Information Mining Over Significant Interval on Historical Data

A Study on World Major Indexes

Sim Kwan Hua
School of Engineering, Computing & Science
Swinburne University of Technology
Kuching, Malaysia
e-mail: khsim@swinburne.edu.my

***Abstract*—Despite the popularity of financial charting software catalyzed by the advancement in computing technology over the past decade, the analysis of financial historical data through charting software remains at the surface of statistical description. Analysis of historical data presented typically on a price chart should be elevated further to information mining that could interpret the fundamental condition on the ground, as an important effort to support a good decision making process. This paper introduces a new way of interpreting historical financial data by calculating the mean value of the historical data over a significant interval, and eventually mining the high intensity price level. Experiment was conducted on historical data of six major world indexes over the period of ten years to assess the competency of the use of mean value over significant intervals in comparison to static interval used in conventional moving averages. The outcome of the experiment reveals the relevancy of the use of mean value over significant intervals on all the six major world indexes. This study institutes and demonstrates a new way of mining fundamental information and insight from a historical data set; the finding stimulates an innovative way on how data can be interpreted to derive information that is crucial in financial decision making process.**

Keywords-Data Analysis; Time Series Analysis; Data Mining; Statistical Analysis; Technical Analysis

I. INTRODUCTION

Over the past one decade, financial charting software has been a primary tool used in analyzing exchange markets such as stocks, futures, commodity and foreign currency exchanges. It is reported that more than 75% of professionals in the exchange markets of various instruments rely on financial charts at some points in making their trading decisions [1]. Therefore, mining the information required from the price chart forms part of the critical components in the decision making process of financial market participants.

The advancement of computing and software technology has made financial charting software widely accessible to institution and individual retailers; some of financial software providers include MetaTrader from MetaQuates, Routers MetaStock, TD Ameritrade, Thinkorswin from Bloomberg and Lauchpad & Charts from TradeStation. It is a very common practice nowadays for brokerage firms and investment banks to integrate real-time financial charting components into their trading software for use by their dealers and clients.

Furthermore, massive volatility in the more recent year in most of the financial instruments worldwide has quite a tremendous impact not only on economics but also on both social and political aspects. It provokes a major challenge confronting everyone ranging from investors, speculators, businesses, and even to the level of governmental policy makers. As a matter of fact, financial markets are affected by many highly interrelated variables such as economics, political and even psychological factors in a very complex manner, making financial time series one of the most difficult analyses among all other time series analysis [2].

Market participants worldwide have always shown keen interest trying to predict the future of the overall market movement in order to maximize their returns, at the same time hedging and mitigating their risks again potential pitfalls. Real-time charts provided by charting software showing the latest price change has become the primary tools in helping them in their day to day decision making process. Nevertheless, price chart is only a graphical representation of historical data from the past. Thus, information mining from the price chart becomes extremely vital to conclude a reliable decision which is in alignment with the future price movement.

Numerous studies have been attempted trying to analyze the financial markets by using various time series analysis techniques, alongside with few other popular time series models and stochastic models used in signal processing. Every single data point on a price chart is a successful transaction recorded at certain point in time, and it can be regarded as a signal reflecting the equilibrium states of all the correlated factors, may it be fundamental, technical or even emotional factor [3].

Nevertheless, most of the approaches introduced so far utilize historical data by performing learning or training on them in order to identify certain patterns, or to compute probabilities, or even optimizations which involve mostly the use of Artificial Intelligent techniques. All these are based on the assumption that the history will repeat itself. But, the

 ISBN: 978-1-61208-227-1

existence of randomness element in financial market claimed by Efficient Market Hypothesis (EMH) is undoubtedly hampering the performance of those approaches.

To date, very little effort has been made to mine information over a selected interval on the price chart. In most cases, the whole set of historical data will be used or fed into a model without any effort of information mining to pre-process the data. Even in technical analysis, which is starting to gain some grounds, majority of the technical indicators remain as a statistical summary of a set of data points in the past.

This paper aims to mine the underlying transaction information from historical data set of a price chart by identifying significant interval. In other words, the amount of data in the past is not the sole determinant in order to mine the required information. Hence, it does not operate on the assumption that history will repeat itself, and essentially eliminating the concern of random element in financial market.

