Homework 1 Solution

Problem 1:

(a) $E(X) = p \cdot 1 + (1-p) \cdot 0 = p$

(b) $E(X^2) = p \cdot 1^2 + (1-p) \cdot 0^2 = p$ and $\text{Var}(X) = E(X^2) - [E(X)]^2 = p - p^2 = p(1-p)$

(c)

$$
\begin{aligned}
E[(X-\mu)^3] =& E[(X-p)^3] = p \cdot (1-p)^3 + (1-p) \cdot (0-p)^3 \\
=& p(1-p)^3 + (1-p)(-p)^3 \\
=& p(1-p)[(1-p)^2 - p^2] \\
=& p(1-p)[1 + p^2 - 2p - p^2] \\
=& p(1-p)(1-2p)
\end{aligned}
$$

It follows that

$$
\text{skew}(X) = E\left[\frac{(X-\mu)^3}{\sigma^3}\right] = \frac{p(1-p)(1-2p)}{[p(1-p)]^{3/2}} = \frac{1-2p}{\sqrt{p(1-p)}}.
$$

(d)

$$
\begin{aligned}
E[(X-\mu)^4] =& E[(X-p)^4] = p \cdot (1-p)^4 + (1-p) \cdot (0-p)^4 \\
=& p(1-p)^4 + (1-p)p^4 \\
=& p(1-p)[(1-p)^3 + p^3] \\
=& p(1-p)[1 + 3p^2 - 3p - p^3 + p^3] \\
=& p(1-p)(1 + 3p^2 - 3p)
\end{aligned}
$$

It follows that

$$
\text{kurt}(X) = E\left[\frac{(X-\mu)^4}{\sigma^4}\right] = \frac{p(1-p)(1+3p^2-3p)}{p^2(1-p)^2} = \frac{1+3p^2-3p}{p(1-p)}.
$$

Problem 2

(a)

$$
\begin{aligned}
\text{skew}(X) =& E\left[\frac{(X-\mu)^3}{\sigma^3}\right] \\
=& \frac{1}{\sigma^3} E[(X-\mu)^3] \\
=& \frac{1}{\sigma^3} E[X^3 - 3X^2\mu + 3X\mu^2 - \mu^3] \\
=& \frac{1}{\sigma^3} \left[E(X^3) - 3\mu E(X^2) + 3\mu^2 E(X) - \mu^3\right] \\
=& \frac{1}{\sigma^3} \left[E(X^3) - 3\mu(\sigma^2 + \mu^2) + 3\mu^3 - \mu^3\right] \\
=& \frac{1}{\sigma^3} \left[E(X^3) - 3\mu\sigma^2 - \mu^3\right]
\end{aligned}
$$

(b)

$$\begin{aligned}
\text{kurt}(X) =& E\left[\frac{(X-\mu)^4}{\sigma^4}\right] \\
=& \frac{1}{\sigma^4}E[(X-\mu)^4] \\
=& \frac{1}{\sigma^4}E[X^4 - 4X^3\mu + 6X^2\mu^2 - 4X\mu^3 + \mu^4] \\
=& \frac{1}{\sigma^4}\left[E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 4\mu^3 E(X) + \mu^4\right] \\
=& \frac{1}{\sigma^4}\left[E(X^4) - 4\mu E(X^3) + 6\mu^2(\sigma^2+\mu^2) - 4\mu^4 + \mu^4\right] \\
=& \frac{1}{\sigma^4}\left[E(X^4) - 4\mu E(X^3) + 6\mu^2\sigma^2 + 6\mu^4 - 4\mu^4 + \mu^4\right] \\
=& \frac{1}{\sigma^4}\left[E(X^4) - 4\mu E(X^3) + 6\mu^2\sigma^2 + 3\mu^4\right]
\end{aligned}$$

