Volatility based Option Trading Strategy Project Report

Sathya Balasubramani - sb9790

December 23, 2024

Contents

Chapter 1

Introduction

Options are financial derivatives that provide traders the right, but not the obligation, to buy or sell an underlying asset at a predetermined price (strike price) before or at a specified expiration date. The two main types of options are:

- **Call Option:** This gives the holder the right to buy the underlying asset at a specified price within a certain time period.
- **Put Option:** This gives the holder the right to sell the underlying asset at a specified price within a certain time period.

This report focuses on applying such volatility-based options trading strategies to the Indian market, specifically the National Stock Exchange (NSE) and its derivatives segment (NFO). A major area of interest is the Nifty 50 index options, which represent the top 50 companies in the NSE, making it a popular and highly liquid index. The strategies will leverage implied volatility trends and historical data to optimize returns, considering the unique characteristics of the Indian market.

Options trading strategies based on volatility involve capitalizing on price movements derived from the implied or historical volatility of the underlying asset. Volatility, a measure of price variation over time, plays a crucial role in determining options premiums and trader strategies. By focusing on this metric, traders can design strategies that profit in markets with high or low volatility.

This report focuses on applying such volatility-based options trading strategies to the Indian market, specifically the National Stock Exchange (NSE) and its derivatives segment (NFO). A major area of interest is the Nifty 50 index options, which represent the top 50 companies in the NSE, making it a popular and highly liquid index. The strategies will leverage implied volatility trends and historical data to optimize returns, considering the unique characteristics of the Indian market.

The underlying currency for all strategies and calculations in this report is **Indian Rupees (INR)**. The exchange rate used for conversions is **1 USD = 83 INR**.

Chapter 2

Strategy

2.1 Strategy Summary

The proposed strategy focuses on leveraging volatility analysis to trade options effectively. The core idea is to identify mispriced options based on the relationship between Implied Volatility (IV), Realized Volatility (RV), and Forecasted Volatility (FV). By grouping options by expiry, ranking them within these groups, and generating trade signals, the strategy systematically evaluates opportunities for short straddle and long straddle trades.

2.2 Key Concepts

2.2.1 Short Straddle

- Simultaneously sell a Call (CE) and a Put (PE) option with the same strike price and expiry.
- Profitable when the market remains stable (low realized volatility).
- **Signal:** IV $>$ RV or IV $>$ FV.

2.2.2 Long Straddle

- Simultaneously buy a Call (CE) and a Put (PE) option with the same strike price and expiry.
- Profitable when the market becomes volatile (high realized volatility).
- **Signal:** IV $<$ RV or IV $<$ FV.

2.3 Analytical Framework

2.3.1 Grouping Options by Expiry

- Organize CE and PE options into groups based on their expiry date.
- Within each group, include relevant columns: strike price, IV, RV, FV, and option type.

- Ensures time consistency for analysis and precise backtesting.

2.3.2 Ranking Within Expiry Groups

- Rank options by their IV within each expiry group.
- Identifies options with the highest deviation from RV/FV, signaling potential mispricing.

2.3.3 Signal Generation

- Compare IV with RV and FV:
  - IV > RV/FV: Short straddle (market overestimates volatility).
  - IV < RV/FV: Long straddle (market underestimates volatility).
- Generate actionable trade signals for each expiry group.

2.4 Steps for Implementation

2.4.1 Data Preparation

- Import options data into a PostgreSQL database.
- Compute IV, RV, and FV for all options.

2.4.2 Cross-Sectional Grouping

- Group options by expiry date.
- Ensure each group contains all necessary columns for analysis.

2.4.3 Ranking

- Rank options within expiry groups based on IV.

2.4.4 Signal Generation

- Generate short/long straddle signals by comparing IV, RV, and FV.

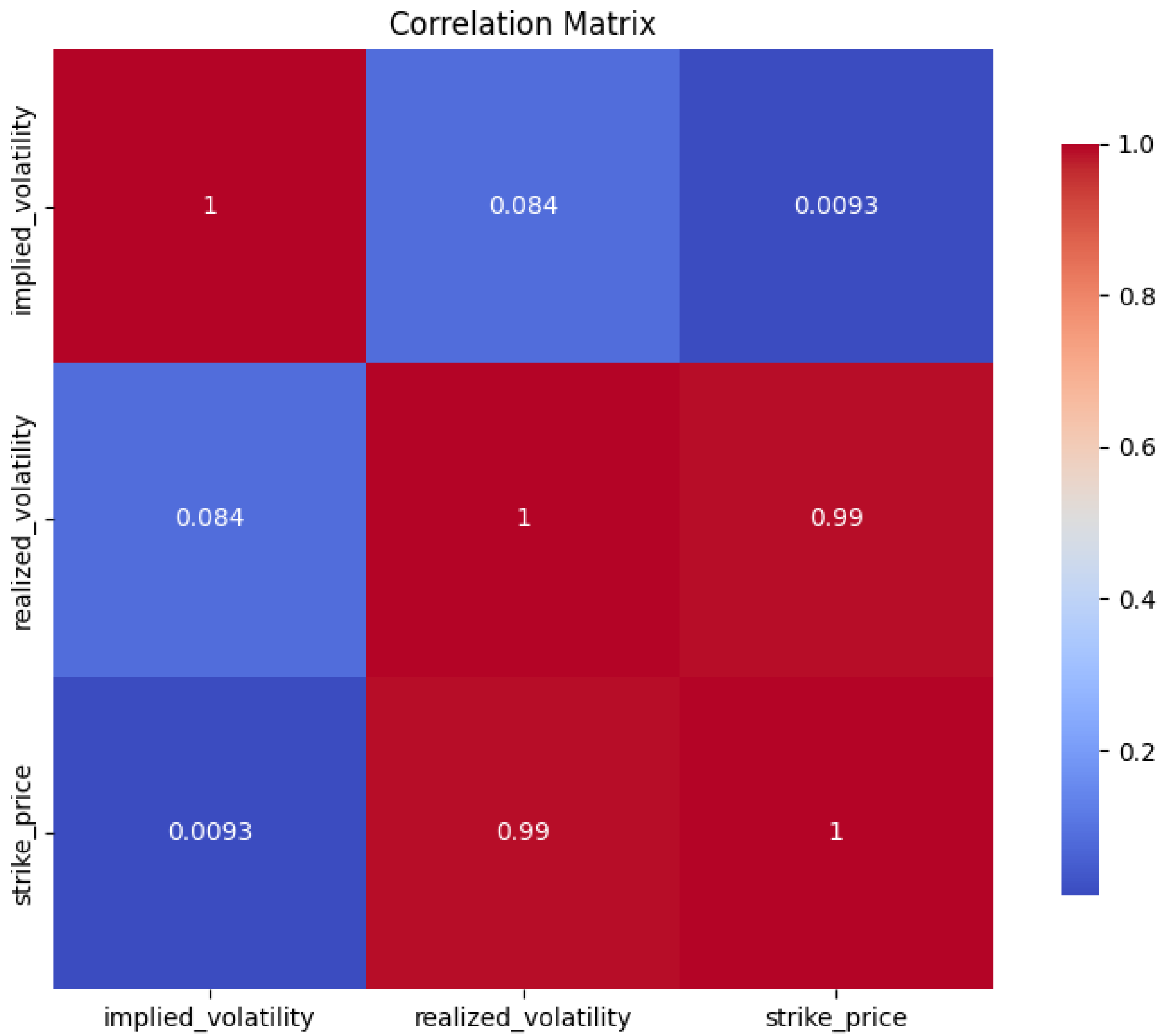
2.4.5 Backtesting

- Simulate trades based on generated signals.
- Evaluate profitability and risk-adjusted returns for each expiry group.

2.4.6 Statistical Analysis

- Use statistical metrics (mean, standard deviation) to validate the results.
- Perform hypothesis testing to check the significance of findings.

2.4.7 Correlation Analysis



The correlation matrix provides a quantitative overview of the relationships between the key variables: **Implied Volatility (IV)**, **Realized Volatility (RV)**, and **Strike Price**. The matrix is visualized using a heatmap, with values ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation). The color gradient from blue to red indicates the strength of the correlation, with red representing a strong positive correlation and blue representing a weak or negative correlation.

