BITCOIN PRICES FROM INCEPTION USING TIME SERIES ANALYSIS

ARIMA MODEL

ARIMA is an acronym for "autoregressive integrated moving average." It's a model used in statistics and econometrics to measure events that happen over a period.

- Why Arima model: ARIMA models are relatively simple and easy to understand, making them accessible to a wide range of users. Furthermore, it is effective when dealing with time series data that can be made stationary through differencing.
- Usefulness: The ARIMA (Autoregressive Integrated Moving Average) model is used for time series forecasting by capturing and modeling the temporal dependencies, trends, and seasonality present in the data. It combines autoregressive (AR) and moving average (MA) components with differencing to achieve stationarity.

- Limitation: ARIMA models may struggle with complex nonlinear relationships and may not effectively capture long-term dependencies or sudden changes in the data, limiting their suitability for certain time series patterns.

- Assumptions: The model assumes that the time series is stationary, implying a constant mean, variance, and autocorrelation over time. If it is not, differencing is applied to achieve stationarity and Invertibility is another assumption, implying that the model's error terms can be expressed as a linear combination of current and past forecast errors.

- Mathematical formulation

$$Y_t = c + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_p Y_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q}$$

Where:

- Y_t is the value of the time series at time t.
- c is a constant term.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients.
- ϵ_t is the white noise error term at time t.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.

SARIMA model

The SARIMA model is designed to capture both the trend and seasonality present in time series data. It is particularly useful when dealing with data that exhibits a regular pattern of variation at fixed intervals, such as daily, monthly, or yearly seasonality.

- Why SARIMA model: the reason why we choose SARIMA is because it is a suitable choice when dealing with time series data that has both a non-seasonal and seasonal component, as it extends the capabilities of ARIMA to handle seasonality.

- Usefulness: SARIMA is valuable for time series forecasting, adept at capturing both trend and seasonality. Its seasonal components make it suitable for data with repeating patterns, providing accurate predictions for applications such as sales and stock prices.

- Limitation: They assume linearity, require stationarity, and may struggle with irregular patterns. Sensitivity to parameter selection, computational intensity, and difficulty with short datasets are additional challenges.

- Assumption: SARIMA assumptions include linearity, stationarity, independence, and normality of residuals. It assumes a constant seasonal pattern and historical data.

- Mathematical Formulation:

$$(1-\phi_1 B-\phi_2 B^2-\ldots-\phi_p B^p)(1-\Phi_1 B^s-\Phi_2 B^{2s}-\ldots-\Phi_P B^{Ps})(1-B)^d(1-B^s)^D y_t$$

$$=(1+\theta_1 B+\theta_2 B^2+\ldots+\theta_q B^q)(1+\Theta_1 B^s+\Theta_2 B^{2s}+\ldots+\Theta_Q B^{Qs})\varepsilon_t$$

where:

- y_t is the observed time series data at time t.
- B is the backshift operator, $B^k y_t = y_{t-k}$.
- $\phi_1, \phi_2, \ldots, \phi_p$ are the autoregressive coefficients.
- $\Phi_1, \Phi_2, \ldots, \Phi_P$ are the seasonal autoregressive coefficients.
- d is the non-seasonal differencing order.
- D is the seasonal differencing order.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average coefficients.
- $\Theta_1, \Theta_2, \ldots, \Theta_Q$ are the seasonal moving average coefficients.
- s is the length of the seasonal cycle (e.g., 12 for monthly data with yearly seasonality).
- ε_t is white noise (error term) at time t.

ARCH Model

ARCH models, or Autoregressive Conditional Heteroskedasticity models, are used in finance to predict changing volatility patterns. They assume that the variance of a time series is time-dependent and influenced by past squared observation.

- Why ARCH model: ARCH models are used to analyze time series data with changing volatility, particularly in financial contexts with volatility clustering. They capture conditional heteroskedasticity, allowing the model to adapt to varying levels of volatility based on past observations.
- Usefulness: ARCH models are crucial in finance for modeling and predicting changing volatility over time, aiding risk management and portfolio optimization. They play a key role in option pricing by providing accurate volatility estimates.

- Limitation: The Autoregressive Conditional Heteroskedasticity (ARCH) model assumes constant conditional volatility and may not capture sudden changes in volatility. It also relies on past squared returns for forecasting, making it sensitive to outliers.

- Assumption: The Autoregressive Conditional Heteroskedasticity (ARCH) model assumes a stationary time series with an autoregressive structure in conditional variance. It requires weak exogeneity of squared residuals, no perfect collinearity, independence of residuals, constant conditional variance, and normality of residuals.

- Mathematical Formulation

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2$$

where:

- σ_t^2 is the conditional variance at time t,
- α_0 is a constant term,
- α_i (for $i = 1, 2, ..., p$) are the coefficients corresponding to the past squared innovations ε_{t-i}^2,
- p is the lag order of the model, representing the number of past observations used in the model,
- ε_t is the white noise error term at time t.

GARCH model

The Generalized Autoregressive Conditional Heteroskedasticity (GARCH) model is a statistical tool used in finance to analyze and forecast volatility in time series data. It captures the changing variance over time by incorporating past squared forecast errors, providing insights into market risk and aiding in risk management and financial modeling.

- Why GARCH model
  The GARCH model is employed in finance to model and forecast changing volatility patterns, crucial for risk assessment. Its flexibility, incorporating both past squared observations and conditional variances, enhances its ability to capture complex volatility dynamics.

- Usefulness: The GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model is valuable in finance for modeling and forecasting volatility. It captures time-varying volatility in financial time series, aiding risk management and option pricing. By considering past volatility in predicting future volatility, GARCH enhances the accuracy of financial models and helps decision-makers assess and mitigate market risks.

- Limitation: The GARCH model has limitations, including its assumption of normal distribution for returns, which may not always hold in financial markets. It also tends to underestimate extreme events and may not fully capture abrupt changes in volatility. Additionally, GARCH models rely on historical data and may struggle with sudden shocks or structural breaks.

- The GARCH model assumes stationary, serially uncorrelated residuals with normally distributed errors. It relies on a linear relationship between past squared returns and current volatility. Homoskedasticity is relaxed, allowing for time-varying volatility.

- Mathematical-Formulation

  $$\sigma_t^2 = \omega + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \sum_{j=1}^{q} \beta_j \sigma_{t-j}^2$$

  In this equation:

  - σ_t^2 is the conditional variance at time t.
  - ω is a constant representing the long-run average of conditional variance.
  - α_i are parameters associated with past squared residuals, capturing short-term volatility.
  - β_j are parameters associated with past conditional variances, capturing long-term volatility.
  - ε_t is the white noise error term with mean zero and constant variance.

State-space Model

A State-Space Model (SSM) is a mathematical framework for modeling dynamic systems. It comprises state and observation equations, representing the system's evolution and its observable output. The state equation accounts for internal dynamics, while the observation equation links the state to observed data.

- Why State Space Model: State Space Models (SSMs) are employed for their flexibility in representing dynamic systems and separating observed data from unobservable states. They excel in modeling processes with latent variables, providing a nuanced understanding of complex dynamics.

- Usefulness: Space State Models are essential in control theory, offering a comprehensive representation of dynamic systems. They excel in handling multivariable and time-varying systems, enabling effective control system design, state estimation, and robust control.

- Limitation: State Space Models (SSMs) face challenges with nonlinear and non-Gaussian systems, limiting their accuracy in representing complex processes. Identifiability issues can impede reliable parameter estimation. SSMs demand a prior knowledge of the system dynamics, making them sensitive to model misspecification.

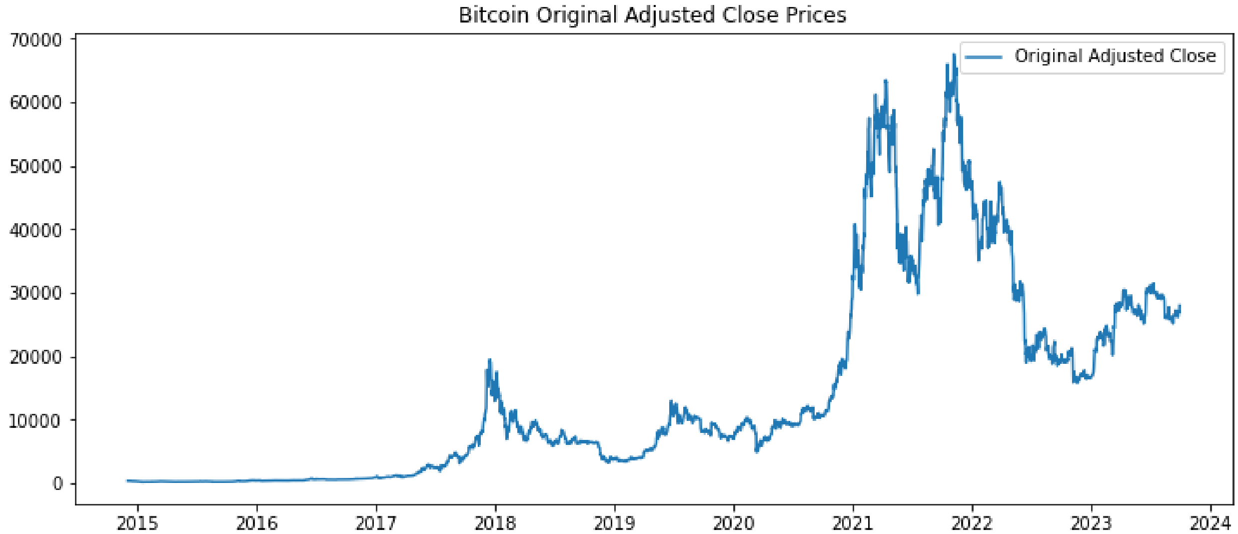
- Assumption: State-space models assume linearity in system dynamics and observation equations, Markovian property for state transitions, noisy measurements following a Gaussian distribution, known initial state, independence of noise over time
- Mathematical Formulation: The general continuous-time linear State-Space Model is:

