Bitcoin (BTC) Time Series Analysis

Vedanshee Patel(202203003)[*], Khushi Prajapati(202201062)[†]
Dhirubhai Ambani Institute of Information & Communication Technology,
Gandhinagar, Gujarat 382007, India
SC475, Time Series and Analysis

***Abstract*—This report presents a time series analysis of Bitcoin (BTC) prices from September 2014 to March 2025 using daily data from Yahoo Finance. Key metrics such as opening, closing, high, and low prices, along with trading volume, are examined to study trends, volatility, and structural changes. Statistical techniques like differencing, log transformations, and decomposition are applied to achieve stationarity and extract seasonality. Autoregressive models and unit circle analysis assess the stability and complexity of Bitcoin's price dynamics. The findings reveal high volatility, strong autocorrelation, and sensitivity to major economic events such as the COVID-19 pandemic and Bitcoin halving.**

I. Introduction

Bitcoin (BTC), the decentralized digital currency introduced in 2009,has been a significant part of the global financial scene. Its price is notoriously volatile because of a combination of macroeconomic factors, technological advancements, and market speculation. The goal of this report is to conduct a thorough time series analysis of the market behaviour of Bitcoin between 2014 and 2025. This report contains the use of statistical and econometric methods to analyse daily price and trading volume data in order to find underlying patterns, identify non-stationarity, and assess model stability.

II. Data Description

The dataset was collected from Yahoo Finance using the `yfinance` Python package. The dataset spans from September 17, 2014, with 3847 daily records. It contains the historical prices of Bitcoin (ticker: BTC-USD), including the following columns:

TABLE I
Description of columns in the dataset(All prices are in USD)

Column	Description
`Date`	Contains the date of the prices for that day in the format: YYYY-MM-DD 00:00:00+00:00.
`Open`	Opening price for that date.
`High`	Highest price during the trading day.
`Low`	Lowest price during the trading day.
`Close`	Final price at the end of the trading day.
`Volume`	Total number of shares or contracts traded.
`Dividends`	Dividends issued on that day, if any (always zero for BTC).
`Stock Splits`	Records of any stock splits (not applicable to BTC).

[*]202203003
[†]202201062

III. Data Interpretation

A. Dealing with Null Values

- The original dataset did not have missing values in important columns such as `Open`, `High`, `Low`, `Close`, `Volume`, `Dividends`, and `Stock Splits`
- Bitcoin does not have `Dividends` and `Stock Splits` because it's a decentralized digital currency, not a company with shareholders. So we can completely remove the columns, as their values were observed to be 0, a flat-lined, straight-line horizontal curve.

B. Data Visualization

- Time series graphs were created for numeric features: `Open`, `High`, `Low`, `Close`, `Volume`
- Line plots were created for:
  - `Open` vs `Year`

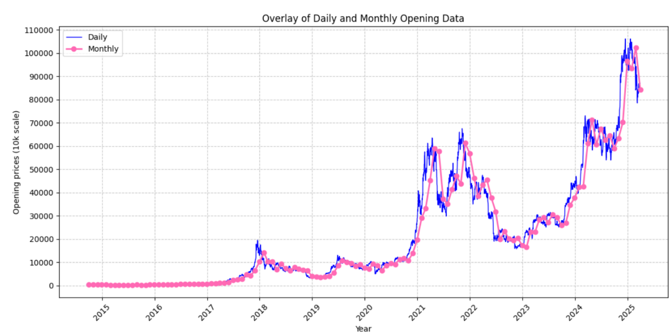

Fig. 1. Plot of the Opening Prices from 2014 to 2025 with Daily and Monthly data overlaying on the same plot from the dataset

