International Journal of Business and Management 1 (2): 226-231, 2017
e-ISSN: 2590-3721
© RMP Publications, 2017
DOI: 10.26666/rmp.ijbm.2017.2.34

Analysis of Time Series Data Mixed With Text

David RAMAMONJISOA[1], Yoshiki SATO[1], Yuki SEGAWA[1]
[1]Faculty of Software and Information Science, Iwate Prefectural University, Japan,

Abstract: In this paper, we present the result of our experiment on analyzing a time series data such as the Nikkei 225 index or foreign currency exchange USD/JPY past data and text corpus related to Japanese economy and finance news or reports. There are several researches reported that forecasting of time series with additional features based on text data can be beneficial rather than relying on time series data history only. Experts on investment are usually making their decision based on those text data as they found patterns on them called the fundamental analysis. Time series analysis based on past history only are called technical analysis. The combination of them should make a better prediction system. The text data also provides an explanation or indication to the trend patterns (uptrend, downtrend, or trend reversal), important moves (spikes, releases) and/or volatility. We collected all Nikkei 225 index and USD/JPY past data on daily average closing prices, all Bank of Japan monthly reports and all related available news text data. We applied on the time series Nikkei 225 data the technical analysis such as simple moving average (SMA), regression, ARIMA models and seasonal return rates, GARCH model combined with text data features such as topic models. We aggregate topics yearly or a long period of 5 to 10 years. Topics are extracted with topic modeling techniques such as Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NMF). Through the observations of those technical and fundamental analyses, we show our results and conclude with some forecasts data.

Key words: *time series, text data, technical analysis, topic model*

INTRODUCTION

Stock and foreign exchange market forecasting is an interesting problem in artificial intelligence where the goal is to make profit in the investment based on available resources and rules. In emerging and developed countries, stocks and foreign exchange are the driver of the economy to grow and to maintain the country private industry or company to run for importing or exporting products. Forecasting the stock price or the foreign exchange currency is based on fundamental and/or technical analysis. Dow theory is an example of technical analysis based on Dow Jones Industrial Average index to predict the stock market. The weak Efficient Market Hypothesis (weak EMH) postulates that the stock or foreign exchange market is unpredictable based on publicly available information and past data. Meanwhile, the increasing exponentially of algorithmic trading in the market has changed the situation and created some more efficient market with less loss comparing to human traders only. This was demonstrated by the early research on trading rules with moving averages by Brock et al. [1] and charting patterns such as head-and-shoulder or inverse head-and-shoulder by Lo et al. [2]. We focus our study to the Nikkei 225 index and the US dollar/Yen exchange rate in order to develop an intelligent agent trader in the future. Past researches also showed a correlation between the time series data (stock and foreign exchange market) and the news data which can be used to improve the accuracy [3], [4], [5]).

In this paper, we present the result of our experiment on analyzing a time series data such as the Nikkei 225 index or foreign currency exchange USD/JPY past data and text corpus related to Japanese economy and finance news or reports. There are several researches reported that forecasting of time series with additional features based on text data can be beneficial rather than relying on time series data history only. Experts on investment are usually making their decision based on those text data as they found patterns on them called the fundamental analysis. Time series analysis based on past history only are called technical analysis. The combination of them should make a better prediction system. The text data also provides an explanation to the trend patterns (uptrend, downtrend, reversal-trend), important moves (spikes, releases) and/or volatility (similar work on causal in time series by Kim et al. [6]). We collected all Nikkei 225 index and USD/JPY past data on daily average closing prices, all Bank of Japan monthly reports and all related available news text data. We applied on the time series Nikkei 225 data the technical analysis such as simple moving average (SMA), regression, ARIMA models and seasonal return rates, GARCH model combined with text data features such as topic models. We aggregate topics yearly or a long period of 5 to 10 years. Topics are extracted with topic modeling techniques such as Latent Dirichlet Allocation (LDA) or Non-negative Matrix Factorization (NMF). Through the observations of those technical and fundamental analyses, we show our results and conclude with some forecasts data.