Problem 3 $X \sim N(\mu, \sigma^2)$

(a)

$$\begin{aligned}
M_X(t) =& E[e^{tX}] \\
=& \int_{-\infty}^{\infty} e^{tx} f(x)dx \\
=& \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)dx \\
=& \int_{-\infty}^{\infty} e^{tx}\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2+\mu^2-2x\mu}{2\sigma^2}\right)dx \\
=& \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{x^2+\mu^2-2x\mu-2\sigma^2tx}{2\sigma^2}\right)dx \\
=& \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[x-(\mu+\sigma^2t)]^2-\sigma^4t^2-2\mu\sigma^2t}{2\sigma^2}\right)dx \\
=& \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[x-(\mu+\sigma^2t)]^2}{2\sigma^2}\right)\exp\left(\frac{\sigma^2t^2}{2}+\mu t\right)dx \\
=& \exp\left(\frac{\sigma^2t^2}{2}+\mu t\right)\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{[x-(\mu+\sigma^2t)]^2}{2\sigma^2}\right)dx \qquad (*) \\
=& \exp\left(\frac{\sigma^2t^2}{2}+\mu t\right)
\end{aligned}$$

We note that the integrand in $(*)$ is a probability density function for nomral distribution $N(\mu+\sigma^2 t, \sigma^2)$ and therefore intergrates to 1.

$$E(X) = \frac{d}{dt}M_X(t)\bigg|_{t=0} = (t\sigma^2+\mu)\exp\left(\frac{\sigma^2t^2}{2}+\mu t\right)\bigg|_{t=0} = \mu$$

(b)

$$\begin{aligned}
E(X^2) =& \frac{d^2}{dt^2}M_X(t)\bigg|_{t=0} = \sigma^2\exp\left(\frac{\sigma^2t^2}{2}+\mu t\right) + (t\sigma^2+\mu)^2\exp\left(\frac{\sigma^2t^2}{2}+\mu t\right)\bigg|_{t=0} = \sigma^2+\mu^2 \\
\text{Var}(X) =& E(X^2) - [E(X)]^2 = \sigma^2+\mu^2-\mu^2 = \sigma^2
\end{aligned}$$

(c)

$$\begin{aligned}
E(X^3) =& \frac{d^3}{dt^3} M_X(t)\Big|_{t=0} \\
=& (\sigma^2 t + \mu)\sigma^2 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) + 2\sigma^2(\sigma^2 t + \mu)\exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) + (t\sigma^2 + \mu)^3 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)\Bigg|_{t=0} \\
=& \mu\sigma^2 + 2\sigma^2\mu + \mu^3 \\
=& \mu^3 + 3\mu\sigma^2
\end{aligned}$$

$$\text{skew}(X) = \frac{E(X^3) - 3\mu\sigma^2 - \mu^3}{\sigma^3} = \frac{\mu^3 + 3\mu\sigma^2 - 3\mu\sigma^2 - \mu^3}{\sigma^3} = 0$$

(d)

$$\begin{aligned}
E(X^4) =& \frac{d^4}{dt^4} M_X(t)\Big|_{t=0} \\
=& (\sigma^2 t + \mu)^2 \sigma^2 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) + \sigma^4 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) \\
& + 2\sigma^2(\sigma^2 t + \mu)^2 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) + 2\sigma^4 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) \\
& + (\sigma^2 t + \mu)^4 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right) + 3(\sigma^2 t + \mu)^2 \sigma^2 \exp\left(\frac{\sigma^2 t^2}{2} + \mu t\right)\Bigg|_{t=0} \\
=& \mu^2\sigma^2 + \sigma^4 + 2\sigma^2\mu^2 + 2\sigma^4 + \mu^4 + 3\mu^2\sigma^2 \\
=& 6\mu^2\sigma^2 + 3\sigma^4 + \mu^4
\end{aligned}$$

$$\begin{aligned}
\text{kurt}(X) =& \frac{E(X^4) - 4\mu E(X^3) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} \\
=& \frac{6\mu^2\sigma^2 + 3\sigma^4 + \mu^4 - 4\mu(\mu^3 + 3\mu\sigma^2) + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} \\
=& \frac{6\mu^2\sigma^2 + 3\sigma^4 + \mu^4 - 4\mu^4 - 12\mu^2\sigma^2 + 6\mu^2\sigma^2 + 3\mu^4}{\sigma^4} \\
=& \frac{3\sigma^4}{\sigma^4} \\
=& 3
\end{aligned}$$

