

RESEARCH ARTICLE

Volatility-informed SPY forecasting: From CGR-SPY analysis to DLSTM prediction

Hong Cao¹, Hao Chen², Yulong Lian^{3*}, Hong-Kun Zhang²

1 Columbia University, New York, New York, United States of America, **2** University of Massachusetts Amherst, Amherst, Massachusetts, United States of America, **3** Carnegie Mellon University, Pittsburgh, Pennsylvania, United States of America

* yulongli@andrew.cmu.edu

Abstract

Forecasting stock returns is a vital and at the same time challenging task in the financial arena, given the market's susceptibility to abrupt swings. In this paper, we propose a strategy that adapts to different volatility regimes: during periods of high volatility, we employ the copper-gold ratio (CGR) as a leading indicator for the S&P 500 (SPY), while in periods of normal volatility, we introduce a differential long-term memory (DLSTM) neural network. The CGR combines the properties of copper (which reflects industrial and economic activity) and gold, a traditional safe-haven asset. In four major economic events, our analysis reveals that sharp movements in the CGR often precede corresponding changes in the SPY, suggesting the ratio's potential as an early warning signal. For more stable markets, we introduce the DLSTM, which extends the standard LSTM architecture through a loss function designed to exploit differences between consecutive price steps. This design increases predictive power and achieves 82% directional accuracy on daily SPY forecasts, outperforming both a baseline LSTM and a binary classification approach. Finally, we validate the trading utility of the DLSTM by simulating intraday trading over one- and three-month periods, demonstrating consistent gains that highlight the practical value of the method. By synthesizing CGR analysis and DLSTM modeling, our approach offers a versatile framework to address diverse market environments and provide new insights for both researchers and practitioners.

OPEN ACCESS

Citation: Cao H, Chen H, Lian Y, Zhang H-K. (2025) Volatility-informed SPY forecasting: From CGR-SPY analysis to DLSTM prediction. *PLoS Complex Syst* 2(8): e0000037. <https://doi.org/10.1371/journal.pcsy.0000037>

Editor: Keith Burghardt, University of Southern California, UNITED STATES

Received: January 17, 2024

Accepted: February 20, 2025

Published: August 5, 2025

Copyright: © 2025 Cao et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All data necessary is included in the supplemental file.

Funding: The author(s) received no specific funding for this work.

Competing interests: The authors have declared that no competing interests exist.

Author summary

Predicting stock movements can be challenging, especially when markets change rapidly. This study addresses the challenge by proposing a two-part approach to forecasting the S&P 500 (SPY) under different market conditions. During periods of high volatility, we use the copper-to-gold ratio (CGR). Copper's value often tracks economic activity, while gold remains stable during financial turmoil, making their ratio a potential early warning indicator of stock price movements. Then, during quieter times, we introduce a specialized neural network called the difference long-short-term memory (DLSTM). The LSTM

is a machine learning tool designed to remember important information over time, which is critical for financial data. Our DLSTM extends the LSTM by focusing on price changes between consecutive days, resulting in stronger predictive power. Tests show that it achieves 82% accuracy in predicting daily price trends, outperforming simpler models. We also validate these predictions using a simulated trading strategy that shows consistent gains over a one- and three-month period. By combining CGR and DLSTM, this study provides an adaptable approach for investors navigating today's unpredictable markets.

1. Introduction

Forecasting stock returns is a crucial task for investors and financial analysts alike. The ability to predict market movements translates directly into the potential for higher investment returns. However, the unpredictable nature of the stock market, particularly during periods of high volatility, poses significant challenges to accurate forecasting. This paper suggests employing the Copper-Gold Ratio (CGR) as a market indicator during turbulent times and leveraging a novel neural network, the Difference Long Short-Term Memory (DLSTM), for more stable periods. The CGR and DLSTM methods combine to create a two-part approach for predicting the SPY.

During times of heightened market fluctuations, we investigate the CGR as a possible precursor to the movements of the S&P 500 (SPY). Copper's significance in the industrial sector often mirrors economic trends, and gold's status as a safe haven during financial instability provides a counterbalance [1,2]. While CGR's role as a leading indicator for the 10-year Treasury Yield is well-documented [3], its application in stock forecasting is less explored. We aim to strengthen the evidence supporting CGR's leading influence on SPY during these volatile episodes.

In periods of normal volatility, we employ DLSTM. However, let us firstly consider some common tools in the realm of neural networks. While traditional feed-forward neural networks are applicable, the Long Short Term Memory (LSTM) model stands out with its ability to capture the temporal dependencies in market data [4]. However, research has been done to show that improvements in structure to LSTM may result in significant accuracy increases in prediction. An example of this is [5] where the authors added an attention mechanism to denoise historical stock data. Furthermore, a limitation of the traditional LSTM is that it strictly uses only the given stock price information at face value (i.e. only adjusted closing price). Thus, we are motivated to extract as much information from the given stock price information by employing a novel loss function which measures the difference between consecutive steps in the network, which pushes us to our ultimate goal of further improving the accuracy which may be achieved by LSTM. In other words, our objective is to beat the baseline LSTM Directional Accuracy with DLSTM. Therefore, we aim to further improve the accuracy which may be achieved by LSTM by employing a novel loss function which measures the difference between consecutive steps in the network; with this, the DLSTM achieves an 82% directional accuracy. We demonstrate the model's practical applications through simulations, demonstrating substantial gains in short-term trading scenarios.

The paper is organized as follows: [Sect 2](#) introduces the CGR-SPY relationship, with [Sects 2.1 to 2.4](#) analyzing this relationship during distinct economic crises. [Sect 3](#) introduces the DLSTM network and its unique loss function. We also present the empirical results, including the DLSTM's predictive performance and its application in simulated trading. Finally, [Sect 4](#) provides a conclusion which summarizes our findings.

2. Correlations of Copper/Gold Ratio (CGR) and SPY during crisis periods

Gold has long been valued as a stable currency, maintaining its worth even during economic downturns, thus often signaling pessimistic market sentiment [6]. Central Banks rely on gold to secure depositor funds, manage external debts, and control inflation. Conversely, copper, widely used in industries ranging from medical applications to renewable energy, reflects global economic growth trends [1]. Its demand is projected to significantly increase by 2050 [7], with its prices offering insights into economic development. While copper shortages are unlikely due to advancements in technology and exploration, its role in infrastructure development, especially in emerging economies, and its price movements, are key indicators for financial markets and policymakers in gauging global economic activity. The Copper-Gold Ratio (CGR) is calculated by dividing the market price of copper by the gold price. Various market participants have noted that the CGR tends to align closely with the 10-year U.S. Treasury yield, serving as a potential leading indicator for the latter [8].