THE NIKKEI 225 AND USD/JPY TIME SERIES DATA

The Japanese Stock Market is the world's second largest because of the government policies of liberalizing capital transactions with other countries and democratization of stock ownership. The 225 companies listed on the First Section of the Tokyo Stock Exchange are forming the Japanese Stock Market. Those 225 listed companies stocks are updated in real-time and are published by the Nihon Keizai Shinbun. They are called Nikkei 225 Average or for short Nikkei225. The Nikkei225 is an economic indicator for the country and surveyed by the government and the central bank of Japan for their monetary policy, financial stability and growth target. The Tokyo Stock Exchange has another indicator called TOPIX (Tokyo stock Price Index) representing all listed companies and the entire market performance but it is hard to follow. The Nikkei225 is the equivalent of the Dow Jones Industrial Average in New York Stock Exchange.

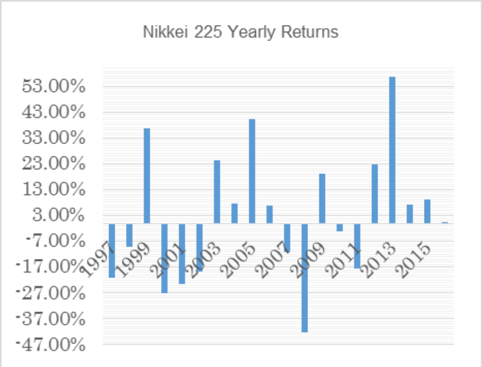

Fig.1 Returns of the Nikkei 225 from 1997 to 2016.

The chart in Figure 1 shows the yearly returns of the Nikkei225 from 1997 to 2015. We can observe the loss effect -41% caused by the financial crisis during 2008, -18% loss on March 2011 disaster and the high return +57% during the financial easing by the central bank of Japan during 2012 until the end of 2016. The year 2016 is flat to +0.42% of return rate. The highest value in the last 30 years is 38915.87 yen on 29th December 1989, the lowest price in the last 30 years is 7054.98 yen on March 10, 2009, and the average value in the last 30 years is 16844 yen.

The question is now concerning the current policy from the Central Bank of Japan, the government policies and changes in US economics on Asia. Will the negative interest rates, treasuries yield curve controls and inflation target 2 percent growth adopted by the central bank and new economic policies make the positive or negative return of the Nikkei225 for the next 5 years?

Similarly, the highest value in the last 50 years of the 1 USD in yen is 358.4 yen during the 1970s and the lowest value is 75.7yen in 2009 during the financial crisis. As shown in the figure 2, the line describes the yearly value of the USD since 1971. The downtrend pattern during 1970s did not stop until the late 1980s. Historical data only can't predict the future exchange rate. It is clear that the Japanese yen has becoming the second most traded currency after the US dollar for a certain period and known as a safe haven.

A study on correlation between the two data is also presented in this paper.

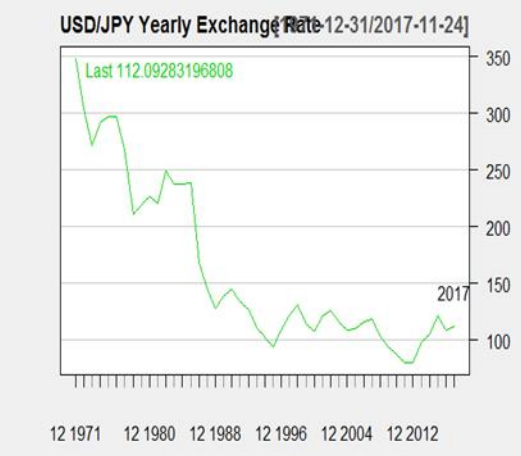

Fig. 2: USD/JPY chart from 1971 to 2017.