Key Observations:

- **Implied Volatility (IV) and Realized Volatility (RV):** The correlation value between IV and RV is 0.084, indicating a very weak positive correlation. This suggests that changes in implied volatility are not strongly related to changes in realized volatility for the observed data.
- **Implied Volatility (IV) and Strike Price:** The correlation value between IV and strike price is 0.0093, which is negligible. This implies that the implied volatility is nearly independent of the strike price.
- **Realized Volatility (RV) and Strike Price:** The correlation between RV and strike price is 0.99, indicating a near-perfect positive correlation. This highlights

that as the strike price increases, realized volatility tends to increase proportionally.

- **Strike Price Correlation:** The self-correlation of strike price with itself is 1, as expected in any correlation matrix.

Insights:

- The near-zero correlation between IV and strike price (0.0093) suggests that implied volatility for options is not influenced by the moneyness or strike levels in this dataset.
- The high correlation between RV and strike price (0.99) indicates that realized volatility is significantly influenced by the strike price, which may be indicative of market trends or option pricing behavior.
- The weak correlation between IV and RV (0.084) highlights the need for careful consideration of these metrics when developing trading strategies, as they behave almost independently.

Color Representation:

- The red blocks in the matrix represent strong correlations (> 0.9), particularly between RV and strike price.
- The blue regions correspond to weaker correlations (< 0.1), such as between IV and strike price, and IV and RV.

2.5 Alignment of IV, RV, and FV

- **Implied Volatility (IV):** Reflects market expectations of future volatility.
- **Realized Volatility (RV):** Calculated from historical price movements over a specified period.
- **Forecasted Volatility (FV):** Predicted using statistical models (e.g., GARCH).

2.5.1 Validation

- Match date ranges for RV/FV with IV (ensure alignment).
- Use statistical measures (e.g., correlation) to verify the relationship between IV, RV, and FV.
- Adjust RV/FV calculations to align with option expiry periods.

Chapter 3

Inspired from Similar Work

3.1 Referred Paper: *Volatility and the Cross-Section of Returns on FX Options*

The paper examines the profitability of volatility-based trading strategies in the foreign exchange (FX) options market by ranking currencies based on their Implied Volatility (IV) and analyzing the returns of long-short straddle positions.

Key Findings: Currencies with low IV generate significantly positive returns when traded with long straddles, while currencies with high IV perform poorly, yielding negative returns on long straddles. A long-short strategy (buying straddles on low-IV currencies and selling straddles on high-IV currencies) achieves an annualized Sharpe ratio of 1.04. Delta-hedged strategies reinforce these findings by exhibiting a negative relationship between IV and straddle returns.

The study highlights the profitability of ranking by IV as a predictor of straddle returns and its robustness across alternative measures, such as decomposed volatility and accounting for transaction costs.

3.2 Application to NSE Nifty 50 Options

Inspired by the paper, the methodology was adapted to the Indian equity options market, specifically the NSE Nifty 50 index options. The following steps were implemented:

Implied Volatility Ranking: Following the paper, Nifty 50 options were ranked based on their IV within expiry groups. IV was calculated using market prices, underlying index value, strike price, time to expiry, and risk-free rate. Options were grouped by expiry date, and ranks were assigned based on IV.

Cross-Sectional Grouping by Expiry: Grouping by expiry ensured consistency in time-to-expiry dynamics. Options data were organized by expiry date and strike price, creating expiry-specific datasets. Realized Volatility (RV) and Forecasted Volatility (FV) were calculated using historical price data and a GARCH model for comparisons with IV.

Signal Generation: Trading signals were derived by comparing IV with RV and FV:

- **Short Straddle:** If IV > RV or IV > FV.

- **Long Straddle:** If IV < RV or IV < FV.

Within each expiry group, overvalued options (high IV) were identified for short straddles, and undervalued options (low IV) for long straddles.

Backtesting: The performance of the strategy was simulated using historical Nifty 50 options data. Metrics such as Sharpe ratio, alpha, drawdowns, and P&L were analyzed for each expiry group to evaluate profitability.

3.3 Inspiration from the Paper

Ranking by IV: The methodology of sorting options by IV and analyzing returns within cross-sectional groups inspired the implementation for Nifty 50 options. **Straddle Strategies:** The adoption of long and short straddles as direction-neutral strategies for capturing volatility mispricing was directly influenced by the paper. **Volatility Analysis:** Emphasis on the relationship between IV, RV, and FV, with a focus on total volatility as a predictor, was another significant takeaway from the paper.

3.4 Summary of Strategy for NSE Nifty 50 Options

- **Rank Options by IV:** Group options by expiry and assign ranks based on IV.
- **Generate Signals:** Compare IV with RV and FV to identify overvalued or undervalued options.
- **Implement Straddle Strategies:** Execute short straddles for high-IV options and long straddles for low-IV options.
- **Backtest and Validate:** Assess profitability using historical data and refine the strategy based on insights.

This adaptation demonstrates the applicability of the paper's methodology to a different market (Nifty 50 options) while maintaining its core principles.

Chapter 4

Describe Data Preparation and Assumptions

Data Preparation and Assumptions

Initial Setup

Obtaining historical Nifty index options data was challenging due to the lack of comprehensive third-party libraries spanning the last decade. To address this, we developed a Python script using REST API calls to systematically extract raw Nifty options data directly from the NSE website. The extracted data spans the years 2015 to 2024, organized by year, expiry, and option type.

Data cleaning was performed using the `pandas` library, handling missing values with forward fill (`ffill`) and backward fill (`bfill`) methods provided by `numpy`. The processed data was then imported into a PostgreSQL database hosted in a Docker container for further analysis.

The API parameters used for fetching data are structured as:

```
    params = {
  "from": start.strftime("%d-%m-%Y"),
  "to": end.strftime("%d-%m-%Y"),
  "instrumentType": instrumentType,
  "symbol": symbol,
  "year": year,
  "expiryDate": expiry,
  "optionType": option
}
```

Risk-Free Rate and Time to Expiry

The risk-free rate was determined using historical government bond yields, with yearly rates as follows:

```
{
  2015: 0.079, 2016: 0.074, 2017: 0.068, 2018: 0.075,
  2019: 0.065, 2020: 0.060, 2021: 0.062, 2022: 0.069,
  2023: 0.070, 2024: 0.068
}
```

Time to expiry was calculated using the difference between the trade date and expiry date, expressed in years.

Calculation of Implied Volatility

We used the `py-lets-be-rational` library to calculate implied volatility (IV) for each option based on the Black-Scholes model. IV serves as a forward-looking measure of market volatility. Invalid data points, such as missing or unrealistic option parameters, were handled gracefully.

Realized Volatility

Realized volatility was computed using a rolling 30-day window of log returns, calculated as:

$$\text{Log Return} = \ln\left(\frac{\text{Price}_t}{\text{Price}_{t-1}}\right)$$

Forecasted Volatility

This table represents the results of a **Constant Mean - GARCH Model** fitted to log-returns. The table includes statistics related to the mean and volatility models, along with their respective coefficients, standard errors, t-statistics, p-values, and confidence intervals.

- **Mean Model (μ)**: The constant mean value is estimated to be -0.00129 with a high p-value (0.678), suggesting it is statistically insignificant.
- **Volatility Model ($\omega, \alpha_1, \beta_1$)**:
  - ω (constant in the variance equation): Estimated at 0.0150, with a statistically significant p-value of 0.013, indicating its importance in the model.
  - α_1 (ARCH term): Estimated at 0.0609, which measures the short-term volatility from past squared residuals. The high t-statistic (5.049) and low p-value (0.000) confirm its significance.
  - β_1 (GARCH term): Estimated at 0.9383, representing the persistence of volatility over time. The high value and significant p-value (0.000) suggest that volatility clusters persist for a long duration.