Problem 4:

(a) We recall the definition of residual sum of squares:

$$\text{SS Residual} = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

To minimize SS Residual, we take derivative for above function with respect to β_0 and β_1 respectively:

$$\frac{\partial \text{ SS Residual}}{\partial \beta_0} = -2\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \tag{1}$$

$$\frac{\partial \text{ SS Residual}}{\partial \beta_1} = -2\sum_{i=1}^{n} x_i(y_i - \beta_0 - \beta_1 x_i) \tag{2}$$

Setting equation (1) to be 0, we have

$$n\bar{y} - n\beta_0 - n\beta_1\bar{x} = 0 \Longrightarrow \beta_0 = \bar{y} - \beta_1\bar{x}. \tag{3}$$

Setting equation (2) to be 0, we have

$$\sum_{i=1}^{n} x_i y_i - n\bar{x}\beta_0 - \beta_1 \sum_{i=1}^{n} x_i^2 = 0$$

Plugging in the value of β_0 from equation (3), we have

$$\begin{aligned}
\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} + n\bar{x}^2\beta_1 - \beta_1 \sum_{i=1}^{n} x_i^2 =& 0 \\
\beta_1 \left[\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right] =& \sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y} \\
\beta_1 =& \frac{\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2} \\
\beta_1 =& \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}
\end{aligned}$$

Therefore, we have

$$\hat{\beta}_0 = \bar{y} - \beta_1\bar{x} \quad \text{and} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

(b) Given that ϵ_i's are iid random variables, i.e., $\epsilon_i \sim_{iid} N(0, \sigma^2)$, $i = 1, \cdots, n$, we have $y_i \sim_{iid} N(\beta_0 + \beta_1 x_i, \sigma^2)$. The likelihood function for observing random sample $y_1, \cdots, y_n$ is

$$\begin{aligned}
L(\beta_0, \beta_1) =& \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \\
=& \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left(-\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right)
\end{aligned}$$

MLEs for β_0 and β_1 are obtained by maximizing the likelihood function. For each fixed σ^2, maximizing likelihood function is the same as minimizing $\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$, therefore, MLEs for β_0 and β_1 are the same as LSEs. To find our the MLE for σ^2, we maximize the log likelihood function

$$l(\beta_0, \beta_1) = \log L(\beta_0, \beta_1) = \text{constant} - n\log\sigma - -\frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}$$

Taking derivative with respect to the parameter σ, we have

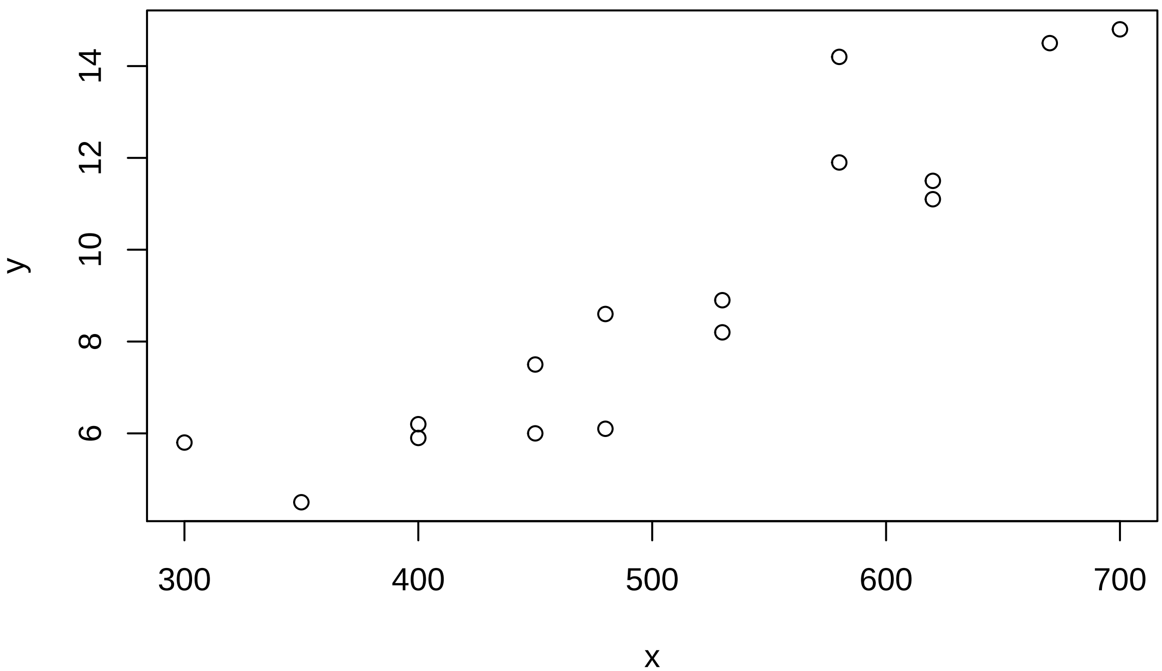
$$\frac{\partial l(\beta_0, \beta_1)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{\sigma^3} \tag{4}$$

