Developing Arbitrage Strategy in High-frequency Pairs Trading with Filterbank CNN Algorithm

Yu-Ying Chen
Institute of Information Management
National Chiao Tung University
Hsinchu, Taiwan, R.O.C
pshkre@gmail.com

Wei-Lun Chen
Institute of Information Management
National Chiao Tung University
Hsinchu, Taiwan, R.O.C
lagame5379@gmail.com

Szu-Hao Huang
Department of Information Management and Finance
National Chiao Tung University
Hsinchu, Taiwan, R.O.C
szuhaohuang@nctu.edu.tw

Abstract*—Pairs trading is a statistical arbitrage strategy, which selects a set of assets with similar performance and produces profits during these asset prices far away from rational equilibrium. Once this phenomenon exists, traders can earn the spread by longing the underperforming asset and shorting the outperforming asset. This paper proposed a novel intelligent high-frequency pairs trading system in Taiwan Stock Index Futures (TX) and Mini Index Futures (MTX) market based on deep learning techniques. This research utilized the improved time series visualization method to transfer historical volatilities with different time frames into 2D images which are helpful in capturing arbitrage signals. Moreover, this research improved convolutional neural networks (CNN) model by combining the financial domain knowledge and filterbank mechanism. We proposed Filterbank CNN to extract high-quality features by replacing the random-generating filters with the arbitrage knowledge filters. In summary, the accuracy is enhanced through the proposed method, and it proves that the integrated information technology and financial knowledge could create the better pairs trading system.

***Keywords*—Deep Learning, Filterbank CNN, Pairs Trading, Statistical Arbitrage, Time series**

I. INTRODUCTION

Arbitrage is the trading strategy to buy low-priced assets and sell high-priced assets to earn the spread which the assets are a group of the same or substantially equal financial assets at different price levels. Since it bears much smaller risk than other trading strategies, the benefits of arbitrage trading are more stable, thus arbitrage trading gradually became the commonly used trading practices in the financial markets. On the other hand, pairs trading, a kind of popular statistical arbitrage strategy used by many financial institutions such as hedge funds and investment banks, observed whether the performances of two historically correlated securities are far away from the rational equilibrium. In academic use, many financial engineering scholars research pairs trading. The expectation-maximization algorithm (EM) with the kalman filter is proposed in 1982 which made the time series smooth [1]. This algorithm is widely used in the practice and many scholars also researched it. In [2], the expectation-maximization algorithm (EM) with the kalman filter is improved and the expectation-maximization algorithm (EM) with the finite-dimensional recursive filter is proposed. This algorithm can reduce the memory demand and have the implementation of parallel computing with multi-core CPU. Apart from the usage in pairs trading, it is also used in multisensory signal enhancement of speech signals and econometric modeling. In [3], it is found that the two correlated securities might have the cointegration when they comply with random walk theory. If the cointegration exists, the distance between the market price and the equilibrium price can be analyzed. Cointegrating regression can verify the cointegration of two securities. A vasicek model is proposed to analyze the stochastic process of the mean-reverting spread in [4].

Besides the statistical analysis of pairs trading, many scholars tried to research the factors of arbitrage. The arbitrage factors in the stock and future market are analyzed and it is found that extremely-high volumes and higher short-term volatilities are common arbitrage factors [5]. In [6], pairs trading strategy is proved to have great performance when the market is bumpy. In the recent global financial crises, pairs trading has the good profitability.

However, with the development of the program trading [7], computers have improved the efficiency of financial trading, which can help monitoring the market trend at any time and place an order in very short time. A great number of trading transactions are generated within milliseconds, thus here comes the high frequency trading [8]. Arbitrage trading also shorten the trading time into microseconds, so that traders need to place orders much faster to catch the opportunity. Therefore, the speed limit can be broke and more profit opportunities can be found if we can capture the signal in advance.

With rapid development of information technology, algorithmic trading is introduced to find out the market trading behavior and the timing of buying and selling. In recent years, many experts use machine learning to help analyze financial data due to its high commercial value. Machine learning can learn financial time series data and outperform compared to traditional methods. Most of machine learning methods in the finance field are introduced to learn the market behavior from a large amount of historical stock prices data and help predicting future market prices or trends [9]. Moreover, with the advent of Robo-advisor [10], auto-investing and how to make Robo-advisor intelligent is an important issue.

