# GSPTSE Directional Forecasting via U.S. Market Signals and Technical Indicator

Beibo Jiang

The Ohio State University, Columbus, OH 43210, USA
jiang.2643@osu.edu

**Abstract.** With the deepening of economic globalization and the international flow of financial capital and assets, the stock markets of different countries tend to be interconnected, and there will be correlations and co-movement among the stock markets globally or regionally. This study aims to predict the daily directional movement, whether upward movement or downward movement, of the Canadian major index S&P / TSX Composite Index (GSPTSE) by using the major stock market indices of the United States and technical indicators. The study used six major US market indices, the volatility index (VIX), the Relative Strength Index (RSI), Moving Average Convergence and Divergence (MACD), and MACD based on historical volatility (MACD-HVIX), and lagging returns for prediction. MACD-HVIX is an indicator based on dynamic volatility adjustment from MACD. Linear model Logistic Regression, and two nonlinear models, Random Forest and eXtreme Gradient Boosting (XGBoost) were applied. The research results showed that nonlinear models performed better than the linear model, and random forest is better than XGBoost. Moreover, the study found that the effect of XGBoost was much better than the other two models when using the MACD-HVIX indicator.

**Keywords**: Machine Learning, Cross-Market Forecasting, MACD-HVIX.

## 1 Introduction

With the development of globalization, international trade and economic exchanges between countries have become increasingly frequent, reaching an indispensable economic cooperation. Financial capital is also accelerating its global market layout. The team behind financial capital needs to analyze a country's stock market index or a specific stock to determine their amount of holding in this asset based on data and other countries' or regions' economic data. The interconnection or co-movement among global markets, or the non-isolation of markets, has long been proven. The working paper of the International Monetary Fund on international stock markets indicates that there is a long-term equilibrium relationship among global indices [1]. For instance, the Shanghai and Hang Seng indices in China show a close correlation [2]. In the North American region, the United States and Canada are also correlated. Similar geographical location, trade and policies, and the flow of financial capital between the


A. J. Moshayedi (ed.), *Proceedings of the 2025 International Conference on Hybrid Commerce, Human Capital, and Economic Dynamics (ICHCH 2025)*, Advances in Economics, Business and Management Research 374,
https://doi.org/10.2991/978-2-38476-585-0_30

United States and Canada, make their stock markets have a certain degree of synergy. In recent research, Hussian and Omrane also discovered that the announcement of the US macroeconomic news affects the stock market in Canada, which provides strong evidence to support this research that there is actually have relationship between those two countries' markets [3].

The S&P/TSX Composite Index (GSPTSE) in Canada is the most representative index of the Canadian stock market. The United States has a larger market size, and its stock indices can more quickly reflect global market sentiments and expectations. Although GSPTSE has its operational logic, it often shows a lag or synchronous response to the trends of the US stock market.

Machine learning can capture the nonlinear and complex relationships as well as the underlying patterns among markets, and by continuously iterating, it can observe the markets from a dynamic perspective. Khaidam et al. used random forests and gradient boosting machines to predict the upward movement or downward movement of stocks, which shows that machine learning can learn complex nonlinear relationships, which is particularly important for the stock market that is difficult to explain with linear logic and proves the rationality of machine learning in predicting the stock market [4]. In addition to the stocks themselves, technical indicators can also be widely applied to prediction. The research of Sami confirmed that Moving Average Convergence and Divergence (MACD), as well as Relative Strength Index (RSI), can provide high-quality signals [5, 6]. The research of Rosillo proved that the Volatility Index (VIX) has a higher accuracy in predicting stock trends than the accuracy without VIX and pointed out that VIX is more accurate for predicting bearish stock markets [7]. On this basis, Wang and Kim found that using MACD calculated by weighting with historical VIX (HVIX) can make the signals more stable [8].

Most previous studies were based on local features and technical indicators for predicting the directional movement of the domestic market, but this study combines US-based market indices and various technical indicators and uses machine learning models to predict the upward movement or downward movement of the Canadian market.

# 2 Methodology

## 2.1 Dataset

The dataset is from Kaggle [9]. The two datasets are combined based on time and exchange symbol to expand the time range, from 2013-04-01 to 2024-04-05, making the ratio of increase/decrease of GSPTSE in the test set as 51%/49%, which is a nearly balanced dataset. The original dataset contains data from over thirty major global exchanges, but here only the major indices in the United States (New York stock exchange composite index (NYA), S&P 500 (GSPC), Dow Jones Industrial Average (DJI), NASDAQ Composite Index (IXIC), Russell 2000 (RUT), NYSE AMEX Composite Index (XAX)) and the COBT Volatility Index (VIX), and the Canadian index (GSPTSE) are used for prediction. The data is filled with forward values for missing values to ensure the dataset's completeness. The training set contained 2448 dates, and the test set contained 613 data points, with a ratio of 4:1.

