"Forecasting the changes in daily stock prices in Shanghai Stock Exchange using Neural Network and Ordinary Least Squares Regression"

AUTHORS	Sunantha Prime ⓘ https://orcid.org/0000-0002-0634-9815

NUMBER OF REFERENCES	NUMBER OF FIGURES	NUMBER OF TABLES
39	0	4

BUSINESS PERSPECTIVES

Sunantha Prime, Ph.D., Associate Professor, Department of Information System and Decision Sciences, Craig School of Business, California State University, Fresno, USA.

Sunantha Prime (USA)

FORECASTING THE CHANGES IN DAILY STOCK PRICES IN SHANGHAI STOCK EXCHANGE USING NEURAL NETWORK AND ORDINARY LEAST SQUARES REGRESSION

Abstract

The research focuses on finding a superior forecasting technique to predict stock movement and behavior in the Shanghai Stock Exchange. The author's interest is in stock market activities during high volatility, specifically 13 years from 2002 to 2015. This volatile period, fueled by events such as the dot-com bubble, SARS outbreak, political leadership transitions, and the global financial crisis, is of interest. The study aims to analyze changes in stock prices during an unstable period. The author used advanced computer sciences, Machine Learning through information processing and training, and the traditional statistical approach, the Multiple Linear Regression Model, with the least square method. Both techniques are accurate predictors measured by Absolute Percent Error with a range of 1.50% to 1.65%, using a data file containing 3,283 observations generated to record the daily close prices of individual Chinese companies. The t-test paired difference experiment shows the superiority of Neural Network in the finance sector and potentially not in other sectors. The Multiple Linear Regression Model performs equivalent to the Neural Network in other sectors.

Keywords	prices forecast, Machine Learning, Chinese stock market, statistical methodology, model evaluation
JEL Classification	F37, C12, C45

INTRODUCTION

The stock market is somewhat volatile and sensitive in various areas such as economic environment and news, political policy, industrial development, market news, and natural factors; therefore, predicting stock prices is a difficult task. The ability to more accurately predict is of high interest to those involved in the investment market. Accurately predicting the stock prices can provide individual investors, stock fund managers, and financial analysts more opportunities to gain profit in the stock exchange.

Forecasting the daily stock prices is challenging due to its being influenced by several factors, as previously mentioned. The challenge is increased further because of the unpredictable high volatility in stock prices of the Chinese markets from 2002 to 2015. During this exceptionally uncertain period, it is expected that stock prices are extremely difficult to predict.

This study focuses on a stock market analysis, applying Neural Network (NN) and Ordinary Least Squares Regression (OLSR) in a

comparison study in the Chinese stock market. Burstein and Holsapple (2008) described BI: "business intelligence (BI) is a data-driven decision support system that combines data gathering, data storage, and knowledge management with analysis to provide input to the business decision process." Han and Kamber (2006) stated that "data mining is extracting knowledge from large amounts of data." Tjung, Kwon, and Tseng (2012) defined Neural Network as follows: "Neural Network is one of these data mining applications and useful in making complex predictions in many disciplines." Their NN is used in this study, and it will be discussed later in section 3.

This work differs from that of Tjung, Kwon, Tseng, and Bradley-Geist (2010), and Tjung, Kwon, and Tseng (2012). That is, Tjung's et al. (2010) models were designed to forecast daily changes in stock prices using data from September 1, 1998 to April 30, 2008, and Tjung, Kwon, and Tseng (2012) forecasted 37 stocks from eight industries over the same periods as in Tjung's et al. (2010). Both of these studies focus on the US stock market, whereas the models are obtained from more updated data from a much larger number of companies in the Chinese stock market.

The daily stock price data for China in this study cover the period from January 1, 2002 to December 31, 2015. With its unparalleled attributes, the Chinese stock market is one of the largest global markets. During this time, a distinct number of important events in China occurred, specifically the technology crash referred to as the dot-com bubble in the latter part of 2002, the party leadership changes in 2002 and 2012, the 2003 outbreak of SARS, the global subprime great financial crisis affecting an economic recession in 2007, the earthquake in Wenchuan and the Beijing Olympic Games, both in 2008.

The study, as described in the methodology section, looks at two models to determine the best outcomes for training and forecasting.

This paper is structured as follows. Section 1 provides a literature review relevant to the Chinese stock market and stock price forecast. Section 2 describes the methodology utilized in the study. Specifically OLSR models and SPSS statistics are applied to forecast changes in daily stock prices in China. Furthermore, Section 2 introduces Alyuda Neural Network (ANN) and discusses how NN can be used to predict stock prices. It details the sources and characteristics of data, variables, and data normalization processes. Section 3 reports the results of t-test pair-wise hypotheses testing, error measurement by forecast methodologies, and other related results. Section 4 provides the relevant discussions of results from this study. The final section closes this paper with the conclusion and future research direction.

