STATS 509 HW3

unique name: tangsw

id: 31975136

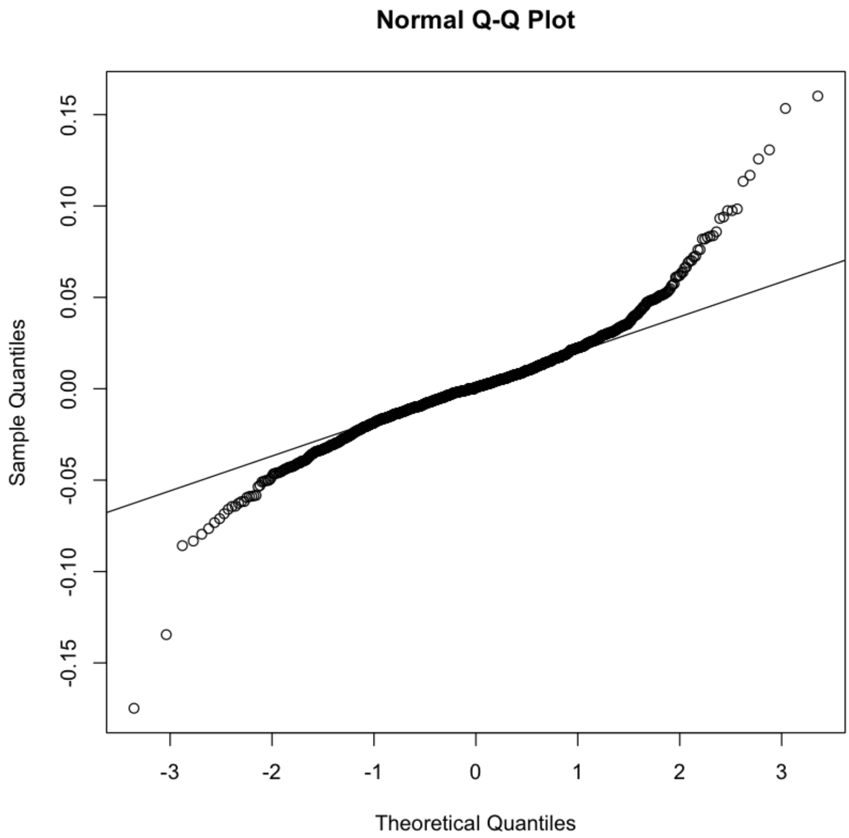
In [1]:
```
library(ggplot2)
source("startup.R")
library('MASS')
library('EnvStats')
library("fGarch")
library('cowplot')
```

```
Attaching package: ‘EnvStats’


The following object is masked from ‘package:MASS’:

    boxcox


The following objects are masked from ‘package:stats’:

    predict, predict.lm


The following object is masked from ‘package:base’:

    print.default


Loading required package: timeDate


Attaching package: ‘timeDate’


The following objects are masked from ‘package:EnvStats’:

    kurtosis, skewness


Loading required package: timeSeries

Loading required package: fBasics
```

Q1

(a)

In [2]:
```
df = read.csv("RecentFord.csv")
# head(df)
# calculate return
n = length(df$Adj.Close)
R = df$Adj.Close[-1]/df$Adj.Close[-n] - 1
qqnorm(R)
qqline(R)
```



The return doesn't look normally distributed; instead it has heavier tails on both sides than a normal distribution.

(b)

In [3]:
```
shapiro.test(R)
```

```
        Shapiro-Wilk normality test

data:  R
W = 0.93151, p-value < 2.2e-16
```

p-value is smaller than 2.2e-16. We can reject the null hypothesis of a normal distribution at 0.01 confidence level.

(c)