In the last few years, the latest and the most popular machine learning method - deep learning - has succeeded in many fields and outperformed the past research results. Deep learning imitates human brain by processing with multi-layers operations. This research tried to build the intelligent arbitrage trading system through deep learning. A gradient-based learning technique was proposed which can be applied to deep neural networks [11]. The researcher compared a lot of machine learning method and found that convolutional neural networks have the best ability of recognition in two-dimension data. Another method that can read millions of text within a day was also proposed as graph transformer networks (GTN), which reached the commercial value and become the milestone. In [12], the faster greedy algorithm-complementary priors algorithm - successfully reduced the demanded memory of training deep belief networks (DBN)

978-1-5386-8180-0/18/$31.00 ©2018 IEEE

and made training much faster. In [13], autoencoder successfully transfers the high-dimension data to low-dimension data by processing with multi-layers which is a brand-new neural network and can be trained by the gradient descent method. Compared to principal components analysis (PCA), it has better reducing dimension power towards non-linear data. Finally, deep learning is highly concerned by the academic community.

In this paper, we used Convolution Neural Network (CNN) which was proposed by LeCun and Bengio in 1995 [14]. In this year, multi-layers neural networks trained by the backpropagation algorithm were the candidate technique in image recognition and speech precognition because of the great learning ability towards more complex, high-dimensional, non-linear data. The previous pattern recognition models need to feed the man-made features, but CNN model doesn't need it. CNN can automatically extract useful features and have outstanding performances in processing image data. In ImageNet Large-Scale Visual Recognition Challenge (ILSVRC), many scholars applied CNN and got high accuracy. These great achievements made CNN popular and attracting. This research attempted to convert time-series price data into image data through a special visualization method. With the ability of capturing high-level features in images, CNN can learn the market knowledge from the time-series visualizations, then detect the arbitrage opportunities early thus earn stable and continuous profits.

II. PROPOSED METHOD

This paper attempted to build the intelligent arbitrage trading system in high-frequency pairs trading. We took Taiwan financial market as an example and researched the arbitrage set of TX and MTX, which was the most frequently traded in Taiwan. First, we collected the tick data and preprocessed these data. This research designed the rule-based strategy as our benchmark and backrested many sets of parameters to find the best parameter set. This parameter set was used to filter the tick data as the experimental data. Second, we proposed the improved time series visualization method, which referenced the moving average visualization [15]. Through this visualization method, we tried to transfer the tick price data into the two-dimension data, which was helpful to predict the arbitrage opportunities. Third, we proposed the improved CNN to construct the intelligent arbitrage trading system. This research compared not only the accuracy but also the profitability for the intelligent trading system.

Moreover, this research proposed filterbank CNN to extract features with expert domain knowledge, which replaced the random-generated filters with our knowledge-based filters.

A. Data and Data Preprocessing

In this research, we filtered the near-month tick data of TX and MTX to construct the intelligent arbitrage trading system in high-frequency pairs trading. The financial database provided by APEX International Financial Engineering Company was used, which contained all Taiwan futures trade books including the trading time, product sign, price and trading volume from 2012 to 2014. The records of the filtered data included trading time, product sign, trading price. The amount of these records was 17 million, 14.5 million and 12.8 million data respectively.

Next, we backrested these filtered data as our experimental data to fit the arbitrage rule and then transferring these experimental data to the images through our time series visualization method. These transferred visualizations were composed of different timeframes volatilities and CNN input.

B. Rule-based Strategy

This research referenced [16] and designed the rule-based strategy as the benchmark. This rule-based strategy had two key parameters, which needed to be backrested and used for filtering the tick data to comply with the arbitrage rule. In Equation (1) and (2), we constructed the Bollinger bands to measure the level of price deviation and find the arbitrage opportunities. R is the division of the TX tick price P_{TX} and the MTX tick price P_{MTX} *at the T tick*. The upper bond is the average of T-tick R added the standard deviation of T-tick R and the lower bond is the average of T-tick R subtracted the standard deviation of T-tick R. $\mu_{R,T}$ is the average of R_1 、 $R_2 \ldots R_{T-1}$ 、 R_T, and $\sigma_{R,T}$ is the standard deviation of R_1 、 $R_2 \ldots R_{T-1}$ 、 R_T, which $R_T = \frac{P_{TX,T}}{P_{MTX,T}}$.

$$\text{Upper Bond} = \mu_{R,T} + multiple \times \sigma_{R,T} \tag{1}$$

$$\text{Lower Bond} = \mu_{R,T} - multiple \times \sigma_{R,T} \tag{2}$$

In this chapter, we evaluated the performance of different parameters with the total trading count, the profit point, the count of gain, the count of loss and win rate. Simulated trading considered the trading cost, which was the tax rate of the Taiwan market maker and comply with the high-frequency traders. We tried to find the best parameters and make our benchmark. After the back test, we decided $\mathrm{T} = 160$ and $\text{multiple} = 2$ as the best parameters.