$$\dot{x}(t) = Ax(t) + Bu(t)$$
$$y(t) = Cx(t) + Du(t)$$

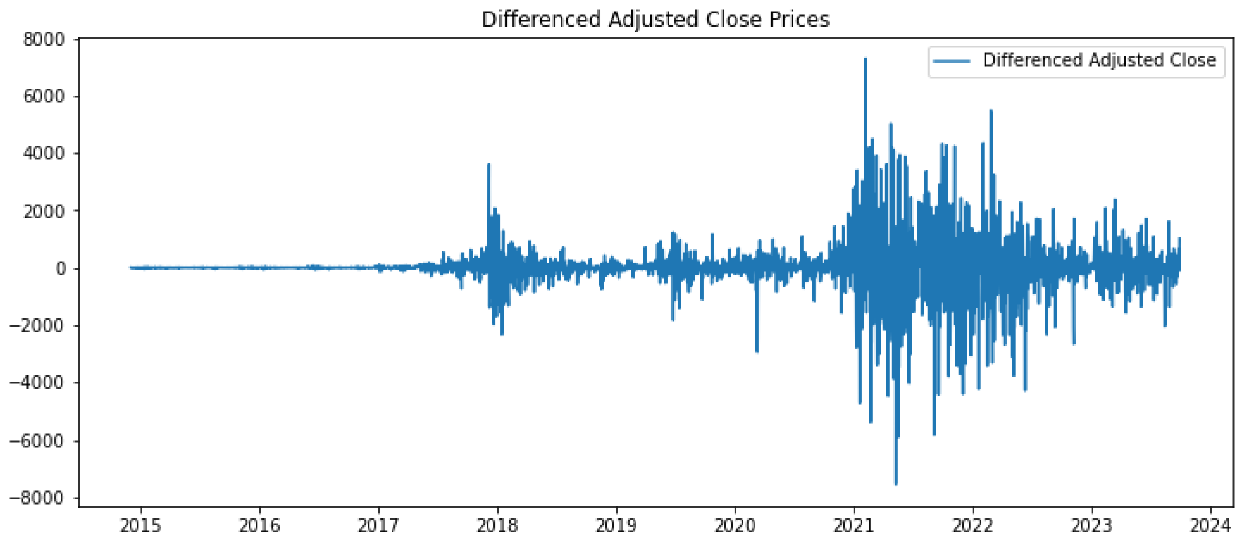
Here:

- $x(t)$ represents the state vector, which contains the system's internal variables.
- $u(t)$ is the input vector, representing the control inputs to the system.
- $y(t)$ is the output vector, representing the observed outputs of the system.
- A, B, C, and D are matrices that define the system:
  - A is the state matrix, describing how the state evolves over time.
  - B is the input matrix, describing how the inputs affect the state.
  - C is the output matrix, describing how the state contributes to the outputs.
  - D is the feedforward matrix, accounting for any direct influence of the inputs on the outputs.

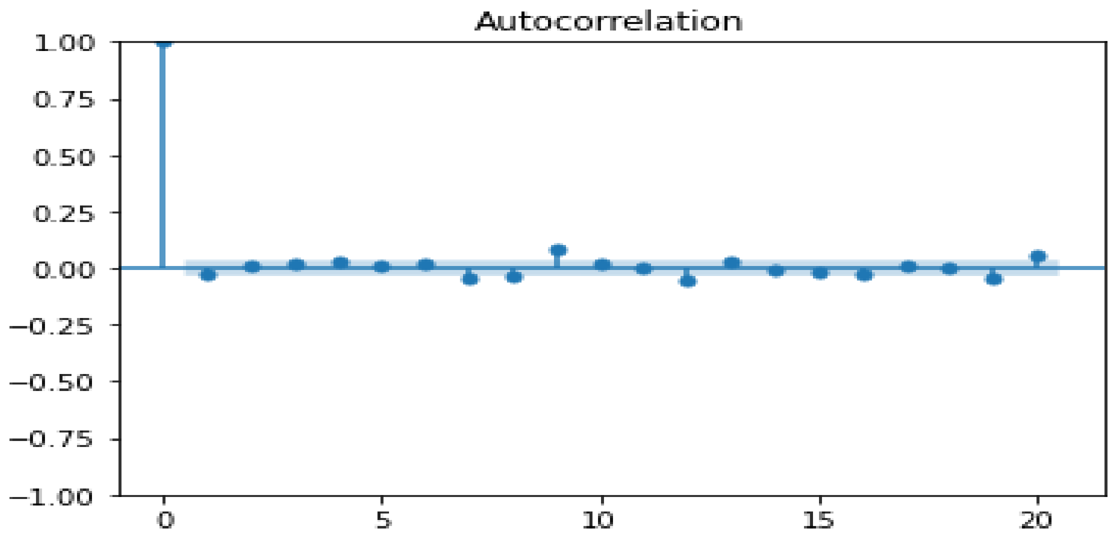
ARIMA MODEL CHART

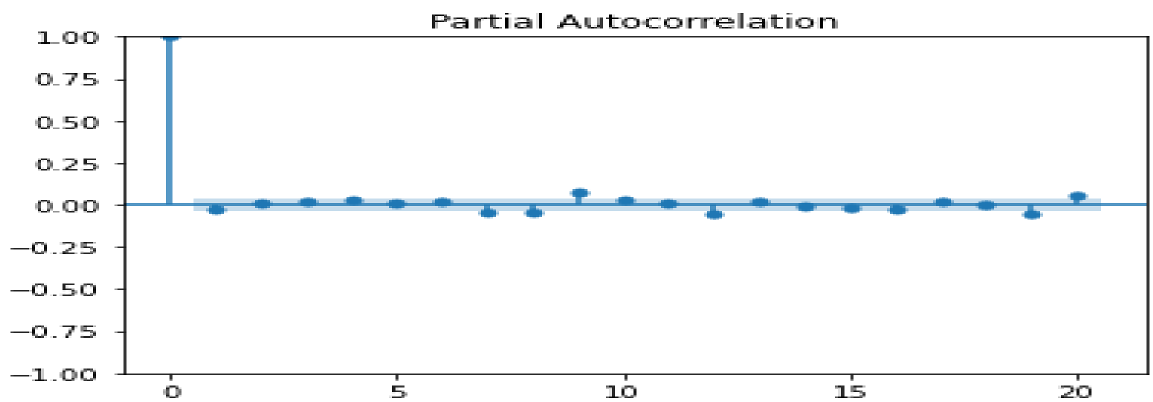


This plot showcases the historical price movement of Bitcoin. The data exhibits a volatile trajectory with significant peaks and troughs, characteristic of cryptocurrency markets. Notably, there is a sharp increase followed by periods of rapid decline, reflecting the market's speculative nature and sensitivity to various factors.

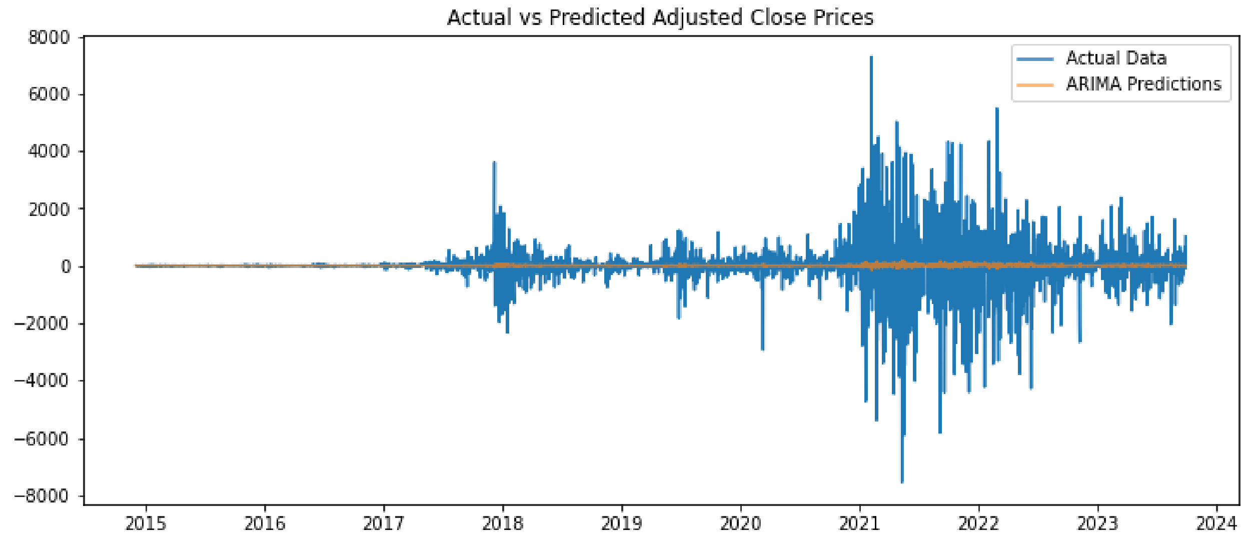


After performing a differencing operation to induce stationarity, the plot indicates the changes in price from one period to the next. The differenced series appears to be mean reverting, with volatility clusters suggesting periods of heightened market activity and period with less volatility suggesting periods of lower intensity of market activity.