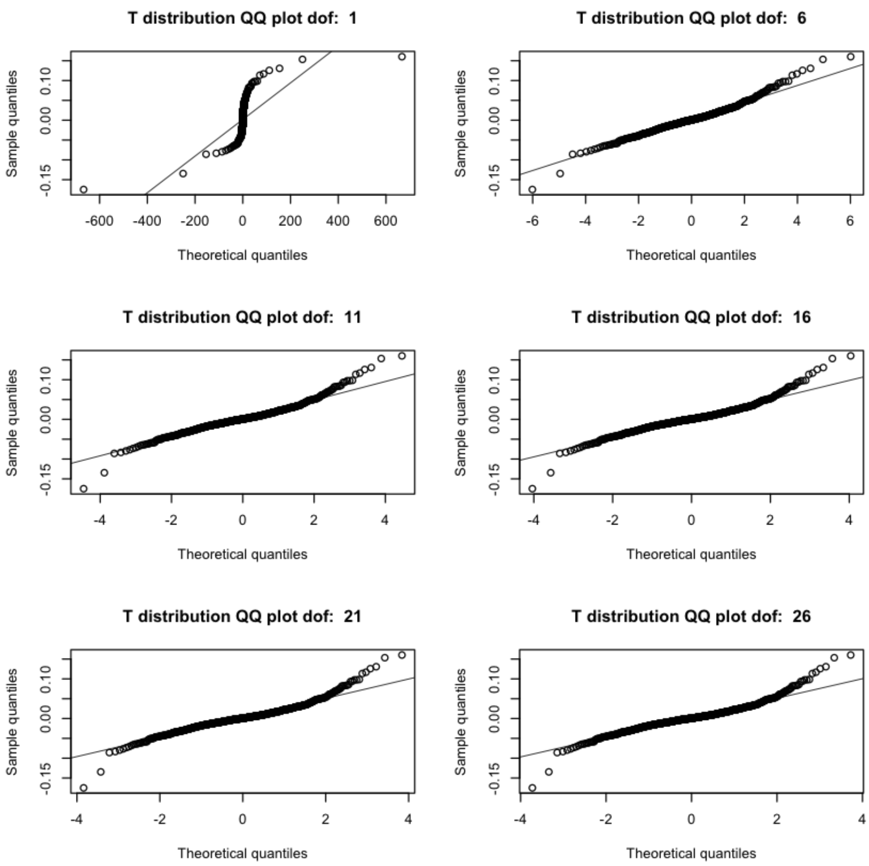
In [4]:
```
params <- as.list(MASS::fitdistr(df$Adj.Close, "t"))
```

```
Warning message in dt((x - m)/s, df, log = TRUE):
“NaNs produced”
Warning message in log(s):
“NaNs produced”
```

In [5]:
```
params
```

```
       m              s              df     
  11.1752019     3.4998282    18.5975864 
 ( 0.1157265)  ( 0.1353128)  (11.2963089)
```

In [6]:
```
df_grid = seq(1,30,5)
par(mfrow = c(3,2))
for (df in df_grid){
    qqPlot(x = R,
           distribution = "t",
           param.list = list(df = df),
           add.line = TRUE,
          main = paste("T distribution QQ plot dof: ",df),
          xlab = "Theoretical quantiles",
          ylab = "Sample quantiles")
}
```



Judged on the MASS distribution fit function, when Dof is close to 18 (dof = 16 in our case), the QQ plot is tend to be as linear as possible.

(d)

Based on the result from (c), the distribution is heavily right skewed hence it has a heavier, longer right tail than its left.

2

(a)

