$$

(c) Derive the mean vector and the variance-covariance matrix of

(d) Despite the weighted variances, derive the ordinary least squares estimates of $\boldsymbol{\beta}$ by minimizing the (unweighted) error sum of squares $Q(\boldsymbol{\beta}) = \sum_{i=1}^{n}\{Y_i - (\beta_0 + \beta_1 X_i)\}^2$ all in matrix terms. Let $\hat{\boldsymbol{\beta}} = (\hat{\beta}_0, \hat{\beta}_1)$ denote the (ordinary) least squares estimates of $\boldsymbol{\beta}$.

(e) Derive the mean vector and the variance-covariance matrix of $\hat{\boldsymbol{\beta}}$.

(f) Draw a connection between $Var(\hat{\beta}_1)$ and $Var(\tilde{\beta}_1)$.

Problem 4

The file 'MortBond.txt' contains monthly effective interest rate z_t for conventional single- family mortgages from Jan. 1973 to Dec. 2003.

(a) Draw threes plots: time plot of the time series, scatter plots of z_{t+1} versus z_t and of z_{t+2} versus z_t.

```
##-- second column is monthly interest rate for conventional single-family mortgage.
Mort = read.csv("MortBond.txt",head = T, sep="");
mortrate = ts(Mort[,2],start = 1937,freq =12);
ts.plot(mortrate,gpars = list(xlab = "Month", ylab = "mortrate"),main = "time series")
```
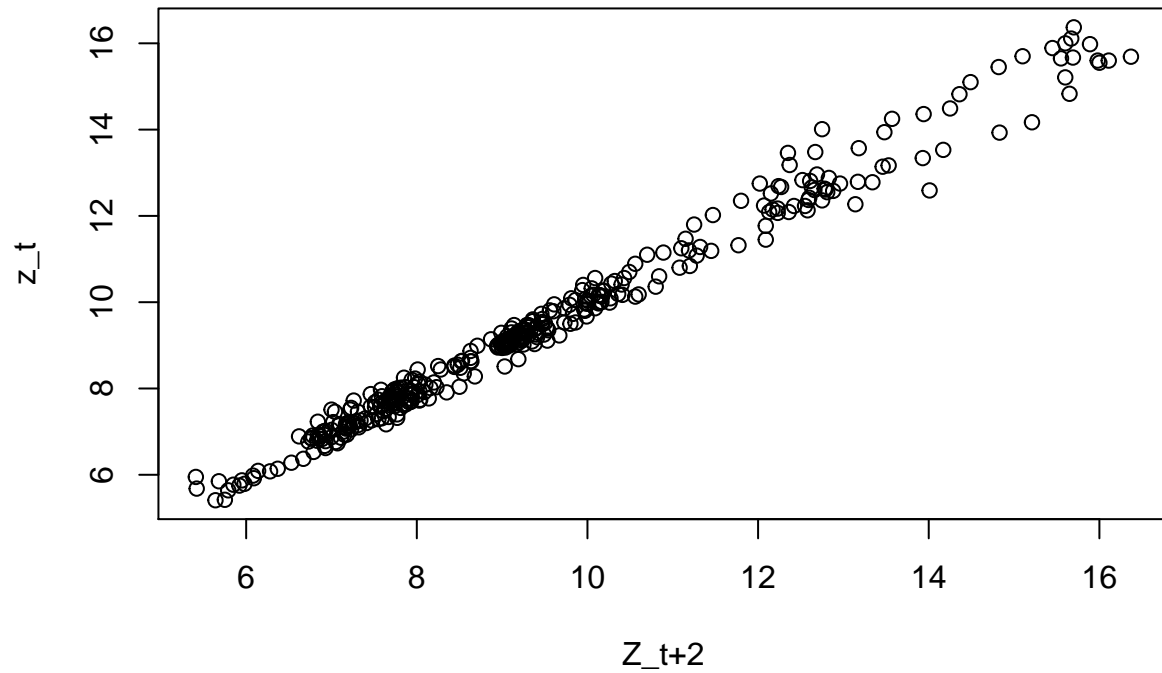
time series
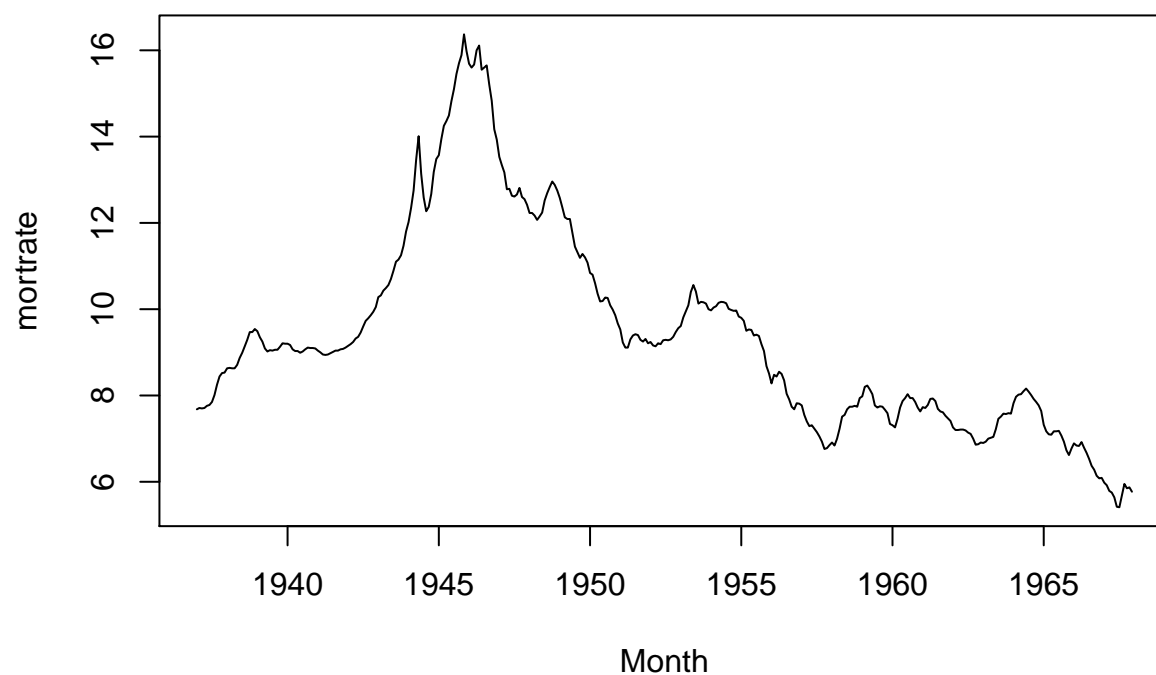

```
z01 = tail(mortrate,-1)
z02 = tail(mortrate,-2)
z1 = head(mortrate,-1)
z2 = head(mortrate,-2)
plot(z1, z01,xlab = "Z_t+1",ylab = "z_t",main="scatter plots of z_t+1 versus z_t ")
```