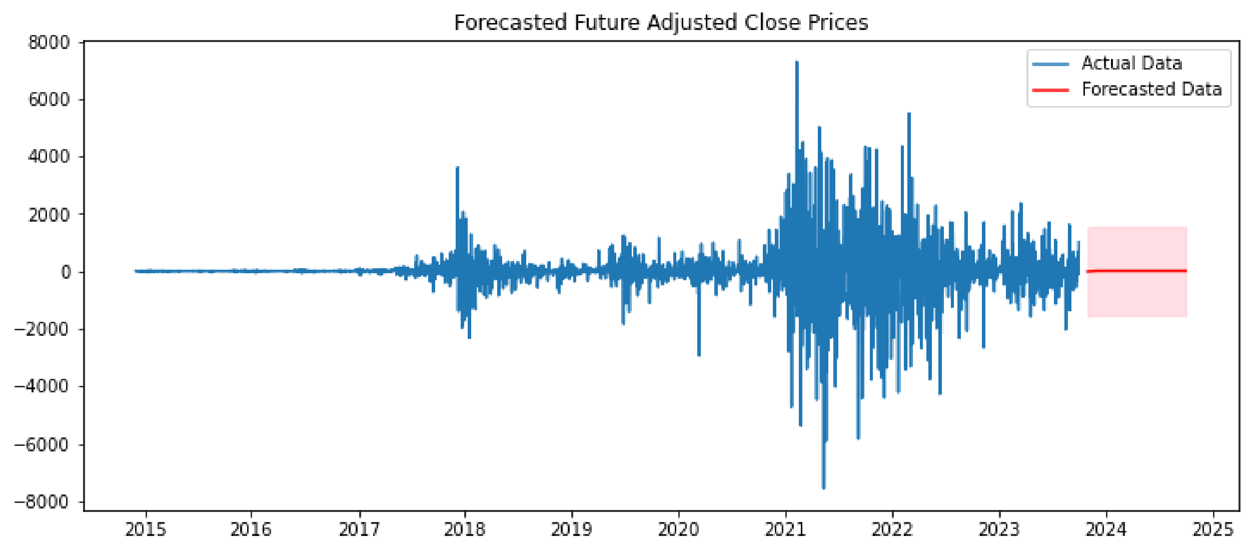




The ACF plot shows the correlation of the series with itself at different lags. A sharp drop after the first lag, which stabilizes around zero, indicates that the differenced series has no significant autocorrelation and that the differencing was effective in removing trends and seasonality. Similarly, the PACF plot displays a significant spike at the first lag and insignificance, thereafter, suggesting that an ARIMA model with a first-order autoregressive term may be appropriate.

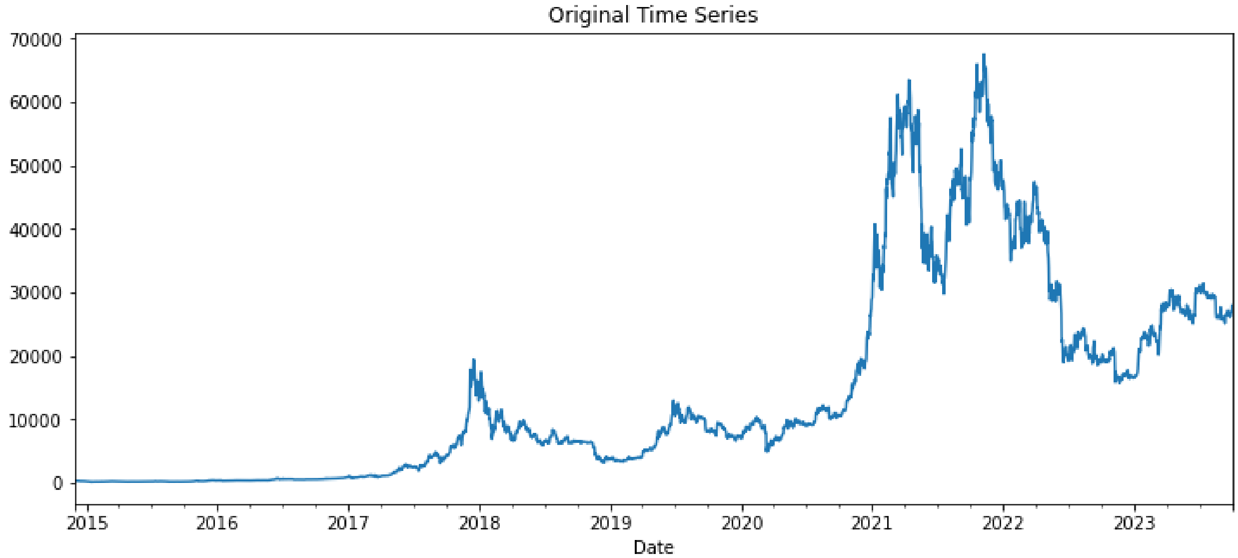


The plot compares the actual differenced data against the ARIMA model's predictions. Ideally, the predicted line should closely follow the actual data to indicate good model fit. However, due to the inherent noise and unpredictability of financial time series like Bitcoin, the model may not capture all market behaviors, especially in a volatile and irregular series such as this.

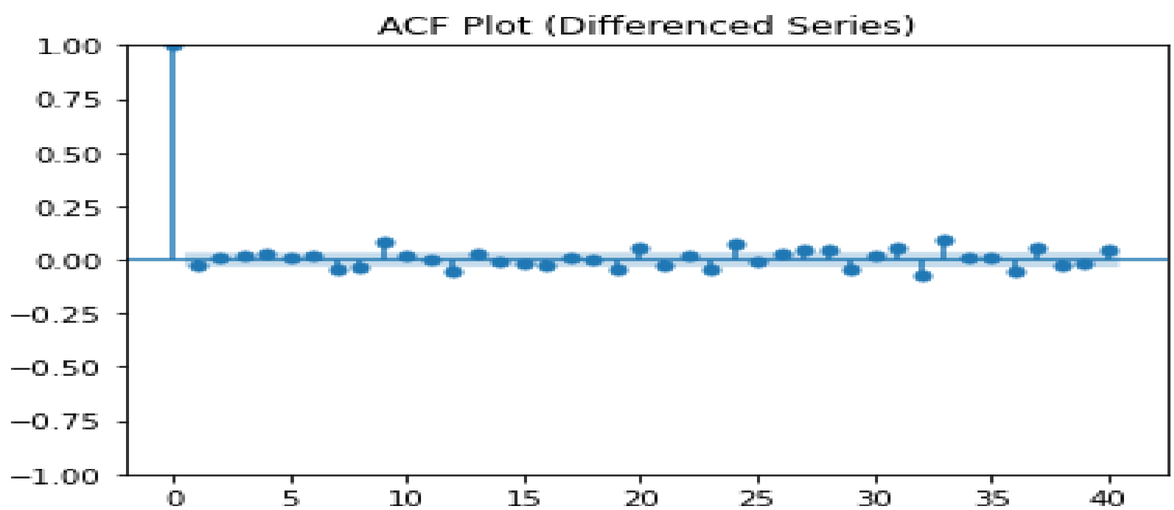


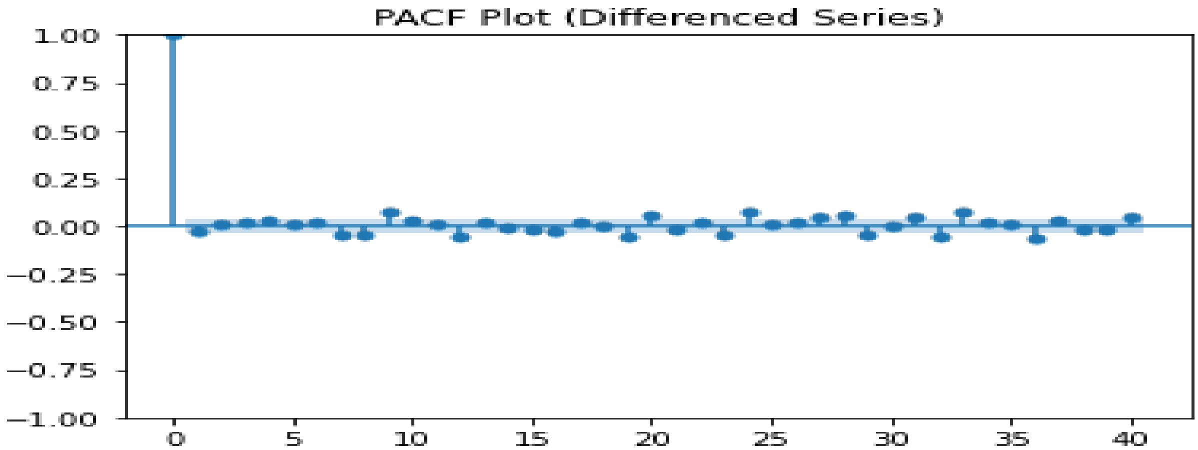
This plot extends the time series into the future, providing a forecast based on the ARIMA model. The shaded area represents the confidence interval for the predictions, illustrating the uncertainty associated with the forecast. In financial markets, especially with assets like Bitcoin, the wide confidence intervals highlight the speculative nature of such predictions.

