Model Uncertainty of Stock Market Based on CNN Predictions

Yunfei Fan

Pratt School Engineering, Duke University, 27704, United States

Email: yunfei.fan@duke.edu

Abstract

The stock market is replete with uncertainty, a factor significantly influenced by investor decisions. In order to assess this uncertainty, China CSI 300 stocks have been gathered, analyzed, and modeled. To quantify market-related uncertainty, CNN models have been constructed and trained using data from various time periods. Each dataset comprises 40 trading days, with each Convolutional Neural Network (CNN) model (investor) being trained to forecast the next day's price. Interestingly, the CNN model based on the peak of the bull market cycle exhibits a considerably larger prediction error compared to the model trained on the lowest point of the preceding market cycle. The 2022 model, characterized by markedly lower uncertainty than its 2010 counterpart, may symbolize the overall growth of the stock market.

Keywords: stock prediction, uncertainty, CNN models, historical data, predicting error.

1. Introduction and Literature Review

The allure of the stock market lies in its potential for high returns, making it a focal point for investors seeking to decipher the underlying principles governing stock price movements, mitigate risks, and enhance profitability. Consequently, the development of models to precisely forecast future stock market trends and ascertain its intrinsic value assumes paramount significance in financial investment research. Stock price forecasting has emerged as a central theme in this realm, given the inherent volatility of stock market prices and the intricate interplay of various factors shaping the market dynamics, including macroeconomic policies, global market trends, and investor sentiments. The multifaceted nature of these influencing factors, coupled with the vast volume of data involved, renders stock price prediction a formidable challenge. The demand for accurate forecasting methods has thus spurred a transition from conventional statistical models to data-driven approaches, notably artificial intelligence models.

The neural network model stands out for its capacity to discern nonlinear relationships within extensive datasets, navigate time series complexities and noise, and delineate the intricate interactions between market influencers and stock prices. This adaptive framework enables autonomous learning and regulation, underscoring its superior intelligence relative to traditional methodologies (Long et al., 2019).

Lapedes and Farber (1987) were pioneers in the application of artificial neural networks to the realm of forecasting, while White (1988) notably ventured into the utilization of neural network models within financial markets. The initial endeavor to construct a neural network model for predicting daily returns on IBM yielded suboptimal results, possibly attributable to the model's susceptibility to local minima. Subsequently, an increasing number of artificial neural networks have been employed for financial market forecasting (Chong et al., 2017). In contrast with traditional methodologies, Schurmann (1993) conducted a comparative analysis utilizing artificial neural network models alongside conventional mathematical and statistical models for stock market prediction. Empirical findings demonstrated the superior forecasting capability of neural networks relative to traditional methods. Similarly, Chen et al. (2006) conducted a performance evaluation comparing Support Vector Regression (SVR), Back Propagation (BP), and Auto-Regressive (AR) models on multiple stock indices, revealing that the artificial intelligence algorithms SVR and BP outperformed the traditional AR models in terms of predictive accuracy.

As a specialized neural network architecture, Convolutional Neural Networks (CNN) exhibit three key characteristics: local perception, weight sharing, and the utilization of multiple convolutional kernels. CNN has found applications in addressing the feature selection requirements of time-series data (Goodfellow et al., 2016) and has been increasingly employed in stock market analysis (Liu et al., 2017). For instance, Gudelek et al. (2017) utilized CNN to analyze two-dimensional images for predicting the time series of 17 distinct Exchange-Traded Funds (ETFs). In their approach, time is represented as the first dimension, while the various types of technical indicators serve as the second dimension. Sezer and Ozbayoglu (2019) leveraged candlestick patterns in finance as input images for CNN to facilitate learning and prediction based on different temporal points within the image. Subsequently, this predictive outcome was harnessed for the development of corresponding algorithmic trading models [1-2].

