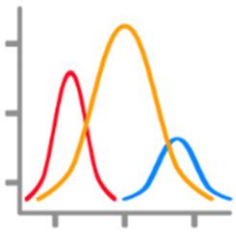
$f(x)$ σ λ $E(X)$ H_0, H_1

QUANT INSIDER



ULTIMATE PROJECT HANDBOOK

Industry oriented projects based on competitions and challenges of top HFT's Hedge Funds

Contents

Project 1 : "News Analytics for Stock Price Prediction"

Can we use the content of news analytics to predict stock price performance? The ubiquity of data today enables investors at any scale to make better investment decisions. The challenge is ingesting and interpreting the data to determine which data is useful, finding the signal in this sea of information

Problem Statement:

In today's data-rich environment, investors have access to a vast amount of information, including news and social media analytics. Can you harness the power of this data to accurately predict stock price performance? The challenge is not only in ingesting and interpreting this data but also in identifying the signals within the noise to make informed investment decisions.

Data:

- Historical stock price data for a set of companies.
- News articles or social media sentiment analysis related to these companies.
- Additional features like market indices, economic indicators, or any other relevant data that might impact stock prices.
- The dataset should be large enough to ensure that machine learning models can be trained effectively, and you should include both training and testing datasets.

Evaluation Metric:

To evaluate the participants' solutions, you can use metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or any other relevant metric to quantify the accuracy of their predictions

Tasks:

Data Preprocessing: Clean, preprocess, and integrate the provided datasets, potentially dealing with missing values, text data, and other challenges.

Feature Engineering: Create meaningful features from the news analytics data and other available information.

Model Building: Build predictive models using machine learning or deep learning techniques. They can experiment with various algorithms and approaches, including time series analysis, NLP, sentiment analysis, etc.

Model Validation: Validate the models using a validation dataset. Cross-validation techniques may be encouraged.

Prediction: Participants need to predict the stock price performance for a given period (e.g., the next day, week, or month) based on the news analytics and other data.

Note:- The above given project idea is based on Problem statement given by Two Sigma

Project 2 - Quantitative Trading Challenge: Maximizing Returns

Problem Statement:

Trading for profit in today's fast-paced and complex financial markets is a challenging task. In this competition, participants are tasked with developing quantitative trading models to identify and take advantage of inefficiencies in a global stock exchange. The goal is to maximize returns using historical market data and mathematical and technological tools.

Data:

- Historical price and volume data for a range of financial instruments (stocks, indices, etc.).
- Relevant market indicators and macroeconomic data.
- Information on potential trading opportunities, including weights and return information for each trade.
- The data should be divided into training and testing datasets.

Tasks:

Data Exploration: Explore the data to understand its structure and relationships between variables.

Model Development: Build predictive models that determine whether to execute a trade for each trading opportunity presented. They can use a variety of machine learning and quantitative finance techniques.

Testing and Validation: Models will be tested against future market returns to evaluate their predictive power.

Utility Maximization: The goal is to develop models that maximize the utility score by selecting the right trades to execute, thus pushing prices closer to "fair" values.

Challenges and Considerations:

Signal-to-Noise Ratio: Develop models that effectively filter out noise in the data to make accurate predictions and decisions.

Redundancy and Feature Correlation: Models should address potential redundancy and strong correlations among features in the data.

Mathematical Formulation: Create mathematical formulations for their trading models that maximize returns.

Note – The Above Project is based on Jane-street problem statement

Project 3 - Nasdaq Closing Price Prediction Challenge

Description:

Develop a model that predicts closing price movements for Nasdaq-listed stocks during the crucial final ten minutes of each trading day. Leverage data from the order book and closing auction to gain insights into supply and demand dynamics, and identify trading opportunities.

Objective:

Create a predictive model that forecasts closing price movements for a diverse range of Nasdaq-listed stocks by using information from the order book and the closing auction.

