# Does Market-Based Probabilities Have Predictive Capabilities for Stock Market?


## Abstract

Market Probability Density (MPD) describes the risk-neutral probability distributions of asset returns and has potential for forecasting future market. In our study, we evaluate the predictive power of MPD data by forecasting the S&P 500's returns, volatility, and reversals using Stepwise Linear Regression, Random Forest, Support Vector Regression, and XGBoost models. Some of models show positive performance, especially in the prediction of reversals. With these insights, we develop a voting strategy and a reversal strategy for trading the S&P 500, each obtaining high excess returns. Specifically, our reversal strategy achieves an annual excess return of 19.18% with a Sharpe ratio of 1.73. The superior performance of the reversal strategy further confirms the predictive power of MPD data regarding future reversals.


## 1 Introduction

The unpredictability of financial assets remains a paramount challenge for both investors and monetary authorities, prompting the adoption of quantitative techniques in decision-making processes. Among these, Market-Based Probability (MPD) measures the probability of the financial market encountering specific events. The concept of MPD was initially explored by Breeden and Litzenberger in 1978. Their groundbreaking work demonstrated that the prices of call options, with minor differences in strike prices, could be used to construct risk-neutral probability distributions. Building on this discovery, the adaptation of MPD rapidly evolved. Shimko, in 1993, introduced a refined approach to MPD by deriving implied volatilities from option prices and employing cubic spline interpolation to yield enhanced risk-neutral distributions. Subsequently, Figlewski (2008) advanced Shimko's methodology by accounting for the market's bid-ask spread and augmenting the risk-neutral density with tails to more accurately represent the S&P 500's probability distribution. Over the past decades, analysts such as Yuriy (2012) and Malz (2014) have extended MPD applications to inflation.

Since 2007, these methodologies have been utilized by the Minneapolis Federal Reserve to periodically update MPD for various assets, thereby providing investors with deeper insights into market dynamics. To rigorously assess the efficacy of MPD in forecasting US Market portfolio behaviors, we study its predictability concerning future returns, volatilities, and reversals. In the latter portion of our paper, we propose direction voting and reversal trading strategies exploiting MPD to showcase its applications by investors for obtaining forward-looking information in the US stock markets.

## 2 Data Analysis

Figure 1 illustrates the movement of the expected MPD returns against the actual returns over the observed period alongside with the 10$^{th}$, 50$^{th}$, and 90$^{th}$ percentiles of MPD. The comparison statistics among them are presented in the Table 1. Specifically, 90.16% of the actual 6-month return and 80.16% of the actual 12-month return falls between the 10$^{th}$ and 90$^{th}$ MPD percentiles. This suggests 10$^{th}$ and 90$^{th}$ percentiles of MPD provide information about possible upper and lower bounds for the actual markets.

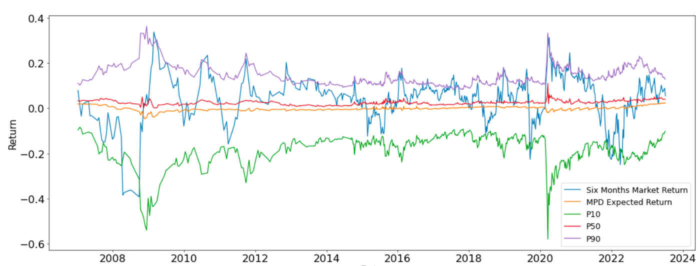


Figure 1: MPD Return Distribution vs. Real Return

Table 1: Comparison of MPD and Real Market

| | < p10 | p10 – p50 | p50 – p90 | > p90 |
|---|---|---|---|---|
| 6 months | 3.38% | 96.62% | 3.49% | 96.51% |
| 12 months | 36.62% | 63.38% | 32.06% | 67.94% |

We conduct likelihood-ratio tests to verify whether the difference between MPD percentiles and actual percentiles is significant or caused by statistical error. The likelihood-ratio test is a statistical test used to compare the goodness of fit of two distributions based on the ratio of their likelihoods. The LR test statistic is given by:

$$LR = -2\log[(1-p)^{n-x}p^{x}] + 2\log[(1-\frac{x}{n})^{n-x}(\frac{x}{n})^{x}]$$

where $p$ stands for the pth percentile we are testing, n is the sample size, and $x$ is the number of actual returns lower than the pth percentile. Under the null hypothesis, the actual percentiles are equal to the MPD percentiles, and the LR test has a chi-square distribution with 1 degree of freedom. The result of the likelihood-ratio test is shown in the Table 2.

Table 2: Likelihood-ratio Test Result

| | 6-month | | 12-month | |
|---|---|---|---|---|
| | LR stats | p-value | LR stats | p-value |
| p10 | 41.4188 | 0.0000 | 38.6037 | 0.0000 |
| p50 | 47.1513 | 0.0000 | 82.9081 | 0.0000 |
| p90 | 10.2084 | 0.0014 | 24.1591 | 0.0000 |

Low p-value of LR test suggests a significant difference between MPD percentiles and actual percentiles. Specifically, the 10th and 50th MPD percentiles of the 6-month return and the 10th, 50th, and 90th MPD percentiles of the 12-month return are underestimated compared with actual market percentiles while the 90th MPD percentiles of the 12-month return are overestimated.