The primary shortcomings of probabilistic-based algorithms in stock markets stem from the inherent irrationality of human behavior. Stock prices can be significantly influenced by emotional factors, such as panic triggered by community discussions or news events. To illustrate the impact of such uncertainty arising from human irrationality, this study involved training several short-term prediction models using CNN to simulate investors across varied historical periods. These simulated investor models from distinct historical epochs generated diverse predictions over time, allowing for the identification of irrational behaviors associated with specific historical contexts. By showcasing the inadequacies of these predictions, the study aimed to capture and model the uncertainties inherent in stock market dynamics influenced by human behavior.

2. Data and Algorithm

The data utilized in this study was sourced from "Tushare," an open-source community dedicated to quantitative analysis. Our analysis primarily focuses on the CSI 300 Index, which is widely regarded as the most representative index in China, effectively reflecting the overall state of the stock market. In order to examine stock market fluctuations over a twelve-year period spanning from 2010 to 2022, we constructed a dataset comprising the combined constituents of the CSI 300 Index in both years. To ensure data integrity, stocks that were inactive for more than six months were excluded from the dataset. The resulting dataset consists of open, close, high, low, and volume data for 349 stocks across 3159 trading days. Given the diverse nature of the stocks in our dataset, it is essential to employ change rates rather than absolute values. Consequently, the dataset incorporates additional derived metrics such as the high-to-close ratio, low-to-close ratio, and open-to-close ratio. These supplemental measures provide valuable insights into the relative changes in stock prices and facilitate comprehensive analysis.

The Chinese stock market has undergone a transition from a market characterized by high trading volume and low stock values to one characterized by lower trading volume and higher stock values. As illustrated in Figure 1, a significant proportion of stocks are clustered on the lower price spectrum (price < 5 CNY), with fewer stocks falling within the low to mid-price range (5 CNY < price < 10 CNY). Moreover, only a minority of stocks are priced above 15 CNY. These findings underscore the prevalence of low-quality, high-volume dynamics that characterized the Chinese stock market landscape in 2010.

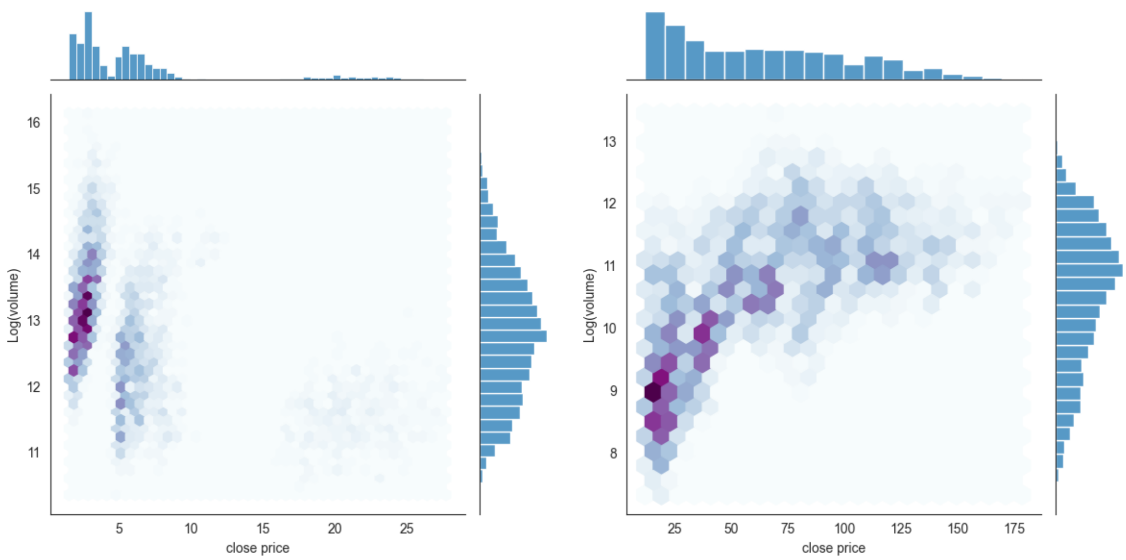

Figure 1 The development of the Chinese stock market is represented by close price-Log(volume) distributions. (Left) Stock market distribution at 2010 Jan. (Right) Stock market distribution at 2022 Dec.