MOVING AVERAGE LINE ANALYSIS

The NIKKEI 225 is a "univariate time series." Given a sequence of the daily price Nikkei data $\{S_t\}_{t=1}^{N}$, the n-moving average is a new sequence $\{MA_t\}_{t=n}^{N+n-1}$ defined from the S_t by taking the arithmetic mean of subsequences of n terms [7].

$$\mathrm{MA}_{\mathrm{t+n}} = \frac{1}{n}\sum_{j=t}^{t+n-1} S_j, \quad (3.1)$$

$$\mathrm{MA}_{\mathrm{t+50}} = \frac{1}{50}\sum_{j=t}^{t+50-1} S_j = \frac{1}{50}(S_t + S_{t+1} + \cdots + S_{t+49})$$

is the 50 days moving average of the Nikkei 225.
We will obtain a sequence

$$\{\mathrm{MA}_{\mathrm{t}}\}_{\mathrm{t=50}}^{\mathrm{N+50-1}} = \{\mathrm{MA}_{50}, \mathrm{MA}_{51}, \dots, \mathrm{MA}_{\mathrm{N+49}}\}.$$

The moving average line is simply the average closing price of the Nikkei 225 and shows the lagging price with time t-n. It smooths out the daily price volatility to indicate actual trend and absorb the unusually high and unusually low closes during the n specified period. In real situation, n is set to 10-day, 25-day, 50-day, 75-day, 100-day, or 200-day. They are used as signals to sell or buy the stock when the daily chart crosses those average lines. When a faster moving average line crosses a slower moving average line, the signals are known by the traders as stronger. Traders called them “death cross” or “golden cross.” The classical death cross is the intersection of a 50-day moving average line and a 200-day moving average line in the down trends as shown in the figure 2. The classical golden cross is the intersection of a 50-day moving average line and a 200-day moving average line in the up trends as shown in the figure 2. We can use a n1-day moving average line and n2-day moving average line to find other death cross and golden cross depending on the forecast period. For example, n1 = 25 and n2 = 75 can give much more crosses than the classical ones and we can use for short term trade. n1 = 100 and n2 = 200 is an example for a long term forecast trade with few crosses.

After a golden cross, when the trend line persists to the up trends for a long period (more than a year), traders call them a bull trend. For example, the period after golden cross at end of the year 2012 is a bull trend until August 2015. In the contrary, after a death cross, when the trend line persists to the down trends for a long period (more than a year), traders call them a bear trend. For example, the period after the death cross at the end of the year 2007 during the Lehman Brothers bankruptcy and bubble burst of the financial market. The great depression during 1929 is comparable to those bear trends in 2007 where the world market crashes simultaneously and the result lasted for years created the real economy slowdown, decreasing growth domestic products and the ending of the gold standard. During the period of 2007 to 2017, there were several recessions where the GDP contracted for consecutive years. The bull trend during 2012 is the result of the financial easing by the bank of Japan and developed countries to relieve the disaster during March 2011. The increase of the consumption tax from 5% to 8% in April 2014 has giving confidence to the investors and created a boost to the bull trend to allow the Nikkei225 to reach the 20000 level for the first time since 1999. The interest rate of the central bank is negative in Japan and nearly 0% for the federal reserve bank of US for facilitating the lending and reviving the economy.

REGRESSION MODEL AND ARIMA MODEL

Regression Model

We use a simple linear regression model to study the relationship between the Nikkei 225 and the forex exchange currency US dollar/Yen rate. The explanatory variable is the currency rate. The linear function to fit the data is as below.

$$y = \beta_1 x + \beta_0 \quad (4.1)$$

y is the Nikkei 225 and x is the USD/Yen rate during the past 10 years.