This study sets out to investigate the statistical properties of CGR using a selection of economic indicators, including Silver, Oil, and the 10-Year Treasury Note (TNX). The data for this investigation were sourced from Yahoo Finance, capturing the daily closing prices or yield rates of the chosen indicators. The dataset spans a 15-year period from January 2007 to July 2022, totaling 3916 data points. Upon assessment, the mean CGR was determined to be 0.0025, with a standard deviation of 0.00079. The trend, seasonality, and residual aspects of the CGR can be visualized in Fig 1. It's evident that it has an almost periodic seasonality. The trend doesn't display a consistently upward trajectory; instead, it exhibits interesting patterns.

To demonstrate that the Gold/Copper ratio (CGR) can serve as an indicator for financial crises, particularly in relation to the S&P 500 Index (SPY), we conducted a rolling correlation analysis involving multiple time series: CGR, S&P 500 Index (SPY), 10-Year Treasury Note Yield (TNX), Crude Oil Prices (Oil), and Silver Prices (Silver).

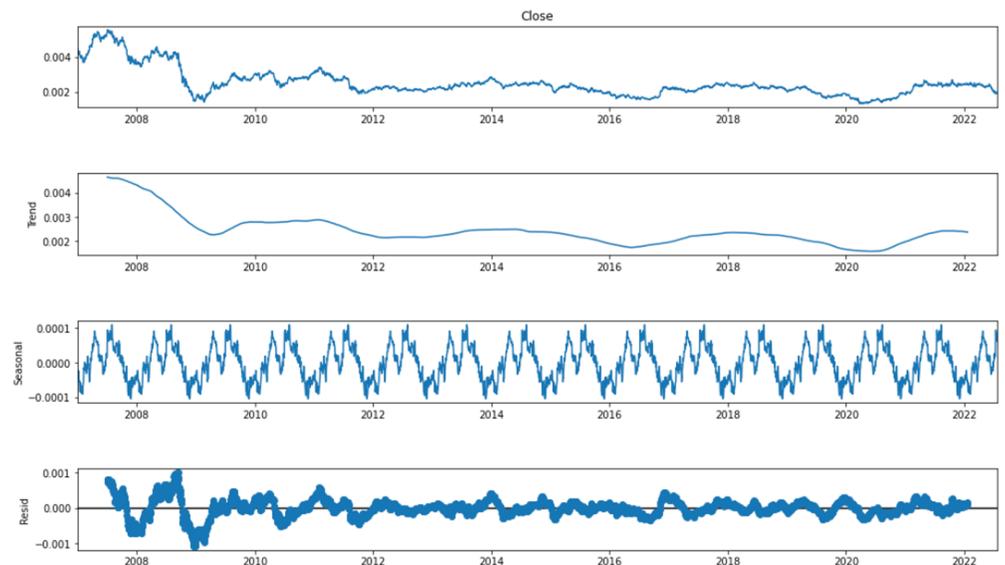


Fig 1. The trend, seasonality and residual of CGR from 2007 to 2022.

<https://doi.org/10.1371/journal.pcsy.0000037.g001>

The rolling correlation provides a dynamic view of how the relationships between these assets change over time. For this analysis, we used a 5-day rolling window, where each point on the plot represents the pairwise correlation between CGR and the other assets over the preceding 5-day period. This approach enables us to capture short-term variations in their correlations, highlighting how the relationships evolve in response to different market conditions.

By analyzing the rolling correlations, we can observe patterns that indicate periods of heightened or diminished alignment between CGR and SPY, which may correspond to times of economic stress or stability.

The following sections delve into specific periods of interest, examining the correlation between the Copper/GOLD ratio and SPY. Our analysis demonstrates that during periods of financial crisis, CGR tends to show stronger correlations with SPY and Oil. This behavior suggests that CGR could act as a leading indicator of market shifts, providing valuable insights into potential crises.

2.1. 2008 Financial crisis (July 1, 2008–October 30, 2008)

The 2008 Financial Crisis, a severe worldwide economic crisis, was precipitated by the bursting of the United States housing bubble, which culminated in the bankruptcy of Lehman Brothers. Banking institutions globally were under immense strain, leading to sharp declines in consumer wealth, severe disruptions in financial markets, and a downturn in economic activity. Fig 2 illustrates CGR's strong performance during the 2008 Lehman crisis, a period marked by extreme market volatility. The positive correlation (0.90) between CGR and SPY highlights CGR's effectiveness in capturing shifts in investor sentiment during crises. During such volatile periods, CGR provides valuable insight into market-wide shifts by reflecting economic trends and investor sentiment. Its utility, however, is more limited during stable periods when macroeconomic fluctuations are less pronounced.

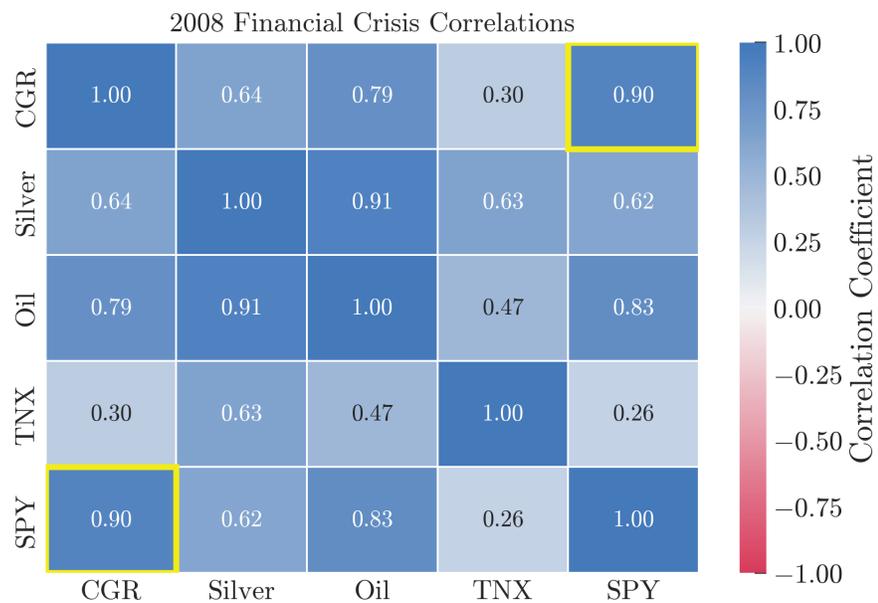


Fig 2. CGR demonstrates a strong positive correlation (0.90) with SPY during the 2008 Lehman crisis, highlighting its ability to capture market sentiment under volatile conditions.

