1 Are stock returns normally distributed?

My Professor, Naveen Srinivasan bought up the question that why do we experience a stock market crash frequently, when the underlying assumption we make about the stock returns to follow normal distribution, events like the Great Recession (1930s), financial Crisis (2007-2008) should occur once in a million years as the probability of experiencing an event that gives a negative 60% return has a very low probability. It is seen as a negative 20-30 sigma (σ) event. Professor Naveen Srinivasan passed the problem to Professor Srikanth Pai, who questioned whether the returns are actually normally distributed and assigned me the task of finding it and suggested the Shapiro–Wilk (SW) test.

Benoît Mandelbrot was the first to challenge the assumption of normality, observing that stock return data exhibit fat tails. This contrasted with the earlier views of Bachelier, Osborne, Kendall, and Moore, where the former two claimed that the data are normally distributed using the central limit theorem, and the latter both, after empirical testing, claimed that the data are approximately normal. Mandelbrot instead proposed that financial returns follow a stable Paretian distribution, which better accounts for the heavy tails observed in real-world data. This hypothesis was later empirically supported by Fama, who confirmed that stock returns are better modeled by stable Paretian distributions.

Hence, I intend to empirically verify whether the S&P 500 stock returns are normally distributed for the period 1928-2025. In-short, answer the following:

1. Is 30, 60, 90, and 120 days rolling daily, weekly, monthly, and quarterly returns distributed normally?
2. Is 30, 60, 90, and 120 days window-window daily, weekly, monthly, and quarterly returns distributed normally?

I will be using MATLAB and python to carry out computations and evaluate whether different data segments follow a normal distribution. The choice of 30-, 60-, 90-, and 120-day windows is motivated by regulatory practices. According to the Basel Committee on Banking Supervision (BCBS) under the Fundamental Review of the Trading Book (FRTB), banks are required to calculate risk measures such as Value at Risk (VaR) and Expected Shortfall (ES) using a minimum of 250 trading days of data, but also need to monitor short-term windows to capture recent market stress and volatility. Common practice often involves the use of 30- to 60-day rolling windows for short-term VaR monitoring, and 60- to 90-day windows are sometimes used for sensitivity analysis and stress testing.

2 Background Study

2.1 Shapiro–Wilk test

Let $\{x_1, x_2, \ldots, x_n\}$ be a random sample data, that is independent and identically distributed, the Shapiro–Wilk test is used to test the null hypothesis that the data is norammlly distributed,

$$H_0 : x_1, \ldots, x_n \sim \mathcal{N}(\mu, \sigma^2)$$

against the alternative that the data do not come from a normal distribution.

The test statistic is defined as

$$W = \frac{\left(\sum_{i=1}^{n} a_i x_{(i)}\right)^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2},$$

where:

- $x_{(i)}$ denotes the i-th order statistic (i.e., the i-th smallest value),
- $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean,
- $\mathbf{a} = (a_1, \ldots, a_n)$ is a vector of constants computed as

$$\mathbf{a} = \frac{\mathbf{m}^\top \mathbf{V}^{-1}}{(\mathbf{m}^\top \mathbf{V}^{-1} \mathbf{V}^{-1} \mathbf{m})^{1/2}},$$

- $\mathbf{m} = \mathbb{E}[x_{(i)}]$ is the vector of expected values of order statistics of a standard normal sample,
- $\mathbf{V}$ is the covariance matrix of those order statistics.

We need not bury ourselves on the mathematical derivation of the test statistic or even knowing how to compute the statistic by hand as we have reliable python libraries, Scipy to carry out the computation. However, we have to pay attention to the assumption that the test makes upon the data set we feed to it; ***independently and identically distributed***.

We call two random variables X and Y are independent if and only if

$$F_{X,Y}(x, y) = F_X(x) \cdot F_Y(y) \quad \text{for all } x, y \in I.$$

Two random variables X and Y are identically distributed if and only if

$$F_X(x) = F_Y(x) \quad \text{for all } x \in I.$$

Two random variables X and Y are i.i.d. (independent and identically distributed) if and only if

$$\begin{aligned} F_X(x) &= F_Y(x) & \forall x \in I, \\ F_{X,Y}(x, y) &= F_X(x) \cdot F_Y(y) & \forall x, y \in I. \end{aligned}$$

Now, the question is are returns independent and identically distributed?

