INTRODUCTION TO SIGNAL PROCESSING AND MACHINE LEARNING IN FINANCIAL ENGINEERING

Chavinpat Naimee

Department of Electronic and Electrical Engineering, University of Strathclyde
16 Richmond St, Glasgow G1 1XQ, United Kingdom
email: chavinpat.naimee.2019@uni.strath.ac.uk

ABSTRACT

This article presents an introduction of financial engineering together with the relation between signal processing and machine learning on financial engineering and shows that machine learning can be applied to financial engineering. Even though It seems that Financial Engineering, signal processing and Machine Learning (ML) are in a different field, but these topics use similar fundamental that base on statistical analysis and modelling of systems. ML can be applied to the financial modelling process, which can help investors generate a compelling financial model and develop their investment strategies. Traditional model (ARIMA) and ML model (Logistic Regression or LR and SVM) that used to predict financial time series are presented. Comparing conventional and ML methods, SVM provided outstanding in price time series prediction, but ARIMA and LR also generated a good performance.

1. INTRODUCTION

Financial Engineering is the use of mathematical methods that are used to solve financial problems. It uses computer science, statistics, economics, and applied mathematics methods and information to tackle current financial problems as well as to formulate new and creative financial products [1].
Recently, in the financial industry, there are many new and innovative investment strategies and tools have been developed for investors and businesses. According to [2], Financial Engineering involves developing new financial instruments, along with innovative customer applications. For instance, new forms of bank accounts, mutual funds, insurance products and mortgages balancing. Moreover, it also includes the development of an application for corporate finance such as derivative products, futures, options, and risk management tools.
Over the last decade, signal processing and machine learning have evolved rapidly in many business and industrial environments [3]. Advertising, real estate, healthcare, e-commerce and many other sectors have been fundamentally changed by modern systems and activities that rely on the processing and analysis of data regarding operations, clients, rivals, emerging opportunities and other market aspects. The financial industry also has adopted a long history of sophisticated approaches and models for evaluating specific data and making wise decisions.
It seems that Financial Engineering, Electrical Engineering and Machine Learning are in a different field, but these topics use similar fundamental that base on statistical analysis and modelling of systems. Both areas are linked on a mathematic. The fundamentals of Financial engineering bases on the mathematical study of computational time series and simulation of financial markets behaviour make forecasts and systematically optimise investment strategies [4]. Likewise, the foundations of electrical engineering based on statistic and modelling of signal processing, for example, wireless communication systems make predictions and systematically refine transmission methods from the modelling of communication channels [4].
Financial Modelling is one of the most crucial section in the financial engineering process. The pricing model can be obtained from financial time series from this step. It is a challenging process for investor and professional analysts to improve their investment strategies that can contribute to gain more profit, which is the aim of investors [5].
In the following, the perspective of Signal Processing and Machine Learning on Financial Engineering will be presented. Then the comparison of financial time series forecasting using the traditional method (ARIMA) and machine learning method (Logistic Regression and SVM) will be discussed. The purpose of this article is to present the introduction to the financial engineering that makes the reader understand the process of financial engineering, the relation between signal processing and machine learning on financial engineering, and how machine learning can apply to financial engineering.
This paper mostly contains the literature reviews of the topic of Signal Processing and Machine Learning on Financial Engineering. Moreover, it also consists of the point of view and discussion of the author. The structure of the paper is organised as follows. Section 2 provided the literature review of Signal Processing and Machine Learning Perspective on Financial Engineering. Section 3 present financial time series modelling and prediction. Finally, Section 4 contains summarise of this paper.

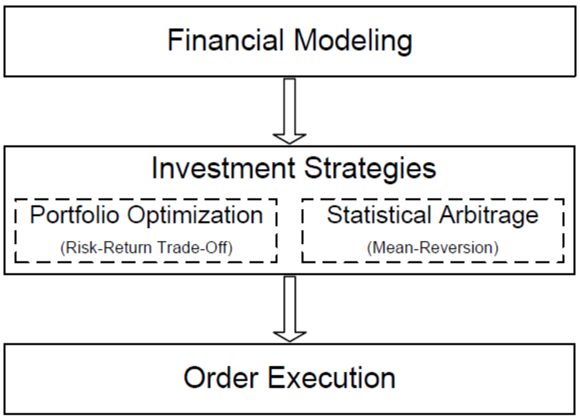

Figure 2.1 Block diagram of quantitative investment in financial engineering [4]

2. SIGNAL PROCESSING AND MACHINE LEARNING PERSPECTIVE ON FINANCIAL ENGINEERING

In this section, signal processing and machine learning perspective on the quantitative finance of equity securities will be presented.