In stark contrast, by the year 2022, a significant transformation has occurred in the Chinese stock market landscape. Stock prices have experienced a remarkable five-fold increase, with all 349 stocks now priced above 20 CNY. The previous segregations among stocks have dissipated, and the distribution of stock prices demonstrates continuous changes (as depicted in the right panel of Figure 1). These shifts in the stock market reflect the evolving development of companies, as the absence of segregation implies greater diversity in the types of companies. From an algorithmic perspective, one approach to assessing the uncertainties within the financial market involves training a predictive model for stock prices within a specific timeframe. By examining the performance of the model, the inherent uncertainties in the market can be gauged.

The predictive model for stock price is structured based on the Convolutional Neural Network (CNN) framework [3-4]. The input data consists of three channels, namely high/close, low/close, and open/close, each comprising 40 consecutive ratios. These ratios encapsulate features that are to be extracted through convolutional operations within the neural network architecture.

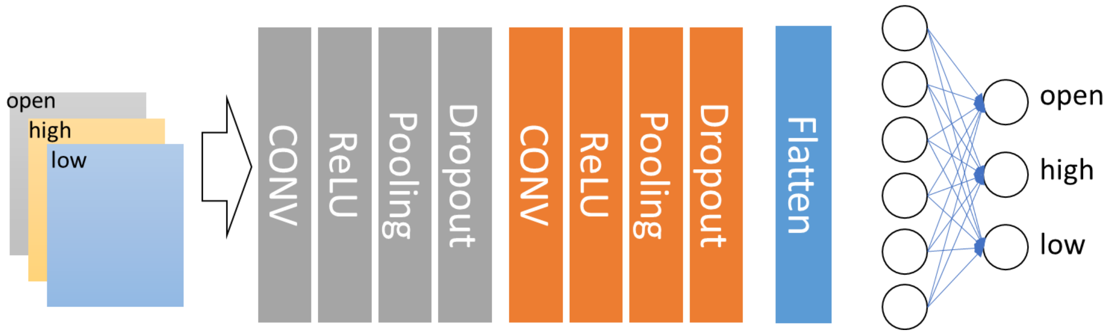

Figure 2 CNN network structure with 3 input channels: open price change rate, highest price change rate, lowest price change rate. The model mainly contains 2 Multilayer Perceptron units (grey and orange complexes of layers), a flatten layer and a fully connected layer.

As depicted in Figure 2, the input channels are subsequently processed by two Multilayer Perceptron units and then pass through a flattening operation and a fully connected layer. The ultimate output of this process is the predicted stock change rate for the following day. The model is trained using the Mean Squared Error (MSE) loss function and the Adam optimizer [5-7].

The initial trained model, referred to as the "2010 spring model," is trained using data from January and February of 2010. Following the division of the data into a 244-stock training set and a 105-stock test set, the model undergoes training to forecast next-day change rates (refer to Figure 3) over the course of 300 epochs.

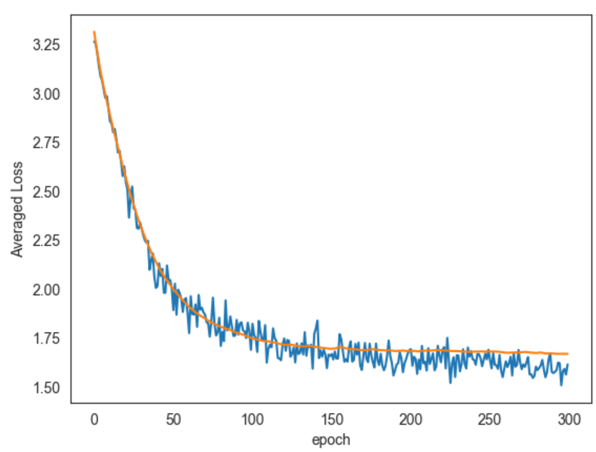

Figure 3 The training process of the CNN model. As the epoch increases, the average loss drops. The orange line represents the test loss while the blue line represents the training loss.