Constraint: $\omega + \alpha_1 + \beta_1 < 1$

The constraint $\omega + \alpha_1 + \beta_1 < 1$ ensures the model's stationarity and stability. It implies that the variance of the model converges over time, avoiding explosive growth in volatility. In this case:

$$\omega + \alpha_1 + \beta_1 = 0.0150 + 0.0609 + 0.9383 = 1.0142$$

Since the sum slightly exceeds 1, this might indicate non-stationarity, potentially leading to long-term divergence in volatility. Further adjustments to the model or constraints could improve its stability.

```
                     Constant Mean - GARCH Model Results
==============================================================================
Dep. Variable:               log_return   R-squared:                       0.000
Mean Model:               Constant Mean   Adj. R-squared:                  0.000
Vol Model:                        GARCH   Log-Likelihood:                -190524.
Distribution:                    Normal   AIC:                           381056.
Method:              Maximum Likelihood   BIC:                           381093.
                                          No. Observations:              88918
Date:                  Mon, Dec 23 2024   Df Residuals:                  88917
Time:                          19:40:26   Df Model:                          1
                                 Mean Model
==============================================================================
                  coef    std err          t      P>|t|       95.0% Conf. Int.
------------------------------------------------------------------------------
mu         -1.2920e-03  3.108e-03     -0.416      0.678 [-7.384e-03,4.800e-03]
                              Volatility Model
============================================================================
                 coef    std err          t      P>|t|      95.0% Conf. Int.
----------------------------------------------------------------------------
omega          0.0150  6.052e-03      2.474  1.335e-02 [3.112e-03,2.684e-02]
alpha[1]       0.0609  1.206e-02      5.050  4.426e-07 [3.727e-02,8.455e-02]
beta[1]        0.9383  1.223e-02     76.704      0.000     [  0.914,  0.962]
============================================================================

Covariance estimator: robust
```

Forecasted volatility was derived from a GARCH(1,1) model, implemented via the `arch` library. The rolling historical data was used for model training and forecasting.

SQL Query

The cleaned dataset was analyzed using the following SQL query:

Listing 4.1: SQL Query for Data Selection

```sql
SELECT
                date,
                expiry,
                option_type,
                strike_price,
                underlying_value,
                settle_price,
```

```
            iv AS implied_volatility,
            COUNT(*) AS total_options
        FROM
            options_data
        WHERE
            expiry IS NOT NULL
            AND underlying_value IS NOT NULL
            AND underlying_value != 'NaN'
            AND settle_price IS NOT NULL
            AND settle_price != 'NaN'
            AND iv IS NOT NULL
            AND iv != 'NaN'
            AND underlying_value ~ '^[0-9]+(\.[0-9]+)?$'
            AND settle_price ~ '^[0-9]+(\.[0-9]+)?$'
        GROUP BY
            expiry,
            option_type,
            strike_price,
            iv,
            date,
            underlying_value,
            settle_price
        ORDER BY
            TO_DATE(expiry, 'DD-Mon-YYYY') ASC,
            strike_price DESC
```

This SQL query is designed to analyze options data by aggregating and organizing relevant fields such as expiry, strike price, implied volatility, and underlying value. Below is a detailed breakdown of its functionality:

- **Selected Fields**:
  - `date`: The date of the options data record.
  - `expiry`: The expiration date of the option.
  - `option_type`: The type of option, such as Call (CE) or Put (PE).
  - `strike_price`: The strike price of the option.
  - `underlying_value`: The value of the underlying asset.
  - `settle_price`: The settlement price of the option.
  - `iv`: Implied volatility of the option.
  - `total_options`: A count of the total number of options meeting the query conditions.
- **Filters and Conditions**: The query ensures data validity and filters out records with missing or non-numeric values:
  - `expiry`, `underlying_value`, `settle_price`, and `iv` must not be `NULL` or contain `'NaN'`.
  - `underlying_value` and `settle_price` must match a numeric pattern, validated using regular expressions.
- **Aggregations and Grouping**: The query groups the data by the following fields:

    - `expiry`, `option_type`, `strike_price`, `iv`, `date`, `underlying_value`, and `settle_price`.

  Additionally, it calculates the `COUNT(*)` of options as `total_options` for each group.

- **Ordering**: The results are sorted by:
    - Expiry date in ascending order, using the `TO_DATE` function to convert the expiry field into a proper date format (`'DD-Mon-YYYY'`).
    - Strike price in descending order within each expiry group.

This query is particularly useful for preparing cleaned and aggregated options data for further analysis, such as volatility studies or options payoff calculations.

Python Packages

The following Python packages were utilized in the data preparation:

- `pandas` for data manipulation.
- `numpy` for numerical operations.
- `arch` for GARCH modeling.
- `py-lets-be-rational` for implied volatility.
- `matplotlib` for plotting.
- `psycopg2` for PostgreSQL database integration.

Conclusion

This section outlines the end-to-end process of extracting, cleaning, and preparing data for analysis. The steps ensure the dataset is robust, reliable, and ready for volatility modeling and strategy backtesting.

Chapter 5

Backtesting

5.1 Backtesting Code Implementation

The backtesting process was conducted to evaluate the effectiveness of the **Long Straddle** and **Short Straddle** strategies by using implied volatility (IV), realized volatility (RV), and the associated payoff calculations. Below, we summarize the results and key insights derived from the backtesting analysis.

Listing 5.1: SQL Query for Data Selection

```sql
SELECT
        expiry,
        strike_price,
        iv AS "IV",
        realized_vol AS "RV",
        option_type ,
        underlying_value,
        settle_price,
        RANK() OVER (
            PARTITION BY expiry
            ORDER BY iv DESC
            ) AS "IV␣Rank",
        CASE
            WHEN iv > realized_vol THEN 'Short␣Straddle'
            WHEN iv < realized_vol THEN 'Long␣Straddle'
            ELSE 'No␣Signal'
            END AS "Trade␣Signal"
    FROM options_data_updated AS main
    WHERE expiry = '{expiry_date}' and NOT EXISTS (
        SELECT 1
        FROM options_data_updated AS subquery
        WHERE subquery.strike_price = main.strike_price
          AND subquery.expiry = main.expiry
          AND (subquery.iv < 0 OR subquery.realized_vol < 0)
    )
    GROUP BY
        expiry,
        option_type,
        strike_price,
        iv,
        realized_vol, underlying_value, settle_price
    ORDER BY expiry DESC, strike_price
```

This SQL query is designed to facilitate backtesting by extracting and organizing relevant options data, with a focus on implied volatility (IV), realized volatility (RV), and their relationship. The query implements ranking and trade signal generation based on IV and RV. Below is a detailed breakdown of its components and functionality:

- **Selected Fields**:
  - `expiry`: The expiration date of the option contract.
  - `strike_price`: The strike price of the option.
  - `iv` (as `"IV"`): The implied volatility of the option.
  - `realized_vol` (as `"RV"`): The historical realized volatility of the underlying asset.
  - `option_type`: The type of option (Call or Put), represented as `"CE"` or `"PE"`.
  - `underlying_value`: The value of the underlying asset at the time of query execution.
  - `settle_price`: The settlement price of the option.
  - `RANK()` (as `"IV Rank"`): The rank of the option's implied volatility within the same expiry group, ordered in descending order.
  - `CASE` (as `"Trade Signal"`): A generated signal indicating:
    * `'Short Straddle'`: Triggered when `IV > RV`, suggesting overvalued options.
    * `'Long Straddle'`: Triggered when `IV < RV`, suggesting undervalued options.
    * `'No Signal'`: Triggered when `IV = RV`.
- **Filter Conditions**:
  - The query focuses on records with valid `expiry` and ensures no invalid or negative values for `IV` or `RV`.
  - The subquery in the `NOT EXISTS` clause filters out options with:
    * The same `strike_price` and `expiry`.
    * Invalid or negative values for `IV` or `RV`.
- **Window Function: `RANK()`**:
  - The `RANK()` function assigns a rank to each option's implied volatility within the same `expiry` group.
  - The ranking is performed in descending order of `IV`, allowing identification of the most volatile options for each expiry.
- **Trade Signal Generation**:
  - A `CASE` statement generates trade signals:
    * `'Short Straddle'` for options with overvalued IV (`IV > RV`).
    * `'Long Straddle'` for options with undervalued IV (`IV < RV`).
    * `'No Signal'` for options where `IV = RV`.

- **Grouping and Aggregation**:
  - The query groups the results by `expiry`, `option_type`, `strike_price`, `iv`, `realized_vol`, `underlying_value`, and `settle_price`.
  - This ensures that the data is aggregated at the level of individual options.
- **Sorting**:
  - The results are ordered by `expiry` in descending order and by `strike_price` in descending order within each expiry group.

This query plays a critical role in the backtesting process by providing a structured dataset for analyzing the relationship between implied volatility and realized volatility. It enables the generation of trade signals, ranking of options by IV, and the preparation of data for payoff calculations and further analysis.