  - `Close` vs `Year`

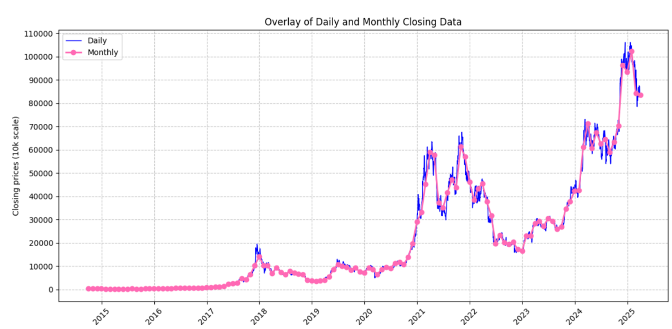

Fig. 2. Plot of the Closing Prices from 2014 to 2025 with Daily and Monthly data overlaying on the same plot from the dataset

In 2020, the economy shut down due to the COVID-19 pandemic. Bitcoin's growth was accelerated by the pandemic shutdown and the government actions that followed, which fuelled investors' concerns about the state of the world economy. Beacuse of this we can see a sudden rise in the price between the year 2020 and 2021.

  - `Volume vs Year`

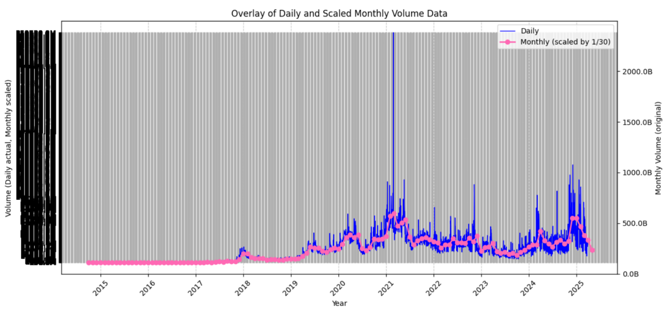

Fig. 3. Plot of the Volume(Amount of the trades carried out) from 2014 to 2025 with Daily and Monthly data overlaying on the same plot from the dataset. Here, the volume of the monthly data has been scaled by 30.30 to have better visualisation of the trend.)

Bitcoins are generated at a specific pace by mining hardware and software. Every four years, this pace is cut in half, which slows down the production of new coins. On April 19, 2024, the most recent halving took place. The term "Bitcoin halving" describes an occurrence that occurs roughly every four years and lowers the block reward by 50 percent. Because fewer bitcoins are entering the market, there is more scarcity, which might lead to an increase in price if market circumstances stay the same. This is the reason why we can see a large spike in the trading volume and Opening and Closing values.

- **Derived Features:** Two features were added:
  - `Difference_high_low = High - Low`
  - `Difference_open_close = Open - Close`

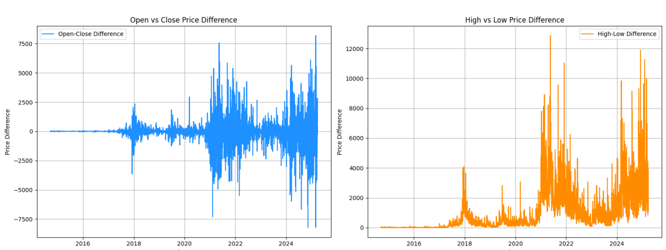

Fig. 4. This graph compares the first differences of open vs. close prices and high vs. low prices of Bitcoin from 2014 to 2024, showing the magnitude of price changes.

From the graphs, we can see that around the year 2021, there were large fluctuations in the opening and closing prices. This can be confirmed

- **Insights from Top Variations:** Top 5 days with largest `Open - Close` and `High - Low` differences were identified

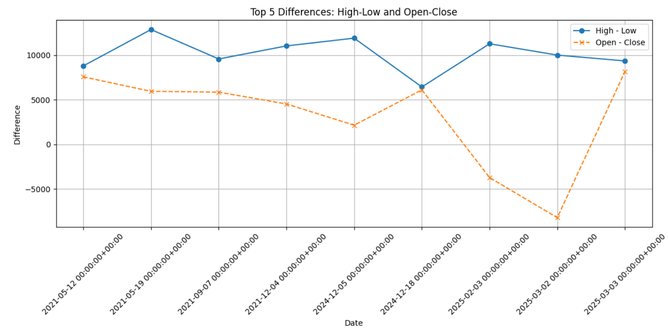

Fig. 5. This graph displays the top 5 differences between high-low and open-close prices of Bitcoin from 2021 to 2025, highlighting the variations over time.