<https://doi.org/10.1371/journal.pcsy.0000037.g002>

During the given interval, we analyze the correlation between the 5-day lagged GCR and SPY, as depicted in Fig 3. A strong positive correlation of 0.90 emphasizes the deep interconnection of global economic sentiments at the time.

Both assets declined in price amid widespread uncertainty. Their simultaneous decline indicates that as confidence in global economic growth fell, equity markets fell. Notably, CGR began its descent on September 12, 2008, nine days before SPY's downturn on September 19, 2008, as illustrated in Fig 3.

The precursory drop in CGR signaled growing investor pessimism about the economic future, possibly prompting a shift from growth assets like copper to safe havens like gold. This CGR decline may have been a warning for imminent market plunges.

2.2. European debt crisis (April 1, 2010–July 31, 2012)

Emerging in the aftermath of the 2008 Financial Crisis, the European Debt Crisis was sparked by high sovereign debts in countries like Greece, Spain, and Portugal. As austerity measures were put in place and bailouts sought, the eurozone grappled with recession. In Fig 4, we observe the correlation of CGR with SPY as compared to the correlation of SPY with other indicators during this period.

2.3. Trade war fears (June 1, 2018–December 31, 2019): Positive correlation

Between June 1, 2018, and December 31, 2019, the copper-gold ratio had a positive correlation with the SPY (S&P 500 ETF) due to market sentiment influenced by the U.S.-China trade war. During this period, trade tensions created uncertainty, affecting global growth expectations. However, positive developments or progress in trade talks led to optimism, driving up both copper prices and the SPY. This alignment in investor sentiment resulted in a positive correlation between the copper-gold ratio and SPY. In Fig 5, we observe the correlation between SPY and CGR as compared to the correlation between SPY and other indicators during this period.

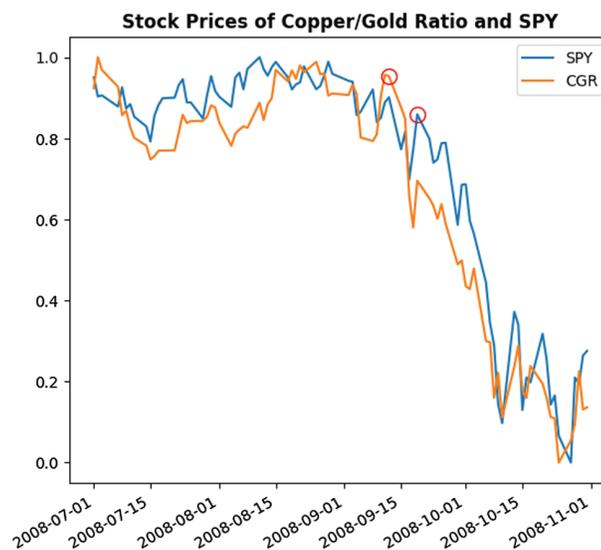


Fig 3. The CGR can be viewed as a prediction indicator for the two sudden falls of SPY in the 2008 Financial Crisis.

<https://doi.org/10.1371/journal.pcsy.0000037.g003>

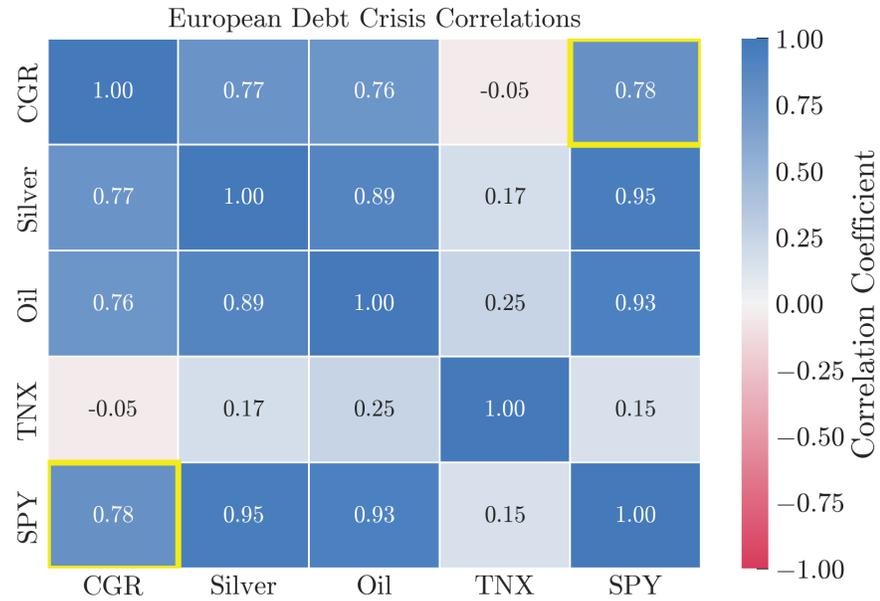


Fig 4. Positive correlation of 5-day lagged CGR with SPY during the 2009 European Debt Crisis.

<https://doi.org/10.1371/journal.pcsy.0000037.g004>

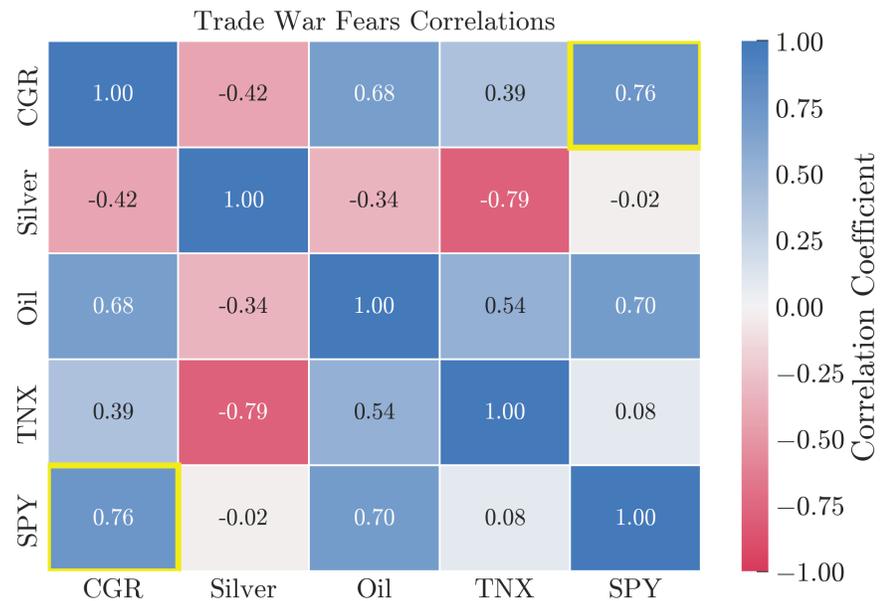


Fig 5. Positive correlation of 5-day lagged CGR with SPY during the Trade War Fears.