1. LITERATURE REVIEW

Kwon, Wu, and Zhang (2016) compare the forecasting performance of different versions of BI models in predicting China's stock prices. They discuss the model's ability to extract and explain vast amounts of data and knowledge and how it all relates to enhancing the process of decision making. They conclude that normalized and denormalized data provide similar results.

Liu and Wang (2011) look at the Independent Component Analysis, the NN model, and the BP model. They demonstrate that the NN model outperforms the other two models in analyzing fluctuations in the Chinese stock markets.

Dai, Liu, and Wang (2012) investigate improved accuracy in predicting Shanghai B shares using a combination of Nonlinear Independent Component Analysis and Neural Network over other models such as LICA-BPN, PCA BPN, and single BPN. With the use of data from Shanghai A and B shares, Nasdaq Composite Index, and Industrial Average Index, predictions and combined time strength functions will possibly increase the accuracy of LeNN.

Cao, Ham, and Lam (2013) report that in the stock market movement, the Back Propagation Network was found to be slightly superior to the Radial Basis Function Network.

According to Cao, Parry, and Leggio (2011), traditional finance models as forecasting techniques are not limited to more recent models. They conduct a cross comparative study between Neural Network models and Dynamic Single-Factor model with Fama's (1970) Neural Network models and two linear models.

The Chinese stock market is sensitive to several factors. Government regulations, corruption, predictability, efficiency, investors' behavior are listed and discussed by numerous scholars as influential factors impacting the Chinese stock and financial markets. Gordon and Li (2003) stated that the banking system in China was required to advance money to enterprises owned by the state, notwithstanding financial performances. This indicates that relevant government mechanisms designate share values. Additionally, in a study by Riedel, Jin, and Gao (2007), as high as 69% of 1,400 public companies' stocks are non-tradable shares. Yao and Yueh (2009) explained that a relatively high fluctuation in market returns is influenced by prescribed pricing by types of share and different types of shares governed by unique trading regulations. This has been a question raised by Firth, Rui, and Wu (2009). The bureaucratic style of regulation of the Chinese Securities Regulatory Commission (CSRC) resulted in declined investor confidence due to their accessibility to information. Sanction Enforcement Information (SEI) delayed information to the public for more than 15 days, causing deterioration of investor confidence.

The consequence of economic transformation periods because of China's corruption during these times results in a highly manipulated market. The competition and economic conditions of the corrupt authority allocate resources based on bribes instead of market efficiency. Knight and Yueh (2008) argue that Chinese entrepreneurs may take advantage of questionable relationships. Additionally, Chinese culture has historically been known to exploit the legal system. Fan, Wong, and Zhang (2007) also report a tradition where companies have political connections to further their motives. They argue that former government officials are CEOs of companies; this results in lower stock returns for companies politically unconnected.

Relative to US markets, Chen (2010) reports less predictability for Chinese stock markets. He found that informative stock prices cause it and less heterogeneously distributed return predictors. Zhang, Wei, and Huang (2014) provide similar findings when predicting markets comparing S&P 500 stocks with Shanghai composite. It was more difficult to predict individual stocks in Chinese markets than to predict markets in the U.S. Jiang (2011) indicated that small capital firms' ownership concentration and stock portfolios aid in the predictability of markets. Industries showing the highest predictability in stock portfolios are insurance, finance, real estate, and services. Others also argue that effective predicting is inefficient in China because of deviations from the standard. The stock market, using market efficiency theory Kang, Cheong, and Yoon (2010) have found that the Chinese stock market is contrary to Market Efficiency. Additionally, Lim et al. (2013) agreed that the Chinese stock market predictions are contrary to the standard.

Wang, Qiu, and Kong (2011) regarding Corporate Social Responsibility (CSR) determine investor reactions occur if there is too high CSR or too low CSR performance of companies. Indications that a higher threshold trend governs momentum behaviors are proposed by Wei, Huang, and Hui (2013). Momentum strategies in the Shanghai Stock Exchange are consistent with the inefficiency compared to mature markets like those in the US, Hong Kong, and Japan. Considerations about herding effects, like Chiang, Li, and Tan (2010) show, both Shanghai and Shenzhen A-share markets are affected, but not B-share markets. Similar results regarding herding behavior are made by Yao, Ma, and He (2014). Li and Tan (2010) agree that investors herding was prevalent in the B share and are mainly in stocks of the largest and the smallest companies. Li (2014) shows that there is positive productivity in the Shanghai Stock Exchange in a sentiment index. Chi, Zhuang, and Song (2012) show that Chinese market returns indicate that investor sentiment is one of the critical factors affecting stock prices.

Li, S. Wang, and X. Wang (2017) examine the impact of social trust, the level of mutual trust among the members of a society, on stock price crash risk. Through the use of a large sample of Chinese A-share

firms listed over the period 2001–2015, firms headquartered in regions of high social trust tend to have a smaller firm-specific stock price crash risk. Thus, social trust is one of the critical predictors of stock price crashes. China's vast social trust diversity originates from fifty-six ethnic groups within thirty-one provinces and more than eighty different native dialects that are not comprehensible to non-native speakers.

