Statistics Assignment

Giampietro Ciancio 965991
Riccardo Valenti 979784

Stock description, data collection and notation

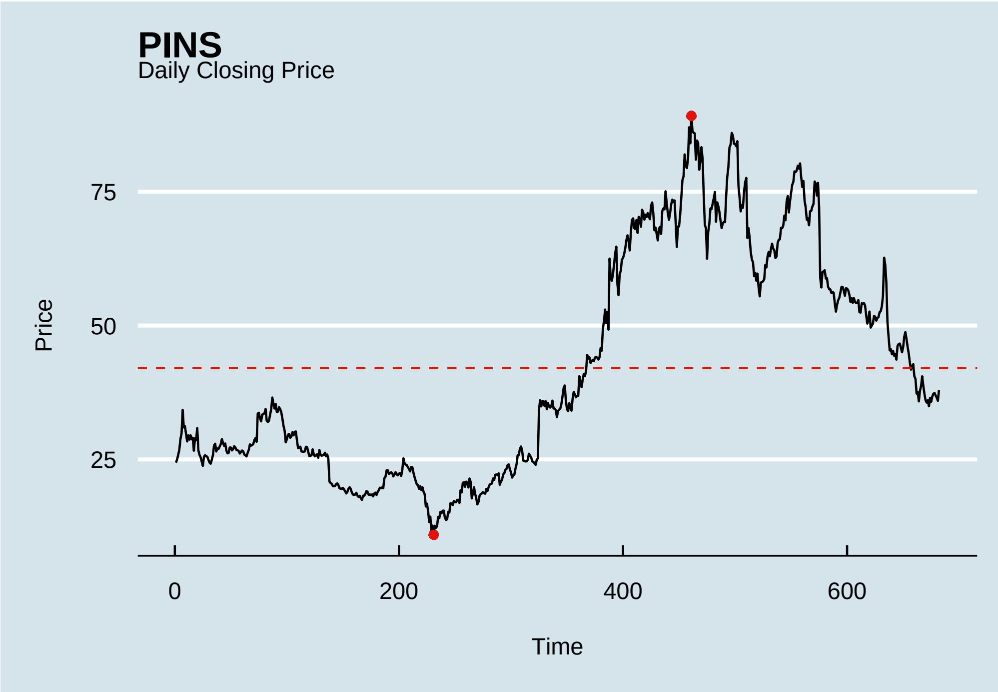
We start collecting data from the `tidyquant` stocks database. We'll be using Pinterest stock closing price from the IPO date (April 18, 2019) until the end of 2021 (December 31, 2021). Pinterest is an image sharing and social media service designed to enable saving and discovery of information on the internet using image. The statistical unit is the day, the variable is the price and the unit of measure of the price is dollar. We define the IPO date as $t = 1$ and the last day of 2021 as T. Hence, the stock closing price at $t = 1$ will be denoted as P_1, the last closing price as P_T and the generic day closing price as P_t. We download the data and pipe it to extrapolate only the `close` column which we assign to vector `price`.

```r
price <- tq_get("PINS", get = "stock.prices", from = "2019-04-18", to = "2021-12-31") %>%
  .$close
head(price)
```

```
## [1] 24.40 24.99 25.85 26.80 28.80 29.85
```