Setting equation (4) to be equal 0, we have

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2}{n}$$

Problem 5:

```r
#(a)#
x<-c(300,350,400,400,450,450,480,480,530,530,580,580,620,620,670,700)
y<-c(5.8,4.5,5.9,6.2,6.0,7.5,6.1,8.6,8.9,8.2,14.2,11.9,11.1,11.5,14.5,14.8)
plot(x,y) # scatterplot
```



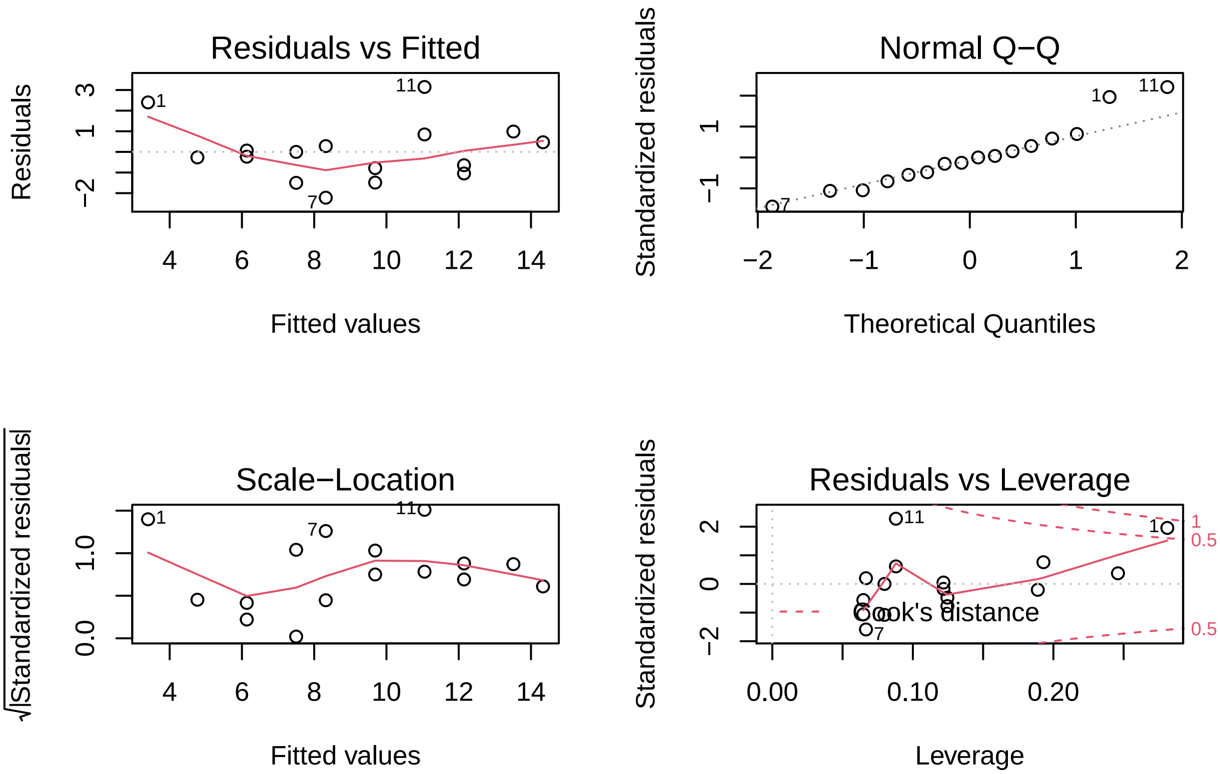

```r
#(b)#
# linear model: estimated intercept=-4.798, estimated slope=0.027
summary(lm(y~x))
```

```
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2.2205 -0.8520 -0.1173  0.5616  3.1464
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -4.79841    1.68750  -2.844    0.013 *
## x            0.02733    0.00324   8.436 7.35e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1.447 on 14 degrees of freedom
## Multiple R-squared:  0.8356, Adjusted R-squared:  0.8239
## F-statistic: 71.16 on 1 and 14 DF,  p-value: 7.346e-07
```

```r
#(c)#
# ANOVA table: result from F-test suggests that there is a strong linear relationship
# between response and explanatory variables. Recall that null hypothesis for F-test is
# H_0: slope=0 and alternative hypothesis is H_a: slope does not equal to 0.
anova(lm(y~x))
```

```
## Analysis of Variance Table
##
## Response: y
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## x          1 148.930 148.930  71.163 7.346e-07 ***
## Residuals 14  29.299   2.093
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
#(d)#
# From model summary, coefficient of determination is 0.8356 or from ANOVA table,
# coefficient of determination is 148.930/(148.930+29.299)=0.8356.
#(e)#
# First of all, note that this is a small data set with only 16 observations and therefore
# it is hard to justify some of the assumptions from residual plots only. Still, QQ plot
# suggests that normality assumption has been well satisfied; since no obvious pattern
# exists in residuals vs. fitted values plot, assumption of constant variance has not been
# violated. Leverage plot suggests that some influential observations exist.
par(mfrow=c(2,2))
plot(lm(y~x))
```