Hafezi, Shahrabi, and Hadavandi (2015) propose the Bat-Neural Network Multi-Agent System (BNNMAS) to predict stock price. To predict eight years of DAX stock prices by quarters, they use (1) BNNMAS in a four-layer multi-agent framework, (2) the quarterly data of 17 national indexes and three international indexes that include oil price, gold price and the exchange rate of the German mark with the US dollar, (3) feature selection and time lag selection for data preprocessing phase, and (4) hybrid bat-neural network (BNN) model are used for the function approximation phase. Their results show significantly more accurate and reliable performance of BNNMAS compared to other models such as GANN, and others. BNNMAS, therefore, is suitable for predicting stock prices over long periods.

Hu, Tang, Zhang, and Wang (2018) use an improved sine cosine algorithm (ISCA-BPNN), the S&P 500 and DJIA Indices data for predicting the direction of opening prices in the S&P 500 and DJIA Indices. They collected data from Google Trends for a period from January 1, 2010 to June 16, 2017. They propose that back propagation neural network models (ISCA-BPNN) improve predicting the direction of stock markets. They compared it with the BPNN model, GWO-BPNN model, PSO-BPNN model, WOA-BPNN model, and SCA-BPNN model. Their results for predicting the direction of the opening price showed that the ISCA-BPNN model is better than all the other models in the study and that Google Trends can help predict future financial returns. Their results indicate that swarm intelligence algorithms can optimize other artificial neural networks' parameters for predictions and classifications.

Wang, Yao, and Yu (2018) extract the event-relevant data from Web news and user sentiments from social media. They analyze their joint impacts on the stock price movements using a coupled matrix and tensor factorization framework. They construct two auxiliary matrices of the stock quantitative feature matrix and the stock correlation matrix and incorporate them to assist the tensor decomposition. A coupled matrix and tensor factorization scheme support heterogeneous information integration and multi-task learning simultaneously. Thus, their prediction was for multiple correlated stocks simultaneously through the utilization of the commonalities among stocks. Using Chinese A-share stock market data and the HK stock market data, their proposed model achieves 62.5% and 61.7%, respectively.

The essence of the problems could be characterized by the sizable market of the Shanghai Stock Exchange (SSE). It is one of the world's largest stock market by market capitalization at USD 5.01 trillion as of May 2019. It is challenged and influenced by government regulations, corruption, predictability, efficiency, investor behavior, and other influences. This has attracted cumulative research work, as described earlier. The previous research found that artificial or business intelligence is a superior predictor of the traditional statistical method. However, this work indicates a contradiction which appears in some industrial sectors of the SSE.

Due to the problem essence and challenge, the author continues this interesting research in the Chinese stock market with the daily stock price data covering the time from January 1, 2002 to December 31, 2015. This period highlights several important events in China, the dot-com bubble in late 2002, 2003 outbreak of SARS, party leadership transitions in 2002 and 2012, the global financial crisis with an economic recession since 2007, the Wenchuan earthquake and the 2008 Beijing Olympic Games.

The analysis aims to compare Neural Network (NN) with Multi-Layer Perceptron (MLP) and Back Propagation Algorithm, and Ordinary Least Squares Regression (OLSR) Models with all independent variables and the stepwise for their performance in predicting changes in daily stock prices in the Shanghai Stock Exchange in China. The accuracy of forecasting methodologies in this study is verified by the t-test paired difference experiment and is measured by Absolute Percent Error (APE).

2. METHODOLOGY

This study research uses normalized data used as input for the NN and OLSR models. It details the sources and characteristics of data and variables to generate and forecast changes in the stock price.

2.1. Data and variables

A shares were downloaded from Yahoo Finance, looking at 151 companies on the Shanghai Stock Exchange from January 2002 to December 2015. These companies are randomly sampled to represent the eight industry-wide sectors: basic materials, conglomerates, consumer goods, financial, healthcare, industrial goods, services, and utilities.

For each company of these 151 firms, a data file consisting of 3,283 records of the daily close prices is generated. The author looks at 21 indicators categorized by Macroeconomic and Microeconomic indicators, indicators of market sentiment, and institutional investors (see Appendix B for the complete list of indicators). They are identified as independent variables to predict the movements of the stock prices in the Shanghai Stock Exchange.

The Dependent Variable is the target company selected from each industry sector. All other companies in the same sector, along with 21 indicators, are arranged as Independent Variables. Both types of variables are employed for model training and forecasting.

For each stock, this comparative study applies to Neural Network and Ordinary Least Square Models that forecast the variability in the Shanghai Stock Exchange. Records of daily closing prices, classified by variable categories, are used as input for the Neural Network and Ordinary Least Square Models. Dummy variables are used as additional input for the model training and forecasting to improve the precision of a forecast. They are added to specify the pre-holiday and post-holiday days. The research applies and uses all-variable OLSM, stepwise OLSM, and Alyuda Neuro Intelligence software to build NN forecasting models.