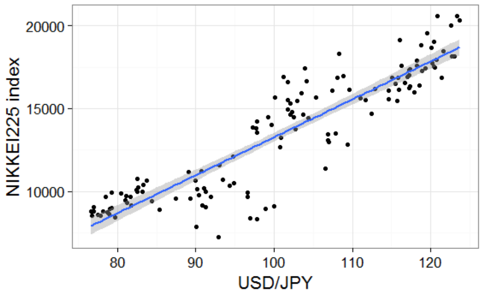

Fig. 3: Nikkei 225 and USD/Yen pair trend

In the figure 3, we obtain the fitted model on those data. Strong yen and weak dollar has a correlation to low Nikkei 225 price and weak yen and strong dollar gives a high Nikkei 225 price. All data points on Nikkei 225 above 20000 in figure 3 correspond to a rate over 120. Similarly, the majority of data points corresponds to a rate below 80 is having a Nikkei 225 below 10000. The Nikkei 225 is composed of corporate companies depending heavily on currency rate for export or import goods. The effect of the weaker yen is a good signal for stock uptrend and inflation on economy. Japan has struggled deflation for more than 30 years since the bubble burst of the Nikkei 225 in December 1989. Beside the central bank policy, the government has raised tax on goods twice since then to increase the public spending. The first tax hike was after the Hanshin earthquake in 1996 and the second was after the Tohoku earthquake and tsunami in 2012. The government has also intervened to foreign exchange market directly to stabilize the market and create some trend reversals [8].

ARIMA Model

ARIMA models are, in theory, the most general class of models for forecasting a time series which can be made to be “stationary” by differencing (if necessary), perhaps in conjunction with nonlinear transformations such as logging or deflating (if necessary). A random variable that is a time series is stationary if its statistical properties are all constant over time. A stationary series has no trend, its variations around its mean have a constant amplitude, and it wiggles in a consistent fashion, i.e., its short-term random time patterns always look the same in a statistical sense. The latter condition means that its autocorrelations (correlations with its own prior deviations from the mean) remain constant over time, or equivalently, that its power spectrum remains constant over time. A random variable of this form can be viewed (as usual) as a combination of signal and noise, and the signal (if one is apparent) could be a pattern of fast or slow mean reversion, or sinusoidal oscillation, or rapid alternation in sign, and it could also have a seasonal component. An ARIMA model can be viewed as a "filter" that tries to separate the signal from the noise, and the signal is then extrapolated into the future to obtain forecasts.

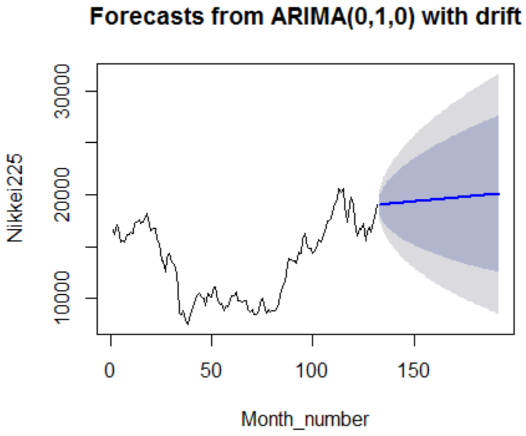

Fig. 4: Forecast from Arima model

The procedure for making the ARIMA model is as follows [9]:

□Make correlograms (ACF and PACF). PACF will indicate AR terms and ACF will show MA terms.

□Fit the model. Find the residuals and do diagnostic tests. If the residuals are autocorrelated, then fitted model is good. Otherwise, repeat the same process.

□Use the fitted model for the forecasting purpose.

The ARIMA forecasting equation for a stationary time series is a linear (i.e., regression-type) equation in which the predictors consist of lags of the dependent variable and/or lags of the forecast errors. Let $\{S_{t+1}, S_{t+2}, \ldots S_{t+n}\}$ a time series with a size n, we can express the value of S_{t+n} with the other values and an error ϵ_t

$$S_{t+n} = \beta_0 + \beta_1 S_{t+n-1} + \cdots + \beta_n S_t + \epsilon_t \quad (4.2.1)$$

the mean is

$$\mu_t = E[S_t] = \mu = \text{constant} \quad 1 \le t \le n \quad 4.2.2)$$

We applied the ARIMA model to the Nikkei 225. We forecast 4 years ahead since 2016. Figure 4 shows the forecast result with the ARIMA model as p=0, d=1, q=0. The blue line is the mean projected until 2020, the area around in blue is the coverage (lower and upper bound) of one standard deviation and grey area is the coverage of two standard deviations. The Nikkei 225 is not a stationary time series so the best fit model is the random walk ARIMA(0,1,0).

$$\widehat{S_t} = \mu + S_{t-1} + \epsilon_t \quad (4.2.3)$$

We can observe with the figure 4 that it is very difficult for the Nikkei 225 index to overcome the price of 20000 yen for long time with the past 20 years data analysis only.