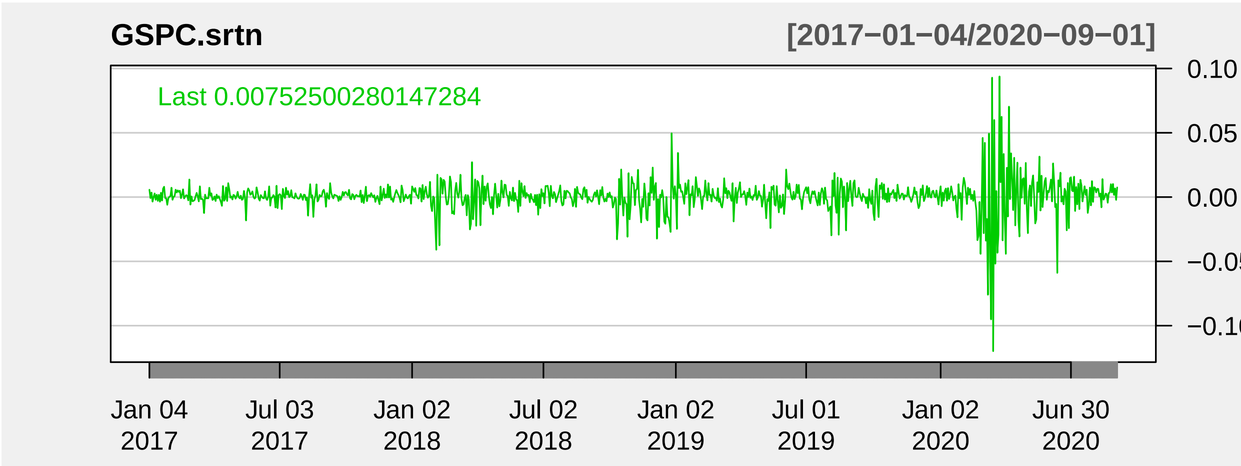
Problem 6:

```r
#(a)#
library(quantmod)
getSymbols("^GSPC",from="2017-01-01",to="2020-09-02")
```

```
## [1] "^GSPC"
```

```r
GSPC.lrtn=diff(log(GSPC$GSPC.Adjusted))[-1,] # log returns
GSPC.srtn=exp(GSPC.lrtn)-1 # simple net returns
```