Figure 2.1 shows the process of quantitative investment that contains three significant steps, which are financial modelling, portfolio design and order execution. Financial modelling is modelling that breaks down a noisy pricing time series into trend and noise components [1]. Portfolio design is developing quantitative investment strategies based on approximate financial models in order to optimise those preferred criteria. Order execution is the proper execution of orders to create or unwind optimally the positions of the portfolio planned.

2.1 Financial Modelling

A log-price series can be broken down into two parts from a signal processing view that is trend and noise components. These components referred to as market and distinctive elements, respectively [4]. Financial modelling or signal modelling has the purpose of decomposing trend components from the noisy financial sequence. Then the primary of the designed financial models can properly plan some quantitative investment strategies for future gains.

This article shows a simple and familiar financial model of the log-price series, which demonstrated in the equation below:

$$y_t = \mu + y_{t-1} + w_t \qquad (1)$$

where y_t is the log-price at a discrete-time t, w_t is a zero-mean white noise series, and μ is the time trend of the log-price y_t From equation (1), the trend of the signal and noise components in the log-prices can be re-written as below:

$$y_t = \mu t + y_0 + \sum_{i=1}^{t} w_i$$

where the term μt is the trend, which $\mu t > 0$ shows uptrend, $\mu t < 0$ shows downtrend and $\mu t = 0$ is no trend.

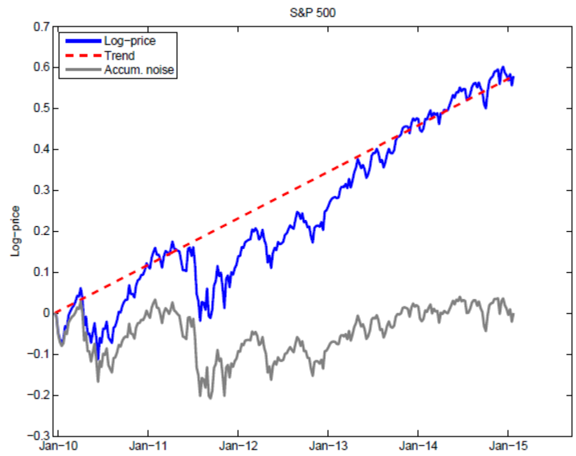

Figure 1.2 The decomposition of the log-price sequence of the S&P 500 Index into time trend component, and the component without time trend (i.e., the accumulative noise)

While the term $\sum_{i=1}^{t} w_i$ represents the accumulative of noise as time increases.

Figure 2.2 illustrates the weekly S&P 500 index log prices from 4 January 2010 to 4 February 2015, which this log-price were shifted down that start from zero, i.e. $y_0 = 0$. From this figure, two patterns are observed. First, there was a dramatic increase since 2010 in the US stock market. Second, the accumulative noise was not steady.

2.2 Investment Strategies

Once the financial model obtained from the financial time series, this can utilise to invest by applying quantitative methods to create investment strategies. From figure 2.2, there are two main elements in monetary series, trend and noise. Accordingly, there are two critical types of quantitative investing approaches that are based on the two components: a trend-based plan, called risk-return trade-off investing; and a noise-based strategy, called mean- reversion investment [4].

The risk-return trade-off investment term appears to optimise the expected portfolio return while keeping the risk low. The variance in portfolio return is taken as a measure of risk, and this approach is referred to as "mean-variance portfolio optimisation" [4]. Interestingly, from the view of signal processing, the concept of a mean-variance portfolio is mathematically similar to the idea of a signal processing filter [4].

While the mean-reversion investment tends to look for profitability based on the noise variable. [4] said that the idea of "pairs trading" or "statistical arbitrage" is an excellent example to understand the mean-revision investment factor. The pairs trading is a trading technique for two stocks with a high correlation that when these stocks are opposed to each other, a trader can short/sell the upwards one and long/buy the downwards one.

2.3 Order Execution

Ideally, the execution should be a smooth part of the process after a prediction and design a portfolio has been generated [1]. However, the process of executing the orders incorrectly affects the original predictions. Interestingly, the problem of

order execution is similar to many other issues in signal processing approach. For instance, [4] suggested that the order execution issue is very close to sensor scheduling in dynamic wireless sensor networks.
Moreover, machine learning can also apply to the order execution step to minimising the execution cost. For example, there are connections between the SVM method and Conditional Value-at-Risk (CVaR) minimisation.