SARIMA MODEL CHART

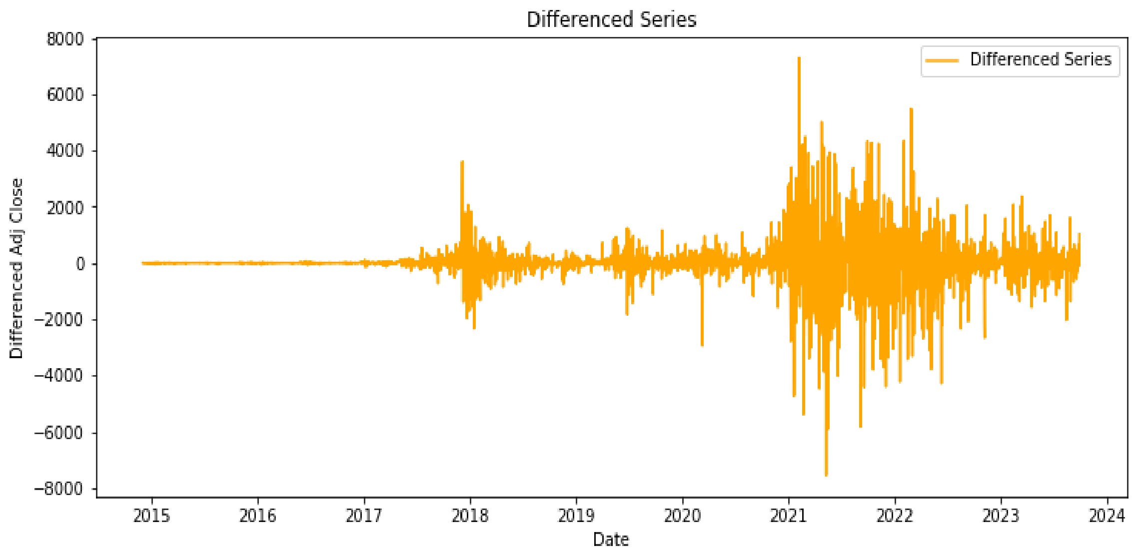


This chart shows the historical data of Bitcoin's adjusted closing price. It indicates a highly volatile market with significant price fluctuations over time, which is typical for cryptocurrencies. Notably, there are periods of rapid growth and sharp declines, reflecting market reactions to various external factors.

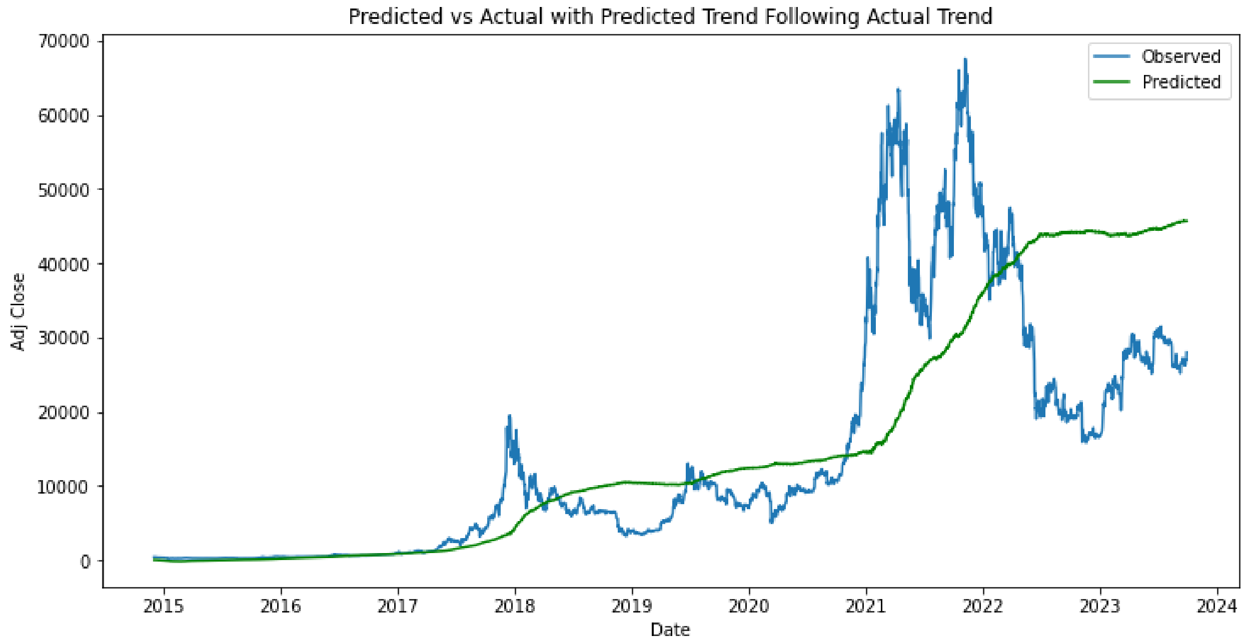




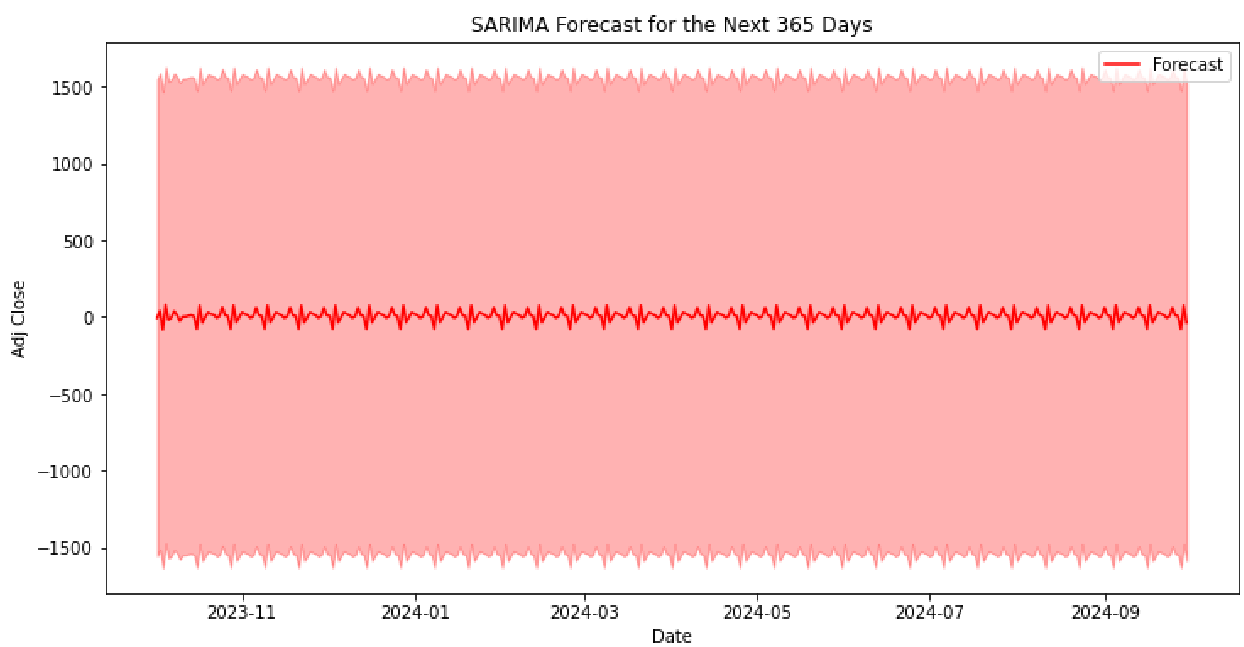
The Autocorrelation Function plot for the differenced series suggests that there is minimal correlation between past and present values after the first lag. This is typical after differencing a series to achieve stationarity. The Partial Autocorrelation Function plot also shows a significant drop after the first lag. This is an indication that the series may be well represented by an ARIMA model with a first-order autoregressive term.



This chart represents the differences between consecutive daily prices to remove trends and seasonality and achieve a stationary time series. The spikes suggest periods of high volatility, which are common in Bitcoin's trading history.

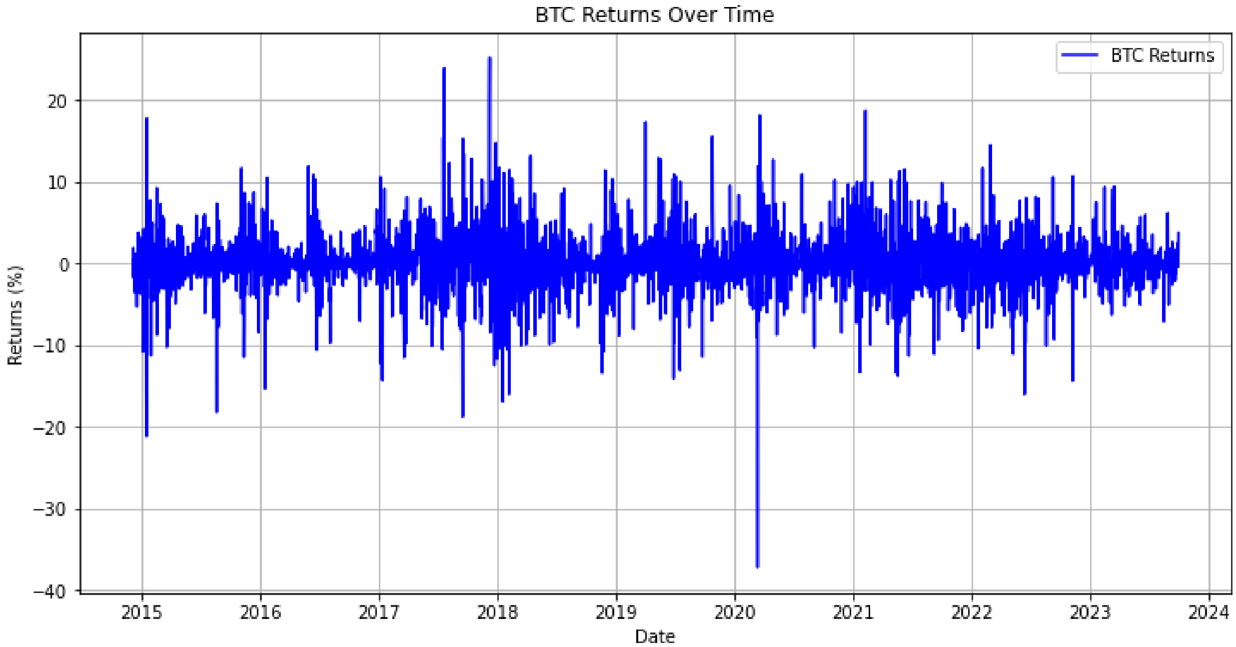


The predicted line (green) attempts to follow the actual observed data (blue) capturing the overall trend. It shows the model's attempt to match the historical data, which is challenging due to Bitcoin's volatility.

