**DS-670-Capestone: Big Data & Business Analytics**

**Assignment 3: Expected Contribution and State of Art**

**February 4th 2017, Madhumita D**

**Literature Survey**

The role of capital market is to allocate ownership of capital stocks in an economy. The ideal way to define is a capital market is a market where companies can make investment decision and investors can make these decisions based under the assumption that prices fully reflect all the information that is publically available. A capital market where prices “always reflect” the information available is known to be an efficient capital market. Efficient capital market can also be described as a market in which people having access to the same information on stock prices try to achieve similar investment goals. The stock market has many profit motivated professionals and private investors who continuously search for mis-valued securities. Both these profit motivated professionals and private investors have similar investment objectives where each is motivated to gain a high profits on returns, certainty in the investment, low risk, and so forth. These goals must adhere to the securities law which states that both parties in a transaction or an investment must have access to the same publically available information.

Efficient market hypothesis (EMH) states that it is not possible to consistently outperform the market—which is usually based on the composite decisions of millions of investors with similar goals and equal access to the same publically available information. In a more general sense EMH explained by Fama [1] states that the stock prices follow a random walk theory and cannot be predicted based on their past behaviour. Here “efficiency” in hypothesis means that the market can quickly digest any new information available on the company, industry, economy, or value of the company and can impound this information accurately into securities prices. Hence in efficient capital markets we can expect people to earn fair returns for the risks taken, no more and no less.

From the hypothesis it does not mean that the investors or participants are denied the profits of investment, it means that for a given risk the benefits from investing in largely competitive markets will be fair (on the average). For instance in an efficient market a news of increase in the earnings of a company will be immediately and accurately assessed by the combined judgement of millions of participants and quickly reflected in the stock prices of the particular company. The result of this efficiency can be seen when you buy the company’s stock before, after, or during release of the news on the earnings. Taking the risk of purchasing or owning the security will yield only in a fair market rate of return in case of efficient market hypothesis.

Efficient market hypothesis is categorized into three forms of market efficiency. The weak form of efficient markets holds when the predicted returns will not reflect values from the historical price data and volume information. In case of semi strong form of market efficiency, all the publically available information and historical price data is considered to reflect the prices in the future and strong form of efficient market holds when both publically available information and private information is used to predict the future stock prices. While it provides to be a useful benchmark in all its forms, semi strong form of EMH is most widely suggested form to predict the future stock prices. Hadi [8] described and tested various forms of efficient markets and observed that generally semi strong form is considered to be efficient in accounting based research.

Sample studies which provide aggregate analysis on the efficient market hypothesis for Australian data are Hogan, Sharpe and Volker [2] and Groenewold and Kuay [3]. Their results are consistent with those found and summarized by Fama [1] for US data which shows the expected returns when measured over short durations like a daily or weekly. Another study by Groenewold, N [4] tests both weak and semi strong forms of EMH and reports consistent results for daily observations from Australia and New Zealand stock exchange based on the log of prices.

Investments which can be used to gain abnormal profits or returns (in general known as anomalies) violate the efficient market hypothesis in both its semi strong and strong form. In finance the most common anomalies include abnormal returns or unfair profits which are relative to unexpected increase in earnings announcements, size of the firm, month of January, day of the week, undue recommendations from the analysts’, impact of the federal budget deficit announcement, and many others. These anomalies are catalogued and extensive literature survey is provided by Raghubir and Das [[5]](http://www.sciencedirect.com/science/article/pii/S016792360100121X#BIB42), and Hong, H., & Stein, J. C [6].