The normalization method (Tjung, Kwon, & Tseng, 2012) is adapted to achieve better forecasting performance. They conduct the forecasting analysis using NN models in the US stock markets and Alyuda Neuro Intelligence. Further, Tjung, Kwon, and Tseng (2012) point out that the NN models generate a lower standard deviation than traditional regression analysis. It is important to note that the normalized data method provides a superior outcome than the non-normalized data in model learning and forecasting.

To normalize the observations for individual companies, the author pinpoints the minimum value (Minimum) of daily stock price changes, then take the absolute value (Absolute), and add 0.1 to lift a zero value to a positive in the normalization dataset. Consequently, the normalization values ((Absolute (Minimum)) + 0.1) are added to every daily change of company data. For example, the observations of daily stock changes for Company Fresno supposing its minimum of –5.03, the normalization value (Absolute (–5.03) + 0.1 = 5.04), is added to every daily stock change of Company Fresno, thus normalizing the observations.

The normalized data are used as input for the Machine Learning software and OLSR models. Alyuda Neuro Intelligence software is used to build NN and forecasting models. 95% of the observations are used for training and validating data, while the remaining 5% for performance testing.

2.2. Neural Network (NN) method

Multi-Layer Perceptron (MLP) and Back Propagation Algorithm are prediction methodologies. There are numerous different Architectures of the Network that were applied in previous research. Multi-Layer Perceptron (MLP) is one of the popular ANN Architectures widely accepted (Alyuda Research, 2006). The Multi-Layer Perceptron (MLP) is a feed-forward neural network with the ability to improve its model errors by iteratively changing its interconnecting weight of the architecture among all connections of the input layer, hidden layer, and output layer (Gardner & Dorling, 1998). The logistic sigmoid function can reduce the effects of the outliers (Hill et al., 1994), and logistic sigmoid function generally works for MLP (Marier & Dandy, 2000). As a result, Multi-Layer Perceptron Architecture with logistic sigmoid function is applied in this research for data analysis.

According to Dawson and Wilby (2001), the Back Propagation Algorithm is popularly used for Multi-Layer Perceptron (MLP) network training. This algorithm can reduce the overall network error between the network output and target value by adjusting the interconnecting weight of the networks iteratively.

This study utilizes former rules from previous research to determine the Numbers of Hidden Nodes (HN) for the Artificial Neural Network. Previous research addresses the relationship among the Numbers of Hidden Nodes (HN), the Numbers of Inputs (I) for input layer, and the Numbers of Outputs (O) for output layer. According to Fletcher and Goss (1993), the numbers of Hidden Nodes (HN) should range from 2I1/2 + O to 2I + 1. Palani et al. (2008) suggest that the HN should range from I to 2I+1 and larger than I/3, and O. Alyuda Research (2006) suggests that the HN should range from I/2 to 4I. Also, Gazzaz et al. (2012) have combined the former three rules for ANN application with Alyuda Neuro Intelligence, which states that the HN should range between I/3 and 4I. As a result, in this study, the Numbers of Hidden Nodes are expected to lay down between 1/3 of Inputs to 4 times of Inputs and larger than Numbers of outputs as well. (⅓ I < HN < 4I ∩ HN > O).

Based on previous research, different criteria can be applied to Alyuda Neuro Intelligence for the best ANN Architecture Searching. Gazzaz, Yusoff, Aris, Juahir, and Ramli (2012) have applied R-squared as model selection criteria to forecast the water quality index. Gaurang et al. (2010) discussed and pointed out the significant efficiency of AIC in searching for neural network architecture. Huang and Chen (2013) used minimum testing error as criteria to select ANN architecture for the Exchange Rate Prediction Model. Considering the huge data Inputs in this research, efficiency is important for our modeling and further practical operation in the future. Hence, AIC is applied as the best ANN Architecture Searching criteria. The Architecture with the highest AIC is selected for the network training. Each company has its own architecture.

There are several training stop criteria from previous papers, Alyuda Neuro Intelligence. Anwer and Watanabe (2010) set the termination of training after 20,000 iterations or Mean squared error (MSE) < 0.000001, and the learning and momentum rate at 0.1 for backpropagation. Gazzaz et al. (2012) applied 0.000001 as the network MSE improvement, 0.01 of the training set MSE, and maximum for 10,000 iterations. Also, Gazzaz et al. (2012) retrain ten times, according to the Alyuda Neuro Intelligence manual. Meng (2008) applied 50,000 iterations and network error (MSE) as 0.01 in predicting the return on IPO in the Chinese stock market. Because of the uncertainty of the training process, more training times for the Artificial Neural Network will have a better chance to achieve better results. The training trials set the iterations to stop when 10,000 are completed with 10 to retrain or stop training at 0.000001MSE improvement or the achievement of 0.01 training error. This training process conducted three trials for each stock, and the Network with the lowest Relative Error is selected.