The experiment and testing are done on six major world indexes on data over the past twelve years, starting from year 2000, and the results will be analyzed and discussed.

This paper begins with Section I as an introduction; Section II concerns the background; Section III elaborates on the mean over significant interval approach; Section IV describes the experiment conducted; at the same time discusses analytical results. Section VI presents the conclusion.

II. BACKGROUND

Generally, there are three schools of thought in financial market analysis; they are generally known as Efficient Market Hypothesis (EMI), fundamental analysis and technical analysis.

Efficient Market Hypothesis believes that no one can achieve above average advantages based on any historical and present data. This is also backed by another prominent, Random Walk Hypothesis, which states that prices of financial instruments wander in a purely random way. Likewise, Efficient Market Hypothesis advocates that all available information is fully reflected on the price itself [4]. Both theories dictate that the previous change in the value of a variable, such as price, is unrelated to future or past changes. As a result, statistical data collection implicitly defines each data point as independent. Based on such contextual assumptions, the data can appear random when the data points are treated as discrete events. However, S. Taylor (1986), Russell and Torbey (2008) provide compelling evidence to reject both of these theories [5].

Fundamental analysis from the second school of thought studies the underlying intrinsic value of a financial instrument by analyzing fundamental factors such as economics, financial, accounting and business environments, and their effects on its future value. Though fundamental analysis possesses the longest and profound history in the world of financial analysis, it is not meant for studying the volatility and fluctuation of prices around the underlying intrinsic value of a financial instrument. The huge volatility in financial and commodity markets lately has denoted the inadequacy of fundamental analysis to attest those huge fluctuations.

The third school of thought is technical analysis with practitioners who analyze primarily upon charts that based solely on market-delivered data such as price and volume. They perform statistical study rather than examine the economics fundamentally driven information in analyzing a financial instrument. Consequently, technical analysis does not receive sufficient level of scrutiny from academic researchers. As such, it served more as a secondary tool in financial market analysis [6].

Recently, researchers have introduced many different approaches in various fields of study as an effort to derive more useful information from historical data, which may include both fundamental data and technical data in order to analyze the financial markets.

Together with some other Artificial Intelligence (AI) techniques, Artificial Neural Network (ANN) has been one of the popular models for predicting financial markets. Cao Qing, et al. [7] and R. M. Rahman, et al. [8] in their respective studies, had used the neural networks to predict the future movements of various financial instruments. The results showed that the performance of ANN technique in forecasting financial instruments was very convincing and outperformed conventional linear models.

Manish Kumar and Thenmozhi M [9] endeavored to the use of Support Vector Machines (SVM) on S&P CNX NIFY; their result showed that SVM outperforms neural networks, discriminate analysis and logic model used in the study. Besides, stochastic model is another prevalent preference; the work of B. Kaushik [10] specified a two-state Markov Chain Model for discredited returns and proposed a measure for efficiency by using the modulus of the second highest Eigen value of the transition matrix and relates it to the speed of convergence of the Markov chain. In a more recent study, S. Vasanthi, et al. [11] explored the capability of Markov Chain Analysis in predicting indexes of emerging markets, and the result showed that Markov model outperforms the conventional moving averages in technical analysis.

Apparently, models were built from various areas of studies over the years by processing a wide range of both fundamental and technical financial data in mining useful information to aid in the decision making process of financial market participants [3]. So far, little attention has been paid to pre-process the data before feeding the data into the model for analysis.

In technical analysis, models have been developed mainly based on historical price data, statistical calculation such as mean, standard deviation and the rate of change will be performed to compute the statistical description required to form an understanding of the latest state of a given data set. Coherently, different data set will derive different statistical description; even selecting two different intervals from the same data set will produce two very different statistical descriptions.

Typically, statistical analysis will be represented graphically through charting software in the form of technical indicators such as Moving Average, Relative Strength Index, Stochastic and Moving Average Convergence Divergence. Analysis is normally done by selecting a technical indicator to be applied on a set of historical data plotted on a price chart.

