Algorithm Trading- Buy/Sell Signal Generator

-By Aryan, Yuvraj, Shatakshi, Shubham

Documentation

Objective: This project investigates an intraday trading approach that blends insights from both daily and real-time (intraday) data. It utilizes a GARCH (Generalized Autoregressive Conditional Heteroskedasticity) model, trained on historical daily data, to **forecast volatility** and identify the daily trend. This volatility prediction is then combined with technical indicators like RSI and Bollinger Bands applied to intraday data to **generate buy and sell signals**. Stock prices are highly volatile in nature, so our aim is to model the volatility in their prices and test a custom-made trading strategy to see how it performs. The performance of the strategy is indicated by the returns it yields over time. A decent return % will indicate that the strategy is working well. It must be noted that this is **not the only strategy that exists** or can be used. There may be different strategies that make use of other indicators.

Data: The data was procured from **yfinance.** The data used in the model is of daily and intraday nature. Initially, our dataset has the following: the opening price of the stock, the closing price of the stock, the highest and lowest price of the stock in a given time frame, the adjusted closing price, the volume of stocks traded and an unnamed column.

Data Preparation:

- The Unnamed column was dropped and we checked for any missing values. The data has 0 missing values.
- The intraday data has 177877 observations whereas the daily day has 3289 observations.
- The variable of interest in the GARCH model is log(yt/yt-1) = log(yt) - log(yt-1), the logarithm of the ratio of this time's value to last time's value. **Why choose log returns?**
  1) Log returns are **more symmetric** as they handle large positive and negative returns more consistently than normal returns.
  2) Log returns are **more stable**. They tend to be less volatile over time compared to normal returns.
  3) Log returns tend **towards normality**. While not perfectly normal, log returns often exhibit a distribution closer to a normal bell curve than normal returns. This is essential for many statistical models and tests that rely on normality assumptions. **Normal returns**, on the other hand, **are often skewed**, especially for large price movements, violating normality assumptions.

Testing for assumption: Before actually fitting the model, we have to first check if our series, on which we'll fit the model, is stationary or not. If not, we have to first make that series stationary and then find the parameters of the GARCH model. Augmented Dickey-Fuller Test and KPSS Test for Stationarity were used. *Results:* The Series is Stationary.

Why the GARCH model?
Before we delve into the nitty-gritty of the model, it is imperative that we clarify why the Garch model is the best in our case.
GARCH processes, because they are autoregressive, depend on past squared observations and past variances to model for current variance and are widely used in finance due to their effectiveness in modeling asset returns and inflation. GARCH aims to minimize errors in forecasting by accounting for errors in prior forecasting and enhancing the accuracy of ongoing predictions.
In the period following the onset of a crisis, however, returns may swing wildly from negative to positive territory. Moreover, the increased volatility may be predictive of volatility going forward. Volatility may then return to levels resembling that of pre-crisis levels or be more uniform going forward. A simple regression model does not account for this variation in volatility exhibited in financial markets and thus, the GARCH model is the most appropriate in such cases.

Obtaining the best parameters for the GARCH Model: To automate the process of computing the values of the parameters of the GARCH model, we'll use the 'auto_arima' function from the 'pmdarima' library and find the best set of parameters that will fit the series.

Fitting the Model:

- We will employ a 180-day rolling window to fit the GARCH model. This will allow us to capture evolving market dynamics. Why choose a 180-day rolling window? In finance, 180 days is often used for various calculations, such as interest rates, bond yields, and derivatives pricing. It is a standard convention in the industry.
- Using the fitted model, we can then forecast the volatility for the next trading day.

Computing daily signals: The daily signal will be used to examine the daily upward or downward trend. For that, we'll calculate the actual variance of the 'log_ret' column on the 180-day rolling window. Previously, we calculated the predicted variance and this time it's the the actual variance. We'll use the predicted variance and actual variance to find the 'premium'. Premium is nothing but the change in predicted variance from the actual variance and is used to indicate the direction of volatility. Now, the question arises, why choose ‘premium’? We chose **premium** as it **determines the direction, magnitude and volatility at the same time.**

We then compare the premium of each day with its standard deviation which is again calculated on a 180-day rolling window. Comparisons are made between premium and standard deviation as they have the same unit.

Based on the comparison between the premium and its standard deviation:

- *Long Position (Signal = 1):* If the premium is significantly higher than its standard deviation, it suggests the market is experiencing higher-than-usual volatility. This might indicate an upward trend, and we assign a "1" signal for a potential long position.
- *Short Position (Signal = -1):* Conversely, if the premium is significantly lower than its standard deviation, it suggests lower-than-expected volatility. This could be a sign of a downtrend, and we assign a "-1" signal for a potential short position.
- *Hold Position (Signal = 0):* If the premium falls within the range of its standard deviation, it indicates the market volatility is behaving as predicted by the model. In this case, we assign a "0" signal, signifying a hold position and no action is taken.