Daily Closing Price

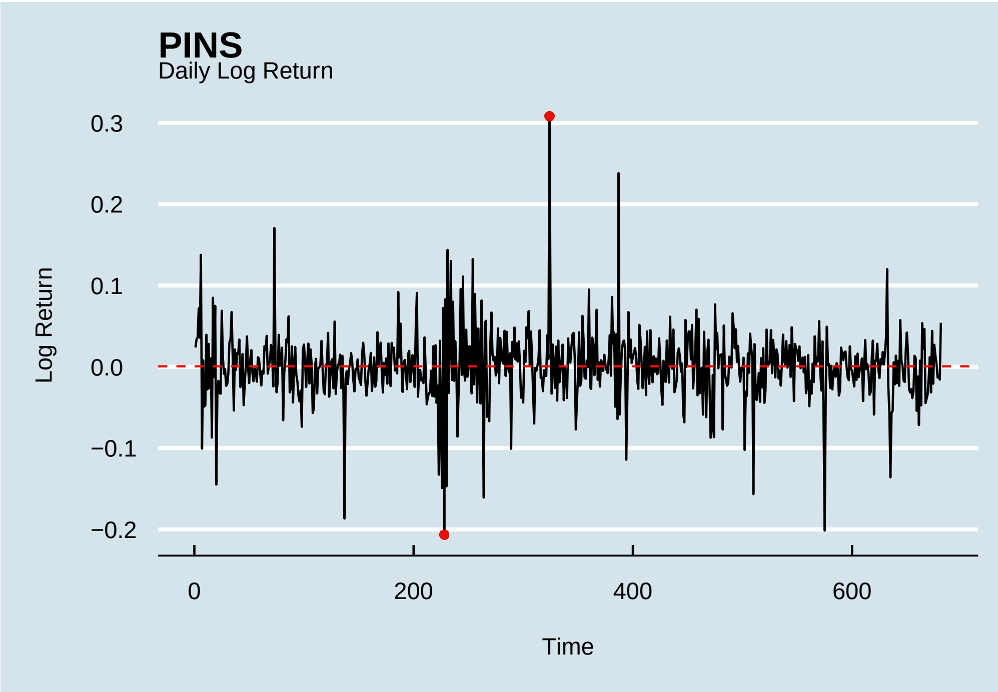
We plot the data as a line having t on the x axis and P_t on y axis.



From the plot we observe the evolution of the stock price within the above said interval. After the IPO, the stock prices moved in the interval between 25$ and 35$, before plunging from the 100th day circa. After day 200, the stock price showed a sign of recovery and then fell again reaching its minimum shortly after. From that point on, the stock prices showed a sharp and constant growth until the 400th day, from which it slowed down for a while and then start growing even faster reaching the maximum price at day 450 circa. The stock then showed a very high volatility, creating both positive and negative spike in an overall decreasing trend. The average closing price of the stock is 40$ circa and we can approximately say that the stock price have been lower than mean in the first half of the period and greater in the second half.

Daily Log Return

We plot the data as a line having t on the x axis and r_t on y axis.



From the plot we can observe the Daily Log Return of the stock. Most of r_t belong to the interval $[-0.1, 0.1]$. However, spikes can reach up to -0.2 and 0.3. We observe that the worst day (in terms of return) is at day 200 circa, around the same time in which the stock price reached its minimum. The stock showed the best return on the 320th day circa, hence contributing in the strong price growth from day 200 to day 400 circa. The Log Return visualization offer a plot similar to the one we can obtain using simple return ($\frac{P_t - P_{t-1}}{P_{t-1}}$). However, Log Return and Simple Return show similar value only if the ratio of prices ($\frac{P_t}{P_{t-1}}$) belongs approximately to the interval $[0.95, 1.10]$ (Appendix A). Given a ratio not belonging in the said interval, Log Return will produce a lower value than simple return if the ratio is positive and greater value (in absolute term) if the ratio is negative. The Log Return visualization do not interfere with the sign of the simple return. Whenever $\frac{P_t}{P_{t-1}} < 1$ both simple and log return will show a negative value; vice versa for $\frac{P_t}{P_{t-1}} > 1$.