Basically, a price chart is composed from price feed supplied by the exchange market, where vertical axis represents the price level and horizontal axis states the time when the transaction is made. In other words, price chart is just a graphical representation of historical data plotted on a chart. Applying a statistical analysis such as an *n*-period Moving Average on a chart simply insinuates the deriving of mean value from the last *n* data points [12]. The average or mean value of a set of price data points derived in the past has very little provision on the future value, and to the direction of the new price value in the future.

This is also supported by Efficient Market Hypothesis and Random Walk Theory which believe that price is a discrete event, thus any information derived from past data, including the mean value provides no influential trace on the possible value of price in the future.

Instead of pondering around the typical vindication of statistical description, this study aims at mining market information from the mean values by focusing on the selection of appropriate intervals used in calculating mean value. Fundamental information can be mined from mean value if the interpretation is based on significant intervals from the highest or the lowest price level recorded on a price chart. Since moving average is the main way of obtaining mean value in financial charting, the investigation will therefore, propagate from Moving Averages in technical analysis.

III. INTERVAL OF MOVING AVERAGES

In statistics, moving average is a type of finite impulse response filter used to analyze a set of data points by creating a series of averages of different subsets of the full data set [12].

Moving average smoothes a data series over the fluctuation for a specified time period to delineate and spot the overall trend of the past data. In financial series analysis, moving average is defined as the average (mean) price of an instrument over a specified time period.

Given a sequence $\{a_i\}_{i=1}^{N}$, an *n*-moving average is a new sequence $\{s_i\}_{i=1}^{N-n+1}$ defined from the a_i by taking the average of subsequences of *n* term.

$$s_i = \frac{1}{n} \sum_{j=i}^{i+n-1} a_j \tag{1}$$

In simple terms, it is calculated by adding the instrument prices for the most recent *n* data points and then dividing by *n* [12].

Over the years, research efforts have been directed to improve analytical power of moving averages through the introduction of various types of moving averages, such as exponential moving average, weighted moving average, accumulative moving average, triangular weighting, double smoothing, triple exponential moving average (TEMA), geometric moving average and many more [13]. Nonetheless, three most commonly used moving averages are simple moving average, exponential moving average and weighted moving average [14].

Exponential moving average (EMA) is a type of infinite impulse response filter that applies weighting factors which decreases exponentially. The EMA for a series *Y* is calculated recursively with:

$$S_t = \alpha \times Y_t + (1-\alpha) \times S_{t-1} \tag{2}$$

Where the coefficient represents the degree of weighting decrease, it may be expressed in terms of *N* interval, where $\alpha = 2/(N+1)$. Y_t is the observation at a time period *t* and S_t is the value of the EMA at any time period *t* [13].

A weighted moving average is any average that has multiplying factors to give different weights to data at different positions in the sample window; it is expressed in the following general form:

$$W_t = \frac{w_1 P_t + w_2 P_{t-1} + \cdots + w_n P_{t-n+1}}{w_1 + w_2 + \cdots + w_n} = \frac{\sum_{i=1}^{n} w_i P_{t-i+1}}{\sum_{i=1}^{n} w_i} \tag{3}$$

This gives the weighted moving average at time *t* as the average of the previous *n* prices (P), each with its own weighting factor w_t [13].

The introduction of a variety of moving average serves as a strong evidence of attempts to improve the meagerness of moving average via mathematical approaches. However, besides optimization techniques to find the most optimal interval, literally no other attempt has been made on the selection of intervals for calculating moving average.

As price chart records data points of prices transacted over a specific period of time, calculating the mean of the data points signifies the average transaction price over that period of time, which brings forth a very important information; the average overall price level of all the participants who involved in the transaction during that interval.

Hence, the selected mean interval is the key in mining useful mean or average transaction price. If the interval used to calculate mean from the highest or lowest price onward, then the portfolio condition of one who has traded during that interval can be inferred easily. In other words, a useful average transaction price level can be mined by selecting a significant interval which starts from a highest or lowest price level.

As the overall average position over a significant interval can be known, decision to be made by majority of the market participants can then be anticipated. According to decomposition effect of Prospect Theory, intensity of reactions and behaviors of market participants will mount when the new price level approaches their overall average position, which unquestionably will affect the position of their portfolios [15].

