# Empirical Finance: Assignment 1


**By Group 15**

| | |
|---|---|
| Lanlan Hou | 2801069 |
| Rick van der Ploeg | 2782774 |
| Sebastiaan van Dijen | 2769505 |
| Wa Chak Sou | 2796840 |

School of Business and Economics


**3rd November 2025**

# QUESTION 1: Stock prices and returns

## a. Prices and returns plots of HSBC stocks

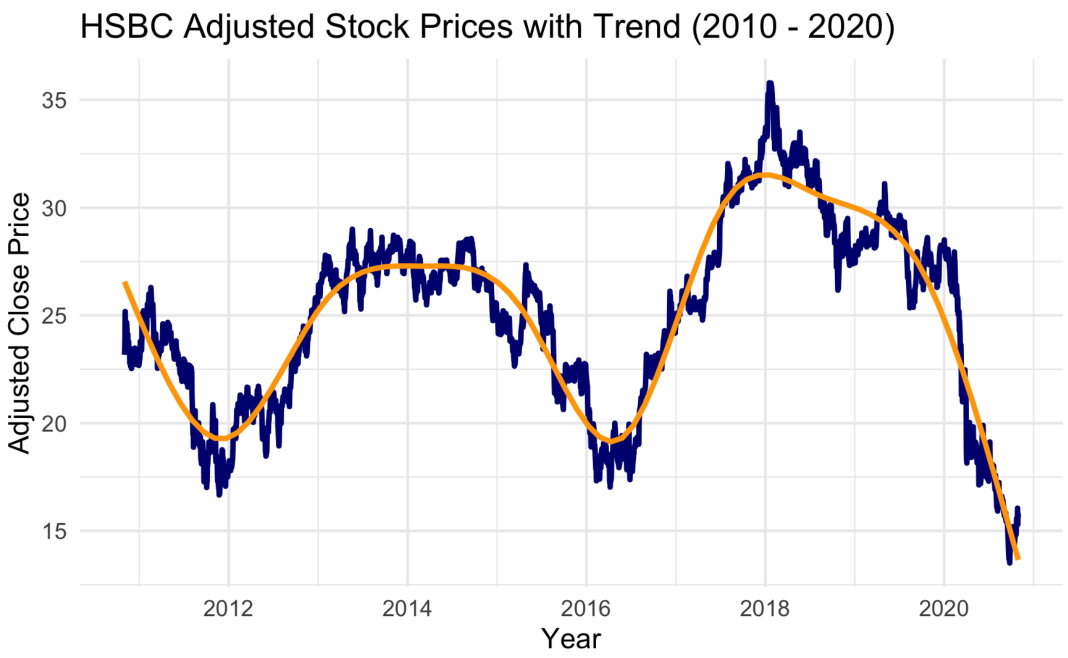


Figure 1.1 Adjusted Close Price of HSBC (November 2010 to November 2020)

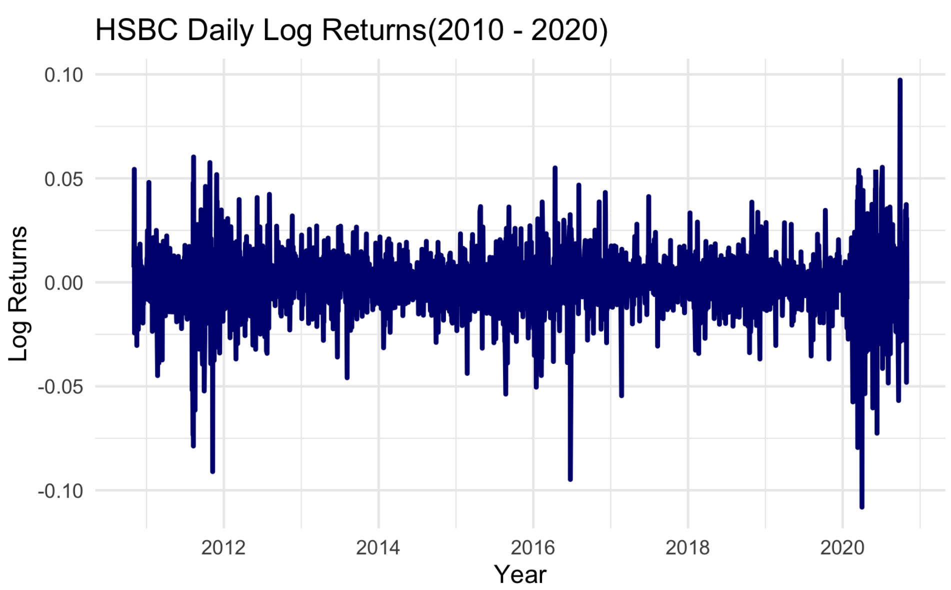


Figure 1.2 Log Returns of HSBC (November 2010 to November 2020)