The performance of the strategies was assessed based on the following key metrics:

- **Trade Signals**:
  - **Long Straddle**: Generated when IV < RV, indicating undervalued options.
  - **Short Straddle**: Triggered when IV > RV, suggesting overvalued options.
- **Payoff Calculations**:
  - **Call Options (CE)**: The payoff was calculated as:

$$\text{Payoff}_{\text{CE}} = \max(0, \text{Underlying Price} - \text{Strike Price}) - \text{Settlement Price}$$

  - **Put Options (PE)**: The payoff was determined as:

$$\text{Payoff}_{\text{PE}} = \max(0, \text{Strike Price} - \text{Underlying Price}) - \text{Settlement Price}$$

The backtesting metrics provided a comprehensive understanding of strategy performance. The **Sharpe Ratio** was calculated to evaluate risk-adjusted returns, defined as the ratio of mean payoff to the standard deviation of payoffs. Additionally, the **Mean Payoff** and **Standard Deviation of Payoff** offered insights into the profitability and variability of the strategies.

The results revealed that the **Long Straddle** strategy performed well in scenarios where realized volatility exceeded implied volatility (RV > IV), suggesting underestimated market volatility. On the other hand, the **Short Straddle** strategy demonstrated effectiveness in stable markets where implied volatility was higher than realized volatility (IV > RV), indicating overvalued options. However, Short Straddle strategies were more prone to significant losses due to unexpected market movements.

The visualizations of **Payoff vs. Strike Price** illustrated the variability in payoffs across strike prices, categorized by IV rank. These graphs highlighted the most profitable strike levels for each strategy and were supplemented with key statistics, such as Sharpe Ratio, Mean Payoff, and Standard Deviation of Payoff, for deeper context. Additionally, a bar chart compared the total payoffs of Long and Short Straddle strategies, emphasizing their relative profitability.

While Long Straddle strategies demonstrated consistent payoffs with manageable risks, the Short Straddle strategies exhibited higher payoff variability, highlighting the risk-reward tradeoff inherent to the latter. Furthermore, analysis of IV ranks revealed that options with higher IV ranks often presented better opportunities for Short Straddle strategies due to their tendency to be overpriced.

The analysis assumed a simulated 10% increase in underlying prices at expiry, which may differ from real-world price movements. Transaction costs and liquidity constraints were not considered, potentially impacting the results in practical applications. Additionally, the backtesting results relied heavily on historical data quality, which may not fully represent future market conditions.

Expiry Dates (2015–2024)

Year	Number of Expiry Dates
2015	12
2016	12
2017	12
2018	12
2019	49
2020	53
2021	52
2022	52
2023	52
2024	52

Table 5.1: Number of Expiry Dates by Year

In options trading, every year comprises a predefined set of expiry dates, which are critical points for the settlement of options contracts. These expiry dates can be grouped into monthly expirations, which occur on the last Thursday of every month, and weekly expirations, which fall on each Thursday of the week, excluding those overlapping with monthly expirations. The table labeled "Table 5.1" highlights the total number of expiry dates for each year from 2015 to 2024. The data demonstrates a steady increase in the number of expiry dates due to the addition of weekly expirations, especially in more recent years, as the options market evolved to offer greater flexibility and liquidity.

For this project, each expiry date is associated with a variety of strike prices across call (CE) and put (PE) options. By organizing and grouping data based on expiry dates, it becomes easier to analyze volatility, implied volatility (IV), realized volatility (RV), and the associated trade signals for each strike price. Grouping expiry dates enables the evaluation of options strategies, such as straddles and strangles, on a weekly or monthly basis, allowing for better risk management and more precise backtesting of strategies.

In the subsequent "Table 5.2," a slice of the year 2023 is presented. This table lists all the weekly and monthly expiry dates for that year, representing a total of 52 expirations. Each of these dates includes data on various strike prices and corresponding metrics,

Year	Expiry Dates
2023	05-Jan-2023, 12-Jan-2023, 19-Jan-2023, 25-Jan-2023, 02-Feb-2023, 09-Feb-2023, 16-Feb-2023, 23-Feb-2023, 02-Mar-2023, 09-Mar-2023, 16-Mar-2023, 23-Mar-2023, 29-Mar-2023, 06-Apr-2023, 13-Apr-2023, 20-Apr-2023, 27-Apr-2023, 04-May-2023, 11-May-2023, 18-May-2023, 25-May-2023, 01-Jun-2023, 08-Jun-2023, 15-Jun-2023, 22-Jun-2023, 28-Jun-2023, 06-Jul-2023, 13-Jul-2023, 20-Jul-2023, 27-Jul-2023, 03-Aug-2023, 10-Aug-2023, 17-Aug-2023, 24-Aug-2023, 31-Aug-2023, 07-Sep-2023, 14-Sep-2023, 21-Sep-2023, 28-Sep-2023, 05-Oct-2023, 12-Oct-2023, 19-Oct-2023, 26-Oct-2023, 02-Nov-2023, 09-Nov-2023, 16-Nov-2023, 23-Nov-2023, 30-Nov-2023, 07-Dec-2023, 14-Dec-2023, 21-Dec-2023, 28-Dec-2023

Table 5.2: Expiry Dates for the Year 2023

forming the foundation for detailed analysis and strategy testing. The weekly expirations ensure frequent opportunities for short-term strategies, while the monthly expirations cater to long-term trading plans. Together, these expiry dates provide a robust dataset for evaluating the performance and reliability of options trading strategies.

The significance of this organization lies in its ability to systematically assess the profitability of strategies under different market conditions. By associating each expiry date with its respective strike prices, the project facilitates comparisons across different timeframes, volatility environments, and risk profiles. This grouping of expiry dates forms the backbone of the backtesting and trading model, ensuring comprehensive coverage of the options market.

The backtest code takes a list of expiry dates for a specific year, such as 2023, and iterates through each expiry date in a loop. For each date, the code retrieves the corresponding call and put options, along with various strike prices and other essential parameters, from the backend database. Using the query designed to rank options based on the magnitude of implied volatility, the options are categorized as either Short Straddle or Long Straddle based on the following condition: if the implied volatility exceeds the realized volatility, the option is labeled as a Short Straddle; otherwise, it is labeled as a Long Straddle.

The backtest code leverages the strike price, underlying asset price, and settlement price (in place of bid/ask prices). The settlement price is used because the historical data from the NSE exchange does not provide bid/ask prices. The payoff for each strike price is then calculated systematically. This approach allows for effective testing and evaluation of strategies over the historical data.

In conclusion, the backtesting process validated that the **Long Straddle** strategy is optimal in environments with underestimated volatility, while the **Short Straddle** strategy can be effective in stable markets with overestimated volatility, albeit at a higher risk. These findings provide actionable insights for refining trading strategies and implementing them in live trading scenarios.

Chapter 6

Results

6.1 Option Chain Table for Expiry: 30-May-2024

The option chain table below represents the data for the expiry date 30-May-2024. It includes detailed information about Call (CE) and Put (PE) options, showcasing their implied and realized volatilities, settlement prices, and corresponding strike prices. Similar computations of implied and realized volatilities, along with other attributes, have been performed for all other grouped expiry dates in the database. The results are systematically stored back in the database for further analysis. The table below is an example randomly selected from the dataset to illustrate the calculated volatilities and related metrics.

The tabular column represents the **option chain data** for the expiry date **30-May-2024**. It contains essential details about the Call (CE) and Put (PE) options, including their implied and realized volatilities, settlement prices, and strike prices.

The **Implied Volatility (IV)** column, expressed as a percentage, represents the market's forecast of the option's future volatility. For instance, for Calls with a strike price of 20300, the implied volatility is 11.48%, while for Puts, it is 12.35%. The **Realized Volatility (RV)** column reflects the historical volatility of the underlying asset, which is critical for comparing against implied volatility to identify undervalued or overvalued options. For example, for the strike price of 20550, the realized volatility is 9.93% for Calls and 10.70% for Puts.

The **Settlement Price** column provides the closing price of the option contract at the end of the trading day or upon expiration. For example, the settlement price for Calls with a strike price of 20300 is 2391.25, while for Puts, it is 17.4. The **Strike Price** column lists the pre-determined price at which the underlying asset can be bought (for Calls) or sold (for Puts). The table covers strike prices ranging from 20300 to 20750 for both Calls and Puts.