## 2.2 Features

The goal is to predict the upward movement or downward movement of the Canadian Stock Exchange Composite Index (GSPTSE) on the next day. In this binary classification problem, 1 represents the upward movement of GSPTSE, and 0 represents the downward movement. To capture short-term returns, the lagged logarithmic returns of the GSPTSE within a maximum of five trading days are recorded. The daily return of the US index is calculated by weighting it with the correlation to the GSPTSE. The technical indicator RSI is calculated as follows:

$$RSI_t = 100 - \frac{100}{1 + \frac{AvgGain_{14,t}}{AvgLoss_{14,t}}} \tag{1}$$

where $AvgGain_{14,\ t}$ means the past 14 trading days' average increase, and $AvgLoss_{14\ ,t}$ means the past 14 trading days' average decrease.

MACD is an indicator used to measure the intensity, direction and momentum of trend changes. This study will compare the ordinary MACD with MACD-HVIX for prediction. The calculation of the ordinary MACD is as follows:

$$\begin{aligned} \text{MACD}_t &= EMA_{12,t} - EMA_{26,t} \\ \text{Signal}_t &= EMA_9(\text{MACD}_t) \\ \text{Histogram}_t &= \text{MACD}_t - \text{Signal}_t \end{aligned} \tag{2}$$

$\text{Signal}_t$ indicates the 9-day exponential moving average (EMA) of MACD serves as the signal line. $\text{Histogram}_t$ is the difference between MACD and the Signal line.

Among them, EMA stands for Exponential Moving Average, and its formula is:

$$EMA_t = \alpha \cdot P_t + (1 - \alpha) \cdot EMA_{t-1}, \quad \alpha = 2/(N + 1) \tag{3}$$

α represents the smoothing factor, Pt is the index value, and N is the period length. In the MACD-HVIX, α undergoes dynamic weighting as:

$$a_t = \frac{HVIX_t}{\sum_{i=t-m}^{t} HVIX_i} \tag{4}$$

m = 12 in this case, as the length of the period being traced back. $HVIX_i$ in here is calculated by a rolling sum ratio of standard deviations based on volatility.

The remaining calculation process remains unchanged. The advantage of this new feature is that it reduces the possible lag of the traditional MACD algorithm. In summary, the indicators used for prediction in this study are the weighted log return rate of the US market, the lagging return rate of GSPTSE itself, the VIX index, RSI, and MACD/MACD-HVIX.

### 2.3 Models

Three models were used: Logistic Regression, XGBoost, and Random Forest. The purpose was to explore the differences in linear and non-linear structures in prediction. Logistic Regression represents a linear structure and can assess the fitting of the linear relationship to the problem. Logistic regression is computationally efficient and interpretable, but its drawback here is that a linear relationship must exist. Random Forest and xgboost are both tree models and can handle non-linear relationships. Compared to Logistic Regression, they are more suitable for non-linear predictions such as stock market predictions. Random Forest is trained using the bagging method, while XGBoost uses gradient optimization to reduce errors. The main evaluation metrics are AUC-ROC and F1-score, and accuracy, precision, and recall rate will also be reported.

## 3 Result

The test set contains 613 samples, among which 343 are downward movement samples and 270 are upward movement samples.

**Table 1.** Test outcome

| Model | Feature Set | Accuracy | Precision | Recall | F1-Score |
|---|---|---|---|---|---|
| Logistic Regression | MACD | 0.506 | 0.514 | 0.506 | 0.507 |
| Logistic Regression | MACD-HVIX | 0.504 | 0.512 | 0.504 | 0.506 |
| XGBoost | MACD | 0.548 | 0.553 | 0.548 | 0.550 |
| XGBoost | MACD-HVIX | 0.577 | 0.588 | 0.577 | 0.579 |
| Random Forest | MACD | 0.593 | 0.607 | 0.594 | 0.595 |
| Random Forest | MACD-HVIX | 0.584 | 0.604 | 0.584 | 0.583 |

As shown in Table 1, the accuracy rates of the three models are all above 50%, which means the models have indeed captured useful signals from the test set. The prediction results of the nonlinear models (XGBoost and Random Forest) also meet expectations and are much better than those of the linear model (Logistic Regression), because the stock market is influenced by multiple dimensions, and it is difficult to explain it solely with a linear model. The accuracy rate and F1-score of Random Forest are both higher than those of XGBoost.