This paper studies and tests the semi strong market efficiency of pricing data on daily and quarterly observations of NYSE stock exchange. The study also tries to relate the same tests based on the sectors of industry to draw conclusions based on the consistency of the predicted stock prices based on each sector. Various studies have been made based on the tests for semi strong form of efficiency in relation to predicting stock prices using various models and theories. Leuthold, R. M., & Hartmann, P. A [7] uses econometric forecasting model to test the semi strong form of efficient markets and reported that Live-Hog future markets are inefficient. The efficiency of Indian Capital Market was test and concluded as efficient in its semi strong form of efficient market hypothesis by Khan, A. Q., & Ikram, S [9] using Correlation Coefficient and Linear Regression analysis. London Metal Exchange market was tested to be efficient in the sense of semi strong form of efficient market hypothesis by Barry A. Goss [10] and results comprised that the future prices for the metals like copper, zinc and lead fully reflect publically available information. Ardiansyah, M., & Qoyum, A [11] paper discussed the concepts of Islamic capital market focussing on Jakarta Islamic Index (JII) to test its market efficiency using market adjusted model and mean adjusted model. Stock split announcements in capital markets were tested for semi strong form of efficiency by Raja, Sudhahar, and Selvam [12] on Indian Stock Market with respect to IT (Information Technology) companies.

Studies also show that some of the capital markets do not hold the semi strong form of efficient markets. This kind of behaviour is shown when the predicted prices do reflect all the information available or we cannot retrieve accurate prices with the data we have. The predictability of expected returns on Athens Stock Exchange was examined using the financial statement information and results indicated that the Greek market does not fully incorporate the information into stock prices, Alexakis, C., Patra, T., & Poshakwale, S [13]. Pele, D. T., & Voineagu, V [14] proposed a model for stock’s return decomposition for testing efficient markets on Romanian Capital Markets and results concluded that the hypothesis of market efficiency cannot be rejected in the weak sense.

All these studies basically review many tests and methods on semi strong form of efficient market hypothesis using the data from various capital markets around the globe. Before we discuss the method used in this paper to predict the expected returns we should know the two approaches which are popular in investment analysis – fundamental analysis and technical analysis. Investment analysis where prediction of the stock prices is based on fundamental factors which are internal or related to the company or its industry, for example product earnings’, competition, management, consumer spending and others is termed as fundamental analysis. A fundamental analyst recommends a purchase for a company which shows a consistent increase in its earnings and believes that the profit will be fair. Therefore, a semi strong form of efficiency opposes the concept of fundamental analysis. Analysis where stock prices can be predicted from historical stock market data which includes changes in prices of stocks and trading volume is termed as technical analysis. A technical analyst believes that all the internal information available on a stock (i.e. fundamental factors) is reflected in the stock price behaviour and suggests a purchase based on the recent prices and volume information. The weak form of efficient market hypothesis opposes the principle of technical analysis.

Our paper uses neural networks model to test the semi strong form of efficient markets by predicting the returns using the pricing information on stocks. The data is categorized mainly based on the quarters and daily stock information. Stock prices and information related to capital markets is time dependent data and forecasting the future prices is majorly dependent on historical prices and all publically available information on the stocks. Pricing information can be considered as a time series which is defined as a sequence of values measured over a period of time (discrete or continuous). Financial information of a company is released publically every year in the form of quarterly and annual statements for a fiscal year. These statements include costs, revenues, expenses, price, earnings and many other factors. Predicting stock prices based on many related factors gives a better result. So we consider a multivariate time series which consists of variables whose values change over time. Hence a multivariate time series analysis is used to study the behaviour of time dependent pricing data and forecast values based on the history of variation of the data.

In general techniques used in time series analysis assume that the relationship among the variables is linear, but data with temporal variables do not exhibit normal regularities and hence are difficult to predict accurately. It seems appropriate to use non-linear modelling and factor analysis like neural networks for temporal data. Neural network approach is a data driven approach and analysis is determined based on the data available. This approach uses the concept of constructing a machine based on the available data. We use this approach to predict the expected returns of the stocks from the multivariate time series based on all the information available publically and historic pricing information. Many studies have used various statistical methods to forecast stock price. Ince, H., & Trafalis, T. B [15] used Principal Component Analysis (PCA) and Factor Analysis in order to identify the most influential inputs for forecasting model with Neural Networks (NN) and Support Vector Regression as the inputs. The results indicate that neural network approach outperformed support vector regression in forecasting the future value of stock price based on technical analysis.