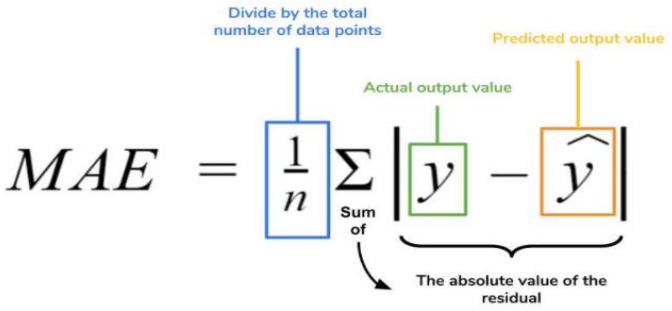
Dataset:

Historical data from the order book and closing auction for Nasdaq-listed stocks. The dataset will include features such as order book dynamics, auction data, and closing price movements.

Evaluation Criteria:

Evaluated based on the Mean Absolute Error (MAE) between the predicted return and the observed target.

Formula for MAE



Challenges and Considerations:

- Combining Order Book and Auction Data: Participants should effectively merge information from both the order book and closing auction data to make accurate predictions.
- Supply and Demand Dynamics: Understanding and modeling supply and demand dynamics is crucial for predicting closing price movements.
- Trading Opportunities: The models should identify potential trading opportunities in the final ten minutes of the trading session.

Note – The above project is based on Optiver's problem statement

Project 4 – Realized Volatility Prediction

Description:

Delve into the world of financial market volatility and develop models that accurately predict short-term volatility for hundreds of stocks across various sectors. Your predictions are critical for the pricing of options and other financial products.

Objective:

Build models that forecast short-term volatility for a diverse set of stocks. These models should provide insights into the expected fluctuations in stock prices over 10-minute periods.

Evaluation Criteria:

Evaluate using the root mean square percentage error (RMSPE) to evaluate the accuracy of predictions.

Challenges and Considerations:

Short-Term Volatility Forecasting: Accurately predicting short-term fluctuations in stock prices over 10-minute periods is a significant challenge.

Pricing Implications: The models developed in this competition have important implications for the pricing of options and other financial products.

Project 5 - Investment Return Rate Forecasting Challenge

Objective:

Build models that forecast the return rate for different investment types. These models should provide accurate predictions of investment returns based on historical price data.

Dataset:

Historical price data for various investment types. This dataset includes time series data that can be used to develop models for forecasting return rates.

Evaluation Criteria:

The competition will be scored based on the mean of the Pearson correlation coefficient for each time ID to evaluate the accuracy of participants' return rate predictions.

Challenges and Considerations:

Time Series Data: Participants will work with time series data, and the challenge is to create models that effectively capture the temporal aspects of investment returns.

Accuracy in Return Rate Forecasting: Accurately forecasting return rates for various investment types is a complex problem, and participants will need to employ sophisticated modeling techniques.

Risk and Volatility: Understanding and accounting for risk and volatility is crucial for predicting investment returns.

Project 6 - "Main/Alternative Exchange Arbitrage Challenge"(CITADEL trading Competition)

Description:

Develop models and algorithms to capitalize on bid-ask spread differences between the main and alternative exchanges. When these spreads cross, use the opportunity to trade, but choose your strategy wisely.

Objective:

Create algorithms that identify and execute arbitrage opportunities when bid-ask spreads cross on two exchanges for the same underlying company's stock. They must choose between market orders for ensured execution or limit orders to control execution price while managing the risk of a partially filled order.

Evaluation Criteria:

Evaluated based on the accuracy of participants' algorithms in identifying and executing profitable arbitrage opportunities.

Challenges and Considerations:

Real-Time Arbitrage: Creating algorithms that can identify and respond to arbitrage opportunities in real-time is a significant challenge.

Risk Management: Effectively manage the risk associated with executing arbitrage trades, especially when using limit orders.

Market Conditions: The bid-ask spreads and arbitrage opportunities can vary depending on market conditions and volatility.