scatter plots of z_t+1 versus z_t
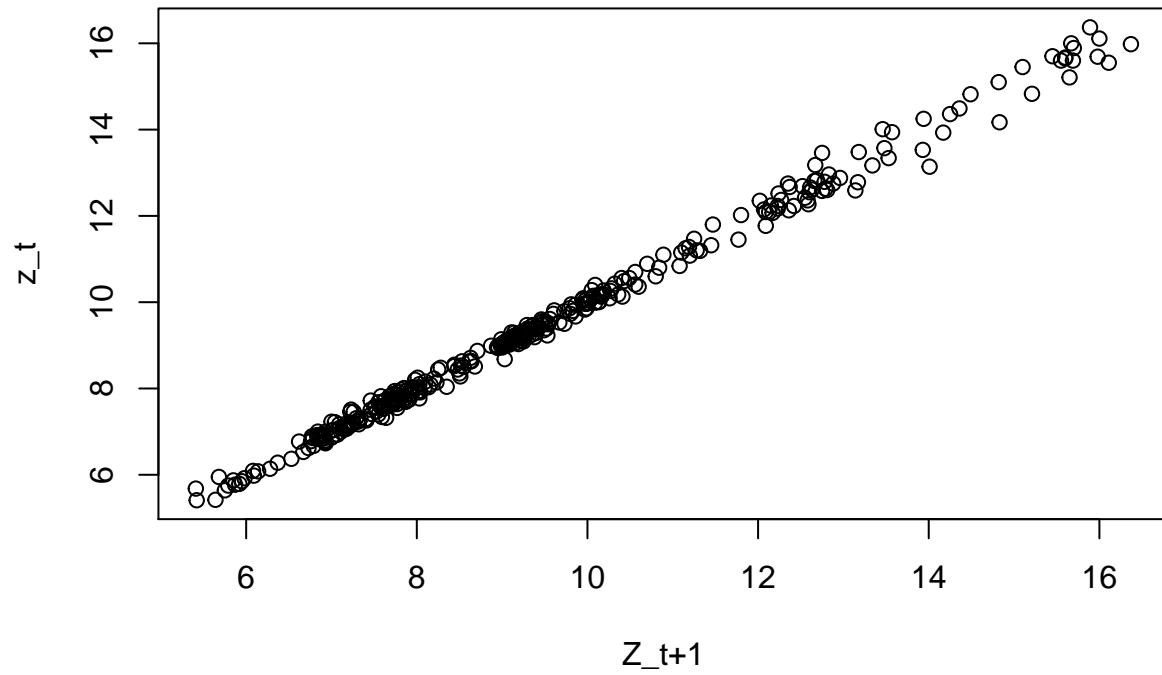

```r
plot(z2, z02,xlab = "Z_t+2",ylab = "z_t",main="scatter plots of z_t+2 versus z_t")
```