<https://doi.org/10.1371/journal.pcsy.0000037.g005>

2.4. 2018 Fall Selloff (September 23, 2018–December 16, 2018): Positive correlation

During the 2018 selloff, a tumultuous period emerged as global markets were shaken by a myriad of challenges. Tightening monetary policies, concerns regarding global economic growth, and escalating geopolitical tensions were at the forefront. Throughout this phase,

the CGR's fluctuations, paralleled the trajectory of SPY. In Fig 6, we observe the correlation between SPY and CGR as compared to the correlation between SPY and other indicators.

It's worth noting that the CGR initiated its decline ahead of SPY, marked by a noticeable dip on November 28, 2018. This precursory trend in CGR was succeeded by a pronounced drop in SPY, commencing distinctly on December 3, 2018, as illustrated in Fig 7.

2.5. Correlation analysis during a stable market period

The period from January 2007 to January 2008 is characterized as a stable market period for the SPY, as evident from its relatively steady price movements and the absence of abrupt

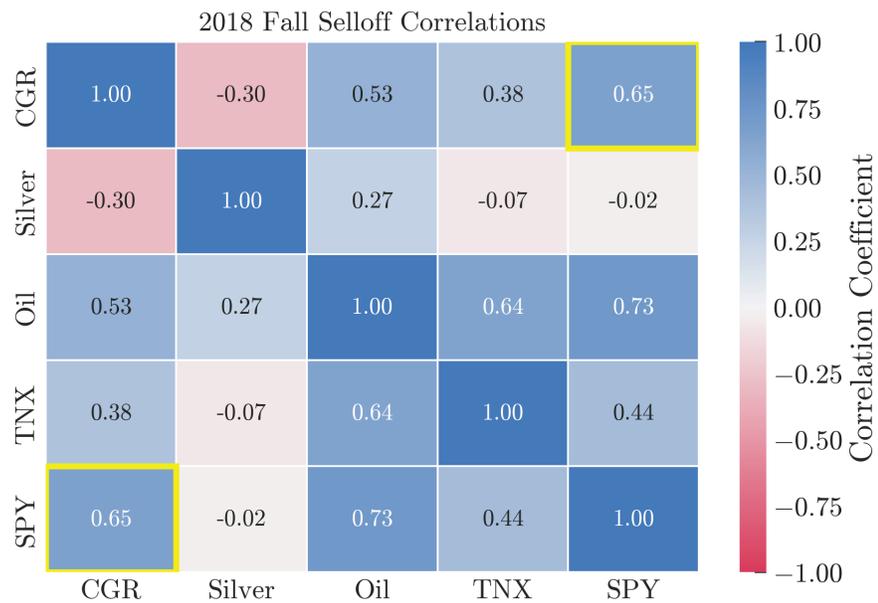


Fig 6. Positive correlation of 5-day lagged CGR with SPY during the 2018 Fall Selloff.

<https://doi.org/10.1371/journal.pcsy.0000037.g006>

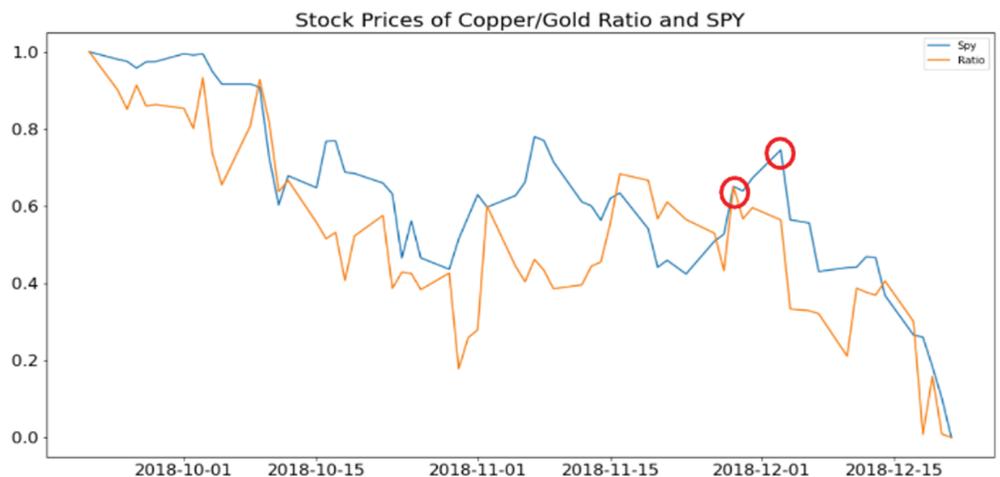


Fig 7. The CGR can be viewed as a prediction indicator for the two sudden falls of SPY in the 2018 Fall Selloff.

<https://doi.org/10.1371/journal.pcsy.0000037.g007>

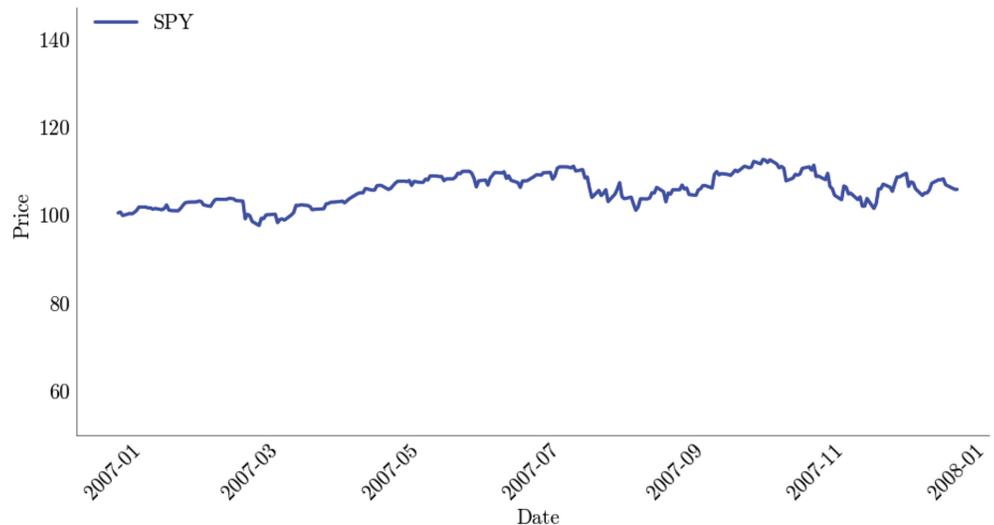


Fig 8. SPY price movements during a stable market period (2007–2008). The steady price trends highlight the absence of significant volatility, establishing this period as a baseline for correlation analysis.