Overall, the table enables a direct comparison between Calls and Puts for each strike price, offering insights into market expectations and volatility. Notably, implied volatility is generally higher for Puts compared to Calls, indicating a potential bearish sentiment in the market. The settlement prices vary significantly between Calls and Puts, reflecting the market's valuation of options at different strike prices. This table is an essential input for analyzing strategies such as Long and Short Straddles by leveraging the implied and realized volatilities alongside settlement prices to compute payoffs.

IV Rank	Option Type	Strike Price	IV	RV	Trade Signal
278	PE	23600	0.0872	0.12823	Long Straddle
277	PE	23700	0.0906	0.08114	Short Straddle
276	PE	24000	0.0914	0.06486	Short Straddle
275	PE	23500	0.0927	0.151742	Long Straddle
274	PE	23400	0.0970	0.18131	Long Straddle
273	PE	23450	0.0972	0.17203	Long Straddle
272	PE	23350	0.0977	0.18131	Long Straddle
271	PE	23300	0.0981	0.17203	Long Straddle
270	CE	23600	0.0999	0.11472	Long Straddle
269	CE	23650	0.1001	0.09932	Short Straddle
268	CE	23700	0.1003	0.09932	Short Straddle
267	CE	23550	0.1005	0.12823	Long Straddle
266	CE	23800	0.1008	0.06486	Short Straddle
265	CE	23750	0.1008	0.05738	Short Straddle
264	PE	23250	0.1013	0.17209	Long Straddle
263	CE	23500	0.1015	0.16222	Long Straddle
262	CE	23450	0.1015	0.17203	Long Straddle
261	CE	23400	0.1021	0.18131	Long Straddle
260	PE	23200	0.1024	0.17203	Long Straddle
259	CE	23850	0.1026	0.06486	Short Straddle

Significance of IV Rank and Explanation of the Table

Significance of IV Rank

The **IV Rank** (Implied Volatility Rank) is a critical metric that compares the current implied volatility (IV) of an option to its historical range. It allows traders to evaluate whether the current IV is high or low relative to past levels, aiding in strategic decision-making for option trading. A higher IV Rank indicates that the IV is elevated compared to its historical levels, suggesting potential overvaluation of options. This typically aligns with premium-selling strategies like *Short Straddles*. Conversely, a lower IV Rank signifies that the IV is at relatively low levels, favoring premium-buying strategies such as *Long Straddles*.

Its Importance

The table provides an overview of options data categorized by their **IV Rank**, **Option Type** (Call or Put), **Strike Price**, and their respective values of **IV** (Implied Volatility), **RV** (Realized Volatility), and the generated **Trade Signal**. Below is a detailed explanation of the columns in the table:

- **IV Rank:** This column ranks the options based on their implied volatility magnitudes, with higher ranks representing options with relatively higher implied volatility compared to others.
- **Option Type:** Options are classified as **Call Options (CE)** or **Put Options (PE)**. Call options give the right to buy, while put options give the right to sell the underlying asset at the specified strike price.
- **Strike Price:** The strike price specifies the price at which the option holder can execute their right to buy (calls) or sell (puts) the underlying asset.
- **IV and RV:**
  - **IV (Implied Volatility):** Represents the market's expectation of future price fluctuations of the underlying asset.
  - **RV (Realized Volatility):** Reflects the actual historical volatility of the underlying asset based on past data.

  The relationship between IV and RV determines the trade signal:

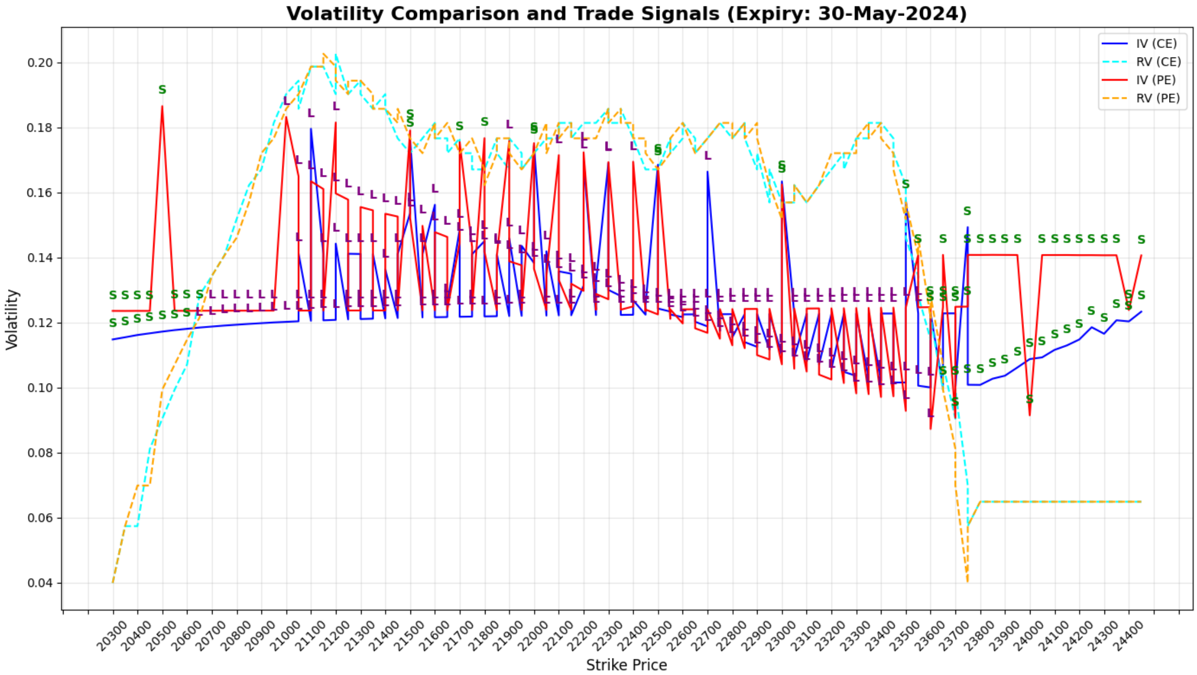
  - When **IV ¿ RV**, the option is labeled as **Short Straddle**, indicating that the implied volatility may revert to lower levels.
  - When **IV ¡ RV**, the option is labeled as **Long Straddle**, suggesting potential undervaluation of options.
- **Trade Signal:** This column provides the trade strategy recommendation for each option based on the comparison of IV and RV. Options are classified as either *Short Straddle* or *Long Straddle*.

Summary of IV Rank

The **IV Rank** organizes options by their implied volatility levels, allowing traders to assess their relative valuation:

- Options with **higher IV Ranks** are likely to have overvalued premiums, making them suitable for *Short Straddle* strategies.
- Options with **lower IV Ranks** may represent undervalued opportunities, favoring *Long Straddle* strategies.

This table serves as a foundational input for backtesting and strategy evaluation. By providing a clear ranking of options based on volatility metrics, it helps traders implement appropriate trading strategies with confidence.



6.2 Volatility Comparison and Signal Generation

The chart titled "Volatility Comparison and Trade Signals (Expiry: 30-May-2024)" presents a detailed analysis of the implied volatility (IV) and realized volatility (RV) for call options (CE) and put options (PE) across various strike prices. This visualization serves as a comprehensive representation of volatility dynamics and associated trade signals for the given expiry date.

The x-axis represents the **strike price**, ranging from 20,300 to 24,400, while the y-axis denotes the volatility values. The plotted lines differentiate between IV and RV for call and put options. Specifically:

- The **blue line** depicts the **IV (CE)**, while the **cyan dashed line** represents the **RV (CE)**.
- The **red line** shows the **IV (PE)**, and the **orange dashed line** corresponds to the **RV (PE)**.

Trade signals are annotated on the graph based on the comparison between IV and RV:

- **Long Straddle (L)** signals are generated when IV is less than RV, indicating that the options are undervalued and suggest a potential buying opportunity.
- **Short Straddle (S)** signals occur when IV exceeds RV, indicating overvalued options and suggesting a selling strategy.