Figure 1.1 demonstrate the daily stock price of HSBC which fluctuates over the decade. It show a cyclical patterns rather than a clear long-term upward or downward trend. It also reflects the market condition like COVID-19 shock at the end of the decade. Figure 1.2 reveals the daily log returns of HSBC stock which fluctuate around 0.1. The high volatility near 2012, 2016 and 2020 while other periods is low volatility, supports the exhibition of volatility clustering with high and low volatilities interchanging.

### b. Relevant summary statistics

| Statistic | Value |
|---|---|
| Mean | 24.994385 |
| Std. Deviation | 4.356907 |
| Minimum | 13.508404 |
| P10 | 18.641351 |
| Median | 25.693600 |
| P90 | 30.548970 |
| Maximum | 35.793488 |
| Skewness | -0.200543 |
| Kurtosis | -0.553233 |

Table 1.2.1: Summary Statistics for Daily HSBC Adjusted Close Price

| Statistic | Value |
|---|---|
| Mean | -0.000156 |
| Std. Deviation | 0.014615 |
| Minimum | -0.108101 |
| P10 | -0.015906 |
| Median | 0.000000 |
| P90 | 0.015354 |
| Maximum | 0.097266 |
| Skewness | -0.551902 |
| Kurtosis | 8.797429 |

Table 1.2.2: Summary Statistics for Daily HSBC Log Returns

From Table 1.2.1 and 1.2.2 we observe that the HSBC's adjusted prices fluctuated between roughly \$13.50 and \$35.79 during 2010-2020, with an average around \$25, suggesting moderate long-term volatility(4.36). The daily log returns show a slightly negative mean and moderate volatility. The negative skewness(-0.55) and high kurtosis(8.8) indicate that the return distribution is left-skewness and leptokurtic - implying more large drops than large gains, and extreme movements are more common than in a normal distribution.

# Question 2: Return distributions

## a. Distribution of the returns

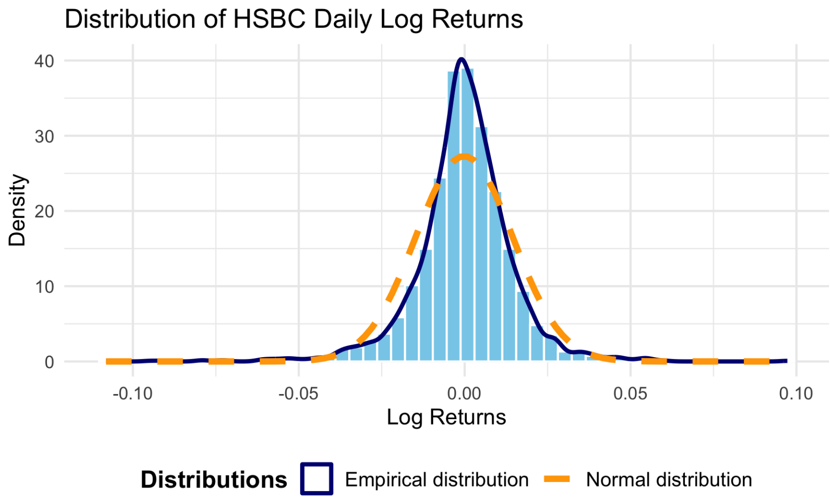


Figure 2.1 Distribution of log returns)

Figure 2.1 shows the distribution of the log returns for HSBC. The histogram of log returns (in *skyblue bar*) is overlaid with the corresponding normal distribution curve (in *orange dashed line*) for comparison. It is evident that the distribution is leptokurtic (more peaked) and exhibits minor skewness, indicating only small deviations from a perfectly normal shape. Overall, the returns appear to be approximately normally distributed, as the normal curve fits the histogram, with no strong evidence of fat tails or extreme outliers.

## b. Normality test of returns

To formally test for normality, the **Jarque-Bera (JB)** test was conducted. The hypotheses for the test are formulated as follows:

$$H_0 : \text{The returns are normally distributed,}$$
$$H_1 : \text{The returns are not normally distributed.}$$

The Jarque-Bera (JB) test statistic, based on the least squares residuals, is given by

$$JB = \frac{T-k}{6}(\hat{S}^2 + \frac{1}{4}(\hat{K}-3)^2),$$

where

- $k$: the number of covariates (regressors),
- $T$: the sample size,
- $\hat{S}$: the sample skewness,
- $\hat{K}$: the sample kurtosis.