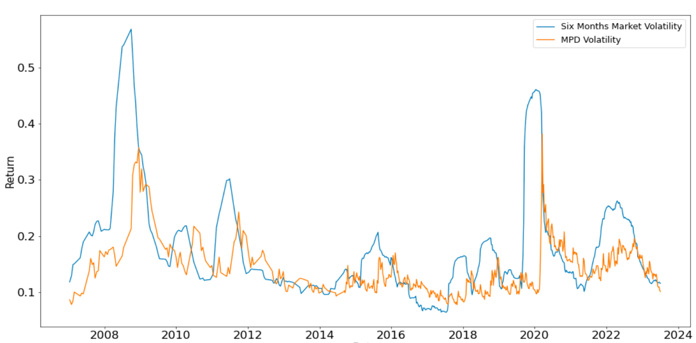

Figure 2: MPD Volatility vs. Real Volatility

For the volatility analysis, the real 6-month volatility basically aligns with the trend of standard deviation in MPD dataset, but they do not always move in same directions, which indicates that they are measuring different aspects of the data or have different sensitivities to market conditions. The real-market volatility line is smoother than the MPD standard deviation line, which could be due to the historical real-market volatility being a rolling measurement that is less sensitive to short-term fluctuations.

# 3 Prediction Models

In this section, we evaluate the predictive power of MPD data for forecasting the S&P 500's returns, volatility, and reversals by building different models. Table 3 is the list of notations we use in the following parts of the paper.

Table 3: Notations

| Symbol | Definition |
|---|---|
| mu | The mean of the MPD |
| sd | The standard deviation of the MPD |
| skew | The skew of the MPD |
| kurt | The kurtosis of the MPD |
| p10/50/90 | The 10/50/90th percentile of the MPD |
| lg_change_prob | The change in the expected return in percentage terms defined as “large” |
| prDec/Inc | The probability of a “large decline/increase” in return as defined by “lg_change_prob” |

## 3.1 Return Prediction

### 3.1.1 Data preprocessing

In the data preprocessing phase, we employ Principal Component Analysis (PCA) to address collinearity issues. Figure 3 highlights a high correlation among *sd*, *p10*, *p90*, *prDec*, and *prInc*, an indication that these variables may be redundant and impair the performance of machine learning models.

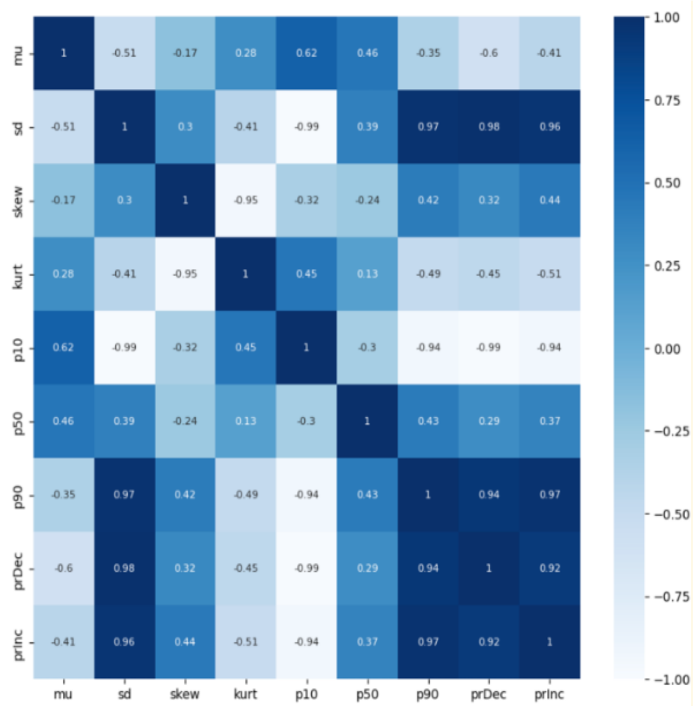

Figure 3: Correlation Matrix

We replace the correlated variables with a more condensed representation to alleviate this. The first principal component extracted through PCA is chosen, as it successfully captures an impressive 96.87% of the total variance within the five variables. In addition to the PCA, interaction terms, are introduced in the regression analysis to capture the potential complexity of the relationships within the data. By adding these terms, our goal is to enhance the efficacy of the baseline linear regression model to extract richer insights from the data and provide helpful feedback for further model development.

### 3.1.2 Model Construction

In the previous section we use statistical methods to test whether MPD percentiles match actual returns. In this section, we explore non-linear machine learning (ML) models to see if the same effect persists. We explore multiple ML models to see if we can get better next-period predictions. However, this approach is accompanied by certain costs. The absence of fixed linear parameters for empirical validation against prior knowledge increases the difficulty of interpreting the results. Moreover, there is a risk of overfitting, which can lead to suboptimal out-of-sample performance. Given the challenges associated with empirically validating these models, we emphasize the importance of meticulous feature selection.