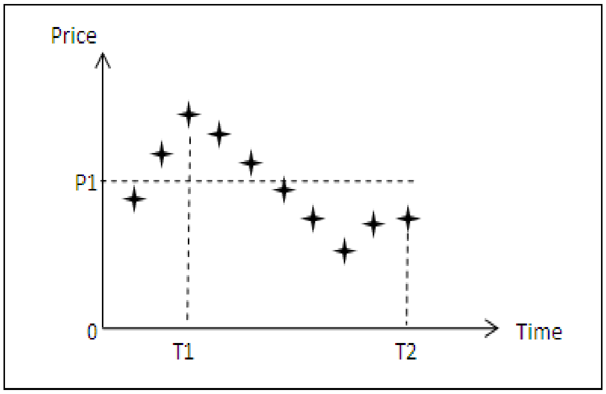

Figure 1. Mean over significant interval.

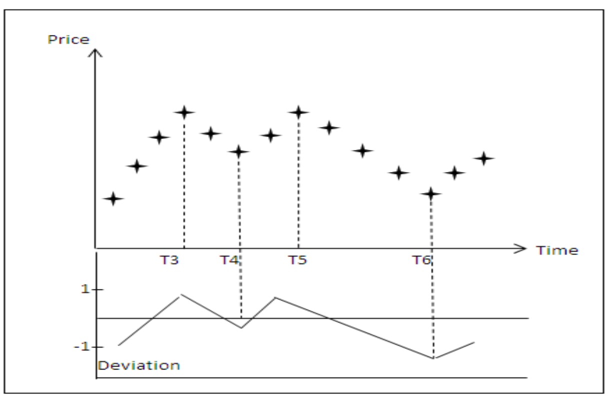

Figure 2. Defining highest and lowest point.

As illustrated in Figure 1, since the mean can be interpreted as the average transaction level over a specific interval, it is critical to mine the mean for the interval from T1 onward, interval from T1 onward can be categorized as significant interval because the highest price is recorded at T1. For that reason, all the participants who entered after T1 will have a negative impact on their portfolios. The mean transaction level for this group of participants is collective at price P1 when time reaches T2. In other words, if the price at T2 were to reach P1, the portfolios of this group of participants will turn from negative to positive, allowing them to reach their breakeven point, and high intensity of reaction can be anticipated in the decision making process of this particular group of people.

However, prices of financial instruments especially those with high liquidity will fluctuate at various degrees of magnitude, causing the highest and lowest price to be very subjective when it comes to identifying significant interval as time moves on. Thus, it is necessary and imperative to exercise objectivity in identifying the highest and lowest price.

Therefore, a significant interval used in this study is defined as the duration of movement of price exceeding one standard deviation away from the 50-period mean, which denotes a high intensity of consensus among market participants has been reached on the forming of a trend [16]. As illustrated in Figure 2, duration from T3 onward will not be considered as a significant interval because price at T4 is less than 1 standard deviation away below 50-period mean. On the other hand, interval from T5 onward is considered as a significant interval after T6, where price movement has exceeded one standard deviation away at T6. So, the mean value over a significant interval should be calculated after T6, and this mean value carries vital information on the overall position of market participants. It is extremely useful in making a favorable decision when price approaches that mean value at any point of time after T6.

IV. Experiement and Evaluation

The experiment was conducted to observe the reaction of price when it approaches the mean value over a significant interval. This experiment was done on six major world indexes over a period of twelve years starting from 2000 to the end of 2011. These major world indexes include Dow Jones Industrial Average (DJI), Deutscher Aktien-Index (DAX), CAC40, FTSE100, Hang Seng Index (HSI) and Nikkei225.

The experiment data used is the daily data of those six major world indexes, downloaded from Yahoo Finance and the experiment was carried up by using Meta Stock's system tester.

The evaluation was performed by comparing the reaction of price between various conventional moving averages with typical static intervals and mean value derived over significant intervals dynamically. Conventional static intervals used for comparison in this experiment range from 10-period to 200-period moving average.

In order to differentiate between the normal price fluctuation and decent price reaction upon approaching high intensity price level, the average daily range of all the indexes needs to be computed. This is to identify and estimate the normal daily range of price fluctuation for all the six major indexes, so that the abnormal price fluctuation can be classified.