C. Time Series Visualization Method

This research referenced the moving average visualization method [15] and attempted to combine different timeframes and different previous ticks historical volatilities to the image. According to our hypothesis, we proposed the volatility visualization method to transfer many historical volatilities to the visualization where $Vol(i,j) = \sqrt{\frac{\sum_{T=t-j-i}^{t-j}(P_T-\bar{P})^2}{i}}$ and p is the future price time series (3). We tried to combine the previous 15 ticks and the 15 timeframes historical volatilities to the 15 x 15 image through our volatility visualization method. The row of the image is different timeframes and the column of image is different ticks. These images are the input data of the intelligent arbitrage trading system.

$$\text{Image} = \begin{bmatrix} Vol(1,1) & \cdots & Vol(i,1) \\ \vdots & \ddots & \vdots \\ Vol(1,j) & \cdots & Vol(i,j) \end{bmatrix}, 1 \leq \mathrm{i,j} \leq 15 \tag{3}$$

Through this transformation equation, we respectively transformed the TX and MTX tick price data to the visualizations. Despite the fact that human cannot find some pattern and trading rules in these images, filters can learn the relationship of some nearby ticks and historical volatilities from these images. Learning filters is just like the real traders who simultaneously consider different timeframes and different previous time technical indicators when trading securities. The visualization method is designed by this trading experience. CNN model can simultaneously consider more different timeframes and more different previous tick historical volatilities than the human brain and try more the relationship of historical volatilities than the human judgment. That's the reason why CNN can usefully capture the arbitrage opportunities, while people cannot know.

D. Filterbank CNN

In order to verify the relationship of long-term and short-term historical volatilities, this research designed the knowledge-based filters learned by knowledge-based filters instead of the random-generated filters. Therefore, we used filterbank CNN and added filterbank layer in CNN model. Filterbank CNN can extract more high-quality features based on the knowledge rule. The structure of filterbank CNN is shown in Fig.1.

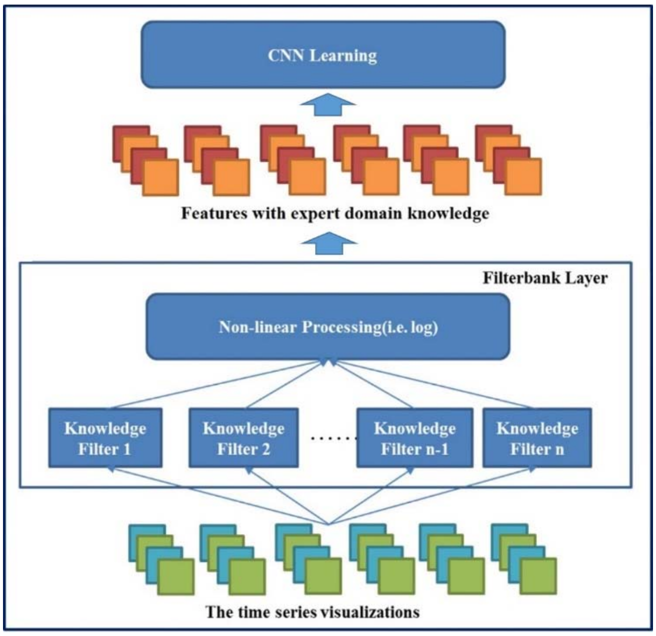

Fig. 1. Overview of Filterbank CNN flow chart

The operation equation of filterbank layer is shown in Equation (4) and (5). Features will be extracted from images through the knowledge-based filters, then processed the non-linearity operation (log layer) to make features become non-linear then finally generated features with expert domain knowledge. In comparison with the features generated by traditional CNN's filters, filterbank CNN can add the professional knowledge in CNN model, tune the knowledge-based filters, capture high-quality features and improve the learning ability.

$$\text{Feature}_{Knowledge} = filter^{T}_{Knowledge} \times Image \qquad (4)$$

$$Log\ Feature_{Knowledge} = log(\,Feature_{Knowledge}) \qquad (5)$$

Our knowledge-based filters are inspired by Haar-like filters, which are composed of 0, -1 and 1. The idea is based on the visualization row and the visualization column and generated 15 kinds of knowledge-based filters. These filters size is 6 x 6. Upper short-term volatilities subtracted the lower long-term volatilities brought out 6 kinds of row-side knowledge-based filters which volatilities were at the same tick. Short-term volatilities subtracted the long-term volatilities brought out 9 kinds of column-side knowledge-based filters and that the volatilities are at the same tick. Then Filterbank CNN used these 15 filters to extract features and adjusted the weight of filters through backpropagation.