scatter plots of z_t+2 versus z_t

(b) Based on the plots in (a), do you think that the series is autocorrelated? Comment on whether you think the time series is stationary.
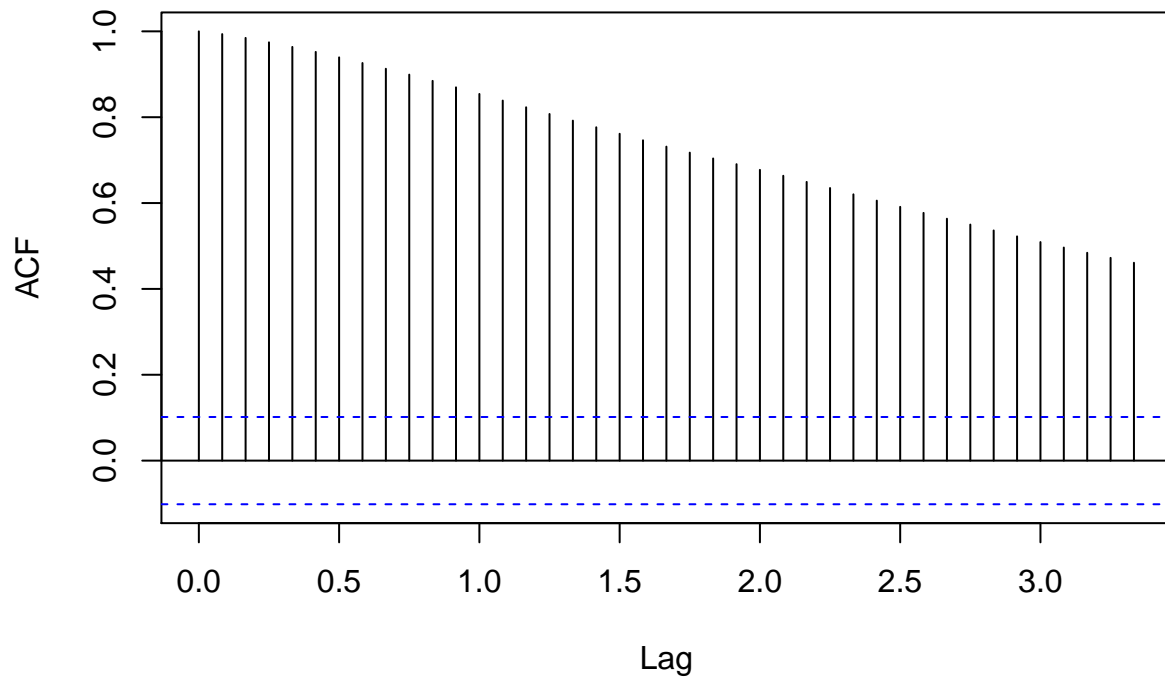
Since the scatter plot is linear, the series is autocorrelated