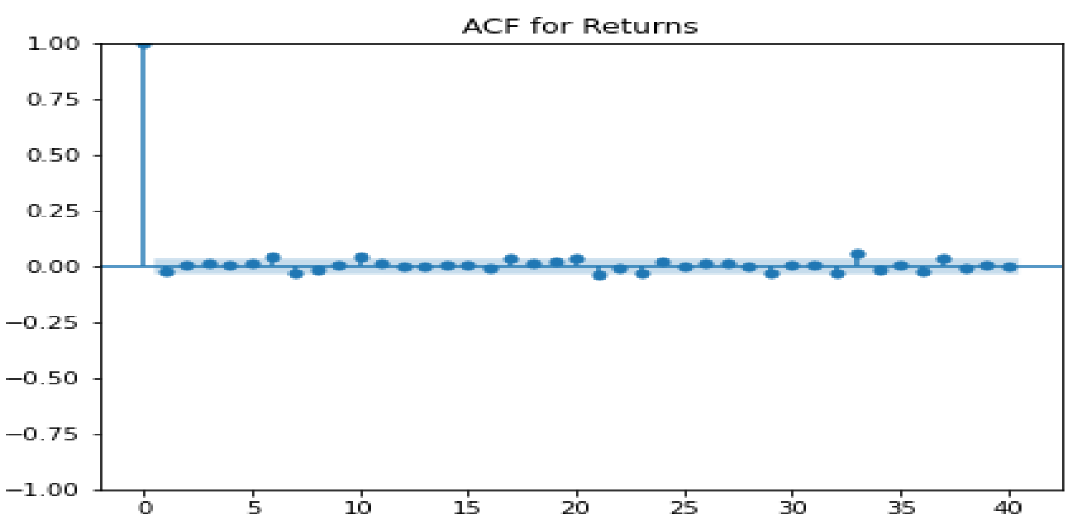


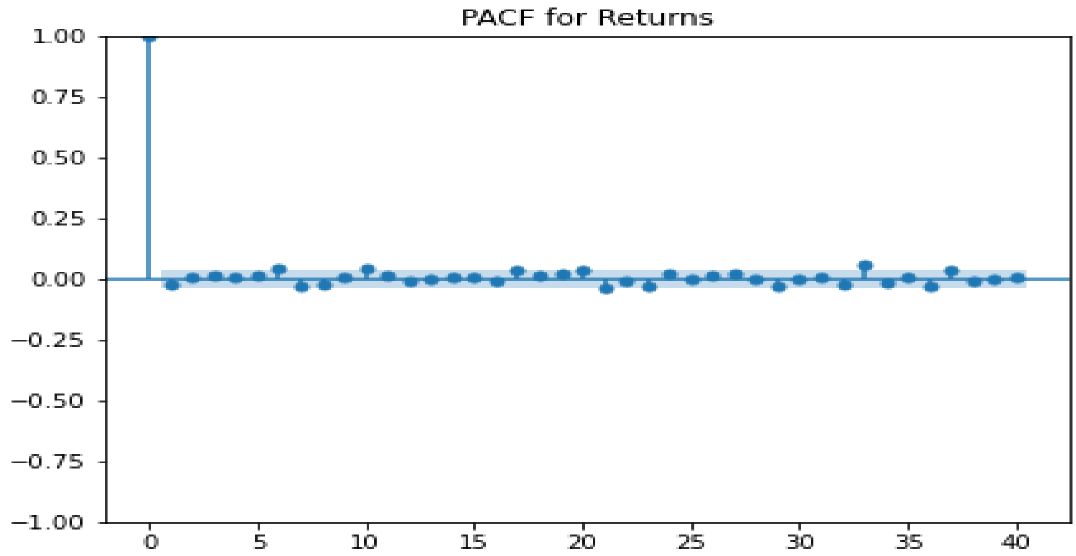
This forecast chart extends the analysis into the future. The shaded area represents the confidence interval, reflecting the uncertainty in the predictions. Given the volatile nature of Bitcoin, the wide confidence intervals indicate that while the model provides a direction, the actual prices could vary significantly.

ARCH MODEL CHART

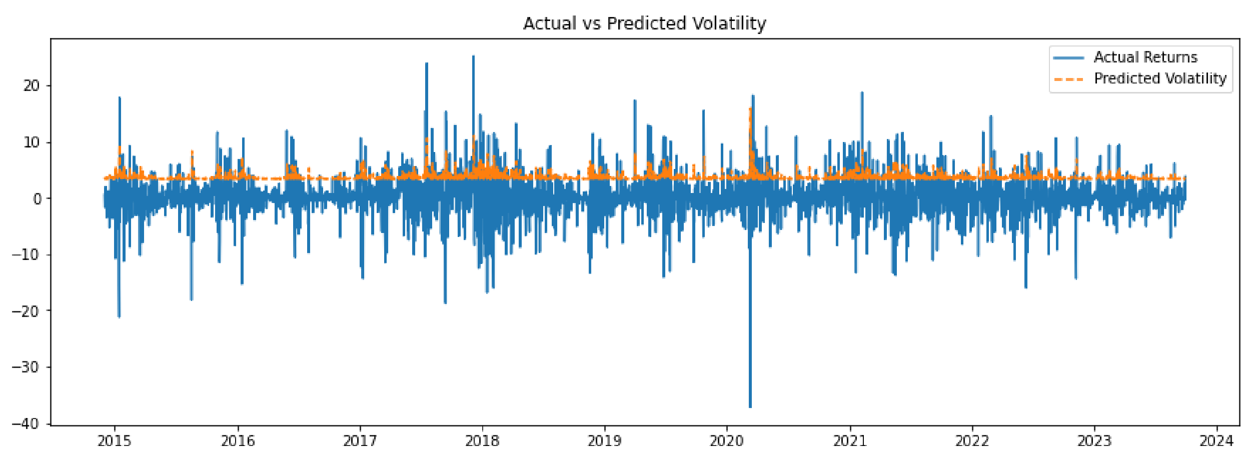


The first chart displays the percentage returns of Bitcoin over time. It highlights the extreme volatility and the presence of large swings in returns, which is characteristic of cryptocurrency markets.

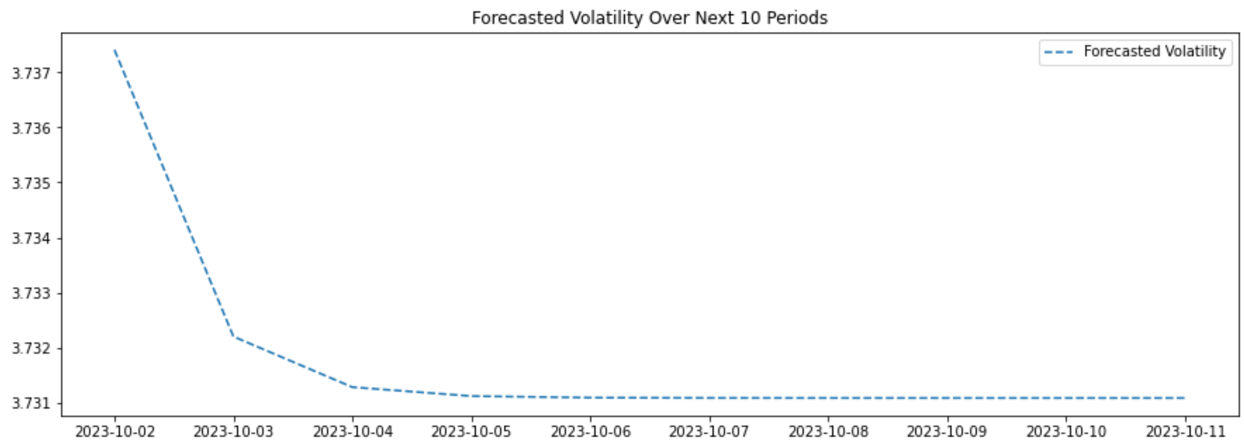




The Autocorrelation Function (ACF) chart suggests that there is a significant initial correlation that quickly diminishes, indicating a potential for a mean-reverting process. The ACF tails off, implying that the returns might be modeled well by a conditional heteroskedastic model such as ARCH/GARCH. The Partial Autocorrelation Function (PACF) chart shows a sharp drop after the first lag and insignificant correlations afterwards, which supports the choice of an ARCH model where only the immediate past volatility impacts current volatility.