Since the historical data used in this experiment is the daily end-of-day data, the average daily range of all six major indexes over twelve years testing period was calculated and summarized in the following table:

 ISBN: 978-1-61208-227-1

TABLE I. AVERAGE DAILY RANGE

Index	CAC	FTSE	HSI	NIKKEI	DJI	DAX
Average Daily Range (%)	1.78	1.63	1.53	1.52	2.35	1.99

It is shown in Table I that the averages of daily range for all the indexes studied are logged within range of 1.52% to 2.35%. This implies that on average, most indexes have a normal daily fluctuation of within approximately two percent of the index itself.

Consequently, in this experiment, a valid price reaction toward high intensity price level is defined as a price reversal into opposite direction for more than five percent from the mean value. It is set to be at least twice of the normal daily fluctuation range, so that the validity of the price reaction can be attested, hence minimizing the possibility that the price reaction is caused by normal a daily fluctuation.

Thus, once the new price data reaches the mean value, a reaction will be captured and recorded if the price reacted or bounced into opposite direction for more than five percent.

However, it will be recorded as no reaction if the new price level continues to move in its preceding direction for more than two percent after hitting a mean value. Likewise, the two percent buffer is to allow the normal daily fluctuation of the indexes.

Simulations were run on twelve years historical data of the six major world indexes individually, and evaluations on the price reaction toward mean value of significant interval along with conventional static moving averages were recorded. Correspondingly, the observation and evaluation of price reaction during the simulation were done for both upward and downward movements.

The outcomes obtained from the experiments are presented in the following tables:

TABLE II. UPWARD MOVEMENT

Interval	CAC	FTSE	HSI	NIKKEI	DJI	DAX
MA10	30.06%	31.68%	28.85%	30.49%	34.65%	31.01%
MA25	33.67%	30.88%	34.44%	25.49%	36.71%	29.69%
MA50	28.95%	28.00%	28.57%	28.57%	32.35%	27.54%
MA100	30.43%	25.81%	36.11%	27.91%	24.00%	26.83%
MA150	30.00%	34.78%	44.00%	25.93%	20.51%	38.71%
MA200	20.83%	30.00%	35.00%	22.22%	12.50%	42.86%
Significant Interval	58.06%	56.00%	60.87%	61.25%	55.56%	55.13%

TABLE III. DOWNWARD MOVEMENT

Interval	CAC	FTSE	HSI	NIKKEI	DJI	DAX
MA10	27.47%	30.28%	30.72%	33.54%	31.72%	27.34%
MA25	27.36%	25.00%	24.00%	39.39%	27.27%	28.81%
MA50	30.67%	33.33%	23.53%	30.00%	26.67%	27.63%
MA100	30.95%	34.48%	25.64%	37.14%	36.36%	27.66%
MA150	32.14%	33.33%	26.09%	45.83%	40.00%	21.21%
MA200	42.86%	42.11%	27.27%	30.43%	42.42%	11.11%
Significant Interval	61.29%	55.93%	65.91%	59.26%	68.29%	59.38%

The result in Table II exhibits that in an upward movement, there is a probability of approximately 30% for price to bounce into opposite direction in response to the moving averages with static intervals ranging from 10 periods to 200 periods. The best probability recorded among static interval moving averages is 44% from a 150-period moving average on HSI, while the lowest probability deeps as low as 12.50% from a 200-period moving average on DJI.

Notably, the mean from significant intervals yields a substantially higher probability of well above 50%; indeed it ranges from 55.13% to 60.87% across all the six major indexes evaluated in this experiment.

Similarly, Table III presents the outcome for downward price movement, and the result obtained is very consistent with the result attained from the upward price movement in Table II. All the conventional moving averages with static intervals have also recorded a probability of around 30%. Again, mean over significant intervals has achieved a probability of well above 50%. Moreover, the highest probability for mean over significant intervals in a downward movement reached as high as 68.29% on DJI.

A promising reaction of price toward mean value over significant intervals has been observed on all the six major world indexes over the twelve years testing period, which obviously cover different phrases of market condition. Conversely, such price reaction fails to be observed on any of the conventional moving averages with static intervals. Clearly, the mean value over significant intervals has the capability to mine a high intensity price level that causes noticeable price reaction.