<https://doi.org/10.1371/journal.pcsy.0000037.g008>

market disruptions, as shown in Fig 8. This is in contrast to volatile periods such as the 2008 Lehman crisis, where sharp declines and heightened uncertainty dominated market behavior. The stability during this period provides a baseline for analyzing the relationship between CGR and SPY under calm market conditions.

Fig 9 presents the correlation matrix of CGR and SPY during this stable period. Unlike during volatile periods, where CGR exhibits a strong correlation with SPY, the correlation during this calm period is notably reduced (0.32). This observation underscores CGR's limited relevance in stable market conditions, where macroeconomic factors play a less prominent role in influencing market behavior. Instead, CGR's utility becomes more evident during periods of heightened market volatility, aligning with its design as a tool for detecting shifts in investor sentiment and market dynamics under turbulent conditions.

2.6. Volatility analysis and correlation

In this section, we analyze the correlation between the SPY (S&P 500 ETF) and CGR (Copper Gold Ratio) during various significant financial events. Additionally, we examine the behavior of the VIX (Volatility Index) during these periods.

The VIX is an index that is computed in real time on each trading day, that is often referred to as the investor fear gauge [9]. It provides a 30-day forward projection of volatility and is a widely used measure of market risk and investors' sentiments. Higher VIX values indicate greater market uncertainty and higher expected volatility [10].

Table 1 summarizes the correlation between SPY and CGR along with the average and maximum values of the VIX during four major financial events. We observe that both the average VIX and the maximum VIX during each crisis period are proportional to the correlation between SPY and CGR.

From the table, it is evident that the highest correlation between SPY and CGR occurred during the 2008 Financial Crisis, which also saw the highest VIX values, reflecting significant market turmoil and investor fear. The European Debt Crisis and Trade War Fears similarly show elevated VIX levels corresponding to high correlations, indicating periods of heightened

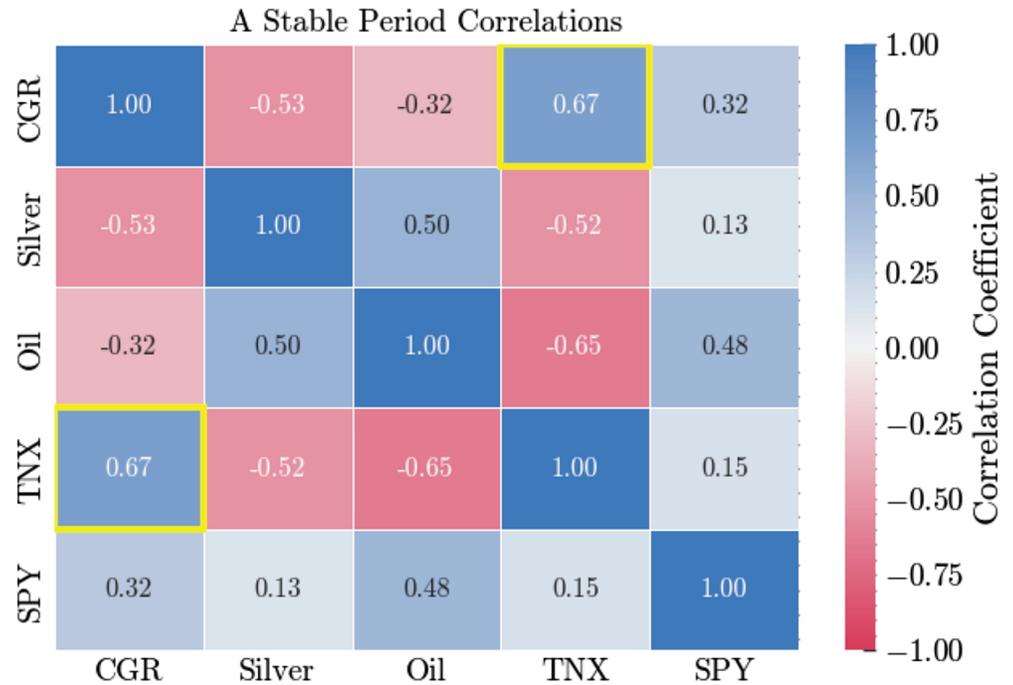


Fig 9. Correlation matrix of CGR and SPY during a stable market period (2007–2008). The reduced correlation highlights CGR's limited applicability in calm market conditions.

<https://doi.org/10.1371/journal.pcsy.0000037.g009>

Table 1. The average and maximum values of the VIX during each crisis period are shown alongside the correlation between SPY and CGR.

Event	Correlation of SPY & CGR	VIX _{ave}	VIX _{max}
2008 Financial Crisis	0.90	33.988	80.06
European Debt Crisis	0.78	22.607	48.0
Trade War Fears	0.76	15.797	36.07
2018 Fall Selloff	0.65	19.035	25.23

<https://doi.org/10.1371/journal.pcsy.0000037.t001>

uncertainty. The 2018 Fall Selloff, while showing a lower correlation and VIX, still aligns with the observed trend that higher correlations are associated with increased market volatility.

This analysis underscores the utility of the VIX as an indicator of market sentiment and its correlation with significant financial stress periods. It also highlights the importance of monitoring correlations between major indices and commodities as part of a broader risk management strategy.

2.7. Comparative roles of CGR and DLSTM

While both CGR and DLSTM are valuable tools for market analysis, they are designed for distinct applications based on market conditions. This section contrasts CGR's role as a macroeconomic indicator with DLSTM's predictive capabilities under stable market conditions.

CGR is particularly effective during periods of high volatility, such as the 2008 Lehman crisis where its ability to capture macroeconomic shifts makes it a robust indicator of market sentiment. CGR's strength lies in its responsiveness to external economic factors, which dominate market dynamics during crises.