To avoid overtrading, we have introduced the hold position as well. The daily signals generated represent our prediction for the market movement on the next day. However, these signals are currently aligned with the day they were calculated. To evaluate the effectiveness of these predictions, we need to shift them forward by one day so they are aligned with the actual price movements they were designed to forecast.

Aligning Daily Signals with Intraday Data:

- *Resampling Daily Data:* We will convert the daily data (which likely has one data point per day) into a higher frequency format, typically with timestamps every 5 minutes. This process, called "resampling," allows us to match the daily signals with the intraday data.
- *Assigning Daily Signal to Intraday Data:* Since the daily signal represents the predicted movement for the entire day, we will assign the same daily signal value (1, -1, or 0) to all the 5-minute time points within that particular day. This essentially replicates the daily prediction throughout the day.

Computing Intraday signals: In our strategy, we will utilize two technical indicators, the Relative Strength Index (RSI) and Bollinger Bands, to generate intraday signals that complement the daily prediction. Here's the logic:

- *Long Signal (1):* If the closing price for a 5-minute interval is higher than the upper Bollinger Band and the RSI value is above 70, this suggests a strong upward trend. We assign a "1" signal for a potential buying opportunity.
- *Short Signal (-1):* Conversely, if the closing price falls below the lower Bollinger Band and the RSI is below 30, it might indicate a downtrend. We assign a "-1" signal for a potential short-selling opportunity.
- *Hold (0):* In any other scenario where the price and RSI values don't meet the criteria for long or short signals, we assign a "0" signal, suggesting holding the current position and taking no action.

Here, we have used RSI and Bollinger bands RSI is a leading indicator and Bollinger band is a lagging indicator, so their combination provides a fairly accurate indication. Another reason behind choosing them is that they are quantifiable and have set industry standards, so rest assured, they are not based on any random assumptions

Final Signal Generation: Once we have generated the daily and intraday signals, now it's time to merge daily and intraday signals to obtain the final "return signal" for trading decisions.

- If daily and intraday signals agree (short-short or long-long), the return signal inherits the common direction (short or long).
- If signals disagree, the return signal is marked as "unclear" (not a number).

The final return signals are stored in a new column "return_signal" in the data frame.

Effectiveness of the Strategy: To assess the effectiveness of our trading strategy, we calculate the daily strategy return. This is achieved by multiplying the daily return (price change) with the corresponding "return signal" (1 for long, -1 for short, 0 for hold). This calculation essentially reflects the profit or loss for each day based on the signals generated by our strategy.

Notes:
1) This strategy focuses on generating a single buy or sell signal for each day based on the combined daily and intraday signals. **Once a signal is triggered (long or short), the position is held throughout the rest of the trading day.** This approach avoids frequent intraday trading and aims to capitalize on the predicted market direction for the entire day.
2) Our model doesn't take data to train the model, instead it **is a strategy testing model, so backtesting is the only viable approach**. Also, backtesting prevents overfitting unlike the traditional train-test split approach when working with time-based data. Instead of randomizing the data, backtesting sequentially increases the size of the training set while maintaining the temporal order of the data. Backtesting applies a model to historical data that wasn't used for training. This exposes the model to market conditions it hasn't seen before and thus, prevents overfitting.
3) In prediction models, we use rmse, mae, etc. as evaluation metrics while in classification models, we use recall, f1 score, precision, etc. as the evaluation metrics, however, our model is not a traditional prediction or classification model so using these evaluation metrics won't be appropriate in our case, the best metric in such models is the return percentage which denotes how well the strategy is working.
4) Common Backtesting Measures:
1. Net Profit/Loss
2. Return: The total return of the portfolio over a given time frame
3. Risk-Adjusted Return: The return of the portfolio adjusted for a level of risk
4. Market Exposure: the degree of exposure to different segments of the market
5. Volatility: The dispersion of returns on the portfolio

Glossary

Augmented Dickey-Fuller Test for Stationarity (ADF Test): The ADF test is a statistical hypothesis test used to determine if a time series is stationary or not. A stationary time series has no trend and exhibits a constant mean and variance over time and a constant autocorrelation (dependence between past and present values) structure over time.
Statistically, The Augmented Dickey-Fuller test can be used to test for a unit root in a univariate process in the presence of serial correlation.

Hypotheses:

H0: The time series is non-stationary. In other words, it has a time-dependent structure and does not have constant variance over time.
HA: The time series is stationary.

Decision Criteria:

If the p-value from the test is less than some significance level (e.g. $\alpha = .05$), then we can reject the null hypothesis and conclude that the time series is stationary.

Auto Arima: Auto ARIMA models usually employ a grid search or stepwise search methodology to identify the optimal ARIMA model structure (p, d, q) based on metrics like AIC or BIC. When to use AIC and BIC?
AIC is often preferred when the primary goal is prediction accuracy.
BIC is more suitable when the goal is to find the true model that generated the data.
AIC and BIC help prevent overfitting and underfitting.
They provide a quantitative basis for comparing different models.
The choice between AIC and BIC depends on the specific goals of the analysis.
Lower AIC and BIC values generally indicate better models.

