Regression and Time Series:HW2

Yaniv Bronshtein

9/22/2021

Import the necessary libraries

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.6     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

1. Calculate the market returns by taking the weighted average of columns 3 to 5, with weights 0.3, 0.4 and 0.3. **Load in the data**

```
data <- read_table("Portfolios_Formed_on_ME_eqW.txt")
```

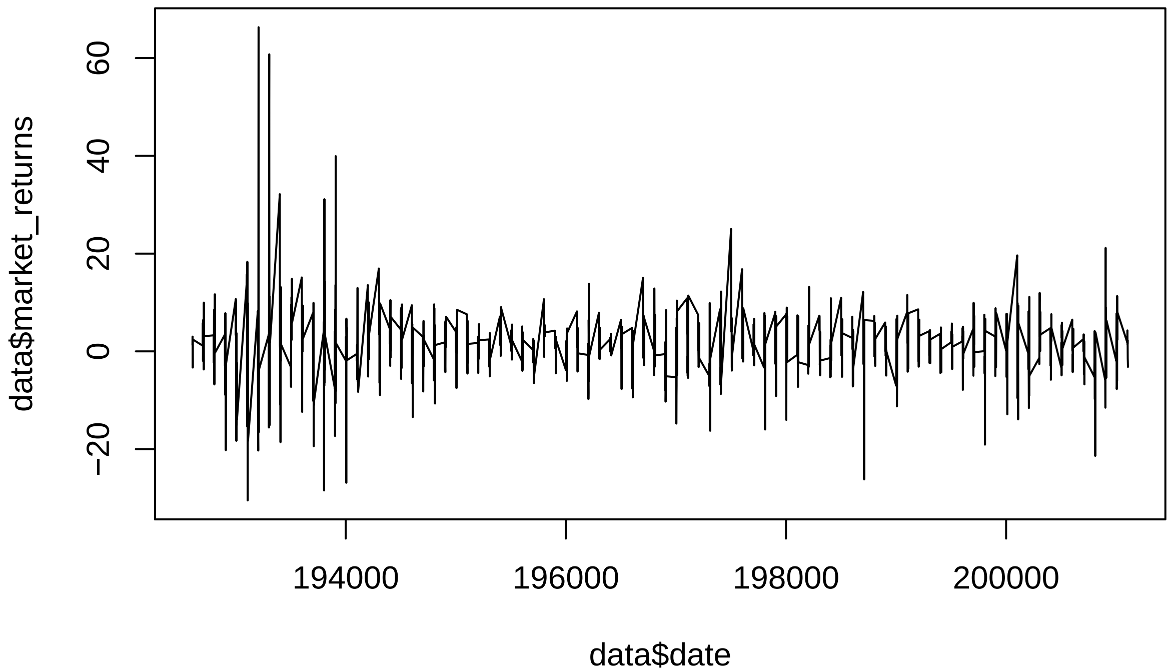
```
##
## -- Column specification --------------------------------------------------------
## cols(
##   .default = col_double()
## )
## i Use `spec()` for the full column specifications.
```

Calculate the weighted average

```
data <- data %>% mutate(market_returns=0.3*Lo30 + 0.4*Med40 + 0.3*Hi30)
```

(a) Plot the market return series (`plot(x,type='l')` will do).

```
plot(x=data$date, y=data$market_returns, type='l')
```



(b) Comment on the features you see in the figure *By setting the x-axis to date and y-axis to market_returns, one can see dramatic Oscillation between prior to 1940 which was the start of WW2 war effort by the U.S Looking at history, this was the time that the U.S came out of the Great Depression In the years that followed, there were key events that caused severe drops, but nothing like the stock market crash of 1929. Notable mentions are Black Monday(1987) and the 2008 financial crisis*

2. Use the data from 1965.01 to 1969.12 (rows 463:522) only to fit a simple linear regression using Column 11 (the small company portfolio) as the response variable and the market return you obtained in (1) as explanatory variable. Answer the following questions:

```r
sub_df <- data[463:522,]
model.fit <- lm(formula = Lo10~market_returns, data = sub_df)
summary <- summary(model.fit)
summary
```

```
##
## Call:
## lm(formula = Lo10 ~ market_returns, data = sub_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3075 -1.5564 -0.2021  1.4379  8.1134
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.59996    0.34120   1.758    0.084 .
## market_returns  1.39472    0.06497  21.466   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.556 on 58 degrees of freedom
## Multiple R-squared:  0.8882, Adjusted R-squared:  0.8863
## F-statistic: 460.8 on 1 and 58 DF,  p-value: < 2.2e-16
```