Strategy Explanation

The "Main/Alternative Exchange Arbitrage" strategy involves identifying and capitalizing on arbitrage opportunities that occur when the bid-ask spreads of the same underlying company's stock cross on two different exchanges. Arbitrage is the practice of buying an asset on one exchange where it's relatively cheaper (ask price is lower) and selling it on another exchange where it's relatively more expensive (bid price is higher) to profit from the price difference.

Detailed Strategy:

- **Monitoring:** The first step is to monitor the bid and ask prices of a specific stock on both the main exchange and the alternative exchange. To do this, you need real-time or near-real-time access to order book data from both exchanges.
- **Identifying Crossing Spreads:** Look for instances where the bid price on one exchange is higher than the ask price on the other exchange (bid-ask spreads cross). This indicates a potential arbitrage opportunity.
- **Assessing Arbitrage Size:** Determine the size of the arbitrage opportunity. Calculate the difference in prices and consider the number of shares or contracts you can trade profitably. You need to assess whether the potential profit justifies the trading costs and risks.
- **Decision-Making:** Depending on the magnitude of the spread crossover, you have two options:

- A. Large Crossover: If the spread crossover is significant, indicating a substantial price difference, you can opt for a market order to ensure immediate execution. This is essential to capture the arbitrage opportunity quickly.
- B. Small Crossover: If the spread crossover is relatively small, sending a market order might not be ideal as it could result in lower profits after transaction costs. In this case, consider using limit orders. A limit order allows you to specify the price at which you want to execute the trade. However, it comes with the risk of not executing the entire order, which may leave you partially stuck in the position.

- Execution: Once you decide on the order type (market or limit), execute the trade on both exchanges simultaneously.
- Profit Calculation: Calculate the profit made from the arbitrage trade after accounting for trading fees and other transaction costs.
- Risk Management: Implement risk management strategies to mitigate potential losses. You need to be prepared for situations where the arbitrage opportunity disappears or reverses before you can execute the trade.

Data Requirements:

- Bid-Ask Spread Data: You need real-time or historical bid and ask price data for the same underlying asset on both the main and alternative exchanges. This data allows you to identify spread crossovers.
- Order Book Data: Detailed order book data, including the depth of the market (level 2 data), is essential for assessing the liquidity and order flow on both exchanges. This data helps you evaluate the size of the arbitrage opportunity.
- Historical Price Data: Historical price data for the underlying asset is useful for backtesting and strategy development.
- Transaction Cost Data: Information about transaction costs, including commissions, fees, and market impact costs, is necessary to calculate the profitability of arbitrage trades accurately.
- Real-Time Market Data Feed: To execute arbitrage strategies in real-time, you need access to a live market data feed for both exchanges.

Project 7 - Index Arbitrage Challenge (CITADEL trading Competition)

Description:

Your mission is to build algorithms that analyze the bid-ask spreads of ETFs and their constituent stocks. Find and capitalize on arbitrage opportunities when the sum of the highest bids of constituents is lower than the ETF ask, or when the sum of the lowest asks of constituents is higher than the ETF bid. Join this competition to showcase your skills in index arbitrage.

Detailed Strategy:

Data Collection: To execute Index Arbitrage, you need access to real-time or near-real-time data for ETFs and their constituent stocks. This data should include bid and ask prices for both ETFs and individual stocks.

Identifying Arbitrage Opportunities: Monitor ETF bid and ask prices alongside the bid and ask prices of the constituent stocks. The strategy involves two potential scenarios:

a. Sum of Highest Bids vs. ETF Ask: Calculate the total of the highest bids of the individual stocks. If this sum is less than the ETF's ask price, it indicates an arbitrage opportunity. You can buy the ETF at its ask price and then sell the constituent stocks at their highest bid prices to profit from the price difference.

b. Sum of Lowest Asks vs. ETF Bid: Calculate the total of the lowest asks of the individual stocks. If this sum is greater than the ETF's bid price, it indicates another arbitrage opportunity. You can buy the constituent stocks at their lowest ask prices and then sell the ETF at its bid price to profit from the price difference.