In contrast, DLSTM leverages historical price patterns to predict future trends. Its reliance on temporal dependencies makes it more suitable for stable markets, where sequential data continuity enhances its predictive accuracy. However, abrupt market disruptions, such as those seen during the 2008 crisis, can reduce DLSTM’s effectiveness.

The complementary nature of CGR and DLSTM reflects their distinct design philosophies: CGR for macroeconomic insights during crises and DLSTM for predictive modeling in stable markets.

3. The DLSTM neural network and SPY prediction

While the CGR serves as a robust predictor during more turbulent times, the stock market also generally undergoes periods of lower volatility where different predictive mechanisms are necessary. In these periods, a different approach is necessary. To address this, this section introduces the Difference Long Short-Term Memory (DLSTM) neural network, a novel architecture which employs a novel loss function that takes into account the difference of consecutive steps in the network. This allows the DLSTM to achieve high levels of accuracy in terms of daily directional changes.

3.1. Design of the DLSTM neural network

The Long Short-Term Memory (LSTM) network, an advanced variant of the Recurrent Neural Network (RNN), was introduced by Hochreiter and Schmidhuber in 1997 [11]. It effectively addresses the vanishing gradient problem common in RNNs through the use of gated cells, which better maintain information over long sequence durations. Yet, the unpredictability of stock market trends calls for further innovation beyond standard LSTM capabilities.

To this end, we have developed the Difference Long Short-Term Memory (DLSTM) network, aimed at improving the prediction of stock prices, specifically for the SPY index, using time-series data. This network enhances LSTM’s data handling by focusing on the closing prices arranged in sequential batches A_{batch} .

To begin, we gather $N = 3018$ Closing Prices of the S & P 500, starting from 2012-1-1 which we use as our training set. Denote our training set D_{train} , and let $D_{train} = \{x_1, x_2, \dots, x_N\}$. Firstly, we fix a sample size $k \geq 1$. The samples in our dataset are grouped into overlapping samples X_i of k consecutive points. Thus, X_i is defined as:

$$X_i = (x_i, \dots, x_{i+k}), \quad 1 \leq i \leq N - k \tag{1}$$

For each sample X_i , we denote the true label as $Y_{true}(X_i) := x_{i+k}$. We randomly group samples X_i into batches, which we denote A_{batch} , for stochastic gradient descent.

Now, recall the traditional Mean Squared Error (MSE) loss function, defined for the batch as:

$$L_{MSE}(A_{batch}) := \frac{1}{|A_{batch}|} \sum_{X_i \in A_{batch}} \|\mathcal{N}(X_i) - y_{true}(X_i)\|_2^2, \tag{2}$$

where $|A_{batch}|$ is the batch size, $y_{true}(X_i)$ is the actual stock price, and $\mathcal{N}(X_i)$ represents the networks prediction given some input X_i .

To capture the temporal progression between stock prices, we introduce the DLSTM Loss function. Firstly, we define Δx_i on some sequence $\{x_i\}$ to be $\Delta\{x_i\} = x_{i+1} - x_i$. We may now

define L_D in the following way:

$$L_D(A_{batch}) = \frac{1}{|A_{batch}|} \sum_{X_i \in A_{batch}} \|\Delta \mathcal{N}(X_i) - \Delta Y_{true}(X_i)\|_2^2 \tag{3}$$

The full loss function thus combines these elements:

$$L_{loss} = L_{MSE} + L_D, \tag{4}$$

3.2. SPY stock prediction using DLSTM

3.2.1. Data collection and preparation The dataset that we use spans from 2012-1-1 to 2024-1-1. We transform the dataset by using Moving Average with a period of 5 days. We use Eq 5 to push all data to a range of between 0 and 1.

$$s_t = \frac{S_t - S_{min}}{S_{max} - S_{min}} \tag{5}$$

Our DLSTM model is based on the typical LSTM with the following layers. The first Layer contains 128 LSTM units and is the first stage of feature extraction from the input sequence. Following a hierarchical approach, where the second LSTM layer with 64 units aims to synthesize the features extracted by the first layer. The reduced number of units might help in focusing on the most relevant temporal features extracted. After the LSTM layers, there is a fully connected (dense) layer with 16 units. This dense layer serves to interpret the features from the LSTM layers and begin the process of outputting a prediction, by mapping the learned representations to a space that reflects the problem’s complexity. The second Dense Layer has 4 units, indicating further compression of learned features into a more refined representation, which is closer to the final output dimension. The hyperparameters for the DLSTM network are summarized in Table 2

In this experiment, the daily closing price of SPY is estimated using DLSTM. The first 90% of the dataset was used as the training dataset, which is 2878 observations, and we use the last 5% of our dataset, which is 140 observations as our testing dataset. The size of the majority class (decreasing) is 79. Batch size is equal to 64. We used Adam to optimize learning rate; we trained the model with the 'lr' parameter of the Adam optimizer set to 0.001.

The test results is showing in Fig 10, which visualizes the predictions compared to the true data.

We compared our results with a baseline LSTM model. Our baseline LSTM model’s first layer contains 128 LSTM units. The second layer has 64 units. The third and fourth layer are Dense layers with 16 and 2 units respectively. The hyperparameters for the baseline LSTM model are shown in Table 3:

Table 2. Hyperparameters for DLSTM network.

Layer Tape	Output Dimension	No. Of Parameters
Input Layer	(6,2)	0
LSTM1	128	67,072
LSTM2	64	49,408
Dense 1	16	1,040
Dense 2	4	68
Output Layer	(2,2)	0

<https://doi.org/10.1371/journal.pcsy.0000037.t002>

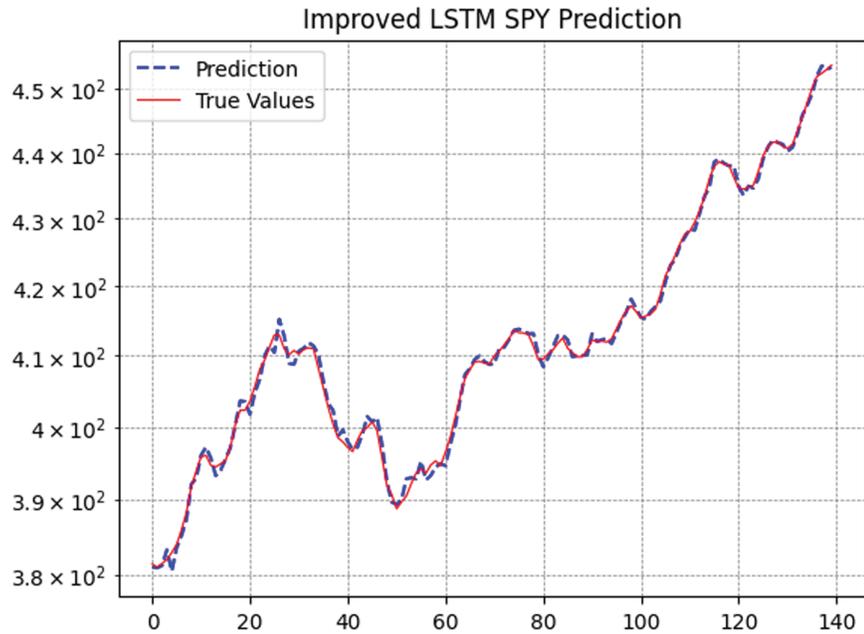


Fig 10. The prediction SPY price on test set.