We use three different machine learning models, Random Forests (RF), Support Vector Regression (SVR), and eXtreme Gradient Boosting (XGBoost). We use stepwise linear regression (LR) as the baseline for the comparison.

Stepwise regression is a variable selection technique used in statistical modeling. It involves iteratively adding or removing predictors from a model based on specific criteria (AIC and BIC in this paper). We compare forward selection, backward elimination, and a combination of the two, verifying our selection with the ANOVA test for a sparse model with relevant predictors. The Random Forests model, an ensemble of decision trees, offers robustness and high accuracy by averaging multiple deep decision trees, reducing the risk of overfitting while handling a large variety of data types. SVR excels in finding the optimal hyperplane in a high-dimensional space, focusing on the vectors that are most informative for constructing the model, thereby enhancing prediction accuracy. Lastly, the XGBoost model leverages an advanced ensemble technique that sequentially builds trees, each one correcting the errors of the previous, hence often delivering superior performance.

We split our data and allocate the initial 80% of the data for the training set and reserve the remaining 20% for the test set. There are no structural breaks in either training or test datasets.

### 3.1.3 Results

We evaluate models by Mean Squared Error (MSE), Mean Absolute Error (MAE), R squared.

Table 4: Return Prediction Model Evaluation

| | 6-month return | | | 12-month return | | |
|---|---|---|---|---|---|---|
| | MSE | MAE | $R^2$ | MSE | MAE | $R^2$ |
| LR | 0.0196 | 0.1037 | -0.6332 | 0.0632 | 0.2202 | -3.0055 |
| RF | 0.0137 | 0.0916 | -0.3081 | 0.0153 | 0.1002 | -0.3152 |
| SVR | 0.0093 | 0.0787 | 0.1153 | 0.0098 | 0.0827 | 0.1047 |
| XG-Boost | 0.0116 | 0.0846 | -0.0964 | 0.0128 | 0.0937 | -0.0110 |

Table 4 shows the out-of-sample results for the machine learning models and the linear regression baseline. Combined with the ANOVA test and stepwise regression, *mu·skew*, *kurt·p50*, *skew·p50*, *p50·p50*, *kurt·kurt*, *skew·kurt*, *skew·skew*, *mu·mu*, *mu·p50*, and *mu·kurt* are selected as predictors based on AIC criteria. The OLS regression results for 6-month return would be:

$$\begin{aligned} \text{return} = 0.17 &+ 35.10\, mu{\cdot}skew - 9.67\, kurt{\cdot}p50 \\ &- 26.59\, skew{\cdot}p50 - 19.70\, p50{\cdot}p50 \\ &- 0.43\, kurt{\cdot}kurt - 2.13 skew{\cdot}kurt \\ &- 2.76\, skew{\cdot}skew - 28.99\, mu{\cdot}mu \\ &+ 412.71\, mu{\cdot}p50 + 9.50\, mu{\cdot}kurt \end{aligned}$$

Except for *mu·mu*, the remaining independent variables are significant at the 95% confidence level, indicating that these variables are statistically significant for return prediction. However, the in-sample $R^2$ score of 0.406 implies that these effects may be moderate, and over time, they are likely to diminish — an observation consistent with the trends observed in the table. Random Forest and XGBoost models exhibit better predictive performance compared to the linear regression model. The out-of-sample $R^2$ score for SVR stands at 0.1153, signifying that the attributes of this MPD hold predictive power for forecasting return.

Table 4 also indicates that the predictive efficacy of the 6-month MPD outperforms that of the 12-month MPD. This could be attributed to the longer time span encompassed by the 12-month MPD, which introduces more variables while potentially unstable factors. Consequently, short-term predictions tend to be more accurate than long-term forecasts due to reduced uncertainties and variabilities over shorter periods.

## 3.2 Volatility Prediction

### 3.2.1 Model Construction

In building models to predict volatility, the predictive features are the same as those used in the return prediction model, consisting of the original features of the MPD along with their interaction and squared terms. The forecast target is the volatility over the next six or twelve months, defined as the standard deviation of daily returns. Although the VIX index is recognized as an effective measure for predicting future market volatility, our research indicates that the VIX's calculation methodology and underlying data are similar to those used in the MPD. The correlation between them is 0.98. Using the 6-month VIX index as a predictive target could lead to data leakage and severe overfitting issues. For the prediction models, we continue to use four different approaches: Random Forest, SVR, and XGBoost, with stepwise regression serving as the baseline for comparison.

