Model Uncertainty of Stock Market Based on CNN Predictions

Yunfei Fan

Pratt School of Engineering, Duke University, 27704, United States

Email: yunfei.fan@duke.edu

Abstract

The stock market is full of uncertainty which is significantly affected by the decisions of investors. To evaluate such uncertainty, China CSI 300 stocks are collected, analyzed, and modeled. To quantify the market-related uncertainty, CNN models are built but trained with data from different periods. Each dataset contains 40 trading days and each CNN model (investor) is trained to predict the price of the next day. The CNN model based on the peak of the bull market cycle makes a much larger predicting error than the CNN model trained based on the lowest point of the previous market cycle. The model from, which exhibits significantly lower uncertainty compared to the model from 2010, may represent the growth of the entire stock market.

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

JEL Classification: G17, C45, C53, C88

1. Introduction and Literature Review

The stock market is always attractive because of its potential high returns. Investors want to find out the inner rules of stock price movements to avoid risks and gain higher profits. Therefore, it is important to build models to accurately predict the future movements of the stock market and determine its value. In the field of financial investment, stock price forecasting has been a major topic of research. Stock market prices are highly volatile and the market is a non-linear system affected by many factors, such as macro-policies, foreign markets, investor sentiments, etc. The various affecting factors and huge amounts of data make the stock price prediction quite hard. The need for accurate forecasting methods keeps growing from simple statistical models to data-oriented models including artificial intelligence models.

The neural network model can extract the nonlinear relationship from a large amount of data, time series process and noise, thus representing the complex relationship between the factors affecting the stock market and the stock price. This approach allows the model to self-regulate and self-learn, which is more intelligent than traditional methods (Long et al, 2019).

Lapedes and Farber (1987) first applied artificial neural networks in the field of forecasting. White (1988) was the first to apply neural network models to financial markets. A neural network model is built to predict the daily return of IBM, but the result was not very good. The reason for the poor prediction may be that the model may fall into local minima. Since then, more and more artificial neural networks are used to make forecasts for financial markets (Chong et al, 2017). In comparison with traditional methods, Schurmann (1993) used artificial neural network models and traditional mathematical and statistical models to forecast the stock market, and the empirical results showed that neural networks performed better than traditional methods in terms of forecasting ability. Chen et al (2006) compared the performance of SVR, BP, and AR(1) on several stock indices, and the artificial intelligence algorithms SVR and BP are better predictors than the traditional AR models.

As a specialized neural network, CNN has three main features, which are local perception, weight sharing and multiple convolutional kernels. CNN is applicable to meet the feature selection demands of time-series data (Goodfellow et al, 2016), and is adopted for stock markets (Liu et al, 2017). CNN is used to process two-dimensional images for the prediction of the time series of 17 different ETFs (Gudelek et al, 2017). In this paper, time is treated as the first dimension, and the type of technical indicators is treated as the second dimension. Sezer and Ozbayoglu (2019) uses candlesticks in finance as input images for CNN to learn and predict the outcome based on the different time points in the image. Based on this prediction result, the corresponding algorithmic trading model is developed.

The major limitations of the probabilistic-based algorithm in the stock market root in the irrationality of human behavior. A stock price can be driven emotionally based on a panic caused by a community discussion or a piece of news. So, in this work, to demonstrate such uncertainty in the irrationality of human behavior, several short prediction models are trained with CNN to simulate investors in different historical periods. These model investors from different historical times give various predictions over time. Thus their ill predictions can be used to indicate their irrational behavior of them to a certain historical time. Thus the uncertainty in the stock market can be modeled.

2. Data and Algorithm

The data used in this work is downloaded from “Tushare”, an open-source community for quantitative analysis. We focus on the CSI 300 Index, which is widely recognized as the most representative index in China and can reflect the overall condition of the stock market. To investigate the stock market changes over the past twelve years (2010-2022), the union of the CSI 300 Index constituents in 2010 and 2022 is considered by removing stocks that are absent from trading for more than six months. The dataset includes open, close, high, low, and volume data for 349 stocks over 3159 trading days. Given the variation in the nature of stocks, the use of change rates is preferred over absolute values. Concretely, the dataset also contains derived values including high/close, low/close, and open/close ratios.