$\mathbb{E}(f_b(x)) = \mathbb{E}(\frac{1}{n}\sum_{n=1}^{n} K_b(X - X_i))$

K_b(X-X_i) is iid to some pdf

$= \mathbb{E}(K_b(X - X_i))$

$= \int_{x=x-w/2}^{x+w/2} K_b(x - x')f(x')dx'$

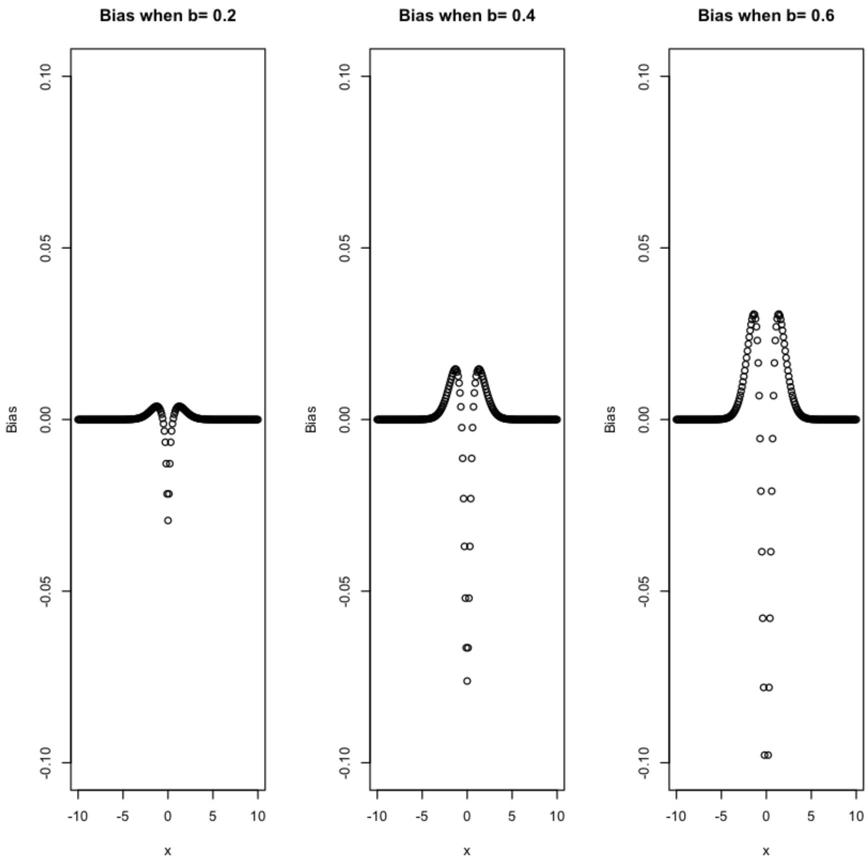
$= \frac{1}{w}(F(x + w/2) - F(x - w/2))$ where w = 3.464b

$= \frac{1}{3.464b}(F(x + 1.732b) - F(x - 1.732b))$

(b)

Bias = $\frac{1}{3.464b}(F(x + 1.732b) - F(x - 1.732b)) - f_X(x)$

In [7]:
```
Bias_func = function(x,b){
    bias = 1/(3.464 * b)*(pged(q = x + 1.732*b, mean = 0,sd = 1, nu
                          pged(q = x - 1.732*b, mean = 0,sd = 1, nu
                        dged(x = x, mean = 0,sd = 1, nu = 1.5)
    return(bias)
}
par(mfrow = c(1,3))
for (b in c(.2,.4,.6)){
    x = seq(-10,10,0.1)
    bias = Bias_func(x, b)
    plot(x = x, y = bias,xlab = "x",ylab = "Bias",main = paste("Bia
}
```



(c)

$Var(f_b(x)) = \frac{1}{n}Var(K_b(X - X_i))$

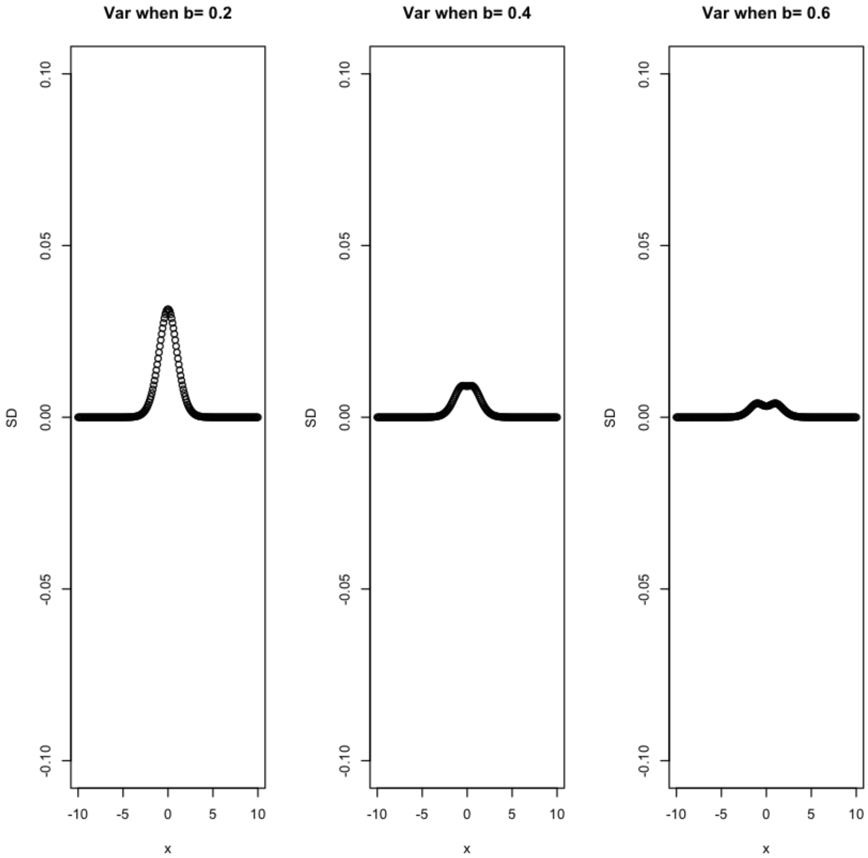
$= \frac{1}{n}[\mathbb{E}(K_b^2(X - X_i)) - \mathbb{E}^2(K_b(X - X_i)]$