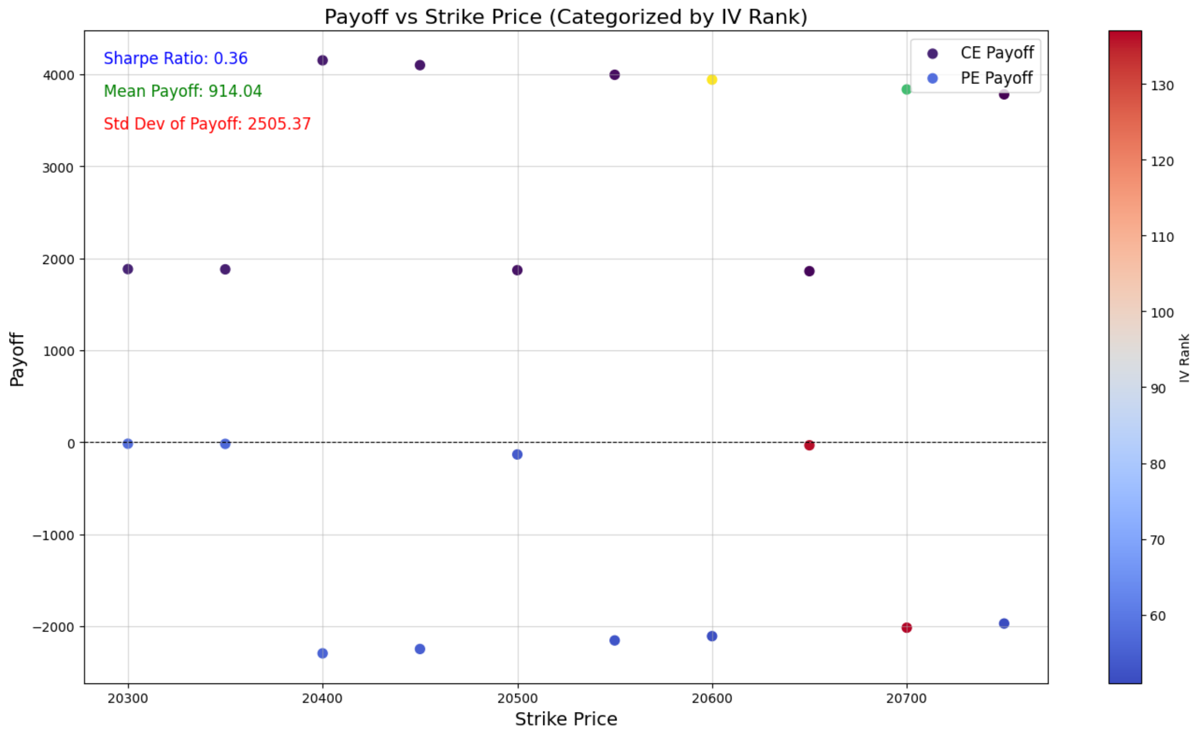
Key observations include:

- At lower strike prices, the trade signals are predominantly **S (Short Straddle)**, reflecting higher IV compared to RV.
- As the strike price increases, **L (Long Straddle)** signals become more frequent, highlighting areas where RV exceeds IV.
- Significant variations in IV and RV for both CE and PE are observed, particularly at extreme strike prices, where the volatility curves diverge.

The annotations for trade signals (**L** and **S**) are placed along the respective curves, providing a clear indication of actionable strategies at different strike prices. The chart effectively conveys the relationship between IV, RV, and trade signals, enabling informed decision-making for options trading strategies.

This visualization underscores the importance of analyzing volatility metrics and their implications on straddle strategies, thereby serving as a crucial component of the back-testing framework for evaluating trading performance.

6.3 Strike Price vs PayOffs vs Ranking the Expriy Group



Payoff vs Strike Price Categorized by IV Rank

The chart titled **"Payoff vs Strike Price (Categorized by IV Rank)"** provides a detailed visualization of the relationship between strike prices, payoffs, and their categorization based on the IV (Implied Volatility) rank. This analysis combines critical metrics such as Sharpe Ratio, Mean Payoff, and Standard Deviation of Payoff to offer a comprehensive evaluation of trading strategies.

Axes and Data Points:

- The **x-axis** represents the **strike price**, ranging from 20,300 to 20,700.
- The **y-axis** depicts the **payoff**, covering both positive and negative values.
- Each data point corresponds to the payoff for a particular strike price, categorized as either **Call Option (CE)** or **Put Option (PE)**.
- The color intensity of each point, as indicated by the color bar on the right, represents the IV rank. Higher IV ranks are shown in red, while lower ranks are displayed in blue.

Key Features:

- **Call Option Payoff (CE)**: Represented by yellow and purple markers, showing the variation in payoffs for calls across strike prices. Higher IV ranks correlate with higher payoffs, emphasizing the predictive strength of IV rankings.
- **Put Option Payoff (PE)**: Represented by green and blue markers, displaying the payoff for puts. Lower IV ranks tend to cluster around strike prices with minimal payoffs.
- **Annotations for Metrics:**
  - **Sharpe Ratio:** Displayed in blue at the top left, indicating the risk-adjusted return of the strategy. In this chart, the Sharpe Ratio is 0.36, reflecting moderate performance.
  - **Mean Payoff:** Shown in green, the mean payoff value is 914.04, highlighting the average profitability of the strategy.
  - **Standard Deviation of Payoff:** Indicated in red, the standard deviation is 2505.37, providing insights into the payoff's volatility.

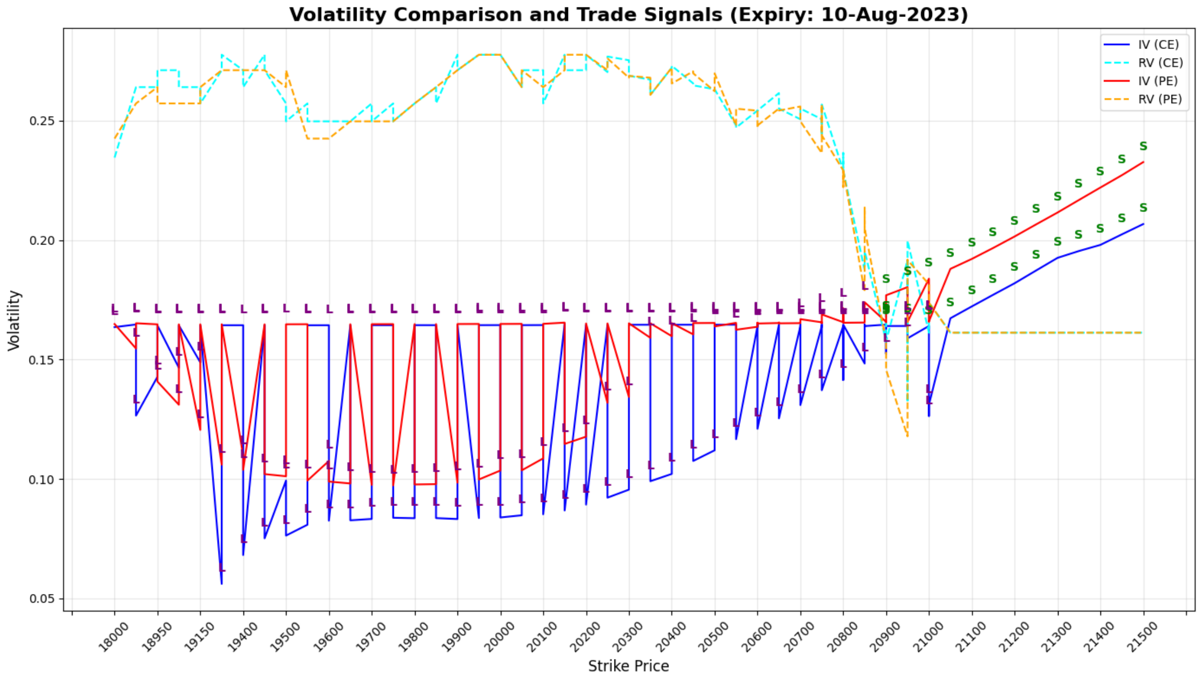
Insights and Observations:

- Strike prices with higher IV ranks (red markers) generally exhibit higher payoffs, underscoring the effectiveness of IV ranking as a selection criterion.
- The distribution of call and put payoffs varies significantly, with calls achieving higher payoffs at specific strike prices, as reflected in the clustering of markers at the upper payoff range.
- Negative payoffs are observed at lower IV ranks, suggesting that options with lower IV ranks are less profitable or potentially loss-making.