Chinese stock market has shifted from a market of high volume, low stock value to a lower volume, higher stock value one. As shown in Figure 1, most stocks concentrate on the low price side (price < 5 CNY), and fewer stocks concentrate on the low-mid price side (5 CNY < price <

10 CNY). Only a small fraction number of stocks have a price larger than 15 CNY. Such results indicate the low quality, high volume features of the Chinese stock market back in 2010.

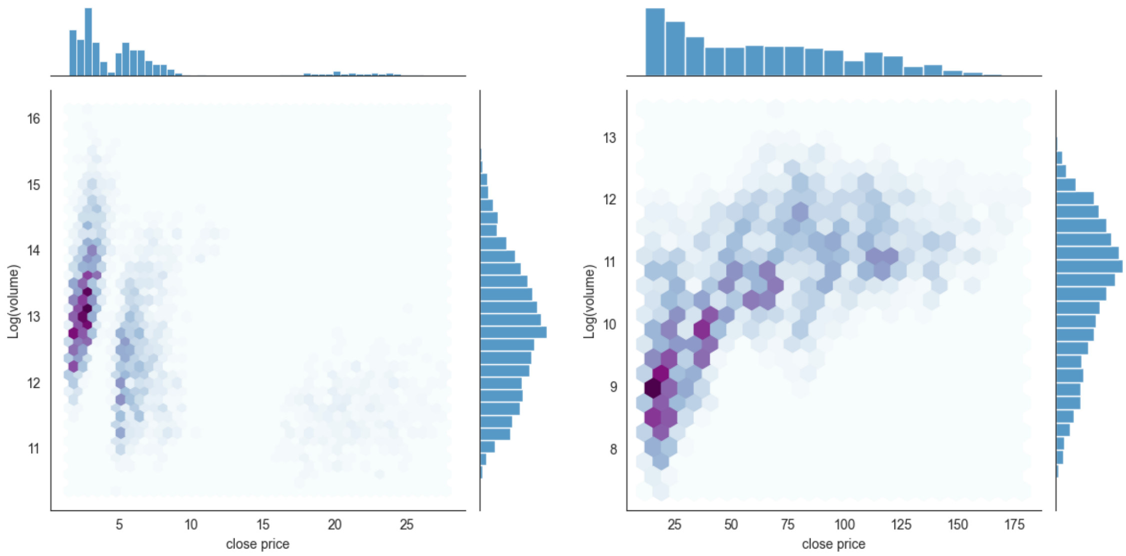

Figure 1 The development of the Chinese stock market is represented by close price-Log(volume) distributions. (Left) Stock market distribution at 2010 Jan. (Right) Stock market distribution at 2022 Dec.

In contrast, when the time comes to 2022, the situation is changed significantly. The stock prices increase by a factor of 5. All 349 stocks have a price higher than 20 CNY. The segregations between stocks vanished and the distribution of stock prices changes continuously (Right panel of Figure 1). Such changes in the stock market represent the shift in the development of companies. No segregation indicates that there are more variances in the types of companies.

On the algorithm side, when one wants to evaluate the uncertainties of the financial market, one can train a predictive model for the stock price within a certain period. Then, the uncertainties can be reflected in the performance of the model.

The predictive model for the stock price follows the framework of Convolutional Neural Network. The input is constituted of 3 channels (high/close, low/close and open/close) of 40 consecutive ratios each of which contains features to be extracted by convolutions.