3. FINANCIAL TIME SERIES FORECASTING

Financial time series forecasting is a big challenge for traders, investor and professional analysts to predict. The stock market forecasting is not easy [5]. The results of prediction to improve their investment strategies that can contribute to gain more profit, which is the aim of investors.
For price movement prediction, the financial model is an essential tool to make a prediction. ARIMA or Autoregressive Integrated Moving Average is a traditional predicting method that can perform well [5]. However, there have been many researched using machine learning and deep learning in this area that could perform better.
In this paper, a traditional technique which is ARIMA and machine learning approaches, Logistic Regressing and Support Vector Machine, will be demonstrated.

3.1 Introduction to Model

3.1.1 ARIMA (Autoregressive Integrated Moving Average)

ARIMA is a modified model of ARMA that combines the AR method and MA method. The non-stationary series can be reduced to stationary series using a sequence of differences. If a time series is stationary, the series data ACF and PACF figures can help determine the order of the sequence ARIMA [5]. The simple equation of ARIMA can be written as below:

$$y_t = \sum_{i=1}^{n} \alpha_i y_{t-i} + \sum_{i=1}^{m} \beta_i u_{t-i} + \mu_t$$

where y_t is the different time series value, α and β are unknown parameters, and μ is independent identically distributed error terms with zero mean. y_t is expressed in terms of its past values and the real and previous error terms values.

3.1.2 Logistic Regression (LR)

Logistic regression is a classification method, which can use to predict an outcome that is a category [5]. It will predict the outcome based on the input variables and works well in many real problems. The simple model can be written as below:

$$y = \frac{1}{1 + e^{-\beta^T Z}}$$

where: $\beta = (\beta_0 + \beta_1 + \beta_2 + \cdots + \beta_n)^T$ is a parameter vector, and $Z = (x_0 + x_1 + x_2 + \cdots + x_n)^T$ is a feature vector. In fact, the function y demonstrates a logistic distribution function, as shown in figure 3.1.

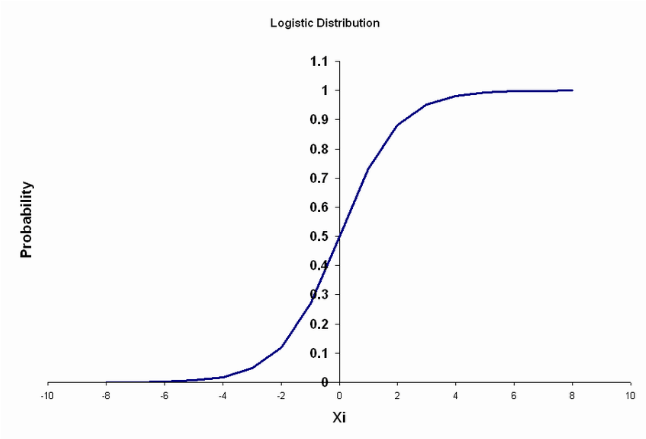

Figure 3.1 The logistic regression distribution

The classification rule can be shown as below:

$$classification = \begin{cases} 1 & if\ y > 0.5 \\ -1 & if\ y \leq 0.5 \end{cases}$$

3.1.3 Support Vector Machine (SVM)

A Support Vector Machine is a versatile and robust algorithm for binary classification in machine learning with both linear and non-linear classification. It uses a kernel trick technique to transform data, and it could find an optimal boundary between the possible outputs based on its transformations [6]. Moreover, it could do utterly complex relationships between data points and then evaluate how to separate the data based on their labels, particularly for non-linear SVM that the model does not have to be the straight line. The decision boundary of an SVM model not only divides the two categories but also remains as far away from the nearest training samples as possible. Figure 3.2 shows that all L1, L2 and L3 can separate the data correctly, but L2 seems to be the best because it remains as far away from the nearest training samples to achieve maximum distances from both clusters [5].
Moreover, SVM is different from several other Machine Learning models who minimises empirical risk. SVM will try to minimise structure risk that can efficiently avoid the overfitting problem [5].