Another study based on trading strategies guided by forecasts of the direction of movement of price on Taiwan Stock Exchange [16] was done using probabilistic neural networks (PNN). The forecasts from statistical performance of the PNN model are measured and compared with the generalized methods of moments. It is observed that the empirical results from using neural networks in investment strategies had higher return values than other strategies in the study. In yet another analysis of forecasting the stock market activity, a hybrid analysis is performed using technique from pattern recognition and artificial intelligence method from neural networks on NYSE data [17]. This hybrid methodology predicted returns that are superior to the rand walk theory. Combining methods (like principal component analysis, genetic algorithms and decision trees) or hybrid techniques [18] have provided better financial factors which are important for stock prediction and future investment decisions rather than using a single statistical technique. Ludvigson, S. C., & Ng, S [19] used dynamic factor analysis approach to model the conditional mean and volatility of stock market returns. This approach allows undermining the limitations from empirical analyses.

Neural Networks was employed in many studies to predict stock price behaviour which yielded in excellent profits [20]. Financial applications like option pricing, stock index trading to currency exchange have used radial basis function networks and neural networks over the decades. Neural Networks are considered to be universal functional approximators which can be used to map a non linear function without actually having any assumption about the data.

We are trying to implement neural network methodology on U.S. stock market data using daily closing prices to compute the returns of each company. The expected returns are computed based on daily and quarterly stock price data. Many studies have used U.S. stock market data to forecast returns implementing many different methods like Gaussian models [26], Bayesian approach, support vector regression [21] and artificial immune algorithms. Giles, C. L., Lawrence, S., & Tsoi, A. C [24] used recurrent neural networks on ‘Chicago Mercantile Exchange’ for financial prediction and White, H. [25] used neural networks to predict the returns of IBM daily stock prices. Stock return predictability measures from regression analysis were used to validate the adaptive market hypothesis [22] which states that returns are subject to market conditions. Observing the non-linearity of the stock market data we have tried to employ neural network factor model to forecast the returns of U.S. stock data.

The study below presents the following sections where data is described, methods used are explained and results are discussed. Section 2 describes the data considered for the analysis and description of every financial indicator used. Section 3 presents the neural network architecture we used in our analysis and the moving average time series model of statistical prediction. The performance and results from the network model are noted in section 4 which is followed by discussion and conclusion of our analysis.

**References**

[1] Fama, E. F. (1991). Efficient capital markets: II. The journal of finance, 46(5), 1575-1617 and Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. The journal of Finance, 25(2), 383-417.

[2] Hogan, W. P., Sharpe, I. G., & Volker, P. A. (1982). Capital market efficiency and the relationship between equity returns, interest rates, and monetary aggregates in Australia. Journal of Economics and Business, 34(4), 377-385.

[3] Groenewold, N., & Kang, K. C. (1993). The semi‐strong efficiency of the Australian share market. Economic Record, 69(4), 405-410.

[4] Groenewold, N. (1997). Share market efficiency: tests using daily data for Australia and New Zealand. Applied Financial Economics, 7(6), 645-657.

[5] Raghubir, P., & Das, S. R. (1999). A case for theory-driven experimental enquiry. Financial Analysts Journal, 56-79.

[6] Hong, H., & Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets. The Journal of finance, 54(6), 2143-2184.

[7] Leuthold, R. M., & Hartmann, P. A. (1979). A semi-strong form evaluation of the efficiency of the hog futures market. American Journal of Agricultural Economics, 61(3), 482-489.

[8] Hadi. “Review of Capital Market Efficiency: Some Evidence from Jordanian Market”. International Research Journal of Finance and Economics, Year: 2006, Issue: 3, pp 13-26.

[9] Khan, A. Q., & Ikram, S. (2010). Testing semi-strong form of efficient market hypothesis in relation to the impact of foreign institutional investors'(FII's) investments on Indian capital market. International Journal of Trade, Economics and Finance, 1(4), 373.