Table 1. List of terms and abbreviations

Terms and abbreviations	Descriptions
Training Set	The data that were normalized and used to train Neuro Intelligence to find the best-fit architecture for model optimization and parameter validation
Query Set	The data that were normalized, with 5% manually extracted and separated, for NeuroIntelligence to use for calculating the estimated price changes
Actual Price Change (APC)	The daily changes of the stock prices ($C_i = P_i - P_{i-1}$) in Chinese currency (Yuan)
Normalized Actual Price Change (NAPC)	Actual Price Change data that were normalized and used by Neuro Intelligence in training the models
Error_ND	The estimation errors due to the use of the normalized data. They are defined as the difference between Normalized Actual Price Change and Actual Price Change

2.3. Ordinary Least Square Regression (OLSR) Models

Ordinary Least Squares Regression Models involve one dependent variable y, more typically dependent variables are functionally related to multiple independent variables $x_1, x_2, x_3, \ldots, x_n$, with the following specification:

$$y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_n x_n + \varepsilon, \quad (1)$$

where y is the actual value, all x_1 are the predictors, and ε. In this study, the actual value that one attempts to predict is the change in daily stock prices. Moreover, therefore, the corresponding predicted values of changes in stock prices $\hat{y}$, as described.

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k. \qquad (2)$$

To analyze data using OLSR, the model involves the following assumptions:

- the dependent variables should be interval or ratio, measured on a continuous scale;

- the observations or residuals are independent;

- there is a linear relationship between (i) the dependent variable (y) and each of the independent variables (xi), and (ii) the dependent variable and the independent variables (x_1, x_2, x_3, …, x_n) collectively;

- data shows homoscedasticity, in which variances along the line of best fit remain similar or approximately equal;

- independent variables contain minimal or no multicollinearity with each other;

- data should not contain significant outliers, high leverage points, or highly influential points;

- the residuals (errors) are normally distributed, or their distribution could be approximated by normal distributions with a mean of zero.

The author runs OLSR for forecasting stock prices in the Chinese markets across all eight industries using SPSS statistics both (a) the method of Enter, including all independent variables of interest, and (b) the stepwise method, including some and potentially not all independent variables, which allow a reduction in the effect of collinearity and multicollinearity.

Multicollinearity may exist in real data. When it exists, the significance of the overall model, tested by F-statistics is not jeopardized. However, the significance of any independent variable on the dependent variable is jeopardized because the impact of a single independent variable cannot be isolated due to collinearity/multicollinearity effects among some independent variables. Thus, one solution to run regression analysis is to use the stepwise method, an iterative modeling procedure that reduces the collinearity/multicollinearity. There are two major forms of stepwise regression:

- Forward Selection Approach: This model starts with a simple linear regression run to determine the significance of individual independent variables and the dependent variable. If some independent variables have the significance of partial regression coefficient, measured by p-value, at or below the desired α levels, the stepwise-in regression selects the variable with the strongest relationship with the dependent variable (or smallest p-value). The process is recursive with each new iteration result in adding the single most significant independent variable to the regression equation. The recursive process starts with the first n iterative of independent variables to find x_1 to add to regression equation, the second iterative of (n–1) independent variables to identify x_2 to be in function on. The procedure stops if none of the partial regression coefficients have the p-value at or below the α level.

- Backward Elimination Approach: This modeling starts with a regression analysis run on the full model, including all independent variables and the dependent variables. If the results from the most recent iterative run show one or more of the partial regression coefficients with the p-value larger than α, the least significant independent variable is removed. The next iteration runs with the exclusion of the removed variables. The process is recursive with each new iteration results in removing the least significant independent variable. The procedure stops if all of the independent variables remaining in the analysis are significant with p-value at or smaller than α.

The author runs the SPSS statistic for both regressions using all independent variables (Regression-Enter) and Stepwise Regression.

Next, the author will discuss the hypotheses testing and their results in section 3 to compare the two models' forecast performance: NN and OLSR.

3. RESULTS

The paired difference *t*-tests typically are used to compare two population means, where the observations in one sample can be paired with observations in the other sample. The observations in the paired difference *t*-tests are defined as the differences between the two dependent samples. This statistical procedure is robust and practical in many circumstances. It has the following assumptions:

- the values of paired difference must be continuous (interval/ratio);
- within the paired difference is independent of one another;
- the paired difference should be approximately normally distributed;
- the sample of the paired difference should not contain outliers.