### 3.2.2 Results

Table 5: Volatility Prediction Model Evaluation

| | 6-month volatility | | | 12-month volatility | | |
|---|---|---|---|---|---|---|
| | MSE | MAE | $R^2$ | MSE | MAE | $R^2$ |
| LR | 0.0072 | 0.07209 | -1.8200 | 0.0052 | 0.2060 | -1.3890 |
| RF | 0.0032 | 0.0479 | -0.0893 | 0.0048 | 0.0579 | -0.1082 |
| SVR | 0.0044 | 0.0583 | -0.4931 | 0.0059 | 0.0653 | -0.5431 |
| XG-Boost | 0.0029 | 0.0478 | 0.0107 | 0.0032 | 0.0495 | 0.0087 |

Table 5 shows the out-of-sample results for the machine learning models and the linear regression baseline. Based on the ANOVA test output, the AIC criteria model using a combination of forward selection and backward elimination turns out to be the most appropriate. The model includes eight features: *skew·kurt, mu·pca, mu·mu, mu·kurt, skew·skew, kurt·kurt, pca·pca, kurt·pca, skew·p50*. The OLS regression results for 6-month volatility would be:

$$\begin{aligned} \text{volatility} = 0.10 &+ 0.89\, skew{\cdot}kurt - 54.37\, mu{\cdot}pca \\ &- 506.30\, mu{\cdot}mu + 3.80\, mu{\cdot}kurt \\ &+ 1.21\, skew{\cdot}skew + 0.18\, kurt{\cdot}kurt \\ &- 1.13\, pca{\cdot}pca + 0.20\, kurt{\cdot}pca \\ &+ 4.74\, skew{\cdot}p50 \end{aligned}$$

Every independent variable is significant at the 95% confidence level, indicating that these variables have significant predictive power for volatility. Random Forest and SVR models exhibit better predictive performance compared to the linear regression model. The out-of-sample $R^2$ value for XGBoost is 0.0107, signifying that the attributes of this MPD hold predictive power for forecasting volatility.

Table 5 also indicates that the features of MPD are less effective at predicting volatility compared to their performance in forecasting returns. Similarly, the 6-month MPD demonstrates better predictive efficacy relative to the 12-month MPD.

### 3.3 Reversal Prediction

#### 3.3.1 The definition of reversals

The technique of directional change (DC) has shown the potential of capturing market trends and contributing to making profits (Adegboye et al., 2017). DC records the changes in a stock price by a pre-specified percentage, a threshold $\theta$ , which was in advance decided by the trader. Under this paradigm, the market is summarized into upturn and downturn trends. Furthermore, each of these trends is supposed to be followed by an overshoot (OS) event, which finishes once an opposite DC event takes place. In other words, each reversal and trend include a DC event and a following OS event.

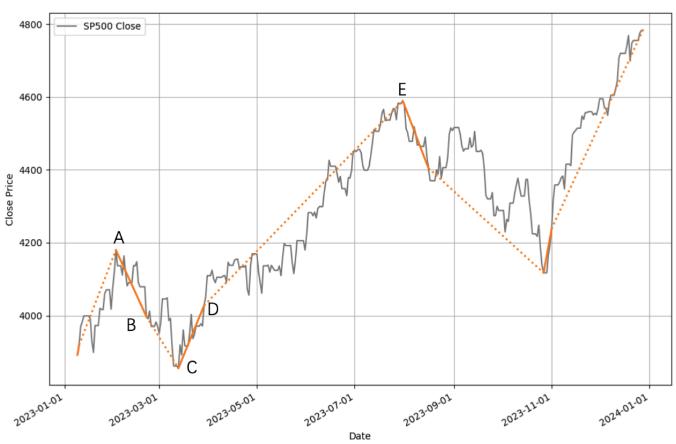


Figure 4: Directional Changes for S&P 500

Figure 4 displays how the curve of S&P 500 close price is dissected into DC and OS events, or upturn and downturn trend. Here, we choose a specific threshold, above which a price change would be considered as a trend. DC events are shown in solid lines, and OS events in dashed lines. As we observe in the figure, a downturn event begins at Point A and continues to Point B. At Point B, a downturn overshoot begins and goes on until Point C. At Point C, the trend changes direction, and an upturn trend begins, lasting until Point D. Between Points D and E, there's an upturn overshoot event, and the pattern continues.

Algorithm 1 demonstrates how we generate DC and OS events.

**Algorithm 1** Pseudocode for generating reversals given threshold $\Delta x_{dc}$.

**Require:** Initialise variables (event is Upturn event, $p^h = p^l = p(t_0)$, $\Delta x_{dc}(Fixed) \geq 0, t_{dc}^0 = t_{os}^0 = t_{dc}^1 = t_{os}^1 = t_0$ )