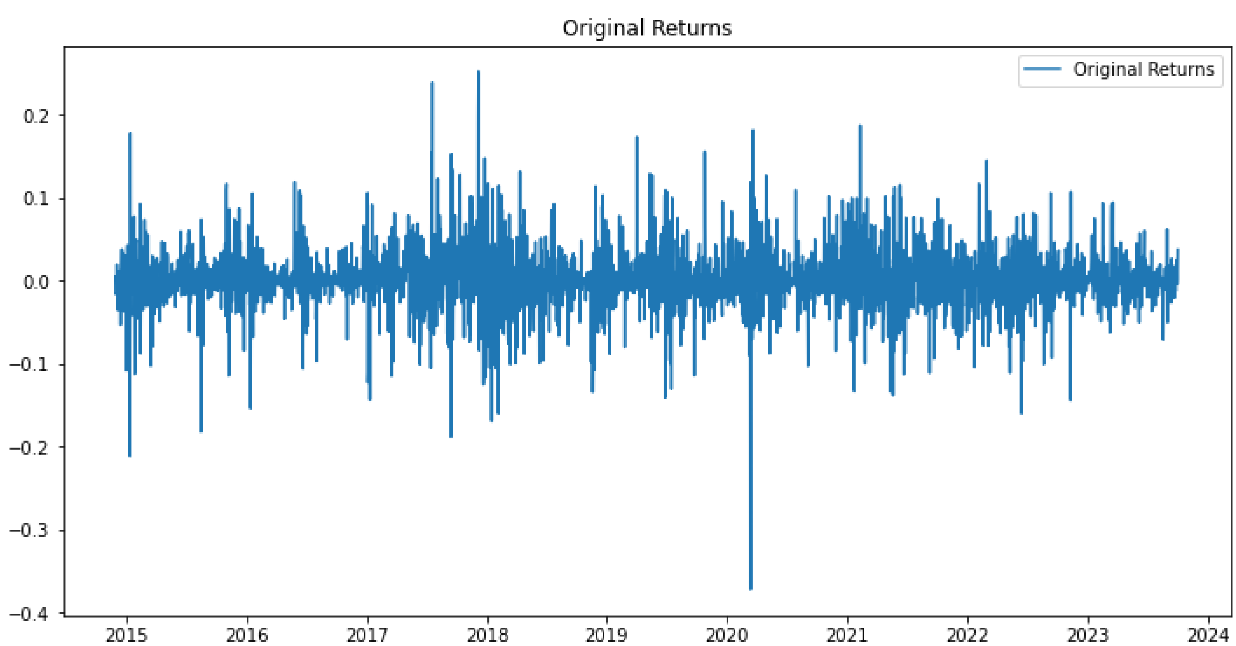


This chart plots the actual returns against the predicted volatility from the ARCH model. The predicted volatility, typically represented as a dashed line, aims to capture the changing variance in returns over time. If the ARCH model is correctly specified, the predicted volatility should match the timing and magnitude of the actual return spikes.

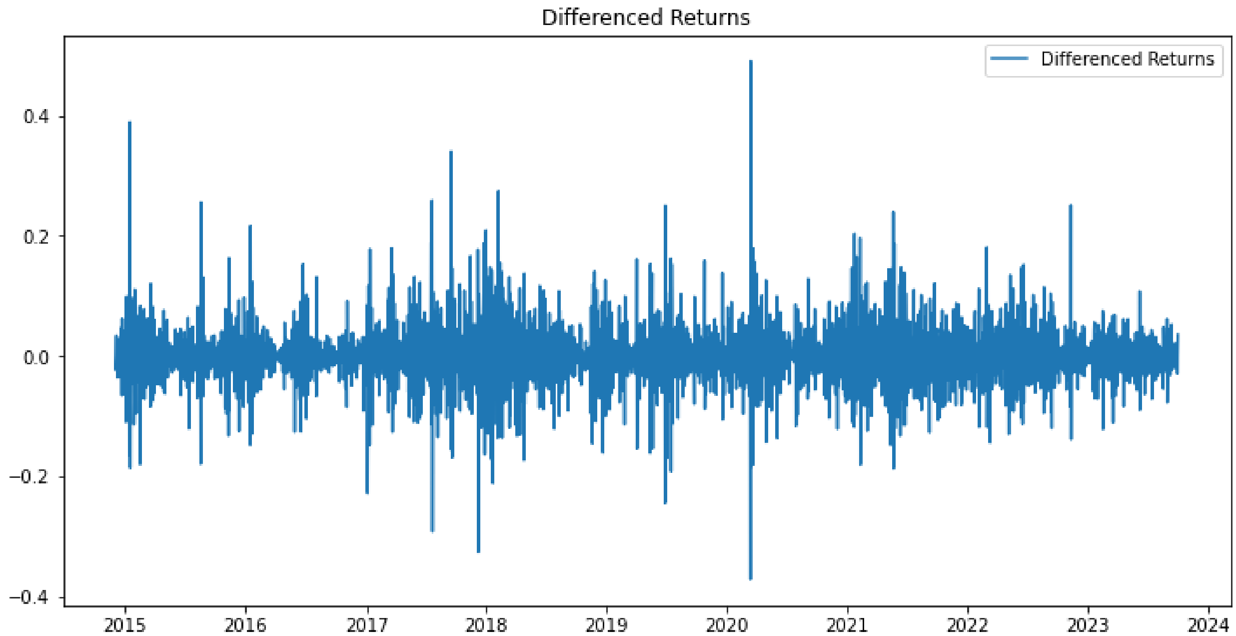


The final chart provides a short-term forecast of the expected volatility over the next 10 periods (which could be days, weeks, etc., depending on the frequency of the data). It shows the model's expectations about future variability in returns. In a well-fitted model, this forecast can inform about the risk associated with Bitcoin returns soon.

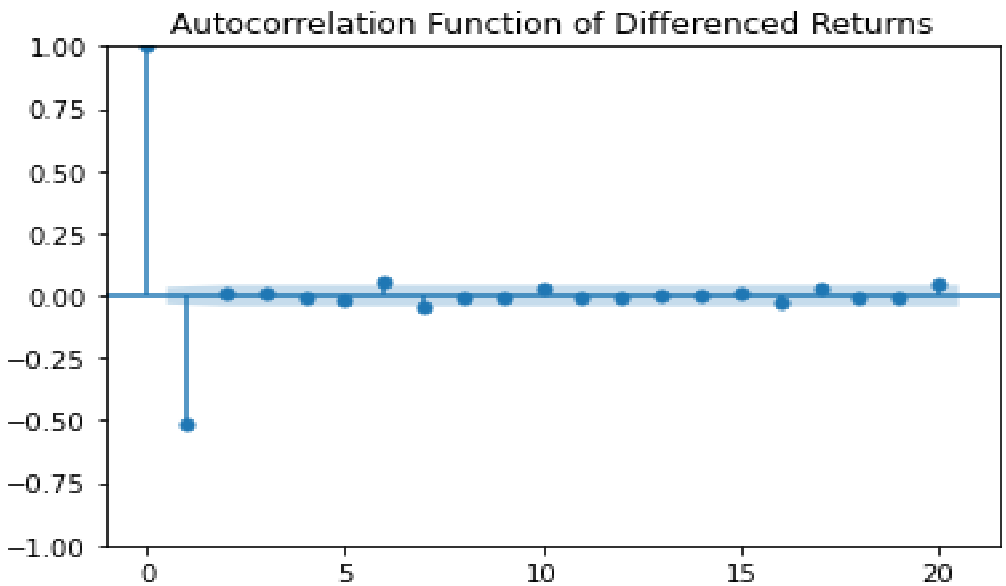
GARCH MODEL CHART

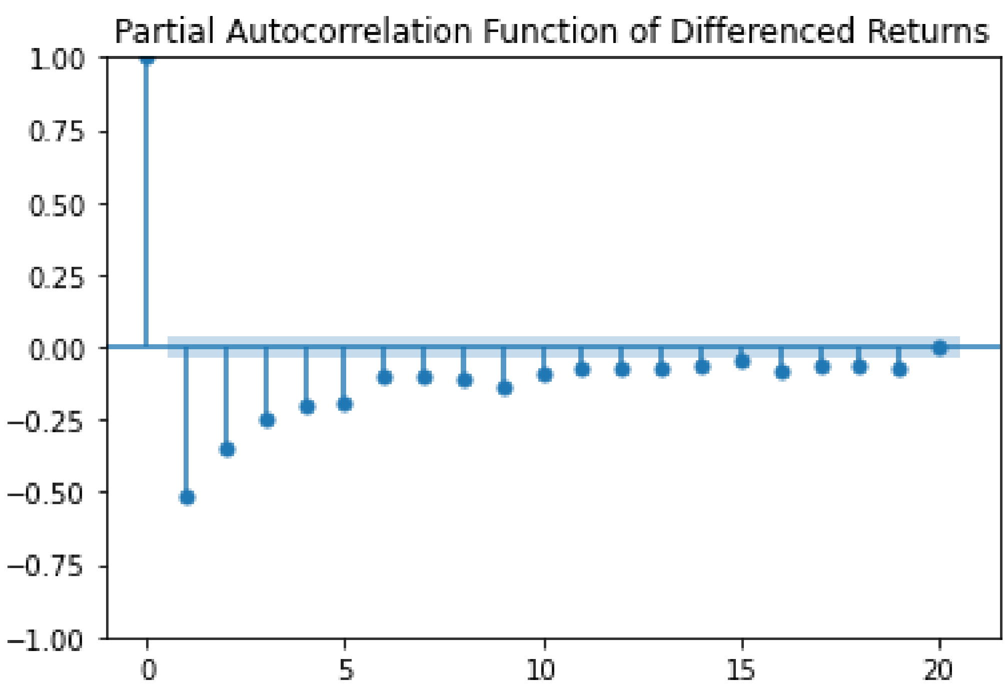


This plot likely shows the raw data of Bitcoin's adjusted closing prices over time. It would typically exhibit a non-linear pattern with peaks and troughs corresponding to market trends and events impacting Bitcoin's valuation.