Under the null hypothesis of normality, the test statistic is asymptotically chi-squared with two degrees of freedom:

$$JB \xrightarrow{d} \chi^2_2 \quad \text{as} \quad T \to \infty.$$

| JB Test Statistic | Critical Value (5% level) | $p$-Value |
|---|---|---|
| 3202.7 | 5.99 | 0.00 |

Table 2.1: Jarque-Bera (JB) normality test for HSBC Log Returns

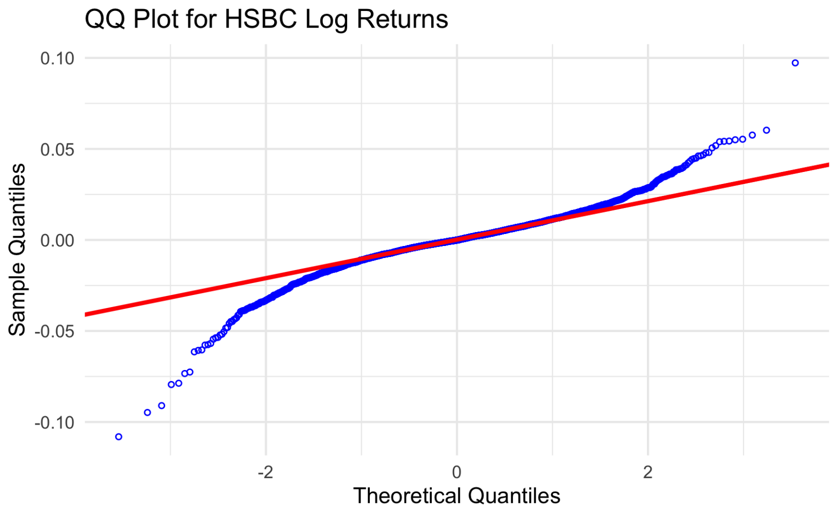


Figure 2.2 QQ Plot for HSBC Log Returns)

Table 2.1 reports the results of the JB normality test. The computed JB statistic (3202.7) is far greater than the 5% critical value (5.99), and the associated $p$-value is 0.00. Hence, we **reject the null hypothesis** of normality at the 5% significance level. The QQ plot shown in Figure 2.2 further supports this conclusion, showing clear deviations from the theoretical normal line, particularly in the tails.

**Conclusion:** The results of both the Jarque-Bera test and QQ plot clearly indicate that the HSBC log returns are not normally distributed. The distribution exhibits excess kurtosis and fat tails, confirming the presence of non-normal behavior commonly observed in empirical financial data.

# Appendix

```
library(openxlsx)
raw <- read.xlsx('student_groups_stocks.xlsx', sheet = 1)

groupNumber <- 15
nameOfStock <- raw$Stock.Name[groupNumber]
startDate <- raw$Start.Date[groupNumber]
endDate <- raw$`End.Date.(+10y)`[groupNumber]

# Extract the daily stock price from source
library(quantmod)
getSymbols(nameOfStock, src = 'yahoo', from = startDate, to = endDate)
date <- index(HSBC)

# Adjusted Close
adjustedPrice <- as.numeric(HSBC[,'HSBC.Adjusted'])

# Log Returns
compoundedReturns <- numeric(length(adjustedPrice) - 1)
for(i in (2:length(adjustedPrice))){
  compoundedReturns[i-1] <- log(adjustedPrice[i] / adjustedPrice[i-1])
}

library(ggplot2)
prices_df <- data.frame(date = date, adjustedPrice = adjustedPrice)
returns_df <- data.frame(date = date[2:length(date)], compoundedReturns
    = compoundedReturns)

# Plot of daily stock price
ggplot(prices_df, aes(x = date, y = adjustedPrice)) +
  geom_line(color = 'navy', linewidth = 1) +
  geom_smooth(color = 'orange', se = FALSE) +
  labs(
    title = paste(nameOfStock, 'Adjusted Stock Prices with Trend (2010
      - 2020)'),
    x = 'Year',
    y = 'Adjusted Close Price') +
  theme_minimal()
ggsave('figures/prices_plot.png', dpi = 300)

# Plot of daily log returns
ggplot(returns_df, aes(x = date, y = compoundedReturns)) +
  geom_line(color = 'navy', na.rm = TRUE, linewidth = 1) +
  labs(
    title = paste(nameOfStock, 'Daily Log Returns(2010 - 2020)'),
    x = 'Year',
    y = 'Log Returns') +
  theme_minimal()
ggsave('figures/returns_plot.png', dpi = 300)
```