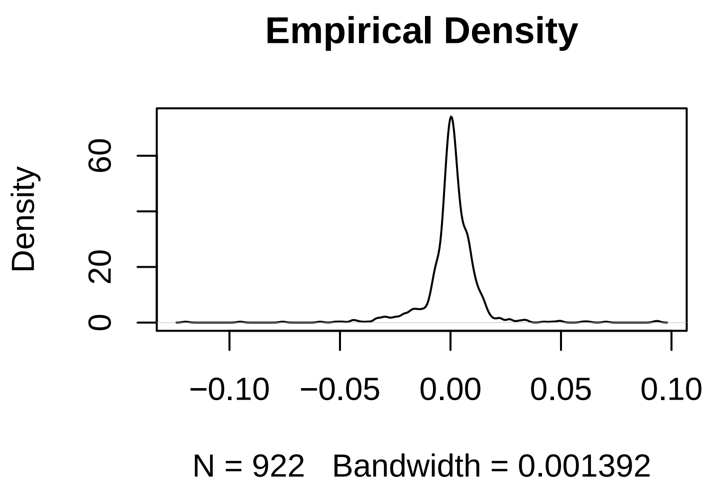
```r
chartSeries(GSPC.srtn,theme="white")
```




```r
#(b)#
library(fBasics)
# sample mean=0.000458; sample s.d.= 0.008280; skewness=-0.540166; excess kurtosis=4.466750;
# minimum=-0.040979; maximum=0.049594.
basicStats(GSPC.srtn)
```

```
##             GSPC.Adjusted
## nobs           922.000000
## NAs              0.000000
## Minimum         -0.119841
## Maximum          0.093828
## 1. Quartile     -0.002776
## 3. Quartile      0.005343
## Mean             0.000569
## Median           0.000816
## Sum              0.524586
## SE Mean          0.000428
## LCL Mean        -0.000272
## UCL Mean         0.001410
## Variance         0.000169
## Stdev            0.013009
## Skewness        -0.730852
## Kurtosis        20.560740
```

```r
#(c)#
plot(density(GSPC.srtn), main="Empirical Density") # empirical density plot
```




```r
# Based on the output below, null hypothesis of normality has been rejected by Jarque-Bera test.
normalTest(as.vector(GSPC.srtn), method="jb")
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 16403.7349
##   P VALUE:
##     Asymptotic p Value: < 2.2e-16
##
## Description:
##  Sat Sep 12 13:29:20 2020 by user:
```

```r
#(d)#
# sample mean=0.000423; sample s.d.= 0.008297; skewness=-0.616535; excess kurtosis=4.488324;
# minimum=-0.041843; maximum=0.048403.
basicStats(GSPC.lrtn)
```

```
##                GSPC.Adjusted
## nobs              922.000000
## NAs                 0.000000
## Minimum            -0.127652
## Maximum             0.089683
## 1. Quartile        -0.002779
## 3. Quartile         0.005328
## Mean                0.000484
## Median              0.000815
## Sum                 0.445944
## SE Mean             0.000431
## LCL Mean           -0.000362
## UCL Mean            0.001330
## Variance            0.000171
## Stdev               0.013087
## Skewness           -1.170368
## Kurtosis           22.075674
```

```r
#(e)#
t.test(GSPC.lrtn) # null hypothesis of zero mean cannot been rejected
```

```
## Warning in tstat + c(-cint, cint): Recycling array of length 1 in array-vector arithmetic is depreca
##   Use c() or as.vector() instead.

## Warning in cint * stderr: Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.

##
##  One Sample t-test
##
## data:  GSPC.lrtn
## t = 1.1222, df = 921, p-value = 0.2621
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.0003621905  0.0013295315
## sample estimates:
##    mean of x
## 0.0004836705
```

```r
#(f)#
# test statistics based on skewness is -7.650997 and the absolute value of it is greater
# than 1.96, null hypothesis of symmetry will be rejected.
skewness(GSPC.lrtn)/sqrt(6/length(GSPC.lrtn))
```

```
## [1] -14.50815
## attr(,"method")
## [1] "moment"
```

```r
#(g)#
# test statistics based on excess kurtosis is 27.84932 and is greater than 1.96, null
# hypotheis of normal tails can be rejected.
kurtosis(GSPC.lrtn)/sqrt(24/length(GSPC.lrtn))
```

```
## [1] 136.8276
## attr(,"method")
## [1] "excess"
```