<https://doi.org/10.1371/journal.pcsy.0000037.g010>

Table 3. Hyperparameters for LSTM network.

Layer Tape	Output Dimension	No. Of Parameters
Input Layer	(5,1)	0
LSTM1	128	66,560
LSTM2	64	49,408
Dense 1	16	1,040
Dense 2	2	34
Output Layer	(2,1)	0

<https://doi.org/10.1371/journal.pcsy.0000037.t003>

In addition, we compare our results with a binary classification model. The binary classification model has three Dense layers; the first two Dense layers have 512 units, and the third Dense layer has one unit. We used Adam for our optimizer, and we set the loss equal to binary_crossentropy. The binary cross entropy function serves to measure the difference between between predicted binary labels and actual binary labels [12]. The reported results for the binary classification model use the same training dataset and testing dataset as the DLSTM model. We report the result of the binary classification model after 700 epochs of training. All models, which include DLSTM, baseline LSTM, and binary classification models use the Tensorflow framework.

3.3. Model’s directional prediction accuracy

An integral part of any predictive model’s evaluation process is its directional accuracy, i.e., its ability to predict the correct direction of change for a particular variable.

We mainly report the directional accuracy of our prediction. Firstly, we define a correct directional prediction. Let S_N be the real price of stock on some day N , and let \hat{S}_N be the predicted price of a stock on some day N . Then, a directional prediction is accurate if $\text{sign}(\hat{S}_{N+1} - S_N) = \text{sign}(S_{N+1} - S_N)$. Then, directional accuracy is the percentage of correct predictions

Directional Accuracy is important because if we predict that a stock price will increase tomorrow, then a profit can be made by buying shares today and selling them tomorrow. Similarly, if we predict that a stock price will decrease tomorrow, then a profit can be made by selling short today. This will be demonstrated in later sections.

We also compare the directional accuracy of three different models: Deep Long Short-Term Memory (DLSTM), Long Short-Term Memory (LSTM), and a Binary Classification model. Directional accuracy is a measure of how well a model predicts the direction of price movements, which is crucial for trading and financial decision-making.

Our results, presented in Table 4, demonstrate the effectiveness of these models in predicting market directions.

The table illustrates that the DLSTM model achieves the highest directional accuracy at 82%, outperforming both the standard LSTM model, which has a directional accuracy of 78%, and the Binary Classification model, which has a directional accuracy of 74%. This improvement can be attributed to the DLSTM model's ability to capture more complex patterns and long-term dependencies in the data, which are often crucial for accurate market predictions.

We also reported the F1 score of all models, which is another metric of predictive accuracy. F1 is the harmonic mean of precision and recall [13]. Since it balances precision and recall, F1 is useful even when there is a class imbalance in the dataset [14]. The table illustrates that the DLSTM model achieves the highest F1 score at 0.8732, outperforming the standard LSTM model which has a F1 of 0.8449.

3.4. Implementation of simulated day trading algorithm

To harness the potential of our model's directional predictive ability, we crafted a simple day trading simulation algorithm. The primary objective of this algorithm is to operationalize our model's predictions into actionable trading decisions, allowing for the real-world applicability of our research. The operational rules of this algorithm are:

Initial Capital: The algorithm starts with an initial budget of \$500.

1. **Positive Prediction without Holdings:** If the portfolio is neutral (no assets held) and the predicted stock price direction for the next day is positive, the algorithm takes a bullish stance and buys one share.
2. **Negative Prediction without Holdings:** In the absence of any holdings and a predicted decline in the stock price, the algorithm adopts a bearish approach, selling one share short, betting on a decline.
3. **Holding through Positive Prediction:** Holding a share and anticipating a price increase, the algorithm remains optimistic, retaining its position to potentially reap benefits from the price rise.
4. **Exit and Short on Negative Prediction:** If a share is in possession and a price decrease is anticipated, the algorithm sells the current holding. Moreover, to capitalize on the predicted decline, it sells another share short.

Table 4. Directional accuracy of DLSTM, LSTM, and binary classification models.

Model	Directional Accuracy	F1
DLSTM	82%	0.8732
LSTM	79%	0.8449
Binary Classification	74%	0.7602

<https://doi.org/10.1371/journal.pcsy.0000037.t004>

5. **Covering and Buying on Positive Prediction:** If a short position exists and an upward price movement is predicted, the algorithm covers the short (buys back the share sold short) and then proceeds to buy a share, positioning itself to benefit from the anticipated price increase.
6. **Maintaining Bearish Stance:** If a share has already been sold short and the prediction still indicates a price drop, the algorithm remains bearish, holding onto its short position, expecting further decline.

In essence, this algorithm translates our model's predictive insights into tangible trading decisions, potentially providing traders and investors with a structured approach to day trading, rooted in data-driven insights.

This performance can be visualized in Fig 11, with red inverted triangles marking local maxima (i.e., sell signals) and green triangles indicating local minima (i.e., buy signals).

Starting with \$500, the we ran the simulation algorithm 10 times. The average profit after 20 days of trading was 21 dollars, equivalent to roughly a month in real time. This corresponds to a 4.2% gain over one roughly one month. After 60 trading days (roughly 3 months in real time), the algorithm made a profit of \$62, representing a 12.4% return on the initial \$500 over the three months, which is consistent with the one-month prediction. This demonstration shows the potential profitability of utilizing the directional predictions of our model in a practical day trading scenario.

4. Conclusion

Concluding our exploration into volatility-informed SPY forecasting, this paper has provided a comprehensive analysis of market behavior across different volatility regimes. Our study highlights the Copper-Gold Ratio's (CGR) capacity as a robust macroeconomic indicator for the S&P 500 (SPY) index, particularly during periods of heightened market turbulence.