$= \frac{1}{n}\{\frac{1}{3.464b}^2[F(x + 1.732b) - F(x - 1.732b)] - [\frac{1}{3.464b}(F(x + 1.732b) - F(x - 1.7$

$= \frac{1}{n}\frac{1}{3.464b}^2[F(x + 1.732b) - F(x - 1.732b)][1 - F(x + 1.732b) + F(x - 1.732b))]$

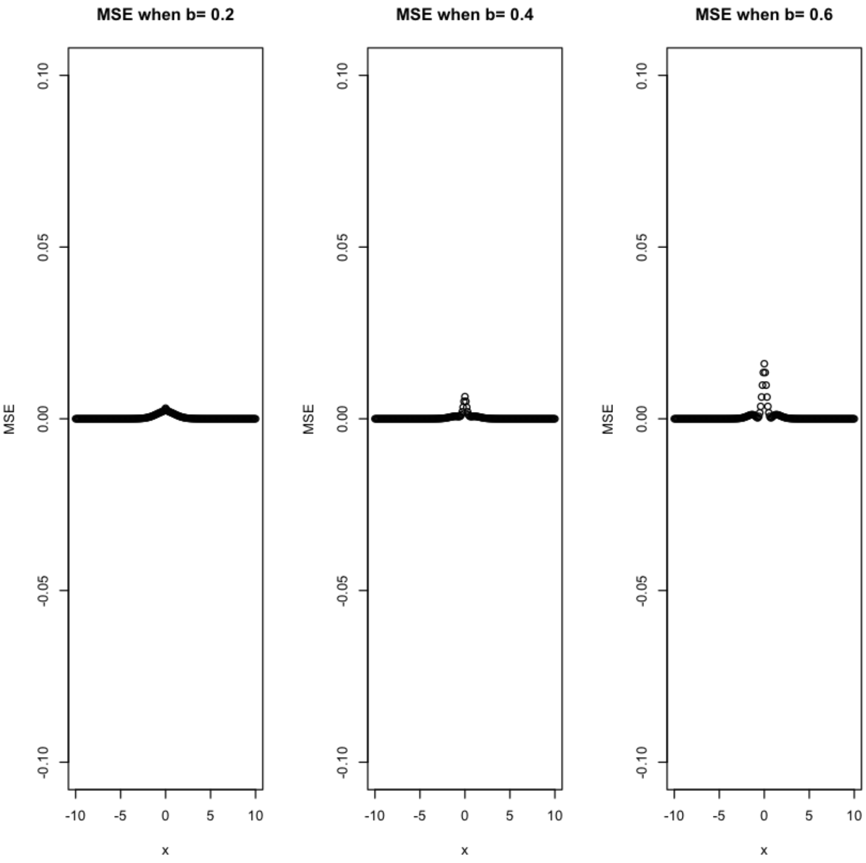
$sd = \sqrt{Var(f_b(x))}$

In [8]:
```
# note that var doesn't include n term
Var_func = function(x,b){
    var = 1/(3.464 * b)^2 *(pged(q = x + 1.732*b, mean = 0,sd = 1,
                             pged(q = x - 1.732*b, mean = 0,sd = 1, nu
                           (1-pged(q = x + 1.732*b, mean = 0,sd = 1, n
    return(var)
}
par(mfrow = c(1,3))
for (b in c(.2,.4,.6)){
    x = seq(-10,10,0.1)
    var = Var_func(x, b)/sqrt(200)
    plot(x = x, y = var,xlab = "x",ylab = "SD",main = paste("Var wh
}
```



(d)

In [9]:
```
par(mfrow = c(1,3))
df = data.frame(x=seq(-10,10,0.1))
for (b in c(.2,.4,.6)){
    x = seq(-10,10,0.1)
    mse = Var_func(x, b)/200 + Bias_func(x, b)^2
    plot(x = x, y = mse,xlab = "x",ylab = "MSE",main = paste("MSE w
}
```