In the analysis, the author uses the *t*-tests for paired difference procedures to compare the error (measured by Absolute Percent Error) from two methods, specifically Neural Networks and OLSR, to determine whether Neural Network is better than OLSR. Both prediction methods attempt to forecast the price changes of the same stocks; this justifies the dependent sample errors. Specifically, the forecast error for price change on a particular day is measured twice: one error from using the Neural Networks and the other error from applying the OLSR, resulting in pairs of differences in errors. Therefore this *t*-test is an appropriate statistical procedure to determine the mean difference of the errors from Neural Networks and OLSR.

Both the regression models with all independent variables and the stepwise regressions serve as the representatives of OLSR and are used in our study. The two designs of the null and alternative hypothesizes are as follows:

Set 1 of hypotheses statements:

Ho: Neural Network Model is not better than the multiple linear regression models (RegAll)

involving all independent variables in forecasting the changes in stock price.

Ha: Neural Network Model is better than the multiple linear regression models (RegAll).

Set 2 of hypotheses statements:

Ho: Neural Network Model is not better than the stepwise regression models (RegStep) in forecasting the changes in stock price.

Ha: Neural Network Model is better than the stepwise regression models (RegStep).

Further specification of hypotheses testing includes the level of significance (α) is 0.05, and if *p*-value at or lower than 0.05, the Neural Networks is better than OLSR. Otherwise, OLSR is as good as Neural Network in their forecasting performance. Due to the ratio of one *t*-tests for one company, 151 *t*-tests were run for 151 sampled companies listed as A share of the Shanghai Stock Exchange, covering eight different industries. The *t*-test sample size for changes in stock prices of each company is above 400 (approximately 5% of 3,283 observations) used for model testing and error computation. Then the *t*-test analysis continues to conduct by industry with sample size (the number of companies in each industry), which varies depending on the industry.

The author evaluated the Neural Network Method and Multiple Regression Models performances for eight industries and summarized them in Table 2.

In the table, the *APE* (Absolute Percent Error) is calculated with the following formula:

$$APE = \frac{|NAPC - NEPC|}{NAPC} \cdot 100\%. \quad (3)$$

Two types of OLSR models are analyzed in this paper. The first is "Regression All" meaning the regression model, which is inclusive of all independent variables or the input data as the independent variables, and the second is "Regression Stepwise," which selects only the significant independent variables.

Table 2. Industry-wide summary for APE (Absolute Percent Error) for daily stock price change and *p*-value of *t*-test for paired differences

APE and *t*-test *p*-values by Industries	Neural Networks (NN) Average APE	Regression All (RegAll) Average APE	Regression Stepwise (RegStep) Average APE	*t*-tests for paired differences (*p*-value)	
				NN vs RegAll	NN vs RegStep
Basic material	1.57	1.36	1.81	0.91	0.28
Conglomerates	2.19	2.17	2.15	0.77	0.96
Consumer goods	1.65	1.66	2.17	0.18	0.08
Finance	1.69	1.71	1.86	0.01	0.17
Healthcare	1.35	1.36	1.34	0.31	0.90
Industry goods	1.42	1.38	1.37	0.80	0.86
Services	1.34	1.22	1.39	1.00	0.36
Utilities	1.23	1.13	1.13	0.98	0.98

4. DISCUSSION

Based on the results shown in Table 2, the prediction performance of Neural Network is better than other methods (Regression with all variables and Stepwise Regression) for the industries of consumer goods and finance using evaluation criterion of industry-wide average APE. The *t*-test paired difference indicates sufficient evidence that NN is superior to OLSM in the finance sector.

For the industries of conglomerates, healthcare, industry goods, the industry-wide average APE show OLSR (RegStep) performance is as good as or even more accurate than Neural Network and Regression with all variables (RegAll). For the industries, including basic material, services, and utilities, the performance of the Regression (RegAll) with all variables is likely to be more accurate than Neural Network and Stepwise Regression. The industry-wide average APE justifies this result.

In conducting *t*-test for paired difference mean μ_d, where $\mu_d = \mu_{NN} - \mu_{RegAll}$ for Set 1 of hypotheses testing, and $\mu_d = \mu_{NN} - \mu_{RegStep}$ for Set 2 of hypotheses testing. μ_{NN} is defined as APE Mean of NN forecast, $\mu_{Reg-All}$, and $\mu_{RegStep}$ APE Means of the forecast using the multiple linear regression models involving all independent variables and stepwise regression, respectively. The *t*-test is the left-tailed test. That is,

$$Ho: \mu_d \geq 0 \quad vs \quad Ha: \mu_d < 0.$$

Therefore, if *p*-value is at or below α, the Neural Network Model is better than the stepwise regression models. Note that MS Excel functions (Data => Data Analysis => *t*-test: Paired Two Sample for Means) were used to create the *t*-test table for each company. All together 151 *t*-test tables of all 151 companies were created. Summary tables of *t*-test for paired difference results for all 151 companies are provided in Appendix A. Besides, the *t*-test is conducted for each industrial section of all eight industries. The summary is available in Table 2.