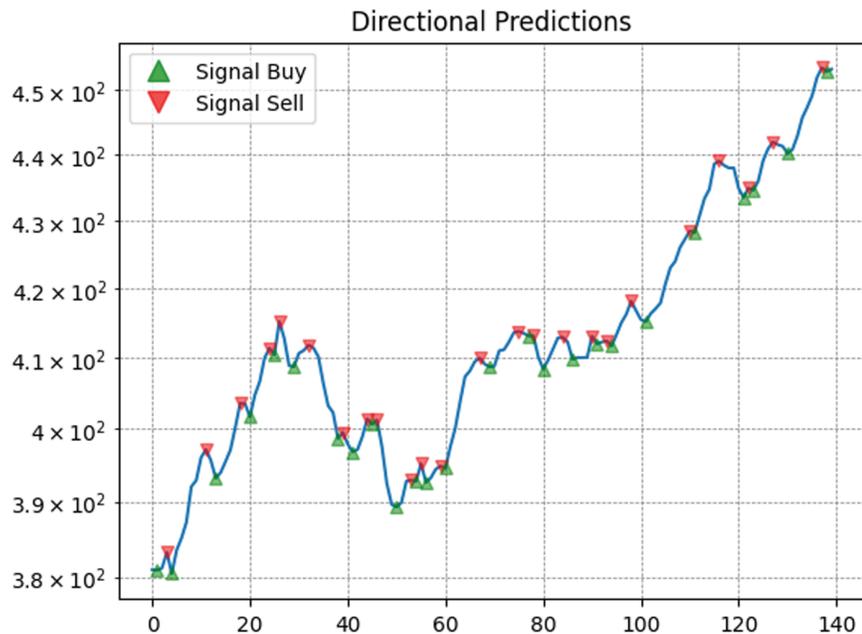


Fig 11. Directional predictions on test set.

<https://doi.org/10.1371/journal.pcsy.0000037.g011>

By investigating CGR's correlation with SPY across four major economic events, including the 2008 Lehman crisis, we have demonstrated its effectiveness in capturing market sentiment and detecting shifts under volatile conditions.

In parallel, we introduced the Difference Long Short-Term Memory (DLSTM) neural network, a data-driven model optimized for stable market environments. Our empirical testing demonstrated DLSTM's ability to achieve 82

The complementary roles of CGR and DLSTM were evident in our findings. CGR excels in volatile markets, where its ability to capture macroeconomic sentiment provides actionable insights during crises, as seen during the 2008 Lehman crisis. Conversely, DLSTM is more effective in stable markets, leveraging sequential patterns in historical data to deliver reliable predictions. Together, these tools form a holistic framework for market analysis, addressing diverse market scenarios with precision and adaptability.

Looking forward, the adaptability and predictive accuracy of DLSTM present opportunities for integration into advanced trading systems, offering investors a robust tool to navigate both volatile and stable market conditions. Future research could further enhance these models by exploring hybrid frameworks that combine CGR's macroeconomic insights with DLSTM's predictive power, potentially broadening their applicability across varied market regimes.

Ultimately, the strategies and insights presented in this paper contribute meaningful advancements to the financial industry, equipping investors and analysts with tools to achieve greater precision and resilience in the face of market volatility. As we refine these models and methodologies, the potential to optimize decision-making processes and enhance financial outcomes becomes increasingly tangible.

Supporting information

S1 Data. The dataset of daily SPY closing prices from January 2012 to January 2024 used in Sect 4.2 analysis.

(CSV)

Author contributions

Conceptualization: Hao Chen, Yulong Lian, Hong-Kun Zhang

Data curation: Hong Cao, Hao Chen, Yulong Lian.

Formal analysis: Hong Cao, Yulong Lian.

Investigation: Hong Cao, Hao Chen, Yulong Lian.

Methodology: Hong Cao, Yulong Lian.

Software: Yulong Lian, Hong-Kun Zhang.

Supervision: Hong-Kun Zhang.

Validation: Hong-Kun Zhang.

Visualization: Hong-Kun Zhang.

Writing – original draft: Yulong Lian, Hong-Kun Zhang.

Writing – review & editing: Yulong Lian, Hong-Kun Zhang.

References

1. Mirafiore A. Copper spot prices and stock market volatility. *The Hedge Fund J.* 2016.
2. Baur D, McDermott T. Is gold a safe haven? International evidence. *J Bank Finance.* 2010;34(1):218–28. <https://doi.org/10.1016/j.jbankfin.2009.09.006>
3. Parnes D. Copper-to-gold ratio as a leading indicator for the 10-year treasury yield. *North Am J Econ Finance.* 2023.
4. Borovkova S. An ensemble of LSTM neural net works for high-frequency stock market classification. *J Forecast.* 2019.
5. Qiu J, Wang B, Zhou C. Forecasting stock prices with long-short term memory neural network based on attention mechanism. *PLoS One.* 2020;15(1):e0227222. <https://doi.org/10.1371/journal.pone.0227222> PMID: 31899770
6. Hergt B. Gold prices during and after the Great Recession. 2013.
7. Dong D, van Oers L, Tukker A, van der Voet E. Assessing the future environmental impacts of copper production in China: implications of the energy transition. *J Clean Prod.* 2020;274:122825. <https://doi.org/10.1016/j.jclepro.2020.122825>
8. Parnes D. Copper-to-gold ratio as a leading indicator for the 10-year treasury yield. 2022.
9. Whaley R. Understanding the VIX. *J Portfolio Manag.* 2009;35(3):98.
10. Ozair M. What does the vix actually measure? An analysis of the causation of SPX and VIX. *J Finance Risk Perspect.* 2014;3:83–132.
11. Hochreiter S, Schmidhuber J. Long short-term memory. *Neural Comput.* 1997;9(8):1735–80. <https://doi.org/10.1162/neco.1997.9.8.1735> PMID: 9377276
12. Wali R. Xtreme margin: a tunable loss function for binary classification problems. 2022. <https://arxiv.org/abs/2211.00176>
13. Derczynski L. Complementarity, F-score, and NLP evaluation. In: *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. Portoroz, Slovenia: European Language Resources Association (ELRA); 2016. p. 261–6. Available from: <https://aclanthology.org/L16-1040>
14. Lipton Z, Elkan C, Narayanaswamy B. Thresholding classifiers to maximize F1 score. *Proc Conf Uncertain Artif Intell.* 2014;30:425–33.