C. Correlation Matrix

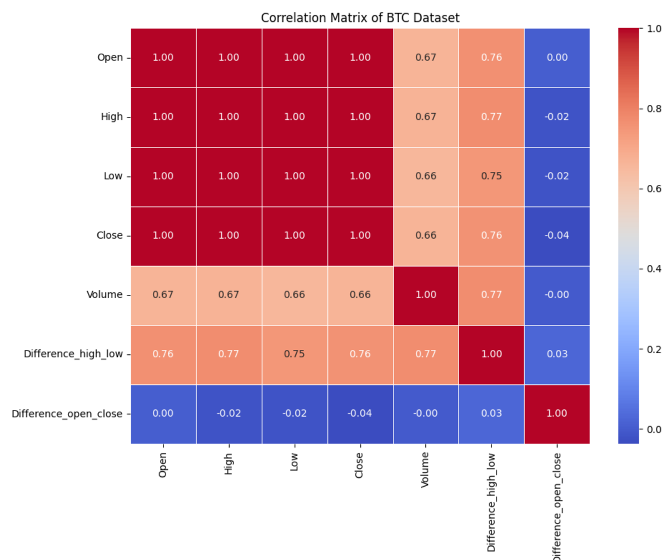

Fig. 6. Pearson correlation matrix of Bitcoin price metrics (High, Low, Close), Volume, and derived difference variables.

- Strong positive correlations were observed between `Open`, `High`, `Low`, and `Close`
- `Volume` showed weaker correlation with price metrics but the difference between the High and Low-value shows had a relatively stronger correlation to the price metrics.

IV. Data Analysis

A. Check Whether the Data Is Stationary

- **Method**: Rolling statistics (365-day window) + Augmented Dickey-Fuller (ADF) test
- **Rolling Mean & Variance**: Here, we have included only the closing price's graph because Opening and Closing show similar trends and have a very small difference

between them. So, from now on, we will continue with our analysis with only closing values.

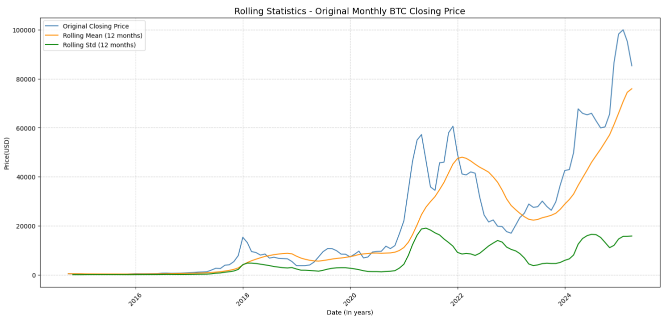

Fig. 7. Rolling mean and standard deviation (12-month windows) for Bitcoin's monthly closing price across a multi-year dataset.

- Non-stationarity can be observed through increasing trends and variable volatility
- Both Open and Close prices showed non-constant mean/variance

- **ADF Test Results (Close column)**:

ADF Statistic	-0.474
p-value	0.8968
Critical Values (1%)	-3.432

- **Conclusion**: Series is non-stationary (p-value > 0.05)

B. Making the Data Stationary

Transformations Applied:

- First-Order Differencing: $\text{Close\_diff} = \text{Close}(t) - \text{Close}(t-1)$
  - ADF p-value: 1.78×10^{-15}

  Similar order of ADF p-values can be seen for the other transformations also, So such low p-values can be approximated to 0.