Once the structure is determined, standard optimization techniques like Maximum Likelihood Estimation (MLE) or Conditional Least Squares are used to estimate the model parameters.

Bollinger Bands: Bollinger Bands are a technical analysis volatility indicator that helps gauge the volatility of stocks and other securities to determine if they are over- or undervalued. These use a moving average and standard deviation to create bands around the price chart. These bands are composed of three lines: a simple moving average (the middle band) and an upper and lower band. The bands widen when a stock's price becomes more volatile and contract when it is more stable. The upper and lower bands are typically two standard deviations above or below a 20-period simple moving average (SMA). Many traders see stocks as overbought as their price nears the upper band and oversold as they approach the lower band, signaling an opportune time to trade.

Daily signal: A daily signal refers to a trading indicator or strategy that generates buy or sell recommendations based on data from the previous day's closing price or other daily metrics.

Forward return: Forward return refers to the expected or estimated return on an investment over a future period. In intraday trading, it represents the projected return on a trade based on the current price and the trader's exit strategy.

GARCH Model: The Generalized AutoRegressive Conditional Heteroskedasticity or simply, the GARCH model is a statistical modeling technique used to help predict the volatility of returns on financial assets. It is usually used to estimate the volatility of returns for strokes, bonds and market indices. GARCH is appropriate for time series data where the variance of the error term is serially autocorrelated following an autoregressive moving average process. GARCH is useful to assess risk and expected returns for assets that exhibit clustered periods of volatility in returns.

Intraday signal: An intraday signal is a trading indicator or strategy that generates buy or sell recommendations based on real-time market data throughout the trading day.

Intraday Trading: The practice of buying and selling stocks within a single trading day

Kwiatkowski–Phillips–Schmidt–Shin (KPSS) Test for Stationarity (KPSS Test): Like the ADF test, the KPSS test is also a statistical hypothesis test used to check for stationarity of a series around a deterministic trend.
Hypotheses:
H0: The time series is trend stationary. HA: The time series is *not* trend stationary.
Decision Criteria:
If the p-value of the test is less than some significance level (e.g. $\alpha = .05$) then we reject the null hypothesis and conclude that the time series is not trend stationary.

Log Returns: Logarithmic returns or Log returns represent the percentage change of a stock price on a logarithmic scale. The advantage of using them is that they are additive over time.
Formula: *Log Return (t) = ln(P(t) / P(t-1))* where P(t) = Stock price at time t and P(t-1) = Stock price at the previous time point (e.g., one day earlier).

LT.NS: ticker for the stocks of Larsen and Toubro, listed on the National Stock Exchange (NSE).

Premium: Here, the premium is used to refer to the variance risk premium calculated as follows:
premium = (predicted volatility - variance) / variance
VRP represents the difference between the implied volatility (derived from stock prices) and the expected future realized volatility of the stock prices.

Relative Strength Index (RSI): The relative strength index (RSI) is a popular momentum indicator (i.e. a technical analysis tool used to determine the strength or weakness of a stock's price trend. Momentum measures the rate of the rise or fall of stock prices). The RSI provides technical traders with signals about bullish (i.e. expectation that prices will rise in the near future) and bearish (i.e. expectation that prices will fall in the near future) price momentum, and it is often plotted beneath the graph of an asset's price. An asset is usually considered overbought when the RSI is above 70 and oversold when it is below 30. The RSI line crossing below the overbought line or above the oversold line is often seen by traders as a signal to buy or sell.

Strategy return: Strategy return refers to the actual return achieved by a specific trading strategy over a particular period. It's calculated by comparing the entry and exit prices of trades executed using that strategy.

References

https://online.stat.psu.edu/stat510/lesson/11/11.1
https://www.linkedin.com/pulse/trade-volatility-using-garch-model-quantra-classroom-goquantra/
https://www.linkedin.com/pulse/aicbic-model-selection-richard-randa
https://stats.stackexchange.com/questions/577/is-there-any-reason-to-prefer-the-aic-or-bic-over-the-other
https://tradingliteracy.com/RSI-Bollinger-Bands-Strategy/
https://trendspider.com/learning-center/rsi-with-bollinger-bands/
https://www.investopedia.com/ask/answers/121014/how-do-i-create-trading-strategy-bollinger-bands-and-relative-strength-indicator-rsi.asp
https://www.linkedin.com/pulse/combining-bollinger-bands-rsi-create-profitable-trading-strategy-vrh6f
https://www.avatrade.com/education/technical-analysis-indicators-strategies/leading-lagging-indicators#:~:text=A%20lagging%20indicator%20is%20an,indicators%20validate%20the%20trade%20setups
https://corporatefinanceinstitute.com/resources/data-science/backtesting/