- The diversity in payoffs and IV ranks across strike prices highlights the need for precise strategy selection based on market conditions and volatility.

This visualization integrates multiple analytical dimensions, including payoff distribution, IV ranking, and strategy performance metrics, making it an essential tool for evaluating and refining trading strategies.

6.4 OverAll Analysis:

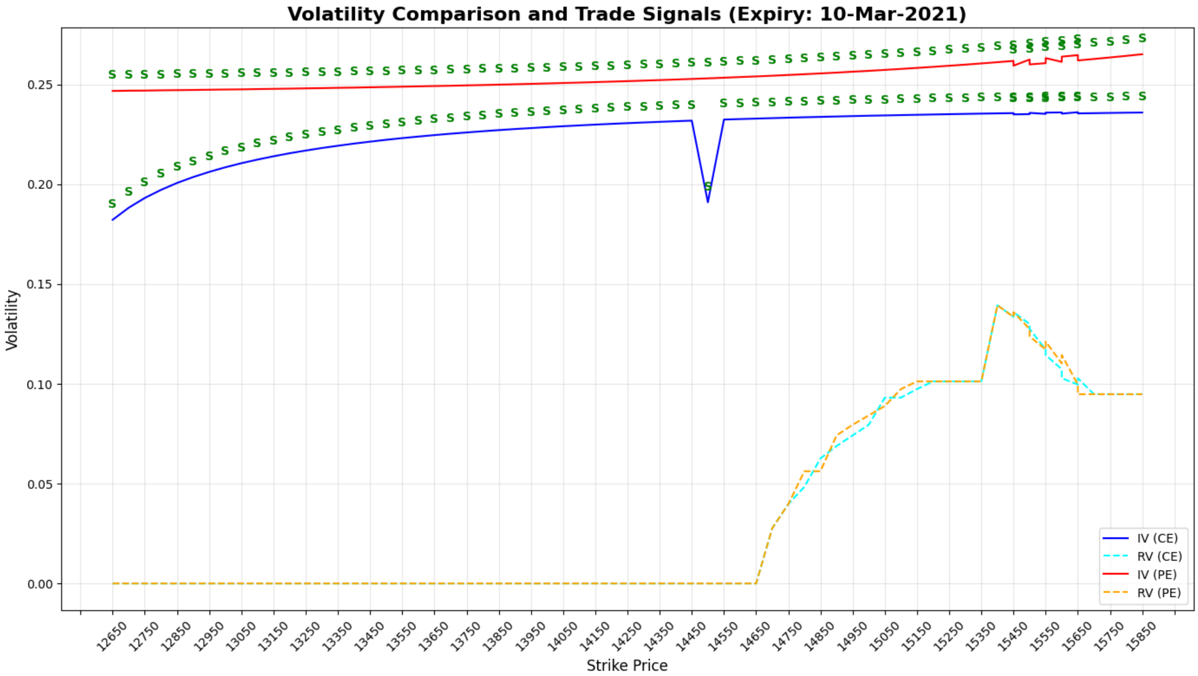


- **File:** `options_chain_table_10_Aug_2023.png`
  - **Description:** The plot showcases the volatility comparison for the expiry date of 10-Aug-2023. The IV and RV curves for both CE and PE options are plotted against the strike prices. Trade signals are represented with "S" (Short Straddle) and "L" (Long Straddle). The IV values exceed RV values in most regions, resulting in a higher concentration of Short Straddle signals, especially at higher strike prices.

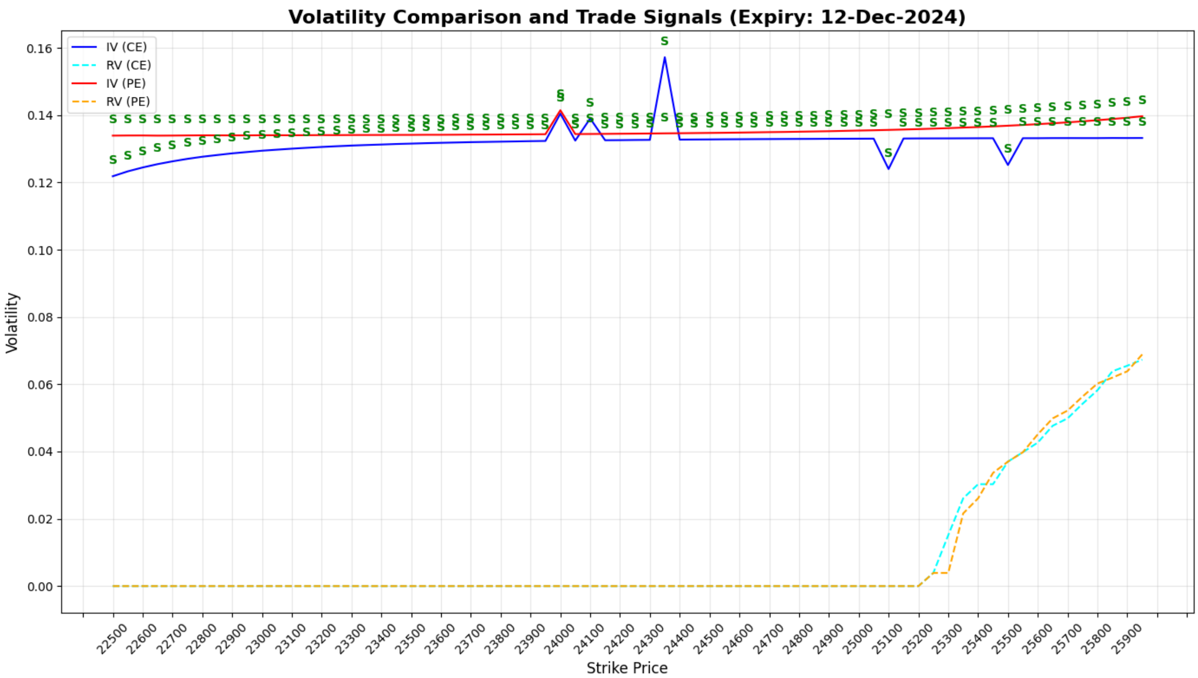
- **File:** `options_chain_table_10_Mar_2021.png`
  - **Description:** This plot illustrates the volatility comparison for the expiry date of 10-Mar-2021. The IV and RV curves for CE and PE options show a stable trend, with IV consistently higher than RV across most strike prices. The trade signals predominantly indicate Short Straddle strategies.

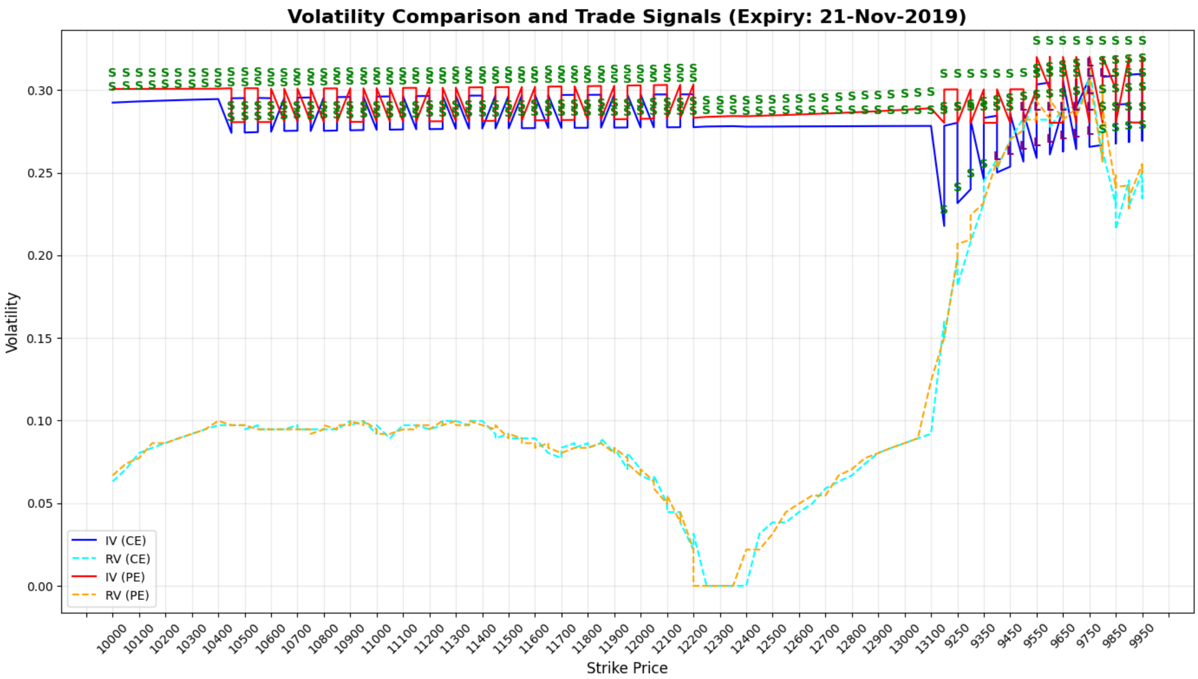
- **File:** `options_chain_table_12_Dec_2024.png`
  - **Description:** For the expiry date of 12-Dec-2024, the plot demonstrates a smooth trend with IV values slightly exceeding RV for both CE and PE options

across strike prices. This consistency is reflected in the uniform Short Straddle trade signals across all strike prices.