**Table 2.** XGBoost comparison with different features

| Evaluation | Class | MACD | MACD-HVIX | Difference |
|---|---|---|---|---|
| Precision | 0 | 0.604 | 0.643 | +0.039 |
| | 1 | 0.488 | 0.517 | +0.029 |
| Recall | 0 | 0.557 | 0.551 | -0.006 |

| | | | | |
|---|---|---|---|---|
| | 1 | 0.537 | 0.611 | +0.074 |
| F1-Score | 0 | 0.580 | 0.593 | +0.013 |
| | 1 | 0.512 | 0.560 | +0.048 |
| Accuracy | | 0.548 | 0.577 | +0.029 |

As shown in Table 2, based on the comparison of the predictions made by the XGBoost model for the MACD and MACD-HVIX feature sets, it can be observed that using the MACD-HVIX feature set has significantly enhanced the model, especially for the classification ability for the upward movement has been improved, with the recall rate of the upward movement increasing by 0.07 and the F1-score increasing by 0.04. This confirms that the dynamically weighted MACD indicator, after being adjusted, to some extent alleviates the lagging problem of the traditional MACD indicator.

**Table 3.** Random Forest comparison with different features

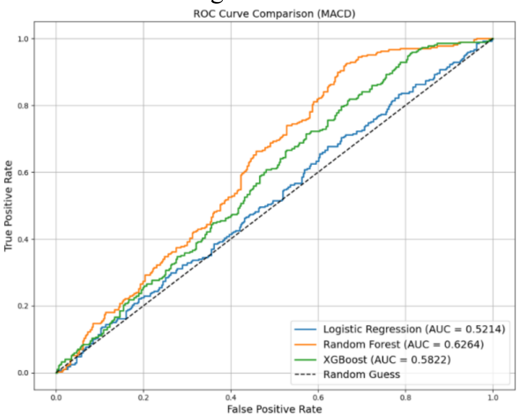
| Evaluation | Class | MACD | MACD-HVIX | Difference |
|---|---|---|---|---|
| Precision | 0 | 0.667 | 0.669 | +0.002 |
| | 1 | 0.532 | 0.521 | -0.011 |
| Recall | 0 | 0.548 | 0.507 | -0.041 |
| | 1 | 0.652 | 0.682 | +0.030 |
| F1-Score | 0 | 0.602 | 0.577 | -0.025 |
| | 1 | 0.586 | 0.591 | +0.050 |
| Accuracy | | 0.594 | 0.584 | -0.010 |

Table 3 is a comparison based on the Random Forest model. The overall accuracy has decreased. At the same time, a performance transfer phenomenon can be observed. It can be found that the recall and F1-score of this model have both decreased significantly when predicting downward movement. Perhaps using new features weakens the capture of the characteristics of the downward movement and enhances its prediction of the rising market. The tree model's responsiveness to new features is not as rapid as the gradient boosting model, and it cannot significantly improve performance like the gradient boosting model. However, the overall accuracy is still higher than that of the XGBoost model.

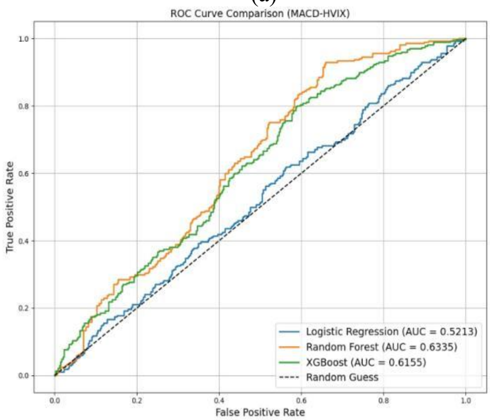
**Table 4.** Logistic Regression comparison with different features

| Evaluation | Class | MACD | MACD-HVIX | Difference |
|---|---|---|---|---|
| Precision | 0 | 0.566 | 0.564 | -0.002 |
| | 1 | 0.447 | 0.445 | -0.002 |
| Recall | 0 | 0.499 | 0.499 | 0.000 |
| | 1 | 0.515 | 0.511 | -0.004 |
| F1-Score | 0 | 0.530 | 0.529 | -0.001 |
| | 1 | 0.479 | 0.476 | -0.003 |
| Accuracy | | 0.506 | 0.504 | -0.002 |

Table 4 shows a comparison based on the Logistic Regression model, with the overall performance slightly deteriorating and all indicators experiencing negative changes. The MACD-HVIX brings about dynamic nonlinear changes, which exceed the capacity of linear models. This once again highlights the significant limitations of linear models in dealing with complex issues. From the data of random forest and XGBoost, the introduction of MACD-HVIX has a more obvious effect on judging bullish trends, which holds certain significance.