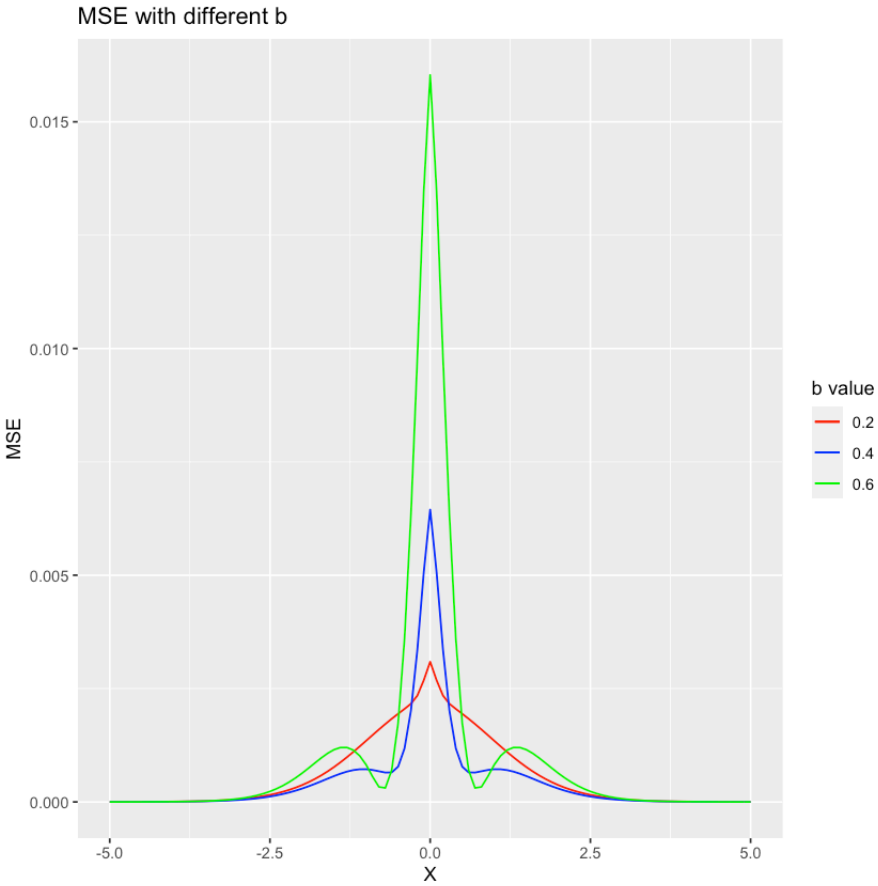
In [10]:
```
df$a = Var_func(df$x, 0.2)/200 + Bias_func(df$x, 0.2)^2
df$b = Var_func(df$x, 0.4)/200 + Bias_func(df$x, 0.4)^2
df$c = Var_func(df$x, 0.6)/200 + Bias_func(df$x, 0.6)^2
head(df)
```

A data.frame: 6 × 4

	x	a	b	c
	<dbl>	<dbl>	<dbl>	<dbl>
1	-10.0	5.140056e-14	5.007790e-14	7.955562e-14
2	-9.9	7.473801e-14	7.239975e-14	1.142647e-13
3	-9.8	1.084647e-13	1.044707e-13	1.637970e-13
4	-9.7	1.571104e-13	1.504573e-13	2.343410e-13
5	-9.6	2.271366e-13	2.162669e-13	3.346066e-13
6	-9.5	3.277409e-13	3.102556e-13	4.768260e-13

In [11]:
```
t = ggplot( data = df, aes(x = x)) +
      geom_line(aes(y = a, colour = "0.2")) +
      geom_line(aes(y = b, colour = "0.4")) +
    geom_line(aes(y = c, colour = "0.6")) +
    xlim(-5,5)+
    labs(x = "X", y = "MSE",title ="MSE with different b")+
    scale_colour_manual(name = "b value", values = c("0.2" = "red",
t
```

```
Warning message:
“Removed 100 row(s) containing missing values (geom_path).”
Warning message:
“Removed 100 row(s) containing missing values (geom_path).”
Warning message:
“Removed 100 row(s) containing missing values (geom_path).”
```