The *t*-test results reveal that NN models are not better than the regression models in some industries, the accountable reason that allows OLSR performance is believed to be as good as NN is the normalization/denormalization process. This indicates that the efficient data normalization allows the benefits of using less costly and less time-consuming traditional methods such as OLSR to be delivered to provide the forecast with a similar level of accuracy to what would be if one uses NN.

Further is the summary of overall average APE by methodologies showing the promising accuracy in their prediction for all the methodologies in our study. Absolute Percent Error measures accurate practical value techniques predictors with a range of 1.50% to 1.65%.

Table 3. Summary for APE (Absolute Percent Error) for daily stock price change across eight industrial sectors

The Overall Average APE Indicator	Neural Networks	Regression All	Regression Stepwise
Average APE	1.56 %	1.50 %	1.65 %

CONCLUSION

The analysis finds that Machine Learning, the NN model known for its unfriendly cost and time consumption, provides superior forecasting in changes of the daily stock price in some industry sectors such as finance and possibly consumer goods. Contradictorily, the traditional statistical approaches, such as OLSR models known for their cost and time efficiency, could deliver similar quality forecasting in many sectors. The accountability for the success of this discovery is the appropriate selection of critical factors and data normalization procedures.

The author looks at the years from 2002 to 2015 and the challenges these years brought to stock market prediction. The collection of considerably highly volatile daily stock price data is what sparked our research interest. The author sets up a system to evaluate and analyze performance outcomes.

From the evaluation criterion of industry-wide average APE and our efficient normalization process, our study concludes that both forecasting methods of NN and OLSR secure accurate outcomes. The practical value is evident, as realized with an outcome range forecast error measured by APE of 1.50% to 1.65%. The hypotheses testing indicates that the prediction performance of Neural Network may not be better than Regression with all variables and/or the Stepwise Regression under some circumstances as described in detail.

Future research could be directed towards an analysis of the US stock market with a cross-comparison between the Chinese and the US stock markets to identify which market is more predictable and whether the two markets' critical factors are somewhat similar or different.

AUTHOR CONTRIBUTIONS

Conceptualization: Sunantha Prime.
Formal analysis: Sunantha Prime.
Methodology: Sunantha Prime.
Validation: Sunantha Prime.
Writing – original draft: Sunantha Prime.
Writing – review & editing: Sunantha Prime.

REFERENCES

1. Alyuda Research. (2006). *Alyuda Neuro Intelligence User Manual.* Alyuda Research, LLC, 2001–2017. Retrieved from https://www.alyuda.com/products/neurointelligence/features.htm

2. Anwer, S., & Watanabe, K. (2010). Predicting future depositor's rate of return applying Neural Network: A Case-study of Indonesian Islamic Bank. *International Journal of Economics and Finance, 2*(3), 170-176. https://doi.org/10.5539/ijef.v2n3p170

3. Burstein, F., & Holsapple, C. (2008). Handbook on Decision Support System 2. In *International Handbooks on Information Systems* (pp. 175-193). Springer Berlinn Heidelbert.

4. Cao, Q., & Parry, M. (2011). The three-factor model and artificial neural networks: Predicting stock price movement in China. *Annals of Operations Research, 185*(1), 25-44. https://doi.org/10.1007/s10479-009-0618-0

5. Cao, W., Han, Y., & Lam, S. (2013). Short-term stock price trend prediction of an emerging market using neural networks. *IIE Annual Conference Proceedings,* 50. Retrieved from https://www.researchgate.net/publication/288441864_Short-term_stock_price_trend_prediction_of_an_emerging_market_using_neural_networks

6. Chen, X., Kim, K., Yao, T., & Yu, T. (2010). On the predictability of Chinese stock returns. *Pacific-Basin Finance Journal, 18*(4), 403-425. https://doi.org/10.1016/j.pacfin.2010.04.003

7. Chiang, T., Li, J., & Tan, L. (2010). Empirical investigation of herding behavior in Chinese stock markets: Evidence from quantile regression analysis. *Global Finance Journal, 21*(1), 111-124. https://doi.org/10.1016/j.gfj.2010.03.005

8. Dai, W., Wu, J., & Lu, C. (2012). Combining nonlinear independent component analysis and neural network for the prediction of Asian stock market indexes. *Expert Systems with Applications, 39*(4), 4444-4452. https://doi.org/10.1016/j.eswa.2011.09.145

9. Dawson, C. W., & Wilby, R. L. (2001). Hydrological modelling using artificial neural networks. *Progress in Physical Geography, 25*(1), 80-108. https://doi.org/10.1177/030913330102500104

10. Fama, E. (1970). Efficient Capital Market: A Review of Theory and Empirical Work. *Journal of Finance, 25,* 383-417. https://doi.org/10.2307/2325486

11. Fan, J., Wong, T., & Zhang, T. (2007). Politically connected CEOs, corporate governance, and post-IPO performance of China's newly partially privatized firms. *Journal of Financial Economics, 84*(2), 330-357. https://doi.org/10.1016/j.jfineco.2006.03.008