There is no test to check whether the data follows i.i.d. But, we can partially test for independence among random variables; test for auto-correlation, Ljung–Box test.

2.2 Ljung-Box test

Ljung-Box test is used to check if a dataset has auto-correlation, that is if the past has a say on the future or is return at time t is dependent on return at time $t-1$?

Usually, applied on residuals of a fitted auto-regressive integrated moving average model (ARIMA) to verify if the ARIMA model fits the data well or not. The Ljung-Box test is defined as the following:

Null and Alternative Hypotheses:

H_0 : The data are not serially correlated (i.e., $\rho_k = 0$ for all $k = 1, \dots, h$).

H_a : The data exhibit serial correlation at one or more lags.

Test Statistic:

$$Q = n(n+2) \sum_{k=1}^{h} \frac{\hat{\rho}_k^2}{n-k}$$

where:

- n is the sample size,
- $\hat{\rho}_k$ is the sample autocorrelation at lag k,
- h is the number of lags being tested.

Asymptotic Distribution: Under the null hypothesis H_0, the Ljung–Box statistic Q is approximately distributed as a chi-squared distribution with h degrees of freedom:

$$Q \sim \chi_h^2$$

The null hypothesis is rejected at significance level α if

$$Q > \chi^2_{1-\alpha,h}$$

where $\chi^2_{1-\alpha,h}$ denotes the $(1-\alpha)$-quantile of the chi-squared distribution with h degrees of freedom.

We will be using $ln(n)$ as number of lags (h), where is n is number of data points. Furthermore, it is important to note that, the test doesn't check for independence, it checks for auto-correlation. Hence, if the data exhibits no auto-correlation doesn't translate independence, it just means the data is not linearly dependent. We will use this test to check for dependencies among the dataset and remove it and work with SW test.

3 Methodology

1. Download the prices data of S&P 500 for 01/01/1928 to 15/03/2025 from Yahoo finance.

2. Prices are transformed into daily returns using the following formula (log differential of prices):

$$R_t = \ln(p_t) - \ln(p_{t-1})$$

3. Weekly returns are calculated using the same formula, where $(t-1) - t = 5$. For monthly, it is 30 and so on.

4. Rolling weekly, monthly, and quarterly returns are calculated by shifting the starting point and ending point by one step forward while maintaining 5, 30, 90 days gap.

5. Window weekly, monthly, and quarterly returns are calculated by shifting the starting point and ending point by 5, 30, and 90 steps.

6. These returns are separated into 30, 60, 90, and 120 data segments and Ljung-Box test is applied over each segment.

7. The segments where the data is dependent is removed and Shapiro-Wilk test is applied.

8. Percentage of times the SW test failed is recorded under each segment to make the final conclusions.

4 Results and Discussion

The Ljung-Box test aided in removing out the datasets that are not independent. As expected the rolling returns are not independent thus, we will work with window returns and carry forward with the Shapiro-Wilk test. Since, Matlab does not have a built in command to carry out the Shapiro-Wilk test, I shifted to python.

Keeping our focus to window returns as rolling returns are not independent, we see, the returns are not normally distributed which is consistent with the findings of Mandelbrot and Fama. More importantly, we find that as the data segment size increases from 30 to 60, 60 to 90, and 90 to 120, it is more and more, not normally distributed. (See the graphs in the next page)

In the future, I would like to investigate the difference between developed and developing countries in this setup. I conjecture that developed countries generally exhibit more stable financial markets, characterized by lower volatility and returns that more closely resemble the normal distribution. In contrast, developing countries often experience higher market volatility, structural inefficiencies, and external shocks, which can lead to return distributions with heavier (fat) tails.