Problem 7:

```r
getFX("USD/CNY",from="2020-04-01",to="2020-9-2")
```

```
## [1] "USD/CNY"
```

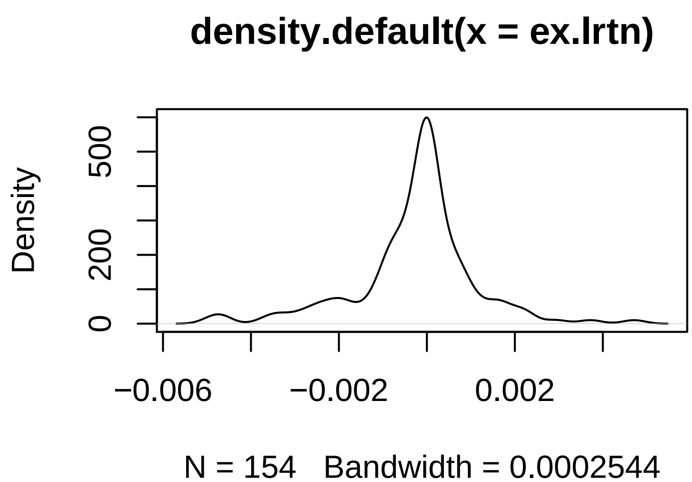
```r
#(a)#
ex.lrtn=diff(log(USDCNY$USD.CNY))[-1,] # compute daily log returns
head(ex.lrtn)
```

```
##                  USD.CNY
## 2020-04-02 -0.0002952694
## 2020-04-03 -0.0005356538
## 2020-04-04  0.0001420136
## 2020-04-05  0.0000000000
## 2020-04-06  0.0001002576
## 2020-04-07 -0.0046378678
```

```r
#(b)#
basicStats(ex.lrtn)
```

```
##                USD.CNY
## nobs        154.000000
## NAs           0.000000
## Minimum      -0.004929
## Maximum       0.004708
## 1. Quartile  -0.000742
## 3. Quartile   0.000295
## Mean         -0.000246
## Median        0.000000
## Sum          -0.037900
## SE Mean       0.000111
## LCL Mean     -0.000466
## UCL Mean     -0.000027
## Variance      0.000002
## Stdev         0.001378
## Skewness     -0.456055
## Kurtosis      2.571355
```

```r
#(c)#
plot(density(ex.lrtn))
```




```r
#(d)#
t.test(ex.lrtn) # null hypothesis of zero mean can be rejected.
```

```
## Warning in tstat + c(-cint, cint): Recycling array of length 1 in array-vector arithmetic is depreca
##   Use c() or as.vector() instead.

## Warning in cint * stderr: Recycling array of length 1 in vector-array arithmetic is deprecated.
##   Use c() or as.vector() instead.

##
##  One Sample t-test
##
## data:  ex.lrtn
## t = -2.2156, df = 153, p-value = 0.02819
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.655441e-04 -2.666455e-05
## sample estimates:
##     mean of x
```

```
## -0.0002461043
```

```r
#(e)#
# Based on the output below, null hypothesis of normality has been rejected by Jarque-Bera test.
normalTest(as.vector(ex.lrtn), method="jb")
```

```
##
## Title:
##  Jarque - Bera Normalality Test
##
## Test Results:
##   STATISTIC:
##     X-squared: 50.3152
##   P VALUE:
##     Asymptotic p Value: 1.186e-11
##
## Description:
##  Sat Sep 12 13:29:21 2020 by user:
```