III. EXPERIMENTAL RESULTS

The experimental results briefly displays the effectiveness of our system by two evaluation methods. Section A shows the accuracy of our classifier and section B describes the accumulated profits of the proposed system in simulated investment environments.

A. The Classification Accuracy Comparisons

Although CNN may be much more computationally complex compared with the moving average visualization method, Filterbank Learning CNN with the improved visualization method has up to 67.03% accuracy and 16.16% improved rate comparing to the rule-based strategy. In the experiment chapter, we compare the volatility visualization method with the moving average visualization method. Normal CNN is designed for testing the visualization method in this experiment which consists of 3 convolution layers, 3 pooling layers and 2 full-connected layers. More details about Normal CNN are not allowed to be listed due to limitations of space. The experimental parameters were referenced by [17] where learning rate equal to 0.001, and Adam optimizer is adopted. In addition, the training step, training period, testing period, and sliding windows are 12000, 6 months, 1 month, and 1 month, respectively. The experimental group is Normal CNN with the volatility visualization method. The control group is Normal CNN with the moving average visualization method [15] and rule-based strategy. First, this research shows the accuracy analysis. Fig.2 is the visualization of classification accuracy in different testing period which longing-TX-and-shorting-MTX model.

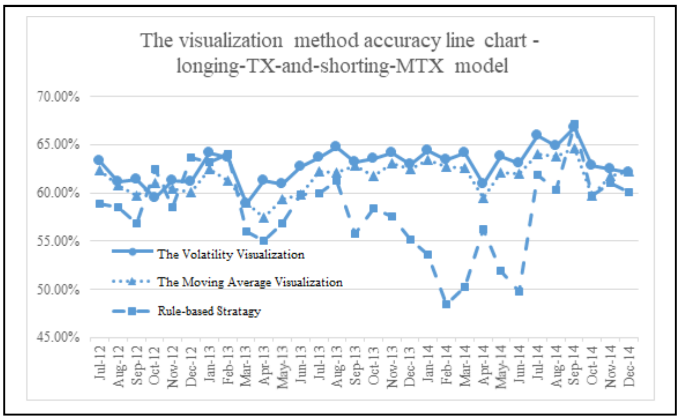

Fig. 2. The classification accuracy comparisons between the proposed system and vatrious models based on longing-TX-and-shorting-MTX strategy

In Fig.2, it can be found that the accuracy of the volatility visualization method is almost the best of three in longing-TX-and-shorting-MTX model. The volatility visualization method has up to 66.82% accuracy in October 2014 and at least 58.92% accuracy. The moving average visualization method [15] also has good predictability, but its performance is inferior to our visualization method. It has up to 64.63% accuracy in September 2014 and at least 57.48% accuracy. Both of visualization methods outperform rule-based strategy. There result proves that the volatility visualization method has the best ability of capturing arbitrage opportunities in longing-TX-and-shorting-MTX model.

B. The accumulated profits comparisons

The second experiment shows the profitability analysis which is the visualization method accumulated profit line chart - longing-TX-and-shorting-MTX model. As figure 3, we can find that the accumulated profit of the volatility visualization method is almost the best in longing-TX-and-shorting-MTX model. The volatility visualization method has the highest accumulated profit from 2014 Q3 to 2014 Q4. It earned 175,540.1 points finally, improved 8,216.8 points towards the moving average visualization method and improved 9,106.8 points towards the rule-based strategy. The moving average visualization method is unstable until 2014. Finally, it earned 167,323.3 points and improved 890 points towards the rule-based strategy. Therefore, the volatility visualization method has the greatest profitability in longing-TX-and-shorting-MTX model.