Since the trend of the time series has a crest beyond 1980, it is not stationary.

(c) Plot the sample autocorrelation for this series up to lag 40. Relate this plot to the plots in (a).

```
acf(mortrate,lag.max = 40)
```

Series mortrate

Problem 5

Consider the daily simple return of CRSP equal-weighted index from January 1980 to December 1999 in the file 'DEWRet.csv'. Use a regression model to study the effects of trading days on the index return. What is the fitted model? Are the weekday effects significant in the returns at the 5% level? Use the Newey-West estimator of the covariance matrix to obtain the t-values of regression estimates. Does it change the conclusion of weekday effect?
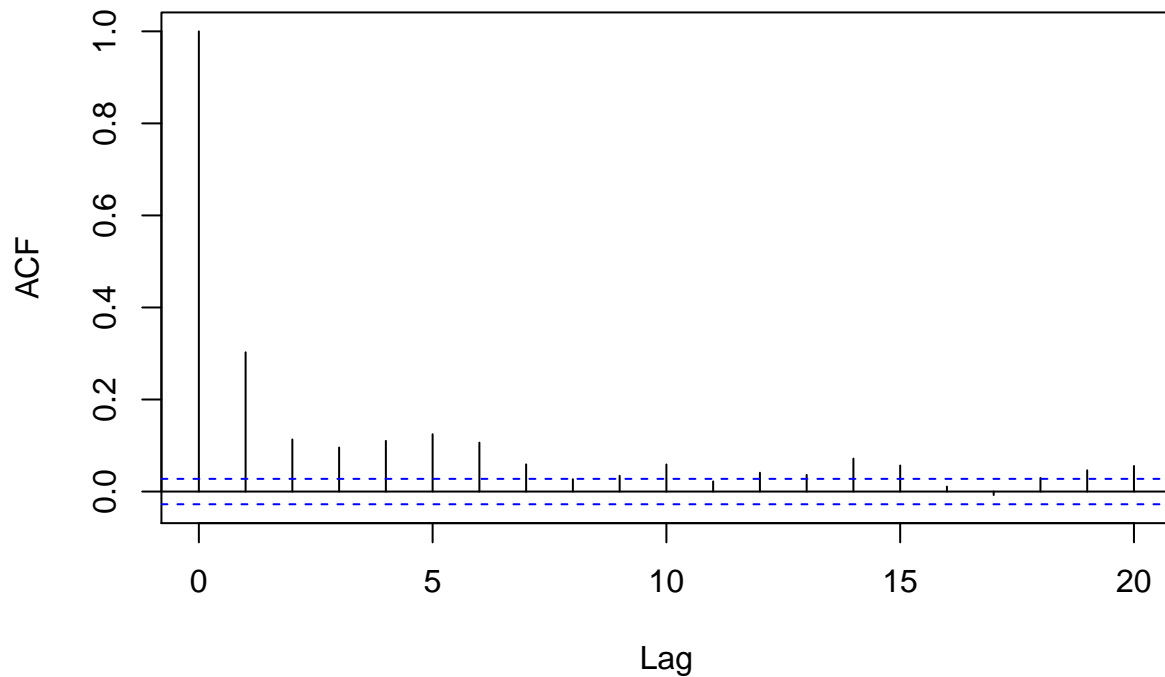
1) OLS Model
2) under 5% 3)it is not change the conclusion of weekday effect

```
crsp = read.csv("DEWRet.csv",skip = 11,header = T)
ret = ts(crsp[,7],start = 1980,freq = 365)
T = length(ret);t=1:T
week0 = 3:5; week1 = rep(1:5,1010);week2 = 1:3
week = c(week0,week1,week2)
w = factor(week)
nreg = lm(ret~t+w)
qqnorm(residuals(nreg))
```