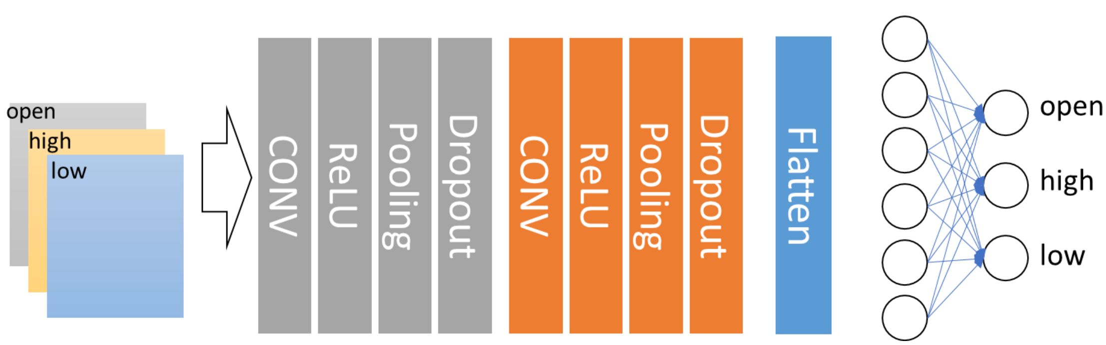

Figure 2 CNN network structure with 3 input channels: open price change rate, highest price change rate, lowest price change rate. The model mainly contains 2 Multilayer Perceptron units (grey and orange complexes of layers), a flatten layer and a fully connected layer.

As shown in Figure 2, the input channels are then loaded by two Multilayer Perceptron units followed by a flatten and a fully connected layer. The output will be the stock change rate of the next day. The model is then trained by using the MSE loss function and the Adam optimizer.

The first trained model is called the “2010 spring model” with data in January and February of 2010. After the data is split into a 244-stock training set and a 105-stock test set, the model can be trained to predict the change rates the next day (Figure 3) after 300 epochs.

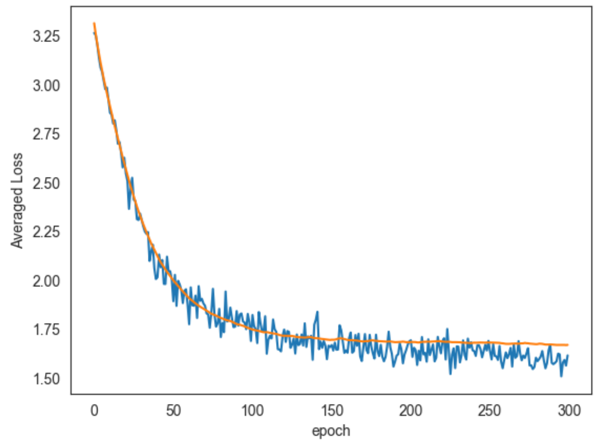

Figure 3 The training process of the CNN model. As the epoch increases, the average loss drops. The orange line represents the test loss while the blue line represents the training loss.

With the same logic, 3 more models are trained with data from different periods. 3 models are the “2015 summer model”, the “2022 summer model” and the “2014 spring” model respectively. The 2010 model corresponds to the past and the 2022 model corresponds to the present. The most significant change is from the 2014 spring to the 2015 summer. From spring 2014 to summer 2015, the Chinese stock market experienced a significant bull market. Therefore, the 2014 model and the 2015 model correspond to the bottom and the peak of the bull market. Each model represents an investor who lives in a particular time with opinions tested by the actual market data. Thus, the emotional part can be reflected by the performance of the investor to some level.

3. Results and analysis

The performance of each model (investor) is shown in Figure 4. We first focus on Figure 4b and Figure 4d as the performances of investors from the 2015 summer and 2014 spring. As is well known, in the summer of 2015, the Chinese stock market reached the highest point of a bull market cycle. Chinese investors celebrated this milestone, and optimism soared to its peak. Such optimistic investors (2015 summer model) may overestimate the performance of many stocks. As a result, the loss function frequently shows spikes over time. Such results indicate that extra passion may lead to poor performance. In contrast, when it comes to 2014 spring when the CSI index reaches its lowest point right before the bull market cycle, the number of spikes of loss function reduces significantly. The 2014 spring model (investor) makes ill predictions in only a few periods, about half the ill predictions from the 2015 summer model (investor). Thus, the market during this time exhibits lower uncertainty.