Execution: Once an arbitrage opportunity is identified, execute the trade in real-time. Buy or sell the relevant assets on the respective exchanges.

Risk Management: Implement risk management strategies to mitigate potential losses. Be prepared for situations where the arbitrage opportunity disappears or reverses before you can execute the trade.

Profit Calculation: Calculate the profit made from the arbitrage trade after accounting for trading fees and other transaction costs.

Data Requirements:

- **ETF and Constituent Stock Data:** Real-time or near-real-time data for ETFs and their constituent stocks is crucial. This data should include bid and ask prices.
- **Historical Price Data:** Historical price data for both ETFs and constituent stocks is useful for backtesting and strategy development.
- **Transaction Cost Data:** Information about transaction costs, including commissions, fees, and market impact costs, is necessary to calculate the profitability of arbitrage trades accurately.
- **Real-Time Market Data Feed:** To execute arbitrage strategies in real-time, you need access to a live market data feed for ETFs and individual stocks.
- **Risk Management Models:** Risk management tools and models are necessary to manage the potential downside risks associated with arbitrage trading.
- **Order Execution Systems:** Reliable and fast order execution systems are essential for executing trades quickly when arbitrage opportunities arise.

Challenges and Considerations:

Real-Time Arbitrage: Creating algorithms that can identify and react to arbitrage opportunities in real-time is a significant challenge.

Risk Management: Effectively manage the risk associated with executing arbitrage trades, especially when using limit orders.

Market Conditions: Arbitrage opportunities can vary depending on market conditions, ETF volatility, and liquidity.

Project 8 - News Response Trading Challenge (CITADEL trading Competition)

Description:

Develop innovative models and algorithms to respond to positive and negative news shocks with precision and speed. Trade swiftly, aiming for timely execution. Demonstrate your ability to respond effectively to news-driven market events.

Objective:

Create models that can effectively respond to news shocks. When a negative shock is informed, participants should initiate a short position and, 3 ticks later, reverse it with a long position of the same amount. The same principle applies when a positive shock is informed.

Strategy in Detail:

News Shocks: The strategy relies on receiving information about news shocks. These news shocks can be broadly categorized into three types:

- **Positive Shocks:** These are news events that are expected to have a positive impact on the price of a financial instrument, such as a company's stock. Positive shocks might include good earnings reports, favorable economic indicators, or positive geopolitical developments.
- **Negative Shocks:** These are news events that are expected to have a negative impact on the price of a financial instrument. Negative shocks might include poor earnings reports, adverse economic indicators, or geopolitical tensions.
- **Small Shocks:** These are news events that are considered relatively insignificant or have uncertain outcomes. Small shocks may not have a clear positive or negative impact on the price and are, therefore, ignored.

Trading Response: When a news shock is received, the strategy involves taking trading positions as follows:

- **Negative Shock Response:**

  Step 1 (Immediate Short Position): If a negative shock is informed, the strategy initiates a short position immediately. This means selling a financial instrument with the expectation that its price will decline.

  Step 2 (Long Position After 3 Ticks): Three ticks later, the strategy reverses the short position by taking a long position of the same amount. This means buying back the same financial instrument.

- **Positive Shock Response:**

  Step 1 (Immediate Long Position): If a positive shock is informed, the strategy initiates a long position immediately. This means buying a financial instrument with the expectation that its price will rise.

Step 2 (Short Position After 3 Ticks): Three ticks later, the strategy reverses the long position by taking a short position of the same amount. This means selling the same financial instrument.

Execution: Both the short and long positions are executed as market orders for timely execution. Market orders are designed to be executed as quickly as possible at the current market price.

Data Requirements:

To implement the "News Response" strategy effectively, the following types of data are required:

- **News Data:** Information about news events, including the nature of the news (positive, negative, or small shock), the impact or significance of the news, and the time of the news release.
- **Real-Time Market Data:** Real-time or near-real-time data, including:

  Bid and ask prices for the financial instruments of interest (e.g., stocks, futures, or currencies).