Results: Window Approach for S&P 500 dataset

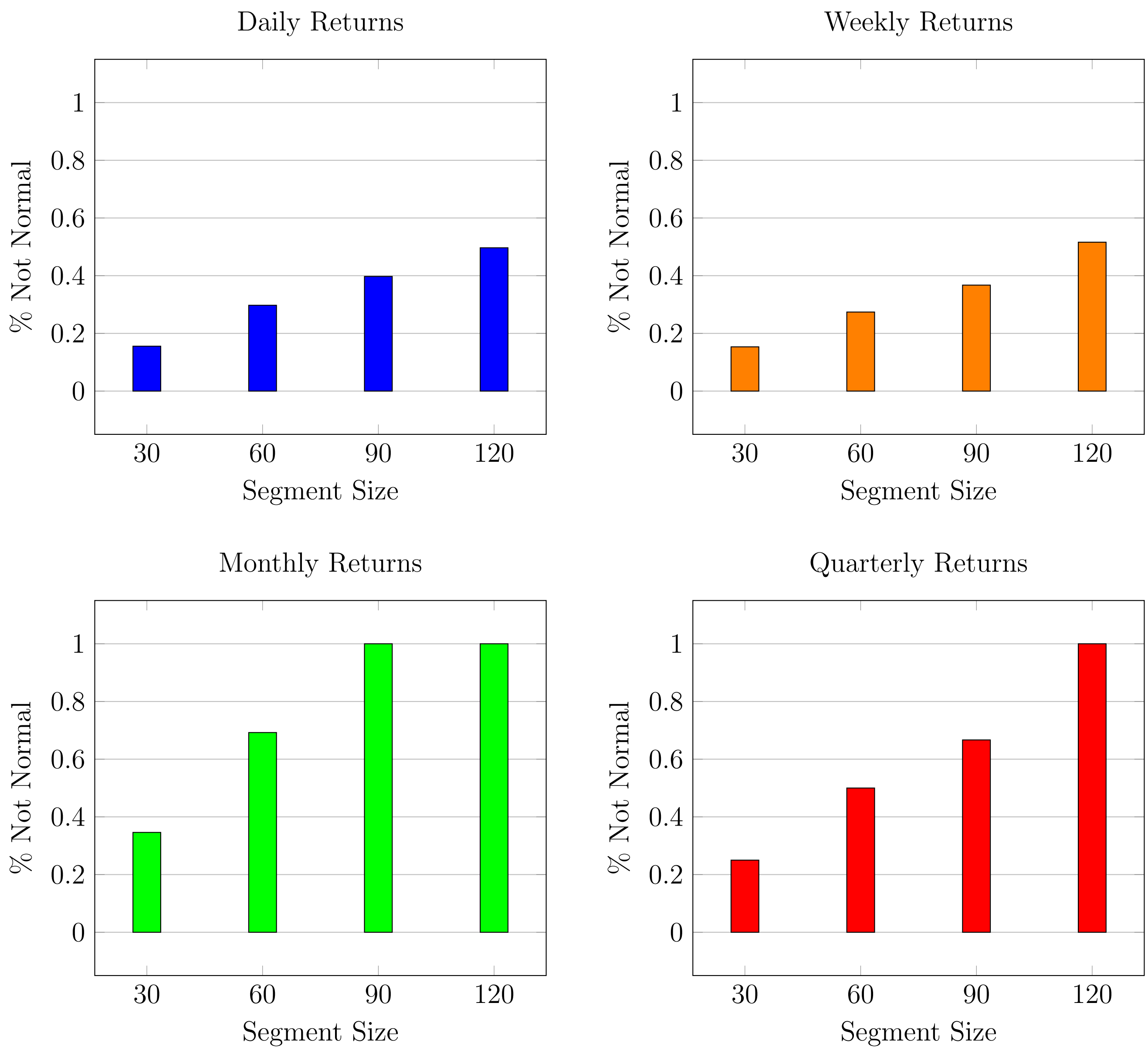

Figure 1: Window Approach - % Not Normally Distributed by Segment Size