Mean

```
m.1
```

```
## [1] 0.0005699206
```

```r
m.2
```

```
## [1] 0.0006144026
```

Both means show small positive values. We can use *Time Additivity* feature and the mean of Log Return to gain interesting insight on the overall return of the stock. *Time Additivity* tells us that it is possible to compute the overall Log Returns from t to $t+k$ just by adding the daily Log Return from t to $t+k$. The `m.1` gives us $\frac{\sum_{t=1}^{n_r-1} r_t}{n_r-1}$. Thus, we can compute the Log Return from $t=1$ to $T-1$ as follow:

```r
m.1*NROW(log.rt.1)
```

```
## [1] 0.387546
```

Given what we said previously, this Log Return is not a good approximation of the Simple Return. Using the formula $R = e^r - 1$ we can recover the Simple Return of Pinterest Stock from the IPO date until the end of 2021.

```r
exp(m.1*NROW(log.rt.1))-1
```

```
## [1] 0.4733607
```

The same reasoning can be applied to `m.2` in order to recover the Log Return and Simple Return from $t=2$ to T.

Standard deviation

```r
s.1
```

```
## [1] 0.04283665
```

```r
s.2
```

```
## [1] 0.04287659
```

Both Standard Deviations present small similar values. This means that the most of the Log Return are concentrated around the mean. However, we need to take in consideration the nature of the data in order to draw useful insight from standard deviation. We showed previously that the most of Log Return belong to the interval $[-0.1, 0.1]$ and both `m.1` and `m.2` are very small values. Given this information, we can say that Standard Deviation (0.05 circa) is quite high relatively to the Mean. We may refer to the Standard Deviation of daily return as Historical Daily Volatility. We can conclude that Pinterest stock presents a relatively high daily volatility.

Correlation coefficient

```r
m.product
```

```
## [1] -0.0001132173
```

```r
cov.log.rt
```

```
## [1] -0.0001135674
```

```r
cor.log.rt
```

```
## [1] -0.06183267
```

The correlation coefficient shows a very weak negative linear dependence. A negative linear dependence means that given a log return r_k at day k, then r_{k+1} will tend to move the stock price in the opposite direction. However, the very low correlation coefficient (in absolute term) tells us that the daily log-return of day $k+1$ cannot be expressed as linear function of r_t.Hence the linear dependence is negligible and we cannot draw any conclusion about the relationship of two consecutive days stock Log Return.

Absolute value

```r
m.1.abs
```

```
## [1] 0.02883113
```

```r
m.2.abs
```

```
## [1] 0.02887561
```

```r
s.1.abs
```

```
## [1] 0.03166774
```

```r
s.2.abs
```

```
## [1] 0.03168204
```

The means of the absolute value of Log Return now present greater values. Trivially, this happens because there are no longer negative values in the computation. Consequently, we observe lower Standard Deviations. We can observe that taking the absolute value do not make any changes in the first moment ($\frac{\sum r_t^2}{n_r-1} = \frac{\sum |r_t|^2}{n_r-1}$) while the second moment gets greater as seen before, thus leading to a lower variance. Taking the square root, being it a monotonous function, do not interfere with the observation.

```r
m.product.abs
```

```
## [1] 0.0009896471
```

```
cov.log.rt.abs
```

```
## [1] 0.0001571308
```

```
cor.log.rt.abs
```

```
## [1] 0.1566142
```

The correlation coefficient is now positive and greater. Computing the correlation between the absolute value of two consecutive days Log Return allows us to focus only on the magnitude of the price variation, without taking into consideration the direction. A positive correlation coefficient may suggest a positive linear dependence between the magnitude of price changes in days t and $t+1$. However, even if the correlation coefficient in this case in greater than the one seen before (in absolute terms) the correlation coefficient is too low to draw any relevant conclusion about the relationship of two consecutive days stock absolute magnitude Log Return. We know that Correlation Coefficient is not sensible to linear transformation. The absolute value, being a non-linear transformation, leads to different Correlation Coefficient.

Source

Giampietro Ciancio, Riccardo Valenti. 2022. "Stats Assignment." 2022. https://github.com/giampietrociancio/StatsAssignment.

Appendix A

Let's denote $x = \frac{P_t}{P_{t-1}}$. We can observe from the plot that $\log(x)$ and $x-1$ have similar values $\forall x \in [0.95, 1.10]$.