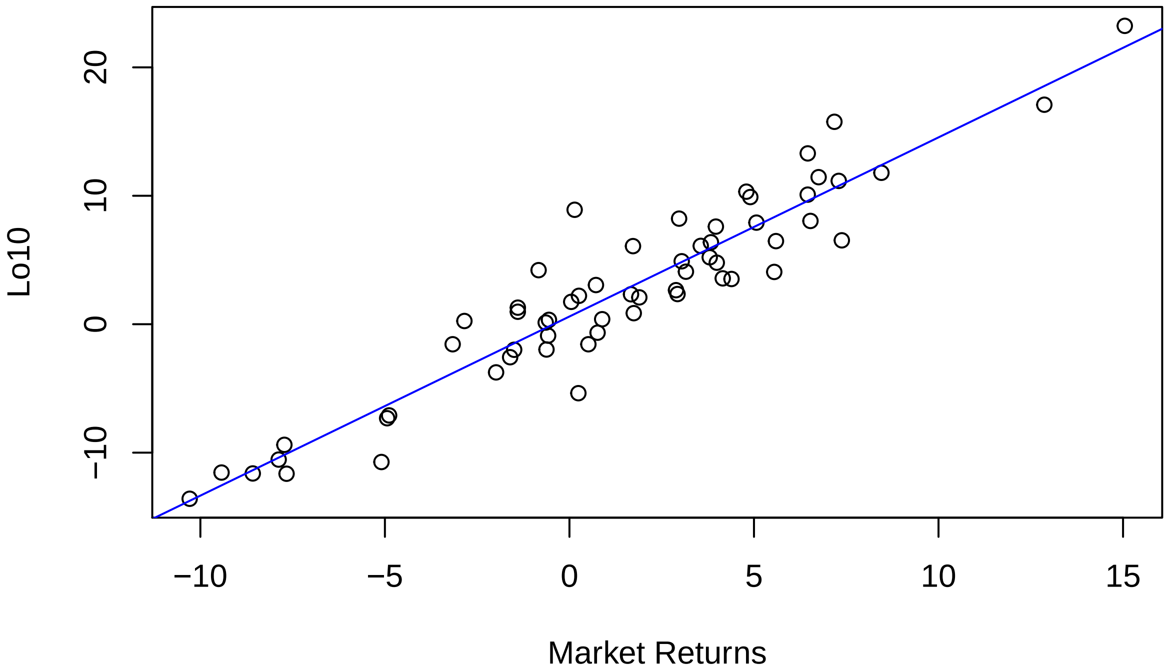
(a) Draw a scatter plot and add the estimated regression line to the plot. **Extract the coefficients**

```
coefs <- summary$coefficients
coefs
```

```
##                 Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)    0.5999598 0.34120231  1.75837 8.396004e-02
## market_returns 1.3947160 0.06497297 21.46610 2.803936e-29
```

Create plot and add abline

```
plot(x=sub_df$market_returns, y=sub_df[[11]], xlab = "Market Returns", ylab = "Lo10")
abline(model.fit, col='blue')
```



(b) What is the standard error of b1 for estimating β_1? *The standard error for estimating β_1 is 0.06497297* (c) Obtain a 95% confidence interval for b1. What does a confidence interval mean in general?

```
confint(model.fit, 'market_returns', level=0.95)
```

```
##                   2.5 %   97.5 %
## market_returns 1.264658 1.524774
```

A confidence interval tells the statistician how stable the estimate is for the variable. A stable estimate means that if the experiment were performed again, a similar value would be computed (d) What is the p-value for testing H0 : $\beta_0 = 0$ vs H1 : $\beta_0 \neq 0$? What is your conclusion?

```
summary$coefficients[,4]
```

```
##    (Intercept) market_returns
##   8.396004e-02   2.803936e-29
```

*Because the P-value is big, we cannot reject the null hypothesis. There is not enough evidence at alpha = 0.05 (e) What is the p-value for testing H0 : $\beta_1 = 0$ vs H1 : $\beta_1 \neq 0$? What is your conclusion?

```
model.fit3 <- lm(formula = Lo10~market_returns, data=sub_df)
summary3 <- summary(model.fit3)
summary3
```

```
##
## Call:
## lm(formula = Lo10 ~ market_returns, data = sub_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -6.3075 -1.5564 -0.2021  1.4379  8.1134
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     0.59996    0.34120   1.758    0.084 .
## market_returns  1.39472    0.06497  21.466   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.556 on 58 degrees of freedom
## Multiple R-squared:  0.8882, Adjusted R-squared:  0.8863
## F-statistic: 460.8 on 1 and 58 DF,  p-value: < 2.2e-16
```

Because the P-value is small, we reject the H0. We can conclde there is a relationship between market return and Lo10 (f) Perform a 5% level test for testing H0 : $\beta_1 = 1$ vs H1 : $\beta_1 \neq 1$? What is your conclusion?

```
model.fit4 <- lm(formula = Lo10~1*market_returns, data = sub_df)
summary_4 <- summary(model.fit4)
summary_4
```

```
##
## Call:
## lm(formula = Lo10 ~ 1 * market_returns, data = sub_df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -16.049  -4.019   0.041   4.338  20.771
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.4590     0.9786   2.513   0.0147 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 7.581 on 59 degrees of freedom
```

```
confint(model.fit4)
```

```
##                 2.5 %   97.5 %
## (Intercept) 0.5007473 4.417253
```

(g) Obtain the R2 of this linear regression model.fit. What does it mean? Comment on it.

```
summary_4$r.squared
```

```
## [1] 0
```

The R^2 of the linear regression model.fit is 0 meaning that it does not explain any variance The residuals are too large