Normal Q–Q Plot
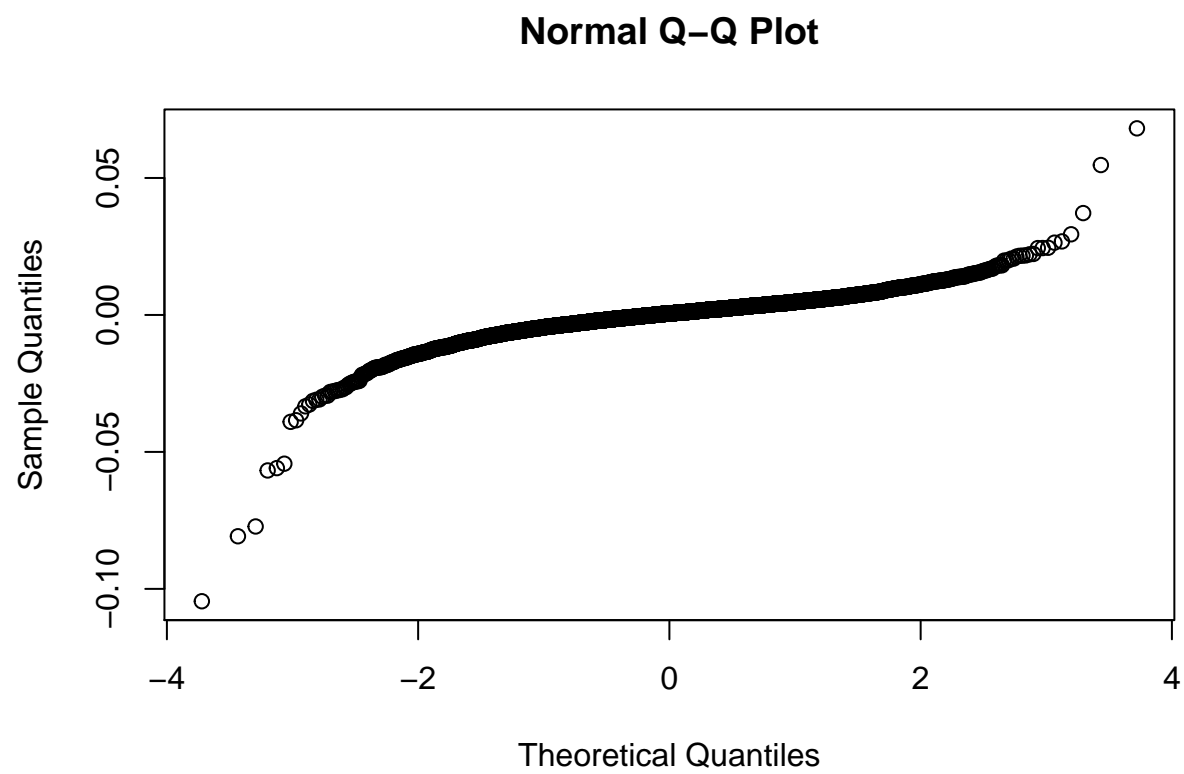

```r
acf(residuals(nreg), lag.max = 20)
```
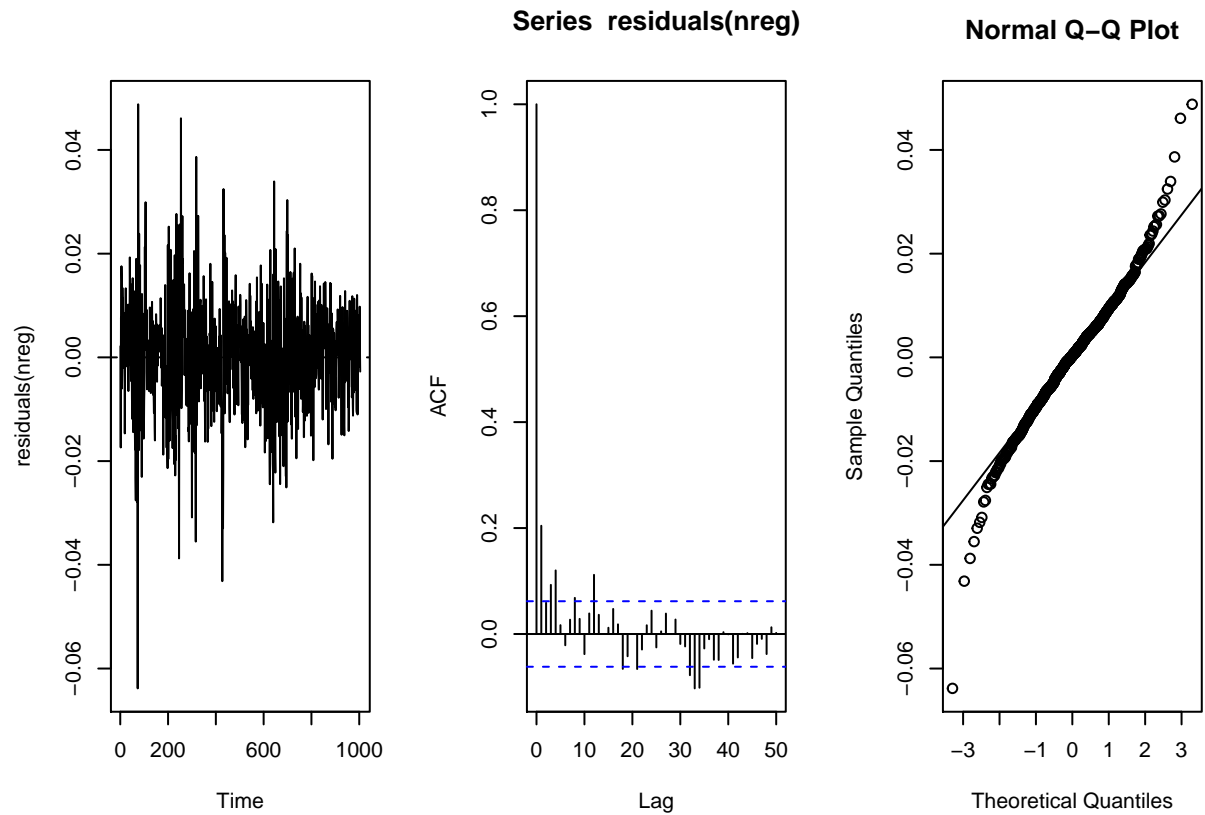
Series residuals(nreg)

Problem 6

Consider the daily returns of S&P composite index from January 3, 2000 to December 31, 2003. The data are in the file 'SP.csv'. Perform all tests using the 5% significance level, and answer the following questions:

(a) Is there a Friday effect on the daily simple returns of S&P composite index? You may employ a simple linear regression model to answer this question. Estimate the model via Newey-West estimator and test the hypothesis that there is no Friday effect. Draw your conclusion.

```
sp = read.csv("SP.csv",skip = 7, header = T)
ret = ts(sp[,2],start = 2000, freq = 1004/4)
T = length(ret); t = 1:T
week1 = rep(1:5,200);week2 = 1:4
week = c(week1,week2)
w = factor(week)
nreg = lm(ret~t+w)
par(mfrow = c(1, 3))
ts.plot(residuals(nreg)); abline(h=0, lty=2)
acf(residuals(nreg), lag.max = 50)
qqnorm(residuals(nreg)); qqline(residuals(nreg))
```

Series residuals(nreg)   **Normal Q–Q Plot**

(b) Check the residual serial correlations. Are there any significant serial correlations in the residuals?

Since almost all lag of the series residuals of nref is about the blue line, there is a significant serial correlation in the residuals.