  Historical price data for backtesting and model development.

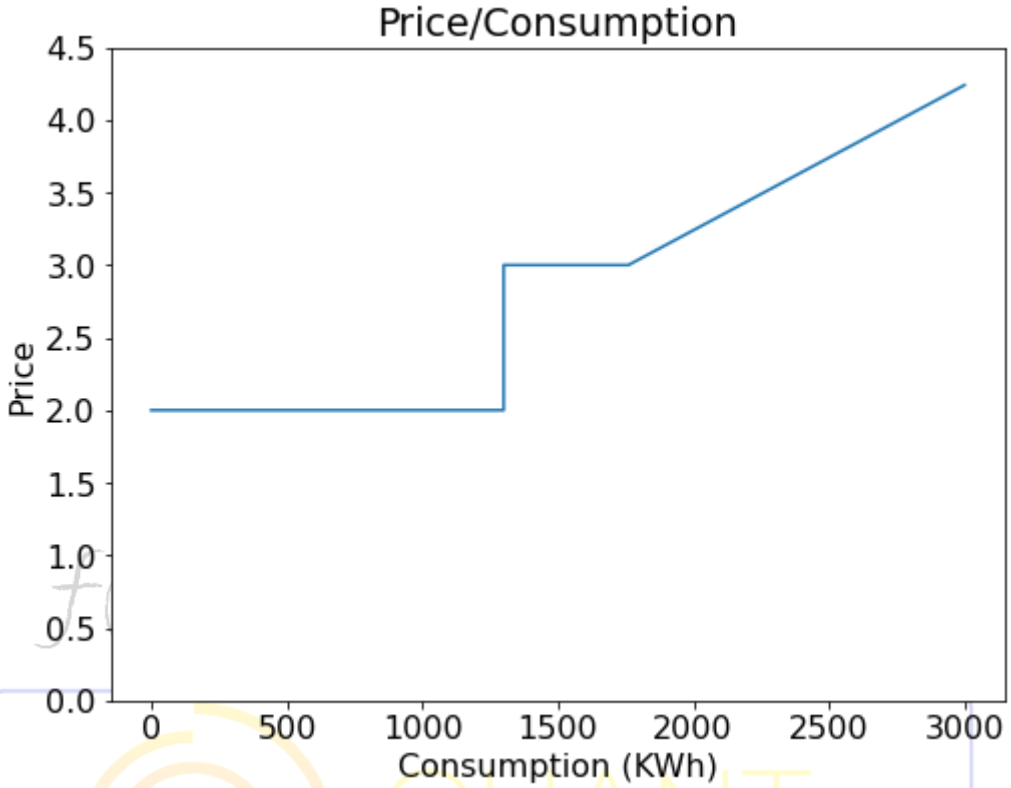
  Tick-level data to monitor price changes and identify the "3-tick" timing for executing the second part of the trade.

  Transaction Cost Data: Information about transaction costs, including trading commissions and fees, to calculate the total costs of executing the trades.

Project 9 – Futures Market making Bot (University of Chicago Trading Competition)

You are a market maker on the Chicago Electricity Futures Market. Real electricity trades daily between consumers and producers on the Chicago Electricity Spot Market, but the futures market is where traders bet on the price of electricity in the future. For instance, an electricity producer could sell 10 MWh of December futures to lock in the price of their electricity in December, thus hedging against the price of electricity. There are 4 futures contracts which expire on the last day of June, August, October and December. These future contracts are cash-settled, so they settle to the current spot price of electricity upon expiry.

Here's how the electricity spot market works. Chicago produces electricity in 3 ways: Natural Gas generators ($2/MWh), coal generators ($3/MWh), and peaker plants. Electricity is consumed in that order, first natural gas, then coal and finally the peaker plants. Each year, the city starts out with fresh capacity of natural gas (around 1300MWh) and coal (around 450MWh), and moves onto the next source once the previous is exhausted. Peaker plants are special - they start at $3.001/MWh and do not run out, but each extra MWh produced costs $.001 more than the previous. This is what the price/consumption graph looks like.