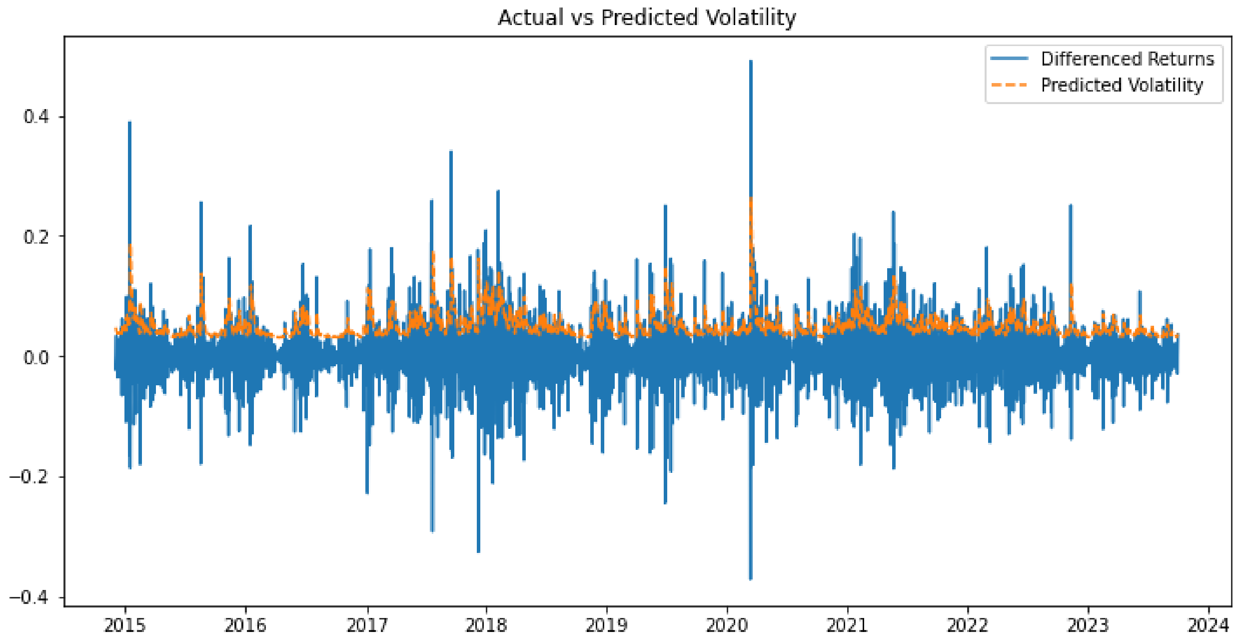


If the data needed to be differenced to achieve stationarity, this plot would show the changes in the adjusted close prices from one period to the next. This transformation is common in time series modeling to remove trends and seasonality, making the data stationary and suitable for ARIMA modeling.

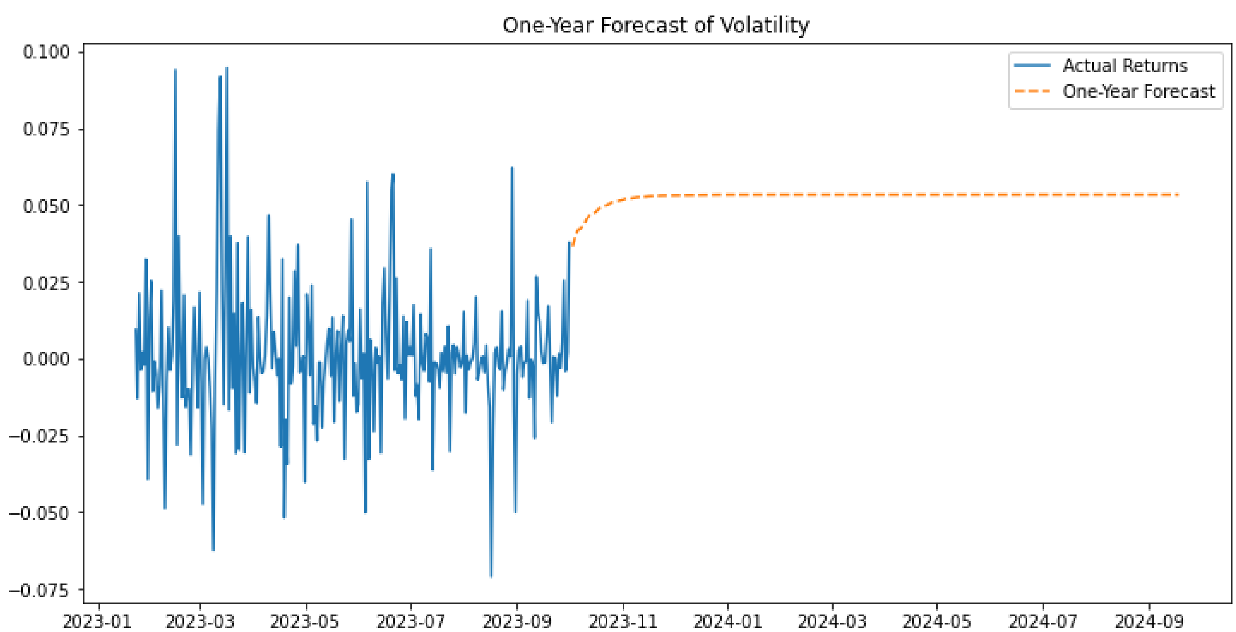




The Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) plots are tools to identify the order of the AR (autoregressive) and MA (moving average) components in an ARIMA model. The ACF plot shows the correlation of the series with its own lags, while the PACF plot shows the partial correlation (correlation of the series with its lags that is not explained by already considered earlier lags).

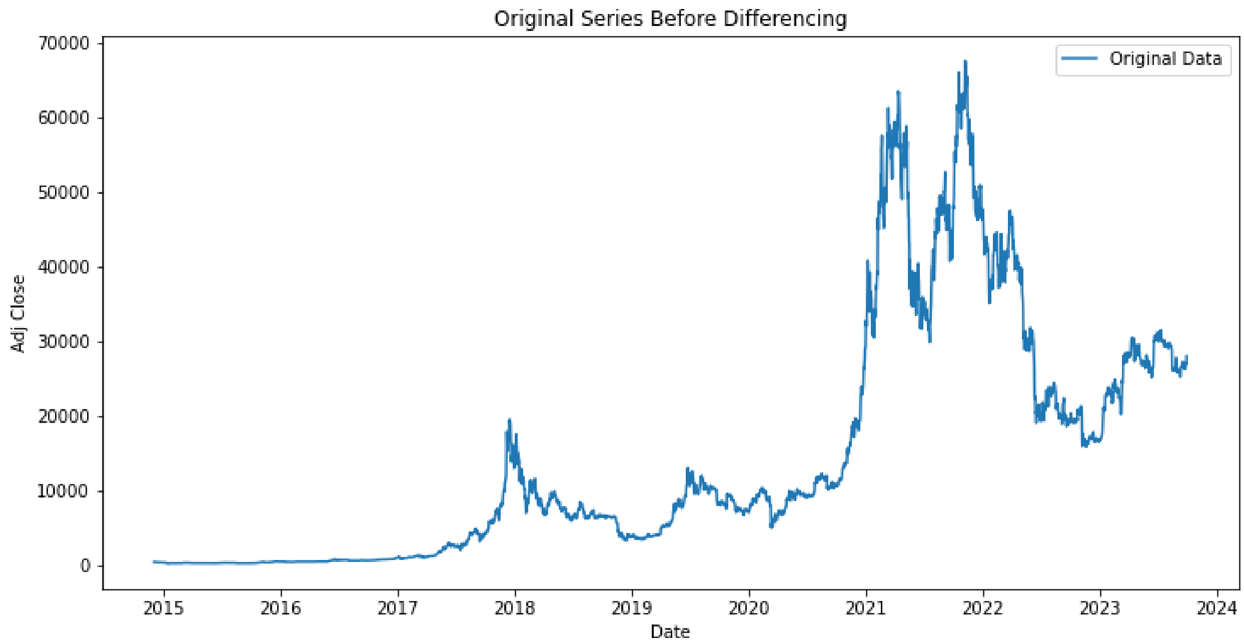


This plot compares the observed values of Bitcoin's adjusted close prices with the values predicted by the ARIMA model. An accurate model will show the predicted values closely following the actual data's trends and cycles.