(a)




(b)

**Fig. 1.** ROC-AUC for different features (Photo/Picture credit: Original).

Fig. 1 shows the ROC-AUC curve of the results. Figure 1-a presents the results of the traditional MACD, and Figure 1-b shows the results of MACD-HVIX. In both cases, Random Forest performed the best. However, XGBoost was the biggest beneficiary, as its AUC increased from 0.5822 to 0.6155, and the curve moved overall towards the upper left corner, widening the gap with a random guess, indicating that MACD-HVIX significantly enhanced the classification ability of the XGBoost model. The Logistic Regression curve fluctuated around a random guess, and the AUC value changed little in the two tests, demonstrating the limitations of the linear model and its inability to utilize volatility information to improve itself.

## 4 Discussion

This study has several limitations. This research only used the US market as a feature for prediction and did not consider the markets of other countries or regions. Other economies or commodity markets can also provide effective signals. The data volume is also questionable. Currently, the test set only has 613 data points, which is relatively small. In the future, more data sets will be appropriately sought to expand the data volume. Also, more technical indicators can be considered. As indicated in the research by Dang and Duong, market sentiment and media articles are also important factors for investors. Incorporating the analysis of financial news through text mining and making predictions can achieve an accuracy rate of 73% [10]. In addition, the existing models can be improved. The improved algorithm FSRF proposed by Feng et al. for Random Forest was improved during the preprocessing, extracting subsets with better classification performance, which can improve the final accuracy [11].

## 5 Conclusion

This article aims to predict the upward movement and downward movement of the Canadian stock market index (GSPTSE) for the next day by using the US stock market index and other technical indicators, with three machine learning models. This article also cites the variant of MACD, MACD-HVIX, to explore the usability of this indicator in this model and its improvement on model performance. Overall, this study reveals that the performance of nonlinear models is significantly better than that of linear models. XGBoost and Random Forest outperform Logistic Regression in all indicators, illustrating that nonlinear relationships play a more significant role in financial time series. XGBoost shows a significant improvement in MACD-HVIX, especially in the bullish aspect, which is of great significance for judging bull markets. The highest accuracy rates of Random Forest and XGBoost are 59.4% and 57.7% respectively, higher than the random guess of nearby 50%, but it is still questionable whether they can be truly applied in reality. The linear model Logistic Regression cannot fully understand the prediction signal, and its accuracy is no different from random selection. This indicates that linear models have certain limitations in complex financial prediction environments. In the future, the model may continue to be improved. Integrating some macroeconomic features, such as interest rate, and using deep learning models such as Long Short-Term Memory (LSTM) could enhance the model's performance.

## References

1. Wu, X., Zhu, S., Bai, Z.: A study on the dynamic linkage between A-shares and H-shares based on the MRS- SJC-Copula model. Operations and Management **29**(1), 9 (2020)
2. Chen, R.S., Lucey, B.M.: Cointegration of international stock market indices. IMF Working Paper, WP/15/284, International Monetary Fund (2015).

https://www.imf.org/en/Publications/WP/Issues/2016/12/30/Cointegration-of-International-Stock-Market-Indices-874
3. Sami, J.: Stock market investment and inflation: Evidence from the United States and Canada. Review of Economic Analysis (2021). https://openjournals.uwaterloo.ca/index.php/rofea/article/view/4047
4. Khaidem, L., Saha, S., Dey, S.R.: Predicting the direction of stock market prices using random forest. arXiv preprint arXiv:1605.00003 (2016)
5. Sami, H.M., Ahshan, K.A., Rozario, P.N., Ashrafi, N.: Evaluating the prediction accuracy of MACD and RSI for different stocks in terms of standard market suggestions. Canadian Journal of Business and Information Studies **4**(6), 137–143 (2022)
6. Anghel, G.D.I.: Stock market efficiency and the MACD. Evidence from countries around the world. Procedia Economics and Finance **32**, 1414–1431 (2015)
7. Rosillo, R., Giner, J., de la Fuente, D.: The effectiveness of the combined use of VIX and Support Vector Machines on the prediction of S&P 500. Neural Comput. Appl. **25**, 321–332 (2014)
8. Wang, J., Kim, J.: Predicting stock price trend using MACD optimized by historical volatility. Math. Probl. Eng. **2018**(PT.17), 9280590.1–9280590.12 (2018)
9. Ahmad, M.: 33 World-wide Stock Market Indices Data. Kaggle (2022). https://www.kaggle.com/datasets/mukhazarahmad/worldwide-stock-market-indices-data
10. Dang, M., Duong, D.: Improvement methods for stock market prediction using financial news articles. In: 2016 3rd National Foundation for Science and Technology Development Conference on Information and Computer Science (NICS), pp. 125–127 (2016)
11. Feng, W., Ma, C., Zhao, G., Zhang, R.: FSRF: An improved random forest for classification. In: 2020 IEEE International Conference on Advances in Electrical Engineering and Computer Applications (AEECA), pp. 173–178 (2020)