**if** event is Upturn Event **then**
  **if** $p(t) > p^h \times (1 - \Delta x_{dc})$ **then**
    event ← Downturn Event
    $p^l \leftarrow p(t)$ ▷ Price at end time for a downturn Event
    $t_{dc}, t_{os}^0 \leftarrow t, t+1$ ▷ End for downturn DC, Start for downturn OS
  **else**
    **if** $p^h < p(t)$ **then**
      $p^h \leftarrow p(t)$ ▷ Price at start of downturn event
      $t_{dc}^0, t_{os}^1 \leftarrow t, t-1$ ▷ Start for downturn DC, End for upturn OS
    **end if**
  **end if**
**else**
  **if** $p(t) > p^l \times (1 + \Delta x_{dc})$ **then**
    event ← Upturn Event
    $p^h \leftarrow p(t)$ ▷ Price at end time for upturn event
    $t_{dc}, t_{os}^0 \leftarrow t, t+1$ ▷ End for a upturn DC, Start for upturn OS
  **else**
    **if** $p^l > p(t)$ **then**
      $p^l \leftarrow p(t)$ ▷ Price at start time for upturn event
      $t_{dc}^0, t_{os}^1 \leftarrow t, t-1$ ▷ Start for a upturn DC, End for downturn OS
    **end if**
  **end if**
**end if**

Figure 5 shows the upturn (green) and downturn (red) trends with a 4.5% threshold, in the period we investigate when there are 153 reversals in total. DC events can be confirmed on the day the threshold is reached, while OS events must be retrospectively confirmed after the start of the next trend. The average period length of DC events is 9.2 days, and that of OS events is 28.4 days.


Figure 5: Trend Reversals Detection Result

Note that we can immediately find the start of reversals and trends by detecting DC events. However, OS events, which decide the duration of each whole reversal period, can only be confirmed retrospectively. Therefore, in this section, we will examine the predictive power of MPD data on the duration of overshoots to determine whether MPD data contains information that is helpful to predict future reversals.

### 3.3.2 Data Preprocessing and Feature Selection

We record the start and end dates of each reversal's DC event and OS event, where the duration of the OS event serves as the target. We select the duration of the DC event, the duration of the previous OS, the dummy variable for trend as conventional features (Adegboye et al., 2017). Additionally, we incorporate the most recent 6-month and 12-month MPD data respectively at the start of the OS event to analyze MPD data. The MPD-related features include *mu*, *skew*, *kurt*, *p50*, and *pca* as mentioned in the previous section. Furthermore, to elucidate the impact of the interaction of the aforementioned features on reversal, this study employs forward selection based on the BIC for the product terms of conventional features and 6-month and 12-month MPD data respectively, selecting product terms with substantial explanatory power as features for the model. For both 6-month and 12-month MPD data, product terms *pca·trend* and *mu·trend* are selected.

### 3.3.3 Model Results

We allocate the first 90% of the data as the training set and the remaining 10% as the test set. We attempt to evaluate the predictive power of 6-month and 12-month MPD data on reversals using three different models: linear regression, SVR, and gradient boosting. Initially, we experiment with models that do not incorporate MPD data and then add 6-month and 12-month MPD data respectively. The performance of these models on the test set is presented in Table 6.

Table 6: Results of Reversal Prediction Models

| Model | Metric | without MPD | 6-month MPD | 12-month MPD |
|---|---|---|---|---|
| LR | MSE | 996.34 | 562.38 | 597.34 |
| | MAE | 22.60 | 19.38 | 19.27 |
| | $R^2$ | 6.47% | 47.21% | 43.92% |
| SVR | MSE | 837.27 | 718.12 | 737.06 |
| | MAE | 24.98 | 20.68 | 21.91 |
| | $R^2$ | 21.39% | 32.59% | 30.81% |
| XGBoost | MSE | 840.07 | 568.08 | 591.56 |
| | MAE | 20.97 | 18.67 | 18.50 |
| | $R^2$ | 21.14% | 46.71% | 44.47% |

Compared to the models that does not incorporate MPD data, both models augment with 6-month and 12-month MPD data exhibit a significant improvement in predictive power. Specifically, there is a notable decrease in both MSE and MAE, alongside a significant increase in the $R^2$ value in the test set, indicating that both the 6-month and 12-month MPD data possess predictive power for future reversals. Among these, the 6-month MPD data shows stronger predictive power, suggesting that, compared to 12-month MPD data, 6-month MPD data contains more information pertinent to forecasting future market conditions. From the perspective of model performance, the XGBoost model exhibits the best predictive efficacy. Specifically, for the linear regression model, the coefficient of the MPD related interaction terms, *mu·trend* and *pca·trend* are both significant in linear regression. This further illustrates that MPD data contains information that is helpful for predicting future reversals.

# 4 Trading Strategies

In previous sections, we have discovered that MPD, at least in the case of reversal identification, has considerable predictive power on the S&P 500 market. To understand how effective the prediction is in the real market, we propose a voting strategy with optimized return predictions and a reversal strategy based on reversal predictions.

## 4.1 Direction Voting Strategy

### 4.1.1 Strategy Overview

In the direction voting trading strategy, the daily return serves as a proxy for market direction and thus forms the basis for our binary classification target variable. The rationale is straightforward: a positive daily return suggests an upward market movement (bullish), while a negative daily return indicates a downward trajectory (bearish). By framing the problem in this binary context, we simplify the complex dynamics of the market into an actionable format.