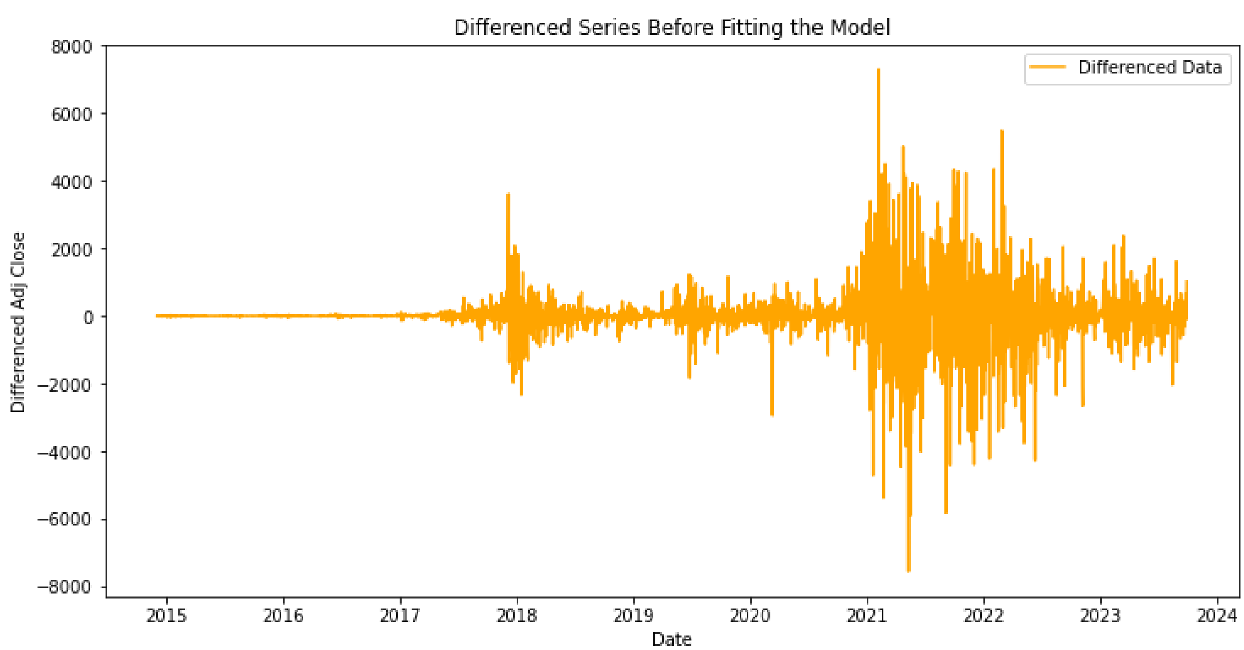


The forecast plot shows that after a period of somewhat variable returns, the model predicts a smoothing out of volatility over the next year. The dashed forecast line indicates a leveling off, suggesting that the model expects the market to become less volatile and more stable as time progresses.

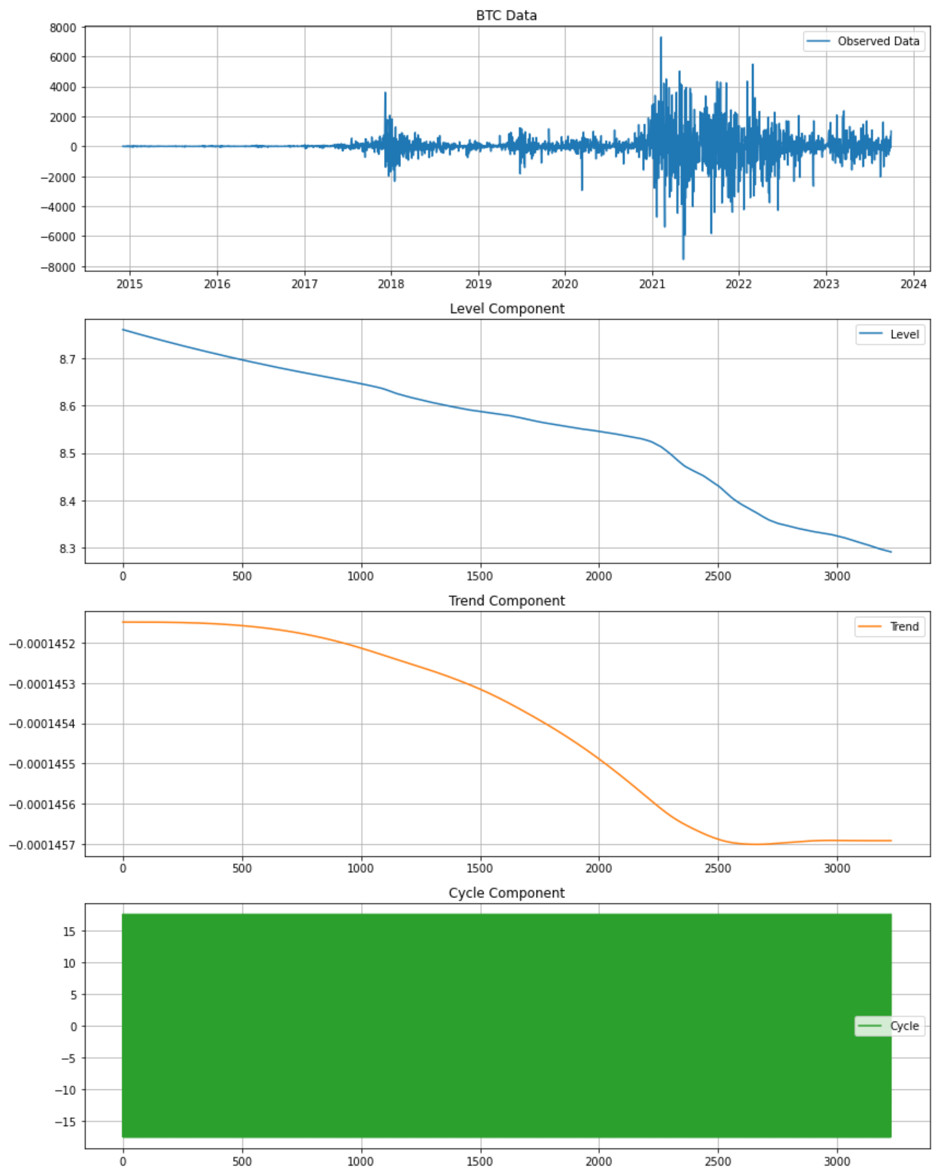
STATE SPACE MODEL CHART



The data appears to be non-stationary with clear trends and possibly some seasonality. The overall trend is upward, but with significant volatility and some noticeable spikes and drops. This non-stationary behavior is typical for financial time series data like stock prices.



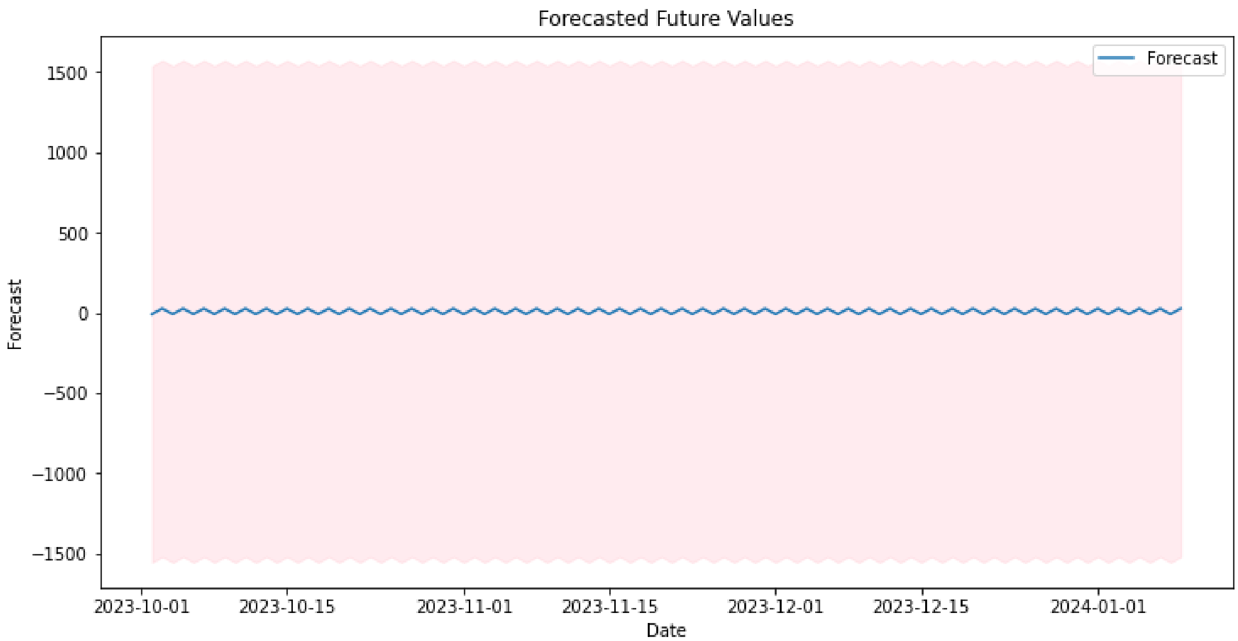
After differencing the data (usually to remove trends and achieve stationarity), the "Differenced Series Before Fitting the Model" chart does not show any clear trend, which is expected after differencing. The spikes in volatility are still present, which could indicate that the series still contains information such as sudden market movements or reactions to external events.



Level Component: The level appears relatively constant throughout the time series, as we would expect from differenced data, which ideally has a mean of zero.

Trend Component: The trend component shows a slight downward trend, indicating that over time, there might be a small persistent downward movement in the differenced series.

Cycle Component: This plot shows the cyclical patterns that the model has identified within the data. The cycle component seems to capture some regular patterns of fluctuation over time. However, without more context or a larger chart, it's challenging to interpret the frequency or amplitude of these cycles.



The forecast appears as a flat line because the model is likely predicting no change from the last observed value in the differenced series. This is typical when forecasting differenced data that's been made stationary; the model predicts that future changes will, on average, be zero.