Following a similar approach, three additional models are developed using data from distinct time periods: the "2015 summer model," the "2022 summer model," and the "2014 spring model," respectively. The "2010 model" signifies the past, whereas the "2022 model" represents the present market conditions. Notably, the most substantial transition occurs between the "2014 spring" and the "2015 summer" models. During this period, the Chinese stock market underwent a notable bull market phase. As such, the 2014 and 2015 models capture the market's bottom and peak during this bullish trend. Each model embodies the perspective of an investor existing within a particular time frame, with their insights validated against real market data. Consequently, the emotional aspect of investing can be partially discerned through the investor's performance within these models [8-9].

3. Results and analysis

The performance of each model, representing different investor perspectives, is visualized in Figure 4. Specifically, we analyze the outcomes depicted in Figure 4b and Figure 4d for the 2015 summer and 2014 spring investors, respectively. During the summer of 2015, characterized by the peak of the bull market cycle in the Chinese stock market, investor sentiment was notably buoyant. Optimism among Chinese investors soared to unprecedented levels, potentially leading to an overestimation of stock performance by the optimistic 2015 summer model. Consequently, the loss function exhibits frequent spikes over time, indicating that excessive enthusiasm can detrimentally impact performance. Conversely, during the 2014 spring period just before the onset of the bull market cycle when the CSI index hit its lowest point, the number of spikes in the loss function significantly diminishes. The 2014 spring model, reflecting a more cautious investor sentiment, makes fewer erroneous predictions compared to the 2015 summer model. This suggests that the market during this period exhibits lower volatility and uncertainty [10].

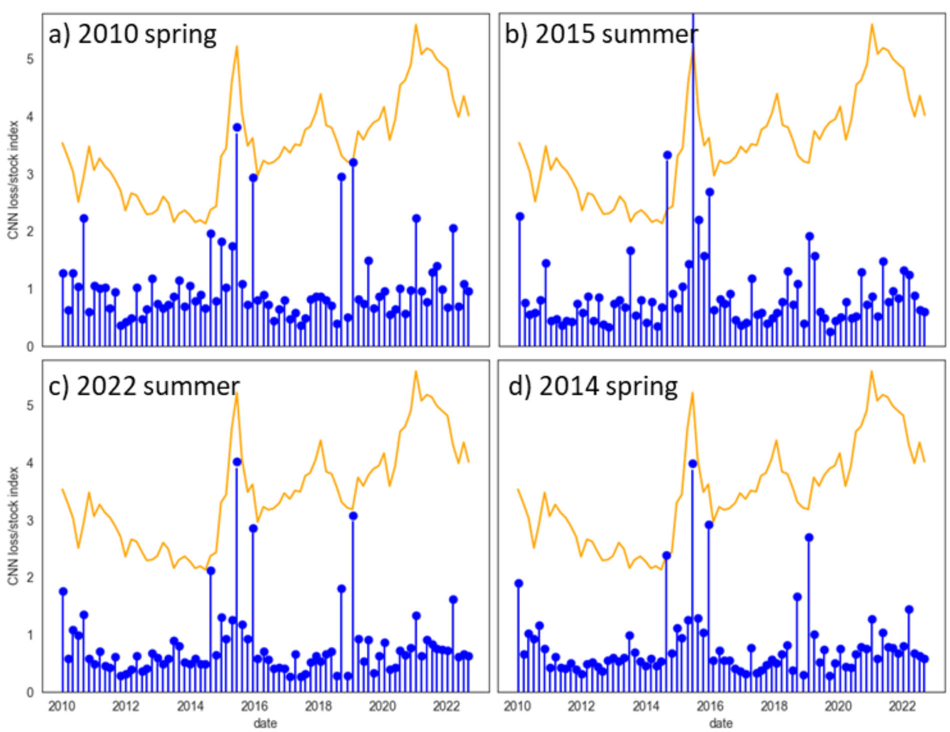

Figure 4 Models' (Investors') predictivity (represented by loss function values, the lower the better) of all dates represented by the blue bars. Models are trained with data from 4 different periods: a) 2010 spring, b) 2015 summer, c) 2022 summer and d) 2014 summer. The scaled CSI index is also provided as the yellow curve.