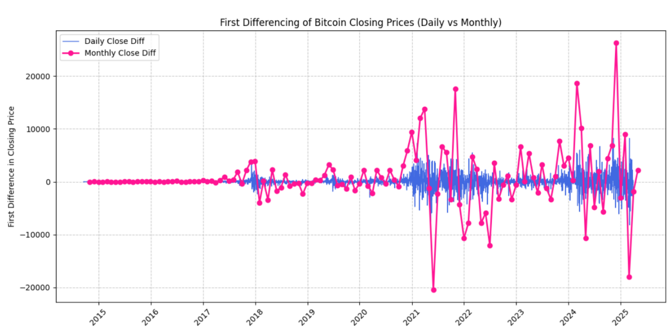

Fig. 8. First differencing of Bitcoin's daily and monthly closing prices over time.

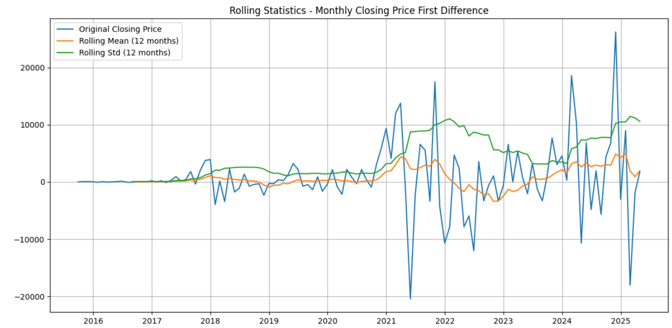

Fig. 9. Rolling mean and standard deviation (12-month windows) for the first differencing of Bitcoin's monthly closing prices.

- Log Transformation: $\text{Log\_Price} = \log(\text{Close})$

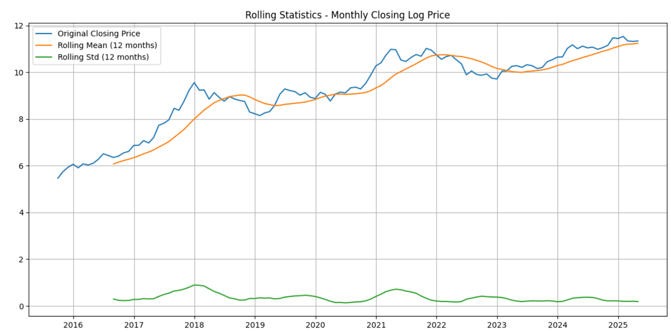

Fig. 10. This graph illustrates the rolling statistics (original closing price, rolling mean, and rolling standard deviation over 12 months) of monthly closing log prices from 2014 to 2025.

- Log Returns:
  $\text{Log\_Return} = \log\left(\frac{\text{Close}(t)}{\text{Close}(t-1)}\right)$

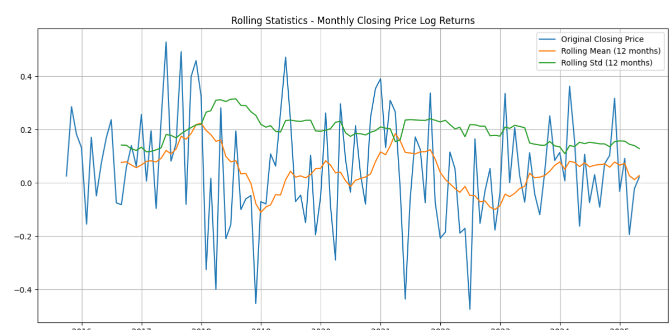

Fig. 11. This graph presents the rolling statistics (original closing price, rolling mean, and rolling standard deviation over 12 months) of monthly closing price log returns from 2014 to 2025.

- Percentage Returns:
  $\text{Pct\_Return} = \frac{\text{Close}(t) - \text{Close}(t-1)}{\text{Close}(t-1)}$ It has the same graph as the log return with similar trends

Stationarity Results:

- All transformed series showed stationarity (ADF p-values = 0.0)
- 30-day rolling statistics confirmed constant mean/variance

C. Seasonal Decomposition

- **Method**: Additive decomposition with period=30 days
- **Components**:
  - Observed - The original time series data as recorded.
  - Trend - The long-term direction or pattern in the data.
  - Seasonal - The recurring, predictable fluctuations within a fixed period.
  - Residual - The random noise or irregularities after removing trend and seasonal components.
- **Key Findings**:
  - Residuals from all decompositions passed ADF test (p-value: 0.0)
  - Log Return and Percentage Return residuals showed minimal seasonality
  - Peaks in the seasonality can be seen during the first quarter of the year, which is also the end of the financial year. Availability or limitation of the liquidity in cash (Supply or Demand) might be said to have affected the market during that time, and other than that, there are 3 other spikes, which denote the other three quarters.