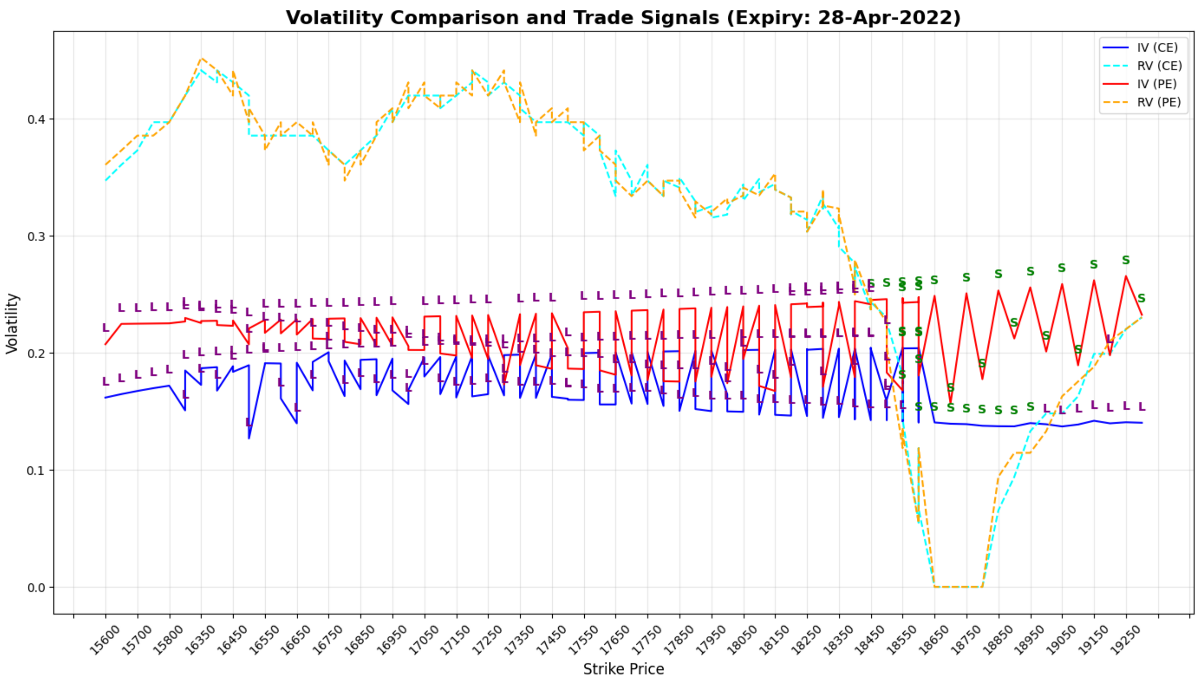


- **File:** `options_chain_table_12_Dec_2024.png`
  - **Description:** For the expiry date of 12-Dec-2024, the plot demonstrates a smooth trend with IV values slightly exceeding RV for both CE and PE options across strike prices. This consistency is reflected in the uniform Short Straddle trade signals across all strike prices.





- **File:** `options_chain_table_21_Nov_2019.png`
  - **Description:** The expiry date of 21-Nov-2019 shows significant IV variation compared to RV, especially for lower strike prices. The trade signals are predominantly Short Straddle for higher strike prices, with a clear distinction between IV and RV.



- **File:** `options_chain_table_28_Apr_2022.png`

– **Description:** The plot for 28-Apr-2022 presents a more dynamic comparison between IV and RV. For CE options, RV surpasses IV at certain strike prices, leading to a mix of Short and Long Straddle trade signals. The PE options primarily show Short Straddle signals due to IV dominating RV.

6.5 Causes for Data Integrity Issues in the NSE Dataset

6.5.1 Incomplete Records

Many entries in the dataset were missing critical fields such as settlement prices, implied volatilities, or realized volatilities. This lack of completeness significantly hindered accurate and reliable calculations for analysis.

6.5.2 Historical Data Limitations (2015–2021)

Older data, particularly from 2015 to 2021, exhibited severe inconsistencies. It is likely that data collection practices during this period were less robust or systematic, leading to gaps in historical records.

6.5.3 Formatting Errors

The dataset contained improperly formatted fields, such as non-numeric values in numeric columns (e.g., `NaN` or erroneous strings). These formatting issues required significant preprocessing and cleaning, often resulting in the removal of potentially useful records.

6.5.4 Free Data Constraints

Since the dataset was obtained from free third-party sources, it lacked the level of rigor and validation typically provided by paid or premium services. The absence of quality checks or comprehensive coverage contributed to the dataset's unreliability.

6.5.5 Data Duplication and Redundancy

Some records contained redundant or duplicate entries, further complicating the cleaning process. Ensuring that only unique and relevant data points were included introduced additional challenges.

6.5.6 Unavailability of Bid/Ask Prices

The absence of bid/ask prices in the historical dataset, particularly for options data, forced reliance on settlement prices. This limitation likely distorted payoff calculations and reduced the accuracy of derived parameters such as implied and realized volatilities.

6.5.7 Gaps in Data Collection Processes

The dataset may have been generated using incomplete data feeds or interrupted data pipelines, leading to missing options series or inconsistencies across strike prices and expiry dates.

6.5.8 Impact of Market Anomalies

Certain market events, such as low liquidity or abrupt volatility spikes, might have caused anomalies within the dataset, resulting in missing or unreliable entries for specific time periods.

6.6 Mitigation Strategies

- **Leverage Premium Data Sources:** Using paid data providers with rigorous validation processes can significantly improve data quality and coverage.
- **Automated Data Validation:** Implementing automated checks during data extraction and cleaning to identify and address missing or inconsistent values more effectively.
- **Interpolation and Estimation:** For minor gaps, statistical methods could be employed to estimate missing values, provided it does not compromise the analysis's integrity.
- **Exclude Problematic Data Periods:** If specific periods, such as 2015–2021, are particularly unreliable, these could be excluded from the analysis to avoid biasing the results.

Addressing these underlying causes would significantly enhance the reliability of datasets used for financial modeling and backtesting.

Chapter 7

Conclusion

7.1 Conclusion

The analysis and modeling efforts were significantly hindered by the data integrity issues inherent in the dataset obtained from the NSE website. Despite implementing extensive cleaning and preprocessing steps, the lack of completeness and consistency in the historical records, particularly from 2015 to 2021, posed substantial challenges. These issues resulted in unreliable derivative parameters such as implied, realized, and forecast volatilities, which are critical for financial modeling and strategy evaluation.

Furthermore, the violation of key model assumptions, such as the stationarity constraint in volatility modeling where the sum of parameters α, β, and ω exceeded 1, rendered the forecasting process ineffective. The optimization process frequently encountered zero or `NaN` values, further impeding convergence and limiting the reliability of the results.

These challenges highlight the importance of high-quality, reliable datasets for robust financial analysis. While the strategy derived from academic research showed potential, its implementation was undermined by the limitations of the available data. Future endeavors should prioritize leveraging premium data sources, improving data validation processes, and considering alternative modeling techniques that are more resilient to incomplete or noisy datasets. Only with reliable data and sound methodologies can accurate and meaningful financial insights be achieved.