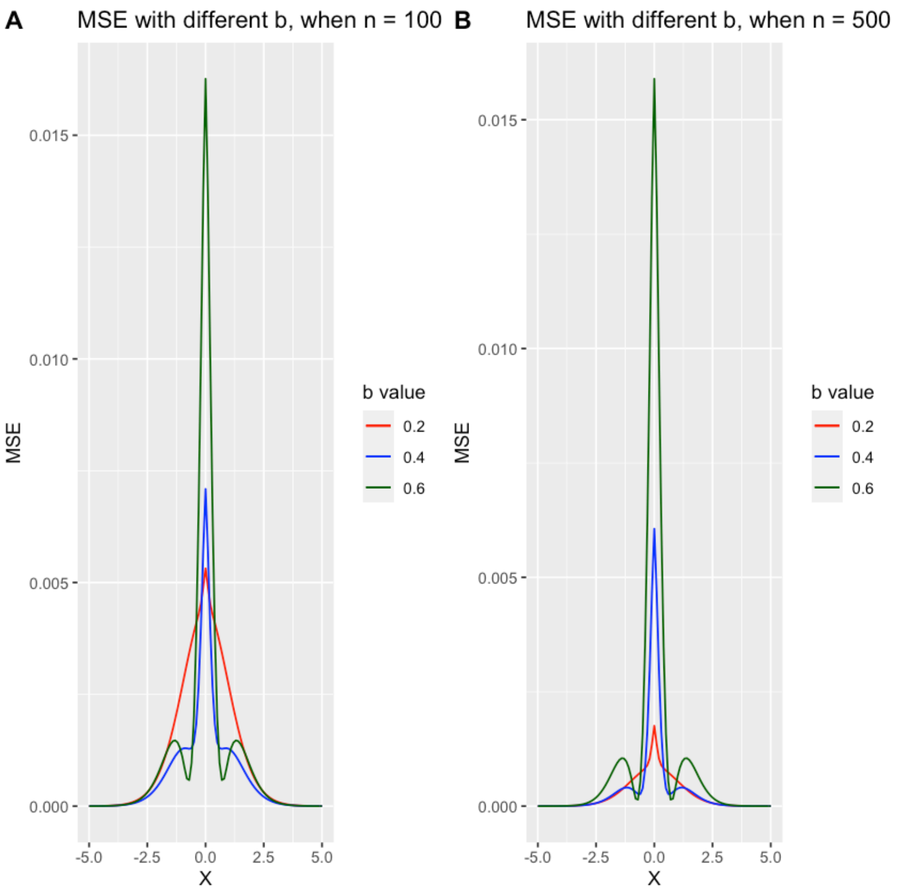
Based on the MSE graph, we should prefer 0.4

(d)

In [12]:
```
df1 = data.frame(x=seq(-5,5,0.1))
df1$a = mse = Var_func(df1$x, 0.2)/100 + Bias_func(df1$x, 0.2)^2
df1$b = mse = Var_func(df1$x, 0.4)/100 + Bias_func(df1$x, 0.4)^2
df1$c = mse = Var_func(df1$x, 0.6)/100 + Bias_func(df1$x, 0.6)^2
df2 = data.frame(x=seq(-25,25,0.1))
```

```
df2$a = mse = Var_func(df2$x, 0.2)/500 + Bias_func(df2$x, 0.2)^2
df2$b = mse = Var_func(df2$x, 0.4)/500 + Bias_func(df2$x, 0.4)^2
df2$c = mse = Var_func(df2$x, 0.6)/500 + Bias_func(df2$x, 0.6)^2
p1 = ggplot( data = df1, aes(x = x)) +
      geom_line(aes(y = a, colour = "0.2")) +
      geom_line(aes(y = b, colour = "0.4")) +
    geom_line(aes(y = c, colour = "0.6")) +
    labs(x = "X", y = "MSE",title ="MSE with different b, when n =
    scale_colour_manual(name = "b value", values = c("0.2" = "red",
p2 = ggplot( data = df2, aes(x = x)) +
      geom_line(aes(y = a, colour = "0.2")) +
      geom_line(aes(y = b, colour = "0.4")) +
    geom_line(aes(y = c, colour = "0.6")) +
    xlim(-5,5)+
    labs(x = "X", y = "MSE",title ="MSE with different b, when n =
    scale_colour_manual(name = "b value", values = c("0.2" = "red",
plot_grid(p1, p2, labels = "AUTO")
```

```
Warning message:
“Removed 400 row(s) containing missing values (geom_path).”
Warning message:
“Removed 400 row(s) containing missing values (geom_path).”
Warning message:
“Removed 400 row(s) containing missing values (geom_path).”
```



when we increase from n =100 to n= 500, we would possiblt change the b to smaller value 0.2, because it has smaller integrated MSE.

In [ ]: