Predicting Aggregate Market Fluctuations:

A Data-Driven Approach to Equity and Bond Allocation

Maddox Roy Pranshu Shrivastava Ryan Jackman Charles Tirendi
Jakcrus Huynh

The Ohio State University

March 2nd, 2025

Team Roles

Jakcrus Huynh - Strategy Development and Implementation

Maddox Roy, Charles Tirendi, Ryan Jackman - Machine Learning

Pranshu Shrivastava - Statistical Modeling and Analysis

Problem Statement and Objectives

Problem Statement: While short term stock price movements cannot be predicted, it is believed that aggregate market movements can be predicted. Is this actually possible, and is it profitable?

Key Objectives:
Given S&P500 index values, bond yields, and calculated market probabilities of an increase/decrease of 20%:

- Find and classify periods of bear and bull markets
- Determine if market based probabilities are accurate
- Create our own model to predict regimes
- Develop a trading strategy using market regimes

Languages, Libraries, and Tools

The strategy itself is written in Python using libraries such as Pandas, NumPy, SkLearn, and Matplotlib.

For some of the visualization, statistical modeling and analysis, R, Tableau, and Excel were used

Understanding the Data

Time Range: 2007-01-08 to 2024-12-31

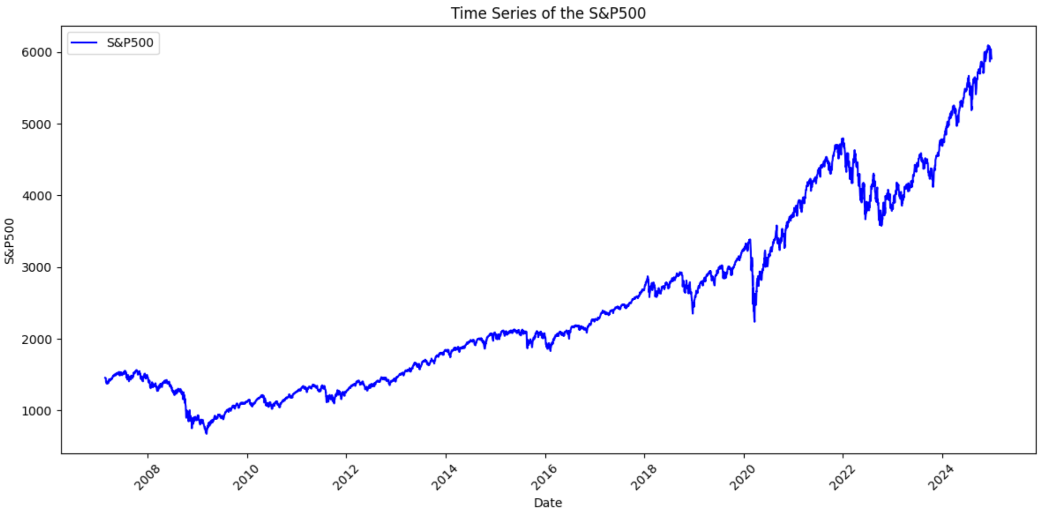
Basic Data Analysis, S&P500:

- Annualized Returns: 8.64%
- Annualized Standard Deviation: 20.64%
- Max Drawdown: -56.77%
- Daily 95% VaR: -2.01%

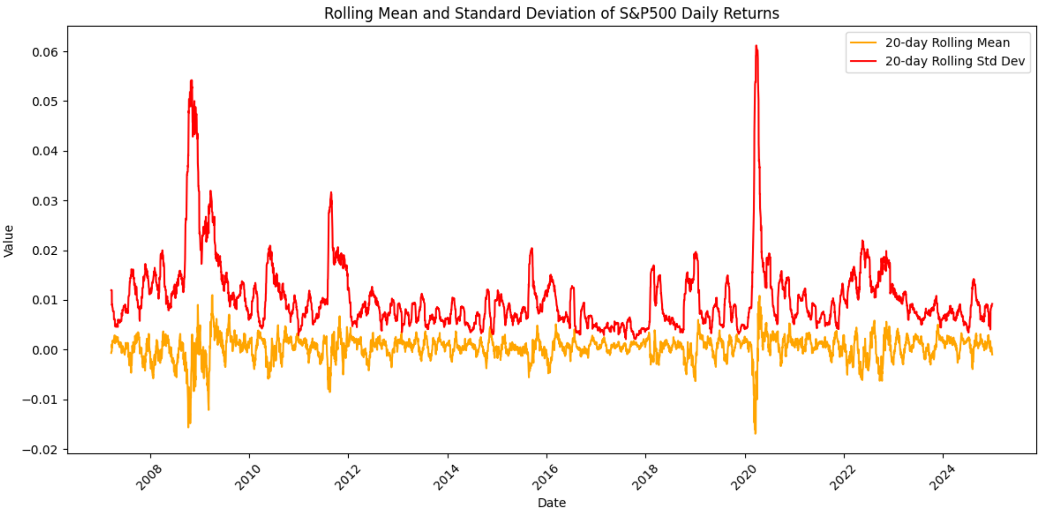
Bond Yields

- Mean Bond Yield: 0.855%
- Max Bond Yield: 5.348% (2024-12-30)
- Min Bond Yield: -0.105% (2007-01-08)

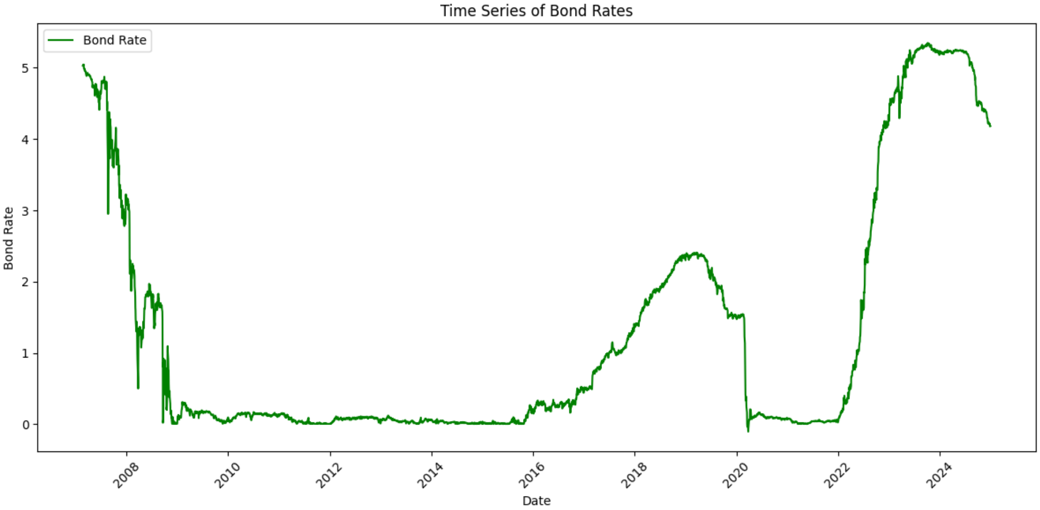
S&P500 Time Series



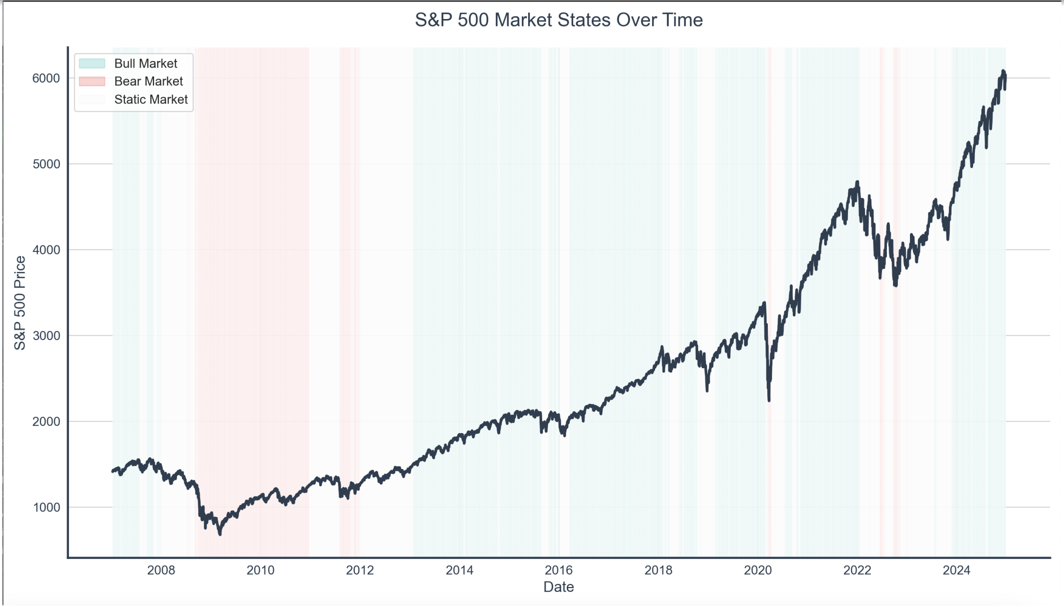
Rolling Mean and Standard Deviation of S&P500



Bond Rate Time Series



Regimes Visualized



Statistical Analysis

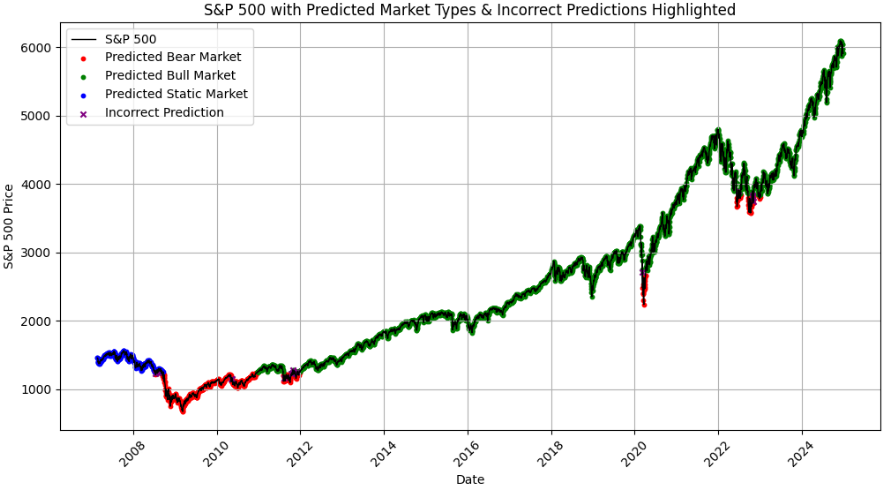
To assess whether the given market-based probabilities are accurately able to predict Bull or Bear markets, we performed a hypothesis test with a significance level of 5%. We use a threshold of probability spread (PrInc - PrDec) of 10% where if the absolute value of the spread is less than 10, it is classified as static, otherwise a positive spread indicates a bull prediction and a negative spread indicates a bear prediction.

- **Null Hypothesis (H_0)**:
  The proportion of correct predictions is $>=$ 55% (suggesting an accurate model)
- **Alternative Hypothesis (H_a):**
  The proportion of correct predictions is $<$ 55% (suggesting an inaccurate model)

Statistical Analysis (Continued)

According to our data, the simulated proportion of successes ($\hat{p}$) was $\frac{203}{681}$ or 29.8%. This corresponds to a Z-score of -12.829 or a p value of ≈ 0. Thus, we have enough evidence to reject the null hypothesis and accept the alternate hypothesis. So, we can conclude that these market-based probabilities are inaccurate.

Random Forest Prediction of Regimes



Risk Management

- **Volatility Targeting:** Adjust equity exposure to target an annualized volatility of 15%. When volatility rises, we lower stock allocation; when it falls, we increase it, ensuring a consistent risk profile.
- **Trailing Stop Loss:** Monitor the peak price of stock holdings and exit if the price falls 10% below this peak, locking in gains and limiting losses.

Position Sizing

- **Regime-Based Allocation:**
  Bull: 90%/10%, Bear: 20%/80%, Static: 60%/40%.
  Let this serve as the base rule for position sizing
- **Monthly Rebalancing:** Adjust every 20 trading days based on rules set on the next slide
- **Logistic Regression Signals:** A LR buy signal immediately increases exposure to stocks by 5% and a LR sell signal immediately decreases exposure to stocks by 5% (max 95% equity).

Extra Indicators

Bond Yield Trend

- Use a 20=day rolling percentage change in bond yields as an indicator. Increasing bond yields indicate tightening monetary policy. Decrease stock position in rising yields and vice versa for falling yields

20-day Cumulative Log Return

- Used as a measure of explosiveness. If cumulative 20-day log returns are greater than 5%, exhibits exponential growth which is unsustainable in the long run. Therefore the position in stocks is reduced if reached. Much simpler of an indicator of explosiveness compared to ADF tests.

Shannon Entropy

- Explained further on the next slide

Shannon Entropy

Given a discrete random variable X with possible outcomes $\{x_1, x_2, \ldots, x_n\}$ and associated probabilities $\{p_1, p_2, \ldots, p_n\}$, the Shannon entropy is defined as:

$$H(X) = -\sum_{i=1}^{n} p_i \log_2(p_i)$$

This quantity represents the average uncertainty or information content of X. In our trading strategy, we compute the entropy of the distribution of daily returns:

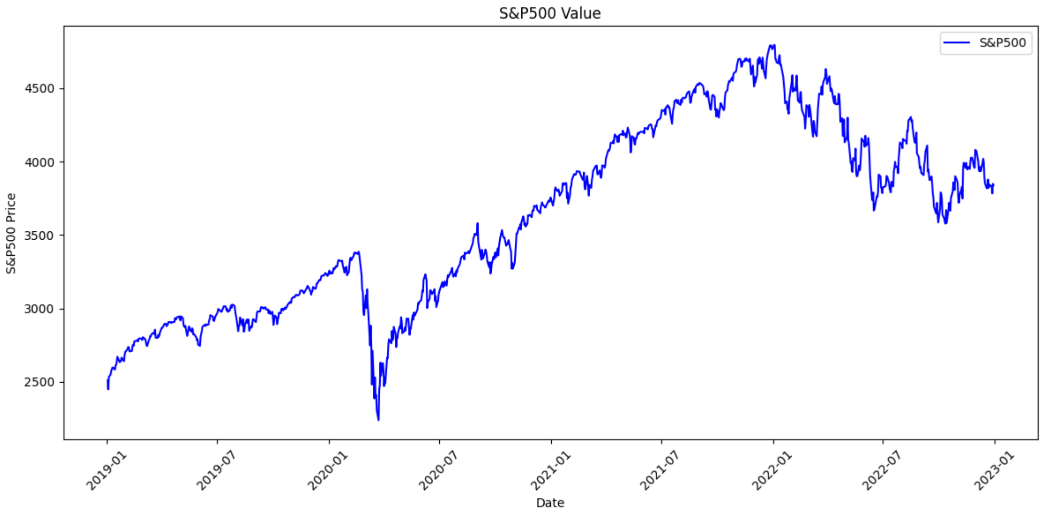
- **Low Entropy:** Indicates a more predictable (trending) market. We increase our position size in stocks during low entropy.
- **High Entropy:** Indicates a choppy, less predictable market. We do not increase nor decrease our position size during high entropy.

Avoiding Overfitting

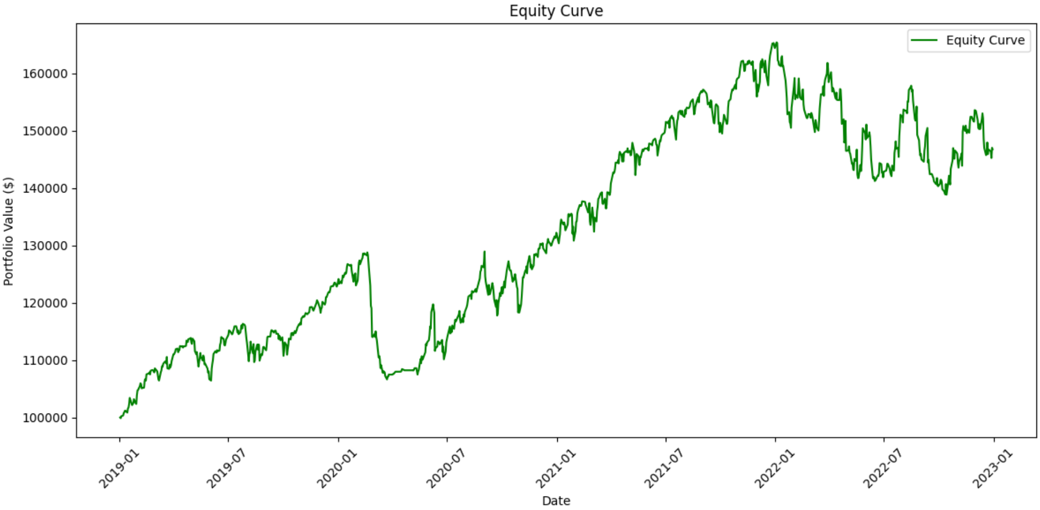
The logistic regression model for trading signals was trained on data before the period of consideration (2019-01-01 to 2022-12-31)

Furthermore, all values used for position sizing and rebalancing were calculated sequentially at each time step using only past data.

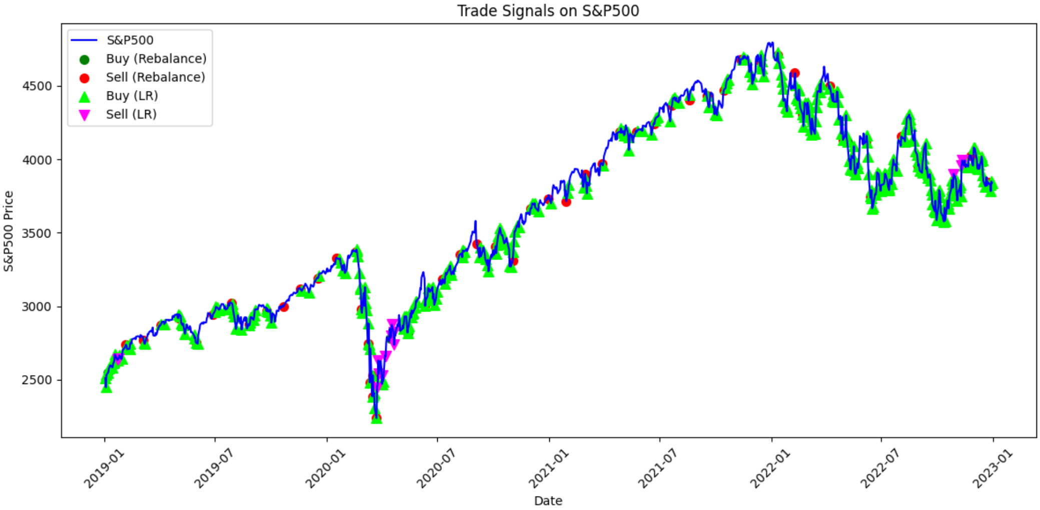
S&P500 Chart on Time Period



Results on Time Period



Strategy Trades and Rebalances



S&P500 Results on Time Period

- **Annualized Return:** 13.28%
- **Annualized Standard Deviation:** 22.91%
- **Max Drawdown:** -33.92%
- **Beta:** 1.0
- **Sharpe Ratio:** .54
- **Total Percent Return:** 52.97%

Results on Time Period

- **Beginning Balance:** \$100,000
- **Annualized Return:** 10.60%
- **Annualized Standard Deviation:** 14.07%
- **Max Drawdown:** -17.20%
- **Beta:** 0.48
- **Sharpe Ratio:** 0.70
- **Total Percent Return:** 46.77%
- **Final Balance:** \$146,783.11
- **Correlation between strategy returns and S&P500 returns:** 0.78

Challenges and Future Improvements

- **Structural Breaks:** The model may not fully account for sudden, external market shocks or structural changes.
- **Trading Constraints:** No short selling or leverage was allowed, reducing the flexibility of the strategy.
- **Restricted Asset Universe:** Only the S&P 500 and short-term bonds were available.
- **Future Improvements:**
  - Expand the data set and include alternative data sources.
  - Explore advanced machine learning and ensemble models for better regime prediction.
  - Incorporate additional risk measures (e.g., dynamic stop losses, drawdown-based adjustments).
  - Broaden the asset universe to include more diversification opportunities.

Conclusion

- Our analysis demonstrates that market probabilities and technical signals can help distinguish market regimes. However, the given calculated market based probabilities were not accurate in predicting bull and bear markets
- The strategy integrating regime detection, volatility targeting, trailing stops, and logistic regression signals—outperforms a static buy-and-hold approach in risk-adjusted terms.
- The model successfully reduces exposure during adverse conditions while capitalizing on bull market trends.
- Future improvements will focus on fine-tuning parameter optimization, integrating additional risk factors, and validating the strategy over diverse market cycles.