Figure 1: Question 1(a)

```
# Compute some relevant summary statistics of Adjusted Close
library(moments)
price_mean <- mean(adjustedPrice)
price_stdev <- sd(adjustedPrice)
price_minimum <- min(adjustedPrice)
price_maximum <- max(adjustedPrice)
```

```
price_median <- median(adjustedPrice)
price_10_percentile <- quantile(adjustedPrice, 0.1)
price_90_percentile <- quantile(adjustedPrice, 0.9)
price_skewness <- skewness(adjustedPrice)
price_kurtosis <- kurtosis(adjustedPrice)

price_summaryStatistics <- data.frame(
  Statistics = c('Mean', 'Std. Deviation', 'Minimum', 'P10', 'Median',
     'P90', 'Maximum','Skewness', 'Kurtosis'),
  Values = round(c(price_mean, price_stdev, price_minimum, price_10_
     percentile, price_median, price_90_percentile, price_maximum,
     price_skewness, price_kurtosis), 6))

price_summaryStatistics

# Compute some relevant summary statistics of Log Returns
logReturns_mean <- mean(compoundedReturns)
logReturns_stdev <- sd(compoundedReturns)
logReturns_minimum <- min(compoundedReturns)
logReturns_maximum <- max(compoundedReturns)
logReturns_median <- median(compoundedReturns)
logReturns_10_percentile <- quantile(compoundedReturns, 0.1)
logReturns_90_percentile <- quantile(compoundedReturns, 0.9)
logReturns_skewness <- skewness(compoundedReturns)
logReturns_kurtosis <- kurtosis(compoundedReturns)

logReturns_summaryStatistics <- data.frame(
  Statistics = c('Mean', 'Std. Deviation', 'Minimum', 'P10', 'Median',
     'P90', 'Maximum', 'Skewness', 'Kurtosis'),
  Values = round(c(logReturns_mean, logReturns_stdev, logReturns_
     minimum, logReturns_10_percentile, logReturns_median, logReturns_
     90_percentile, logReturns_maximum, logReturns_skewness, logReturns
     _kurtosis), 6))

logReturns_summaryStatistics
```

Figure 2: Question 1(b)

```
single_logReturns_df <- data.frame(compoundedReturns =
   compoundedReturns)

# Distribution Plot of daily log returns with equivalent normal
   distribution curve
ggplot(single_logReturns_df, aes(x = compoundedReturns)) +
  geom_histogram(aes(y = ..density..),
                 bins = 50, fill = 'skyblue', color = 'white') +
  geom_density(aes(color = 'Empirical distribution'), size = 1) +
  stat_function(fun = dnorm,
                args = list(mean = mean(single_logReturns_df$
                   compoundedReturns, na.rm = TRUE),
                            sd = sd(single_logReturns_df$
                               compoundedReturns, na.rm = TRUE)),
                aes(color = 'Normal distribution'),
                linewidth = 1.5, linetype = 'dashed') +
  scale_color_manual(
    name = 'Distributions',
    values = c('Empirical distribution' = 'navy',
               'Normal distribution'    = 'orange')) +
  labs(title = 'Distribution of HSBC Daily Log Returns',
```

```
        x = 'Log Returns', y = 'Density') +
  theme_minimal() +
  theme(
    legend.position = 'bottom',
    legend.title = element_text(size = 12, face = 'bold'),
    legend.text = element_text(size = 10))
ggsave('figures/returns_distribution.png', dpi = 300)
```

Figure 3: Question 2(a)

```
# Perform Jarque-Bera normality test
library(tseries)
jb_test <- jarque.bera.test(compoundedReturns)
jb_test

# Show the results of the JB test
jbTest_summaryStatistics <- data.frame(
  JB_Test_Statistic = round(jb_test$statistic, 2),
  Critical_Value = round(qchisq(0.95, 2), 2),
  p_Value = round(jb_test$p.value, 2)
)

jbTest_summaryStatistics

# QQ Plot of daily log returns
ggplot(single_logReturns_df, aes(sample = compoundedReturns)) +
  stat_qq(shape = 1, color = 'blue', size = 1) +
  stat_qq_line(color = 'red', linewidth = 1) +
  labs(title = 'QQ Plot for HSBC Log Returns',
       x = 'Theoretical Quantiles', y = 'Sample Quantiles') +
  theme_minimal()
ggsave('figures/returns_qqplot.png', dpi = 300)
```

Figure 4: Question 2(b)