Each day, consumers purchase a random amount of electricity. Each day, you also know how much they used the previous day before you start trading. Note that you cannot buy or sell electricity on the spot market itself, you can only observe it.

As a market maker (MM), your job is to facilitate trading of futures contracts by reducing the spread in the market, which lower costs of transactions for traders. Take a market with bids for $1.50 and offers at $2.00. A 'fair' price of the contract would be $1.75, the mid price. However, the $.50 spread between the bid and offer makes it impossible to trade at the fair price. You (a MM) could come in and bid for $1.60 and offer at $1.90, decreasing the spread from $.50 to $.30. Someone who buys a stock pays $1.90 for it instead of $2.00, saving $.10, while you make the difference between $1.90 and the 'fair' mid price of $1.75. If someone else then sells a contract to you at $1.60, you have made a clean profit of $.30 without a resulting position in the contract, and everyone's happy. But it's not that simple: If the contract price rises to $2.00 after your sale, then you have made a loss of $.10. Also, another market maker could come into the market to bid for $1.70 and offer at $1.80, taking away your profits.

To make money market making, you have to be smart - know the value of the contract by predicting the price of electricity in the future. You also need to be aggressive - provide your best price to the market to make more profitable trades. Therefore, a successful bot requires 3 components: a good guess for the 'fair' price of the contract, a bot which trades aggressively but minimizes risk taking, and a way to augment the dumb bot with human intuition.

Things to Consider

- Running multiple threads for market making so that our orders update much faster
- Fooling other players by putting bogus bid/asks towards the beginning
- Using aggressive trading: penny in + levels
- Finding a way to maximize trades while not going over position limit (being able to close out of large positions)

Relevant Links (https://tianyi.io/post/chicago1/) (https://github.com/John-Trager/UChicago-Trading-Competition)

Project 10 – Upset Prediction in Sports Challenge (Based on CITADEL Datathon)

Description:

Dive into the exciting world of sports analytics and test your skills in identifying upsets in sports matches. Upsets occur when underdog teams triumph over well-renowned opponents, and your task is to predict these surprising outcomes. Whether you're a data scientist, sports enthusiast, or betting expert, this competition invites you to explore the factors that lead to upsets and develop predictive models to identify them.

Objective:

Build predictive models that can identify matches where underdog teams are likely to beat well-reputed opponents. The competition will focus on several hypotheses related to team strategies, player ratings, betting bias, and anomalies in the datasets that may contribute to upsets.

Problem Statement - Analysing and trying to identify upsets from the dataset on matches. Upsets are matches when an underdog team beats a well reknowned team.

Dataset:

- Historical match data, including teams, scores, and match outcomes.
- Betting data, including odds and betting predictions.
- Team attributes and strategies.
- Player ratings and popularity data collected through web scraping.

Methodology

- Web scraping for additional data on player popularity
- We first tried finding the upsets using the betting data given to us for the matches.
- Then we performed exploratory data analytics and tried gaining insights and testing several hypothesis on the way
- We used neural networks, anomaly detection and regression modelling concepts to test out the hypotheses.

Hypotheses and Challenges:

- Team Strategies: Participants will explore the strategies and attributes of winning teams in upset matches. This may involve analyzing possession, defensive strategies, and counterattacking approaches to detect commonalities among underdog victories.

- Betting Bias: Investigate whether betting odds from platforms like Bet365 exhibit a bias toward popular teams and how popular opinion influences their analysis.

- Player Ratings: Explore the correlation between the average rating of players on a team and their ability to win matches.

- Anomalies: Analyze the datasets for anomalies in terms of team skill levels and other factors that may lead to upsets.

Contact Details

Linkedin – Quant Insider (https://www.linkedin.com/company/quant-insider/)

Email - contact.quantinsider@gmail.com

Website - https://topmate.io/quant_insider

Phone Number – 834948921, 9559022860