In summary, the results reveal that the mean value over significant intervals has absolute higher probability of getting the price to bounce into opposite direction compared to all the conventional static intervals of moving averages. The obvious differences signify that information mining from calculating mean of significant intervals deserve a serious attention.

V. CONCLUSION

This study explored a new dimension of using technical analysis to mine information that represents the underlying condition, and not just based purely on statistical description. It is believed that the mean calculated from a

 ISBN: 978-1-61208-227-1

significant high or low can be interpreted as a high intensity average price level of majority of the market participants. The experiments and preliminary results suggest that mean over significant intervals demonstrate a very promising accomplishment compared to the other conventional moving averages with static intervals. This has also initiated a new epoch of how technical analysis can be interpreted to mine information from a price chart, and to elevate the performance and superiority of charting software in financial market analysis. Last but not least, future research should explore the possibility of extending this concept of mining fundamental information right from price chart to other technical indicators, with an expectation for more precise information mining in financial decision making.

REFERENCES

[1] T. Gehrig and L. Menkhoff, "Extended Evidence on The Use of Technical Analysis in Foreign Exchange", International Journal of Finance & Economics, Vol 11, Issue 4, 2006, pp. 327-338.

[2] G. Boetticher, "Teaching Financial Data Mining using Stocks and Futures Contracts", *Journal of Systemic, Cybernetics and Informatics*, Vol 3, no 3, 2006, pp. 26-32.

[3] J. V. Hansen, J.B. McDonald, and R. D. Nelson. "Some Evidence of Forecasting Time-series with Support Vector machines", Journal of Operational Research society, vol. 57, no.9, 2006, pp. 0153-1063.

[4] M. G. Kendall and H. Bradford, "The Analysis of Economic Time-Series- Part1: Prices". Journal of the Royal Statistical Society, 116(1), 1953, pp. 11-34.

[5] W. M. Martin, "Technical Anaysis: The interface of Retional and Irrational Decision Making" Business Review, Cambrige, 11. 2, 2008, pp. 48-54.

[6] K. P. Hanley, "Scientific Frontiers and Technical Analysis", Journal of Technical Analysis, vol. 64, 2006, pp. 20-33.

[7] Q. Cao, K. B. Leggio, and Schniederjans, "A Comparison between Fama and French's model and Artificial Neural Networks in Predicting the Chinese Stock Market", Computer and Operations Research 32, 2005, pp. 2499-2512.

[8] R. M. Rahman, R. K. Thulsiram, and P. T. Thulasiraman, "Performance Analysis of Sequential and Parallel Neural Network Algorithm for Stock Price Forecasting", International Journal of Grid and High Performance Computing, 3(1), 2011, pp. 45-68.

[9] K. Manish and M. Thenmozhi, "Predictability and Trading Efficiency of S&P CNX Nifty Index Returns Using Support Vector Machines and Random Forest Regression", Journal of Academy of Business and Economics, Vol.7(1), 2007, pp. 15-23.

[10] B. Kaushik, "A Measure of Relative Effieciency of Financial Markets from Eigen value based Mobility Indices", Finance India, Vol.XVI, No.4, 2002, pp. 1419-1425.

[11] S. Vasanthi, M. V. Subha, and Thirupparkadal Nambi, "An Empirical Study on Stock Index Trend Prediction using Markov Chain Analysis", Journal of Banking Financial Services and Insurance Research, Vol 1, 2011, pp. 72-91.

[12] M. F. Triola, "Essientials of Statistics, 4th Edition", Addison-Wesley, Longman Inc, 2011, pp. 112-126.

[13] P. J. Kaufman, "Trading Systems and Methods", John Wiley & sons, 1998, pp. 66-88.

[14] L. Stevens, "Essential Technical Analysis: Tools and Technique to Spot Market Trend", John Wiley and Sons, 2002, pp. 218-238.

[15] M. Massa and N. W. Geotzmann, "Disposition Matters: Volume, Volatility and Price Impact of a Behavioral Bias", Yale ICF Working Paper, No. 03-01, 2003.

[16] K. H. Sim, "Recognizing the Formatioin of Trend: A standard Deviation Approach", Proceeding of 4th International Conference of Interation Sciences: IT, Human and Digital Content", IEEE press, 2011, pp. 136-142.

 ISBN: 978-1-61208-227-1