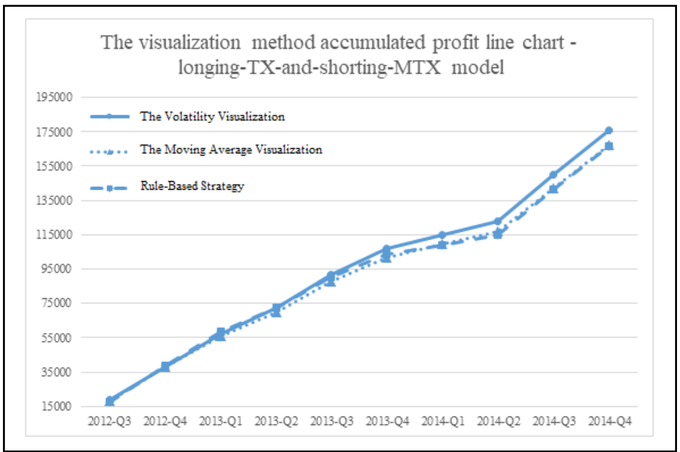

Fig. 3. The accumulated profits comparisons between the proposed system and vatrious models based on longing-TX-and-shorting-MTX strategy in simulated investment

IV. CONCLUSION

Through the time series visualization method and CNN, our intelligent trading system can successfully capture the arbitrage signal and thus enhance the profitability. Filterbank CNN replaces the random-generated filters with the knowledge-based filters and successfully improves the predictability. It has up to 67% accuracy and improves up to 15% accuracy in the comparison of traditional rule-based strategy. In the profitability, it can earn 360,362 points within 2.5 years. These results proves that filterbank CNN is better than CNN model in the intelligent pairs trading system. Also, combination of information technology and financial knowledge can create more profits.

Although this research has proved that our method has the better performance, we still think it can be improved in some aspects as follows:

A. Hybri deep learning model

This research proposed filterbank CNN to capture the arbitrage opportunities. The successors can try to deepen CNN model to test whether the deeper CNN model has the more accuracy and the high profitability. Moreover, deep learning has various model, which can be used in solving financial problems.

B. Tests on various arbitrage models

This research only experimented the future arbitrage of TX and MTX in the Taiwan financial market. However, there are many different combinations of arbitrage practices, so we hope that in the future the research can head to more different arbitrage models.

ACKNOWLEDGEMENT

The authors would like to thank the Ministry of Science and Technology of the Republic of China, Taiwan, for partially financially supporting this research under contract number MOST 106-3114-E-009-014 and MOST 106-3114-E-009-015. This work was also supported by the FinTech Innovation Research Center, National Chiao Tung University.

REFERENCES

[1] Shumway, R. H., & Stoffer, D. S. (1982). An approach to time series smoothing and forecasting using the EM algorithm. Journal of time series analysis, 3(4), 253-264.

[2] Elliott, R. J., & Krishnamurthy, V. (1999). New finite-dimensional Filters for parameter estimation of discrete-time linear Gaussian models. IEEE Transactions on Automatic Control, 44(5), 938-951.

[3] Vidyamurthy, G. (2004). Pairs Trading: quantitative methods and analysis (Vol. 217). John Wiley & Sons.

[4] Elliott, R. J., Van Der Hoek*, J., & Malcolm, W. P. (2005). Pairs trading. Quantitative Finance, 5(3), 271-276.

[5] Merrick, J. J. (1987). Volume determination in stock and stock index futures markets: An analysis of arbitrage and volatility effects. Journal of Futures Markets, 7(5), 483-496.

[6] Do, B., & Faff, R. (2010). Does simple pairs trading still work?. Financial Analysts Journal, 83-95.

[7] Grossman, S. J. (1987). An analysis of the implications for stock and futures price volatility of program trading and dynamic hedging strategies.

[8] Aldridge, I. (2009). High-frequency trading: a practical guide to algorithmic strategies and trading systems (Vol. 459). John Wiley and Sons.

[9] Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & Operations Research, 32(10), 2513-2522.

[10] Fein, M. L. (2015). Robo-advisors: A closer look. Browser Download This Paper.

[11] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11), 2278-2324.

[12] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. Neural computation, 18(7), 1527-1554.

[13] Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. Science, 313(5786), 504-507.

[14] LeCun, Y., & Bengio, Y. (1995). Convolutional networks for images, speech, and time series. The handbook of brain theory and neural networks, 3361(10), 1995.

[15] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, An-Pin Chen (2016). Financial Time-series Data Analysis using Deep Convolutional Neural Networks. Proceedings of the 7th International Conference on Cloud Computing and Big Data (CCBD2016).

[16] Jacobs, B. I., & Levy, K. N. (1993). Long/short equity investing. The Journal of Portfolio Management, 20(1), 52-63.

[17] Jou-Fan Chen, Wei-Lun Chen, Chun-Ping Huang, Szu-Hao Huang, An-Pin Chen (2016). Financial Time-series Data Analysis using Deep Convolutional Neural Networks. Proceedings of the 7^{th} International Conference on Cloud Computing and Big Data (CCBD2016).