Fig. 12. This graph shows the monthly time series decomposition of closing prices and their derivatives from 2014 to 2024, broken into observed data, trend, seasonal, and residual components.

D. Auto-correlation Analysis

ACF/PACF Observations:

- **Original Closing price (Top Row):**
  - ACF shows a slow and gradual decay, indicating **non-stationarity**.
- **First Difference of Closing Prices (Middle Row - Close_diff):**
  - ACF and PACF decay quickly after lag 1, indicating that the differenced series is **stationary**.
  - A significant PACF spike at lag 1 supports a potential **AR(1)** process.
- **Log Returns of Closing Price (Bottom Row):**
  - Both ACF and PACF plots show minimal autocorrelation, implying that log returns have **white noise characteristics**.

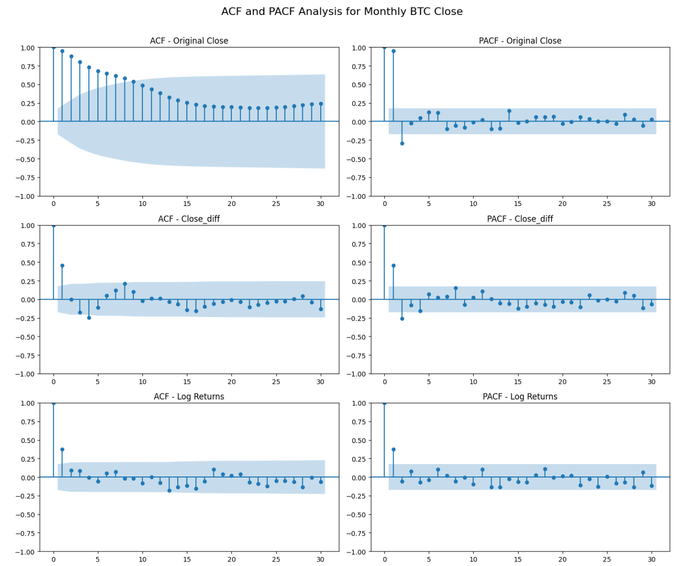

Fig. 13. This graph presents the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) analysis for monthly BTC closing prices, including original close, close differences, and log returns, up to a lag of 30 periods.

E. Frequency Distribution of Log Returns

You can see outliers on both the left and right tails (especially on the left side, toward -0.4). These represent rare but extreme market moves — potentially crashes or sudden jumps. The distribution seems slightly left-skewed — more extreme negative returns than positive ones. This suggests that downward moves might be more sudden or severe than upward ones.

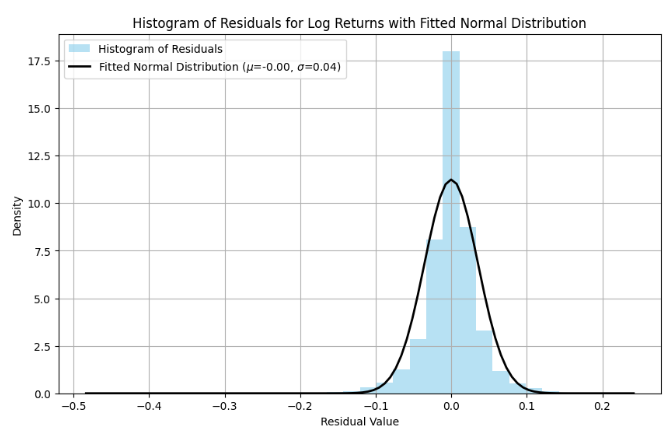

Fig. 14. This is a distribution of residuals of daily log returns with a sharp peak near zero with fat tails suggesting frequent small changes with rare but intense price shocks. Fitted to it is a nearest fitting Normal distribution with mean 0 and standard deviation 0.04

To check whether it is perfect, we found out its 3rd Moment (Skewness) and 4th Moment (Kurtosis), and they have values of -0.6522684368865438 and 10.79813601390954, which are not equal to 0. Therefore, it is not normal.

