Project proposal

Research question

We aim to forecast the S&P 500 closing index price for the period from Monday, February 24th, to Friday, February 28th. Our objective is to analyze patterns in historical price movements and leverage forecasting techniques to predict short-term fluctuations in the index.

Specifically, we will explore the following aspects:

- The impact of recent market trends on short-term price predictions
- The effectiveness of different forecasting models in predicting daily closing prices
- The role of external factors, such as economic events and market sentiment, in influencing forecast accuracy

Our goal is to evaluate whether short-term forecasting methods can reliably capture movements in the S&P 500 and assess how different modeling approaches compare in terms of predictive performance.

Data

For this project, we have collected daily historical price data for the S&P 500 index (SPX) spanning a five-year period from February 24, 2020, to February 24, 2025. The dataset has been sourced from the Nasdaq website, which provides publicly available financial market data.

The dataset includes the following key attributes for each trading day:

- Date: The trading day recorded in MM/DD/YYYY format
- Close/Last: The closing price of the S&P 500 index
- Open: The opening price of the index on a given trading day
- High: The highest price reached during the trading session
- Low: The lowest price recorded during the trading session

This dataset serves as the foundation for our forecasting analysis, enabling us to examine historical price movements and apply predictive modeling techniques to estimate future index values.

Code

The group has selected Python as the primary programming language for this project due to its extensive capabilities in data analysis, statistical modeling, and forecasting. We will leverage various Python libraries, including statsmodels, pandas, NumPy, and scikit-learn, to conduct data preprocessing, statistical analysis, and predictive modeling. These libraries provide a comprehensive toolkit for implementing robust forecasting methodologies and evaluating model performance.

Empirical design

For this project, we will employ ARIMA and GARCH models to forecast the S&P 500 closing index price. ARIMA (AutoRegressive Integrated Moving Average) will be used to model the time series' linear trends and autocorrelations, while GARCH (Generalized Autoregressive Conditional Heteroskedasticity) will be applied to capture volatility clustering and time-dependent variance in the data.

The models will be estimated using Python libraries such as statsmodels and sklearn. We will optimize the model parameters by iterating through different lag orders and hyperparameter configurations. The selection process will involve minimizing Mean Squared Error (MSE) on a validation dataset to identify the best-performing model specifications.

The predictive performance of the models will be assessed using multiple evaluation metrics:

- R^2 (coefficient of determination) to measure how well the model explains variance in the data.
- Root Mean Squared Error (RMSE) to quantify the average prediction error in the same units as the index price.
- Additional diagnostic checks, such as residual analysis and backtesting, will be performed to ensure model robustness and reliability.

This empirical approach will allow us to compare the effectiveness of different models and determine the most suitable forecasting method for short-term S&P 500 price movements.

Contribution

Each team member has been assigned specific roles to ensure a structured and efficient workflow. The contributions and responsibilities are as follows:

- All Members → Data retrieval and preprocessing. Responsible for ensuring data accuracy and consistency.
- Olav Nikolai Meli Nymo → Conducting empirical analysis, implementing models, and testing. Leads statistical modeling and analysis.
- Nicolai Harvik → Writing and structuring the final report. Ensures clarity, coherence, and organization of the document.
- Olav Berger → Preparing and recording the PowerPoint presentation. Responsible for effectively communicating the project's findings.
- Ole Ekern → Peer-reviewing the report and presentation. Ensures quality, accuracy, and refinement of deliverables.

These roles serve as a guideline, but all members will collaborate and provide feedback to enhance the overall quality of the project.