SEASONAL MONTHLY RETURN RATES

Our motivation to find some patterns in the seasonal monthly return rates comes from the paper by Ben Jacobson and Cherry Zhang [10]. These authors have studied worldwide stock market historical data to produce seasonal monthly return rates and concluded that the well-known seasonal effect such as "Sell in May and Go away" (May through October) is a general pattern for over 300 years in all stock prices except of an island in Indian Ocean.

We studied the monthly seasonal return rate of the Nikkei 225 for the past 20 years. We obtained the histograms in figure 5. The Nikkei 225 does follow the general pattern of stock market as negative or neutral return from May until October. The negative return on January is also well understood as January effect and the result of the Christmas and New Year holidays. The chart shows the 10 years and 20 years monthly return rates. June and July have different opposite outcome that we consider neutral in June and negative return in July. August has the most negative return rate because it represents the holiday season during the summer in Japan. November and December are generally very profitable months.

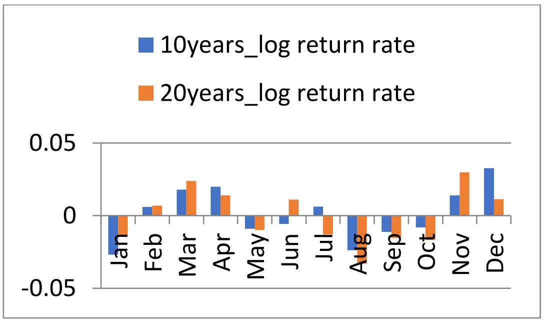

Figure 5: Seasonal monthly log return rates for Nikkei 225 stock index.

GARCH MODEL

The generalized autoregressive conditional heteroskedasticity (GARCH) model is tested for the data Nikkei 225 daily price during 2006/1/1 to 2017/9/15 to predict the next 50 days. It is shown in the figure 6 that the forecast price is lower than the actual price. The fundamental analysis is necessary to obtain more meaningful results.

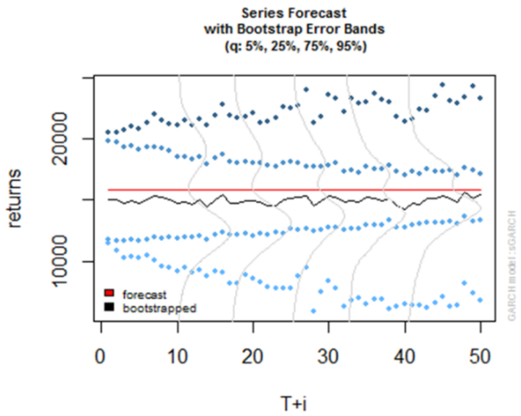

Figure 7: 50 days forecast for the Nikkei225

TOPIC MODEL

We have used topic model before in other project [11],[12],[13]. We collected text data from two financial indicator webpages as the Japan Central Bank monthly report and the Japan Research Institute (Limited Company) economic monthly report and USD/JPY foreign exchange monthly report in Japanese language. We extracted topics from those text data every 4 years by using the non-negative matrix factorization (NMF). We set the number of topics to 3 and for each topic there are 20 keywords. The text data was started from 2000 and ended in 2015. We can observe in the figure 7 that during the 2008 to 2012 financial crisis, the keywords in Japanese mean “crash” and “disaster” with the USD/Yen down (near 75). During 2012 to 2015, the main keyword is the “tax hike” with the USD/Yen up from 75 to 120. There are keywords such as “abenomics”, “financial stimulus”, “negative interest rate”, “monetary easing” and now “tapering” that we should find in the topic keywords in the financial news [12].