E. Unit circle Analysis

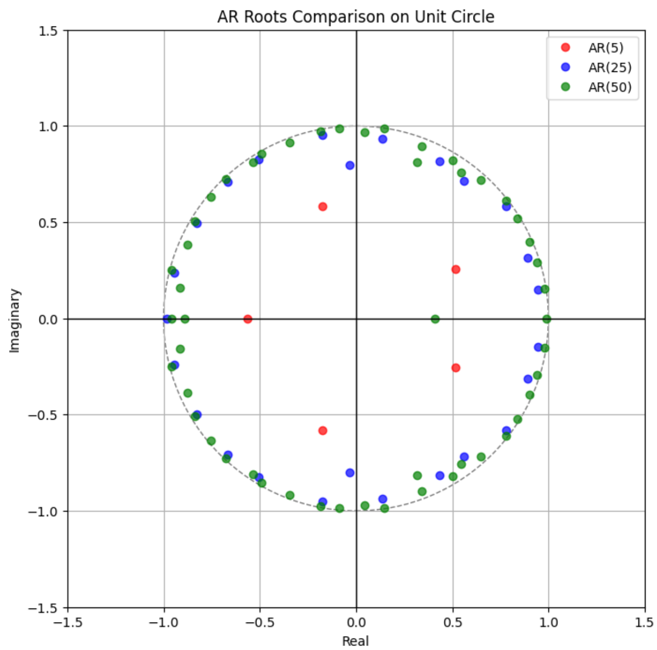

Fig. 15. These are the roots of the AR(p) for p = 5, 25 and 50 plotted together in the same unit circlein the complex plane

- **Unit Circle Significance:** The unit circle (dashed) serves as a stability boundary. For an autoregressive (AR) model to be stable, all characteristic roots must lie inside the unit circle.
- **Root Distribution:**
  - Red (AR(5)) roots are sparsely placed and some lie very close to or inside the unit circle, indicating potential marginal stability or instability.
  - Blue (AR(25)) roots show a denser distribution, generally clustering just outside the unit circle, which implies better model stability and spectral richness.
  - Green (AR(50)) roots are even more tightly packed around the unit circle but remain mostly outside, suggesting high model order with potentially better frequency resolution.

  All AR roots appear in complex-conjugate pairs, which is expected for models with real coefficients.
- **Model Complexity vs. Stability:**
  - AR(5) exhibits a simpler structure but with higher risk of instability due to proximity of roots to the unit circle.
  - AR(25) and AR(50) models offer increased stability and finer resolution, as evident from the roots clustering near the unit circle without crossing into the unstable region.
- **Interpretation:** Increasing the model order (from AR(5) to AR(50)) allows for capturing more complex dynamics in time series data. However, it also demands careful regularization to prevent overfitting and numerical instability.

V. Conclusion

This analysis of Bitcoin's price data from 2014 to 2025 highlights its highly volatile and non-stationary nature. Through statistical transformations and model diagnostics, key patterns such as seasonality, structural shifts, and autocorrelations were identified. The unit root analysis of AR models further revealed insights into the stability and frequency characteristics of the time series. By design, only 21 million Bitcoins will ever be created. The closer Bitcoin gets to its limit, the higher, theoretically, its price should be, assuming all other factors remain the same.

Link

Code : link

References

[1] M. Shell, "How to use the IEEEtran LaTeX class," 2023.
[2] Yahoo Finance API documentation, 2025.
[3] Dickey, D. A. & Fuller, W. A. (1979). "Distribution of the estimators for autoregressive time series..." *JASA*.