### 4.1.2 Data Preprocessing and Feature Selection

The trading target of this strategy is the daily price direction of the S&P 500 index. This section selects data from January 2007 to July 2020 as the training set and data from August 2020 to January 2024 as the test set, conducting strategy backtesting within this period. To manage gaps in daily MPD data, we employ forward filling, using the last available value to fill subsequent days until the next update.

A binary classification model is employed to forecast whether a particular stock, or in this case, the S&P 500 index, will experience a price increase (up) or decrease (down) over a specified future period, categorizing potential outcomes into two distinct classes:

$$\text{RET}=\begin{cases}1, & \text{if daily return} < 0\\ 0, & \text{if daily return} \geq 0\end{cases}$$

$$\text{Signal}=\begin{cases}\text{Long}, & \text{if RET} = 1\\ \text{Short}, & \text{if RET} = 0\end{cases}$$

This binary signaling aligns with a common trading philosophy: to 'buy low and sell high'. In essence, it encapsulates the quintessential goal of timing the market—capturing gains by purchasing assets ahead of expected rises and divesting prior to anticipated declines.

We optimize predictions with these technical indicators: Simple Moving Average (SMA), Exponential Moving Average (EMA), Bollinger Bands, MACD, Stochastic Oscillator, RSI, Donchian Channels and Keltner Channel, OBV.

Given the complexity and the potentially redundant nature of combining numerous features and indicators. Based on the backward selection method, we select a comprehensive set of features to ensure a robust model. These include MPD features such as *mu*, *skew*, *kurt*, and *p50*. To capture the interactions between variables, we incorporate interaction terms like *mu·kurt*, *mu·p50*, *mu·pca*, and 21 other interaction terms. Additionally, we include technical indicators, such as the 50-day Exponential Moving Average, 3 Bollinger Bands, along with 9 other technical indicators, to further enrich our feature set.

### 4.1.3 Ensemble Voting Models

Ensemble voting models integrate various binary classification models, each with its own unique strengths and predictive capabilities. The involved models are K-Nearest Neighbors (KNN), Logistic Regression, Random Forest, SVM, Gradient Boosting Decision Tree (GBDT), XGBoost, Naive Bayes (NB), AdaBoost.

The hard voting strategy leverages the collective wisdom of multiple classifiers, reducing the likelihood of misprediction and enhancing the overall predictive performance. By aggregating the discrete predictions of each model, the ensemble effectively mitigates individual model biases and variances, aiming to deliver a more accurate and robust forecast of future price directions. Our trading strategy goes long in the Bull market and short in the Bear market.

### 4.1.4 Strategy Performance and Analysis

Table 7 describes different machine learning models' performance. The voting ensemble method outperforms most individual models. it achieves the highest total return (66.21%), annual return (16.32%), and Sharpe Ratio (94.71%). This suggests that combining the different models captures more nuances in the data than any single model alone. Logistic Regression and SVM also show strong results, particularly in total and annual returns, indicating their predictions may be more aligned with profitable trades. However, models like Random Forests and AdaBoost underperform in terms of returns, with Random Forest yielding a negative Total and annual return, which is suboptimal for investment strategies. The Sharpe Ratio across most models, except for the voting method, suggests a varying degree of risk-adjusted return, with some models showing negative values indicating a riskier or less consistent performance. Overall, the ensemble approach seems to provide a more robust prediction resulting in superior risk-adjusted returns.

Table 7: Models' Performance

| Model | Accuracy | Total Return | Annual Return | Sharpe Ratio |
|---|---|---|---|---|
| KNN | 53.85% | 13.86% | 3.94% | 30.74% |
| LR | 54.60% | 36.96% | 9.80% | 61.94% |
| RF | 49.71% | -4.50% | -1.36% | 1.06% |

| | | | | |
|---|---|---|---|---|
| SVM | 51.43% | 37.75% | 10.00% | 62.90% |
| GBDT | 49.42% | 8.40% | 2.43% | 22.48% |
| XGBoost | 50.70% | 17.17% | 5.45% | -22.91% |
| GNB | 50.41% | 0.22% | 0.07% | 54.61% |
| AdaBoost | 49.54% | -19.23% | -6.16% | -27.17% |
| Voting | 53.58% | 66.21% | 16.32% | 94.71% |

To quantify the strategy's effectiveness, we calculate the cumulative returns of both the market and the devised strategy. This involves applying the model's signals to the market's daily returns and adjusting for signal shifts to avoid lookahead bias. Figure 6 illustrates the buy and sell signals for the S&P 500 as determined by the directional voting strategy. Figure 7 delineates the performance of our strategy relative to the market. The strategy achieves an impressive total return of 166.68% during the back testing period. The annual return of 16.32% signifies the average yearly return if the total return is compounded annually. The excess annual return over the S&P 500 is mentioned as 6.65%, which indicates that the strategy not only surpasses the benchmark return but does so by a significant margin.