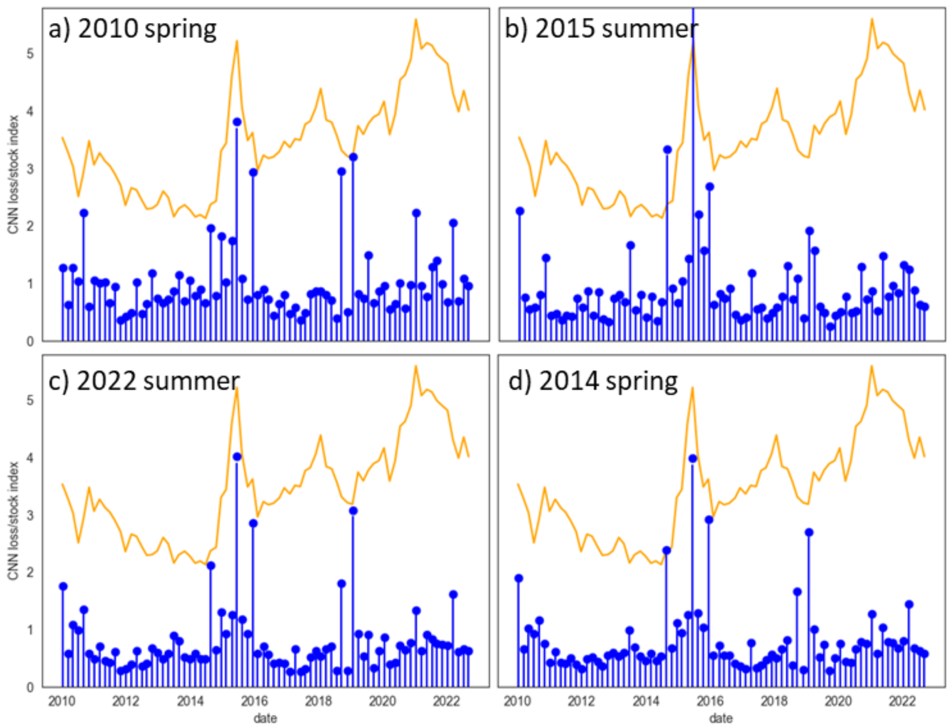

Figure 4 Models' (Investors') predictivity (represented by loss function values, the lower the better) of all dates represented by the blue bars. Models are trained with data from 4 different periods: a) 2010 spring, b) 2015 summer, c) 2022 summer and d) 2014 summer. The scaled CSI index is also provided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

The results from the 2010 spring and 2022 summer models are also provided in Figure 4a, c to show the historical change of the investors along the development of China in the past 12 years. The 2022 summer model makes fewer predictions with large losses. Thus, the 2022 summer model (investor) outperforms the 2010 spring model (investor). Such outperformance indicates the development of the investors. The uncertainty in the market may be reduced when there are more mature investors.

4. Summary

In this work, the uncertainty of investors from different historical times is simulated and evaluated by using a trained CNN investor. Due to the historical limitation, each investor can observe stock data within a time window of two months to make a limited prediction. The model investors, though show no emotions, are affected by the bull or bear patterns of the market. Thus, the uncertainty can be represented and compared. Based on the simulation, the 2014 model shows the fewest predicting error which indicates the low market uncertainty during that period. In contrast, the 2015 model shows the largest predicting error due to the affection of the bull market with very large uncertainty. When one compares a 2010 model and a 2022 model, the uncertainty reduces with the reduction of predicting error, which indicates the development of the whole market.

This work modeled and showed the uncertainty of the stock market with a data-based algorithm. Future works may involve exploring other types of markets such as bonds, funds, futures, derivatives, etc. The spreading of uncertainty may be another interesting topic to investigate. Besides CNN, other machine learning frameworks may be adopted to achieve higher precision.

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.