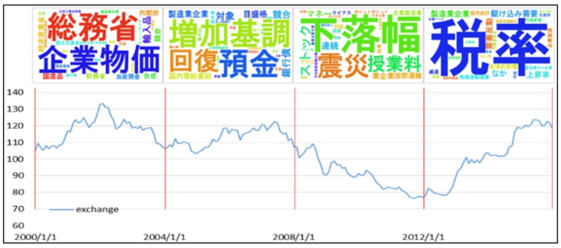

Figure 7: time series USD/JPY data and topic clouds

CONCLUSION AND FUTURE WORKS

We presented different tools for analyzing the fluctuation of the time series data such as Nikkei 225 stock and USD/JPY time series data. We added the text data analysis to the time series for the explanation and better the results. We are building a prediction system for estimating the future price of the stock and currency for short term or long term.

For the future, combined machine learning techniques can be used to improve the accuracy such as in [14] and [15] by using intensively some sentiment analysis from important social sites for traders. We will validate the accuracy of the result with a dataset for different market time frames and deal with new currency such as cryptocurrency.

REFERENCES

[1] Brock, W. et al., 1992. Simple technical trading rules and the stochastic properties of stock returns, Journal of Finance, 157(5), pp 1731-1764.

[2] Lo, A. et al., 2000. Foundations of Technical Analysis: Computational Algorithms, Statistical Inference, and Empirical Implementation, Journal of Finance, vol. LV, No 4, pp 1705-1765.

[3] Tak-chung Fu et al. (2008). Discovering the Correlation between Stock Time Series and Financial News. In proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology (WI-IAT '08) - Volume 01, pp. 880-883, 2008.

[4] Hagenau et al. (2013). Automated news reading: Stock price prediction based on financial news using context-capturing features. In Journal of Decision Support Systems, Volume 55, Issue 3, June 2013, Pages 685-697, Elsevier publisher, 2013.

[5] Nguyen et al. (2015), Sentiment analysis on social media for stock movement prediction. In Journal of Expert Systems with Applications, Volume 42, Issue 24, Pages 9603-9611, December 2015.

[6] Kim, H. D. Et al. (2013). Mining causal topics in text data: iterative topic modeling with time series feedback. In Proceeding of the 22nd ACM international conference on Information & Knowledge Management (CIKM '13 Proceedings), pp. 885-890, 2013.

[7] Ruey S. Tsay 2005. Analysis of financial time series. John Wiley & Sons Publishers, New Jersey, USA.

[8] Dominguez K.M.E., 2006. When do central bank interventions influence intra-daily and longer-term exchange rate movements? In Journal of International Money and Finance 25, pp 1051-1071.

[9] Schmueli, G. and Lichtendahl, K.C., 2016. Practical Time Series Forecasting with R. Axelrod Schnall Publishers, USA.

[10] Jacobsen, Ben and Zhang, Cherry Yi, 2014. The Halloween indicator, 'Sell in May and go Away': An Even Bigger Puzzle (October 1, 2014). Available at SSRN: https://ssrn.com/abstract=2154873 or http://dx.doi.org/10.2139/ssrn.2154873.

[11] Ramamonjisoa, D. et al. (2015). Comments Analysis and Visualization Based on Topic Modeling and Topic Phrase Mining. In Proceedings of the Third International Conference on E-Technologies and Business on the Web, pp. 1-6, Paris, France 2015

[12] Ramamonjisoa, D. et al. (2016). Analysis of the Japanese Central Bank Monthly Reports and Nikkei 225 Index Monthly for Future Prediction. In International Journal of Applied Business and Economic Research, vol.14 (2016) Issue 5 pp. 3059-3069.

[13] Ramamonjisoa, D. et al. (2014). Topic modeling on Users's comments. In 3rd ICT international conference. Bangkok, Thailand. 2014.

[14] Patel J., et al. (2015). Predicting stock market index using fusion of machine learning techniques. In Expert Systems with Applications, Volume 42, Issue 4, Elsevier Publisher, pp. 2162-2172.

[15] Seker S.E., et al. (2014). Time Series Analysis on Stock Market for Text Mining Correlation of Economy News. In International Journal of Social Sciences and Humanity Studies, Vol.6, No.1.

.