[10] Barry A. Goss (1983) The semi-strong form efficiency of the London Metal Exchange , Applied Economics, 15:5, 681-698, DOI: 10.1080/00036848300000044

[11] Ardiansyah, M., & Qoyum, A. (2012). Testing the Semi-strong Form Efficiency of Islamic Capital Market With Response to Information Content of Dividend Announcement: A Study in Jakarta Islamic Index. Journal of Modern Accounting and Auditing, 8(7), 1025.

[12] Raja, M., Sudhahar, J. C., & Selvam, M. (2009). Testing the semi-strong form efficiency of Indian stock market with respect to information content of stock split announcement: a study in IT industry. International Research Journal of Finance and Economics, 25(2009), 7-20.

[13] Alexakis, C., Patra, T., & Poshakwale, S. (2010). Predictability of stock returns using financial statement information: evidence on semi-strong efficiency of emerging Greek stock market. Applied Financial Economics, 20(16), 1321-1326.

[14] Pele, D. T., & Voineagu, V. (2008). Testing market efficiency via decomposition of stock return. Application to Romanian Capital Market. Romanian Journal of Economic Forecasting, 3(2008), 63-79.

[15] Ince, H., & Trafalis, T. B. (2007). Kernel principal component analysis and support vector machines for stock price prediction. IIE Transactions, 39(6), 629-637.

[16] Chen, A. S., Leung, M. T., & Daouk, H. (2003). Application of neural networks to an emerging financial market: forecasting and trading the Taiwan Stock Index. Computers & Operations Research, 30(6), 901-923.

[17] Leigh, W., Paz, M., & Purvis, R. (2002). An analysis of a hybrid neural network and pattern recognition technique for predicting short-term increases in the NYSE composite index. Omega, 30(2), 69-76.

[18] Tsai, C. F., & Hsiao, Y. C. (2010). Combining multiple feature selection methods for stock prediction: Union, intersection, and multi-intersection approaches. Decision Support Systems, 50(1), 258-269.

[19] Ludvigson, S. C., & Ng, S. (2007). The empirical risk–return relation: A factor analysis approach. Journal of Financial Economics, 83(1), 171-222.

[20] Kimoto, T., Asakawa, K., Yoda, M., & Takeoka, M. (1990, June). Stock market prediction system with modular neural networks. In Neural Networks, 1990., 1990 IJCNN International Joint Conference on (pp. 1-6). IEEE.

[21] Chiu, D. Y., Shiu, C. Y., & Lin, Y. S. (2011, June). USA S&P 500 stock market dynamism exploration with moving window and artificial intelligence approach. In Networked Computing and Advanced Information Management (NCM), 2011 7th International Conference on (pp. 341-345). IEEE.

[22] Kim, J. H., Shamsuddin, A., & Lim, K. P. (2011). Stock return predictability and the adaptive markets hypothesis: Evidence from century-long US data. Journal of Empirical Finance, 18(5), 868-879.

[23] Leigh, W., Frohlich, C. J., Hornik, S., Purvis, R. L., & Roberts, T. L. (2008). Trading with a stock chart heuristic. IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 38(1), 93-104.

[24] Giles, C. L., Lawrence, S., & Tsoi, A. C. (1997, March). Rule inference for financial prediction using recurrent neural networks. In Computational Intelligence for Financial Engineering (CIFEr), 1997., Proceedings of the IEEE/IAFE 1997 (pp. 253-259). IEEE.

[25] White, H. (1988). Economic prediction using neural networks: The case of IBM daily stock returns.

[26] Petelin, D., Šindelář, J., Přikryl, J., & Kocijan, J. (2011, September). Financial modeling using Gaussian process models. In Intelligent Data Acquisition and Advanced Computing Systems (IDAACS), 2011 IEEE 6th International Conference on (Vol. 2, pp. 672-677). IEEE.