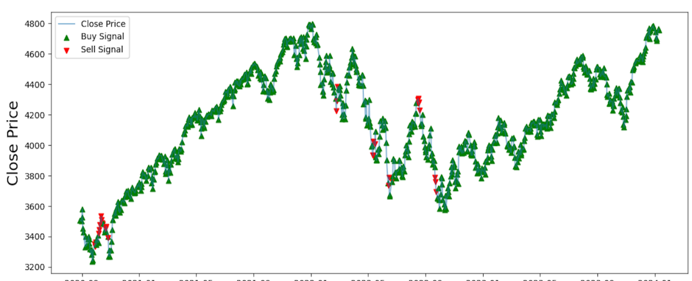


Figure 6: Buy and Sell Signals


Figure 7: Performance of Direction Voting Strategy Based on Added Information

Figure 8 delineates the trading strategy solely based on MPD as features for a voting model to predict market signals, exhibits suboptimal overall performance, with certain erroneous predictions leading to a strategy return curve that fell below actual market returns.

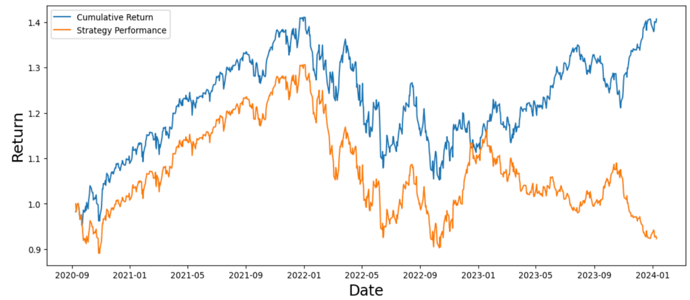


Figure 8: Performance of Direction Voting Strategy Based on MPD

Therefore, using the same voting model but combining MPD with other trading indicators significantly enhances the strategy's overall performance. This demonstrates the importance of integrating diverse data sources to improve the robustness and effectiveness of trading strategies.

## 4.2 Reversal Strategy

### 4.2.1 DC-based reversal strategy overview

Directional-change-based trading strategy is a classical quantitative trading strategy widely used in stock and FX markets. The basic idea of DC-based trading strategy is to detect potential reversals (directional change) and exploit the momentum and trends (overshoot) after reversals for profit. The core of the DC-based trading strategy is to construct a model predicting the duration of OS events following DC events. As demonstrated in the previous section, MPD data possesses explanatory power for the duration of future OS events. Therefore, in this section, we will develop a DC-based trading strategy by combining conventional features with MPD data.

The strategy's process and rules are as follows:

Step 1: Detect potential directional change and establish positions. The rules for detecting potential directional changes (DC events) are the same as previously described, with the threshold set at 4.5% to maintain consistency with the prior section. Establish long positions if the current reversal shifts to an upward trend and short positions if it shifts to a downward trend on the nest trading day.

Step 2: Predict the duration of OS events and hold positions. Use ensembled machine learning models combined with both conventional features and MPD data to predict the duration of the following OS event and maintain the position for the predicted duration.

Step 3: Close positions at the end of the predicted OS event duration. Moreover, if the next DC event is detected during the predicted duration, close positions immediately to control losses. After closing positions, repeat the above steps until the strategy concludes.

### 4.2.2 Model and Methodology

The trading target of this strategy is the S&P 500 index. This section selects data from January 2007 to June 2020 as the training set and data from July 2020 to January 2024 as the test set, conducting strategy backtesting within this period.

Constructing a model to predict OS events is the core of this strategy. The accuracy of the model's predictions will directly affect the strategy's performance. Specifically, we construct two predictive models with different features for two reversal strategies: the MPD Reversal Strategy and the Augmented Reversal Strategy.

In the MPD Reversal Strategy, the predictive model solely utilized features related to the 6-month MPD, including *mu*, *skew*, *kurt*, *p50* and *pca*.

In the Augmented Reversal Strategy, we also add conventional features including the duration of the DC event, the duration of the previous OS, the dummy variable for trend (Adegboye et al., 2017) Then we perform forward selection based on BIC to select interaction features with strong explanatory power on the training set.

The selected factors are categorized into MPD features, conventional features, and their interaction. The MPD features include the *mu*, *skew*, *kurt*, *p50*, and *pca*. The conventional features list trend, duration of the DC event, and duration of the last OS event. The interaction between these features is represented by trend·duration of the DC event, trend·duration of the last OS event, trend·*pca*, and trend·*mu*.

To enhance the model's predictive power, we employ Model Averaging for ensemble learning. Specifically, linear regression, SVR, and gradient boosting are trained as weak learners, and their prediction results are averaged to obtain the outcome of the ensemble learning model. Given that the duration of DC events is defined non-negative, all negative model prediction values are adjusted to zero. The model with both MPD and conventional features has a better predictive performance on test set as shown in Table 8. The overall high error is due to the model's difficulty in learning from historical data the prolonged upward trend from December 2020 to August 2021.