The outcomes of the 2010 spring and 2022 summer models are showcased in Figure 4a and Figure 4c, illustrating the evolution of investor behavior over the past 12 years alongside China's development trajectory. Remarkably, the 2022 summer model displays a tendency towards making fewer predictions resulting in significant losses compared to the 2010 spring model. This superior performance of the 2022 summer model implies an enhancement in investor capability over time. The reduced frequency of erroneous predictions by the 2022 summer model suggests a more refined decision-making process, indicative of investor growth and development. Consequently, the presence of more seasoned investors may contribute to mitigating market uncertainty and fostering a more stable investment environment.

4. Summary

This study employs a CNN-based investor simulation to assess and compare the uncertainty levels among investors from different historical periods. Due to temporal constraints, each investor is restricted to analyzing stock data within a two-month window for making predictions. While the model investors do not exhibit emotions, they are influenced by the prevailing market conditions, be it bullish or bearish, resulting in varying degrees of uncertainty in their predictions. Through the simulation, it is observed that the 2014 model exhibits the fewest prediction errors, indicating relatively low market uncertainty during that timeframe. Conversely, the 2015 model displays the highest prediction errors, attributed to the heightened uncertainty stemming from the influence of a bullish market trend.

Comparing the 2010 and 2022 models reveals a reduction in uncertainty accompanied by a decrease in prediction errors over time, reflecting the overall maturation and development of the market. This study elucidates the market's uncertainty dynamics through a data-driven algorithmic approach, paving the way for future research to explore uncertainties in other financial markets such as bonds, funds, futures, and derivatives. Additionally, investigating the propagation of uncertainty presents an intriguing avenue for further exploration. Beyond CNN, the integration of alternative machine learning frameworks holds promise for enhancing predictive accuracy in future endeavors.

Bibliography

[1] W. Long, Z. Lu, and L. Cui, 2019. Deep learning-based feature engineering for stock price movement prediction. Knowledge-Based Systems, 164, 163–173.

[2] A. S. Lapedes and R. F. Farber,1987. Nonlinear signal processing using neural networks: Prediction and system modelling. In 1. IEEE international conference on neural networks, San Diego, CA, USA, 21 Jun 1987.

[3] H. White, 1988. Economic prediction using neural networks: the case of IBM daily stock returns. In IEEE International Conference on Neural Networks, 2, 451–458.

[4] E. Chong, C. Han, and F. C. Park, 2017. Deep learning networks for stock market analysis and prediction: Methodology, data representations, and case studies. Expert Systems with Applications, 83, 187–205.

[5] M. Schumann and T. Lohrbach, 1993. Comparing artificial neural networks with statistical methods within the field of stock market prediction. In Proceeding of the Twenty-Sixth Hawaii International Conference on System Sciences, 4, 597–606.

[6] WH. Chen, JY. Shih and S. Wu, 2006. Comparison of support-vector machines and back propagation neural networks in forecasting the six major Asian stock markets. Inderscience Publishers.

[7] I. Goodfellow, Y. Bengio, and A. Courville, 2016. Deep learning. MIT press.

[8] W. Liu, Z. Wang, X. Liu, N. Zeng, Y. Liu and F. E. Alsaadi, 2017. A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11–26.

[9] M. U. Gudelek, S. A. Boluk, A. M. Ozbayoglu, 2017. A deep learning based stock trading modelwith 2-D CNN trend detection. 2017 IEEE Symposium Series on Computational

Intelligence.

[10] O. B. Sezer, A. M. Ozbayoglu, 2019. Financial trading model with stock bar chart image time series with deep convolutional neural networks. arXiv:1903.04610.