3.2 Result Comparison

In this section, the comparison of each model for price time series prediction will be discussed.
According to [5], the experiment used S&P, Dow 30 and Nasdaq as data that were collected from Yahoo.finance from 1 February 2012 to 26 December 2016, which includes open, close, high, low price and the trading volume for the day. It was implemented in Python programming and ran it on PyCharm. ARIMA model can be found in Python's open-source tool, "statsmodels". While machine learning model is provided in "Sklearn" library. Other parameters are used in this experiment can be found in [5]. [5] and [6] used a hit ratio to measure the accuracy of the forecasting that the equation is shown in equation (2):

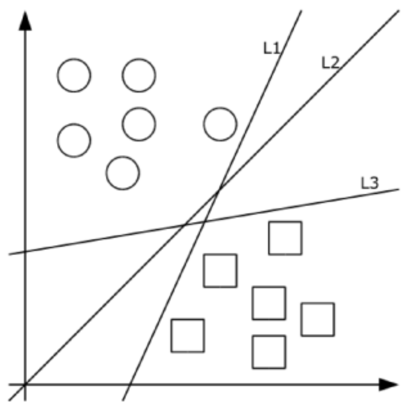

Figure 3.2 The support vector machine diagram

$$hit\ ratio = \frac{\sum_{i=1}^{N} prediction_i}{N} \qquad (2)$$

where, $$prediction = \begin{cases} 1 & P > 0 \\ 0 & else \end{cases}$$

and $$P = prediction_i \times real_i$$

From [5], they obtain the results that the best hit ratio of ARIMA, LR, and SVM are 0.593, 0.623, and 0.642, respectively, see figure 3.3.

3.3 Discussion

From the results, SVM provided outstanding in price time series prediction, while ARIMA and LR also gave good accuracy. ARIMA only can extract information from financial time series, but machine learning can take other features to train the model that can also use feature aside from price time series.
In this paper, only the traditional and machine learning method are presented. However, there are many machine learning and deep learning approaches that can be tried in financial time series' forecasting such as variants SVM (v-SVM), MLP, and LSTM. Especially, LSTM tends to perform outstanding model because it uses entire sequences of data as input of the next calculation that might suitable for time series prediction. Besides, [3] suggested that v-SVM also provided excellent performance in this kind of problem.

4. CONCLUSION

This paper presents an introduction of financial engineering together with the relation between signal processing and machine learning on financial engineering and shows that machine learning can be used to financial engineering.
Even though It seems that Financial Engineering, signal processing and ML are in a different field, but these topics use similar fundamental that base on statistical analysis and modelling of systems.

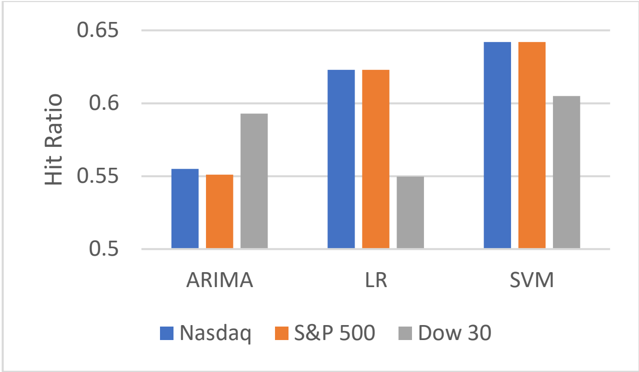

Figure 3.3 The results of different prediction in each market

Machine Learning can be applied to the financial modelling process, which can help investors generate a compelling financial model that could assist them in developing their investment strategies. Traditional model (ARIMA) and ML model (Logistic Regression or LR and SVM) that used to predict financial time series are presented. The results show SVM gave the best accuracy, while ARIMA and LR also generated an excellent performance.
Future research will include other applications of financial modelling that use additional machine learning and deep learning techniques such as v-SVM, MLP, and LSTM. Another topic of study may focus on financial engineering in portfolio optimisation and execution order using machine learning and deep learning. Besides, another technique, stock trend prediction using sentiment analysis is a fascinating topic in these days.

REFERENCES

[1] A. N. Akansu, *A primer for financial engineering : financial signal processing and electronic trading*. London, UK : Academic Press, 2015.
[2] J. Finnerty, "Financial Engineering in Corporate Finance: An Overview," *Financial Management,* vol. 17, no. 4, p. 14, 1988, doi: 10.2307/3665764.
[3] A. N. Akansu, S. Kulkarni, D. Malioutov, d. Ieee Xplore, and p. Wiley, *Financial signal processing and machine learning*. West Sussex, United Kingdom : Wiley : IEEE Press, 2008.
[4] F. Yiyong and D. P. Palomar, "A Signal Processing Perspective on Financial Engineering," ed. Boston, USA: now Publishers Inc., 2016.
[5] Q. Xin-Yao, "Financial Series Prediction: Comparison Between Precision of Time Series Models and Machine Learning Methods," *arXiv.org,* 2018.
[6] K. J. Kim, "Financial time series forecasting using support vector machines," no. 28 February 2002, pp. 307 – 319, 2003.