Table 8: Performance of Two Predictive Models

| | MPD features only | MPD and conventional features |
|---|---|---|
| MSE | 3545.99 | 3352.13 |
| MAE | 31.62 | 30.41 |
| $R^2$ | -0.0152 | 0.0402 |

### 4.2.3 Strategy Performance and Analysis

The performance of two reversal strategies is shown in the Figure 9. Both reversal strategies generate stable return and outperform S&P500 and the Augmented Reversal strategy performs better relatively.

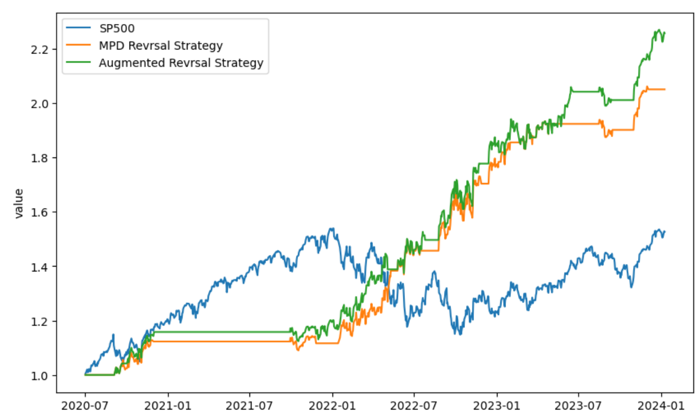

Figure 9: Reversal Strategy Performance

Specifically, the Augmented Reversal strategy achieves an annualized return of 26.04%, which is an excess annualized return of 13.40% over the S&P 500. This strategy boasts a Sharpe ratio of 1.73, with a maximum drawdown of only 6.27%. Specifically, due to the ensemble learner's failure to predict the prolonged upward trend from October 2020 to September 2021, this reversal strategy maintains a long-term no-position state during this interval.

Table 9: Method Comparison

| | Total return | Annual return | Sharpe ratio | Maximum drawdown |
|---|---|---|---|---|
| SP500 | 51.96% | 12.64% | 0.77 | -25.43% |
| MPD reversal strategy | 105.01% | 22.65% | 1.58 | -6.27% |
| Augmented reversal strategy | 125.68% | 26.04% | 1.73 | -6.27% |

Both reversal strategies generate relatively lower return during phases of long-term stable upward or downward trends due to fewer trading opportunities while they perform better in market intervals with higher volatility, benefiting from more trading opportunities. As we can see from Figure 9, both reversal strategies outperform S&P500 significantly after 2022 when the market experienced a period of higher volatility and more frequent reversals. The strategy's volatility and drawdown are low, indicating a high overall stability.

# 5 Conclusion

In this paper, we explore the predictive power of MPD on future markets and try to develop trading strategies based on MPD. First, we make comparison of MPD, and real market based on likelihood-ratio test. The test result shows that the $10^{th}$ and $50^{th}$ MPD percentiles of the 6-month return and the $10^{th}$, $50^{th}$, and $90^{th}$ MPD percentiles of the 12-month return are underestimated compared with actual market percentiles. In contrast, the $90^{th}$ MPD percentiles of the 12-month return are overestimated.

Then we construct multiple machine learning models including linear regression, random forest, SVR, and XGBoost to check the predictive power of MPD data on future returns, volatility, and reversals. For data preprocessing, we conduct feature engineering including PCA and feature selection on MPD data and feed it to the machine learning models. The out-of-sample result shows relatively strong predictive power on future reversals but low predictive power on future returns and volatilities.

Based on our findings of MPD's predictive power, we develop two strategies: direction voting strategy and reversal strategy. In the direction voting strategy, we construct voting models to classify the direction of the market and establish positions accordingly. The performance of the direction voting strategy is not very stable due to the low predictive power of MPD on future returns. In the reversal strategy, we detect direction changes and predict the duration of OS event based on both MPD data and conventional features to exploit reversals for profit. The reversal strategy generates higher and stable return, yielding 19.18% excess annual return over SP500 and Sharpe ratio of 1.73. The superior performance of the Reversal strategy further corroborates the predictive power of MPD data regarding future reversals.

# References


[1]. Adegboye, A., & Kampouridis, M. (2021). Machine learning classification and regression models for predicting directional changes trend reversal in FX markets. Expert Systems with Applications, 173, 114645.

[2]. Adegboye, A., Kampouridis, M., & Johnson, C. G. (2017). Regression genetic programming for estimating trend end in foreign exchange market. 2017 IEEE Symposium Series on Computational Intelligence (SSCI).

[3]. Breeden, D. T., & Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. The Journal of Business, 51(4), 621–651.

[4]. Federal Reserve Bank of Minneapolis. (n.d.). Current and historical market-based probabilities. Retrieved from https://www.minneapolisfed.org/banking/current-and-historical-market--based-probabilities.

[5]. Shimko, D. C. (1993). Bounds of probability. Risk, 6(4), 33-37.

[6]. Skouras, S. (2001, January 23). Risk neutral forecasting. SSRN. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=254291.