12. Firth, M., Rui, O., & Wu, X. (2009). The timeliness and consequences of disseminating public information by regulators. *Journal of Accounting and Public Policy, 28*(2), 118-132. https://doi.org/10.1016/j.jaccpub-pol.2009.01.001

13. Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: an application using bankruptcy data. *Information Management, 24*(3), 159-167. https://doi.org/10.1016/0378-7206(93)90064-Z

14. Gardner, M. W., & Dorling, S. R. (1998). Artificial neural networks (the multilayer perceptron) – a review of applications in the atmospheric sciences. *Atmospheric Environment, 32*(14-15), 2627-2636. https://doi.org/10.1016/S1352-2310(97)00447-0

15. Gazzaz, N. M., Yusoff, M. K., Aris, A. Z., Juahir, H., & Ramli, M. F. (2012). Artificial neural network modeling of the water quality index for Kinta River (Malaysia) using water quality variables as predictors. *Marine Pollution Bulletin, 64*(11), 2409-2420. https://doi.org/10.1016/j.marpol-bul.2012.08.005

16. Gordon, R., & Li, W. (2003). Government as a discriminating monopolist in the financial market: The case of China. *Journal of Public Economics, 87*(2), 283-312. Retrieved from https://www.nber.org/papers/w7110

17. Hafezi, R., Shahrabi, J., & Hadavandi, E. (2015). Bat-neural network multi-agent system (BNNMAS) for stock price prediction: Case study of DAX stock price. *Applied Soft Computing, 29,* 196-210. https://doi.org/10.1016/j.asoc.2014.12.028

18. Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques* (2nd ed.). Morgan Kaufmann.

19. He, Z. (2000). Corruption and anti-corruption in reform China. *Communist and Post-Communist Studies, 33*(2), 243-270. https://doi.org/10.1016/S0967-067X(00)00006-4

20. Hu, H., Tang, L., Zhang, S., & Wang, H. (2018). Predicting the direction of stock markets using optimized neural networks with Google Trends. *Neurocomputing, 285,* 1-8. https://doi.org/10.1016/j.neucom.2018.01.038

21. Huang, H. C., & Chen, C. C. (2013). A Study on the construction of the prediction model for exchange rate fluctuations. *International Journal of Intelligent Information Processing, 4*(4), 63-74.

22. Kang, S., Cheong, C., & Yoon, S. (2010). Long memory volatility in Chinese stock markets. *Physica A: Statistical Mechanics and Its Applications, 389*(7), 1425-1433. https://doi.org/10.1080/17538960903529543

23. Kwon, O., Wu, Z., & Zhang, L. (2016). Study of the forecasting performance of China stock's prices using Business Intelligence (BI): Comparison between normalized and denormalized data. *Academy of Information and management Sciences Journal, 20*(1), 53-69. Retrieved from https://www.researchgate.net/publication/301869519_Study_of_the_forecasting_performance_of_China_stocks'_prices_using_busi-ness_intelligence_BI_Compari-son_between_normalized_and_denormalized_data

24. Li, X., Wang, S. S., & Wang, X. (2017). Trust and stock price crash risk: Evidence from China. *Journal of Banking and Finance, 76,* 74-91. https://doi.org/10.1016/j.jbankfin.2016.12.003

25. Lim, T., Huang, W., Yun, J., & Zhao, D. (2013). Has stock market efficiency improved? Evidence from China. *Journal of Finance & Economics, 1*(1), 1-9.

26. Liu, F., & Wang, J. (2012). Fluctuation prediction of stock market index by Legendre neural network with random time strength function. *Neurocomputing, 83,* 12-21. https://doi.org/10.1016/j.neu-com.2011.09.033

27. Liu, H., & Wang, J. (2011). Integrating Independent Component Analysis and Principal Component Analysis with Neural Network to Predict Chinese Stock Market. *Mathematical Problems in Engineering,* 1-15. https://doi.org/10.1155/2011/382659

28. Maier, H. R., & Dandy, G. C. (2000). Neural networks for the prediction and forecasting of water resources variables: a review of modelling issues and applications. *Environmental Modeling and Software, 15*(1), 101-124. https://doi.org/10.1016/S1364-8152(99)00007-9

29. Meng, D. (2008). A neural network model to predict initial return of Chinese SMEs stock market initial public offerings. In *2008 IEEE International Conference on Networking, Sensing and Control* (pp. 394-398). https://doi.org/10.1109/ICNSC.2008.4525247

30. Palani, S., Liong, S., & Tkalich, P. (2008). An ANN application for

water quality forecasting. *Marine Pollution Bulletin, 56*(9), 1586-1597. https://doi.org/10.1016/j.marpolbul.2008.05.021

31. Riedel, J., Jin, J., & Gao, J. (2007). *How China grows: Investment, finance, and reform.* NJ: Princeton University Press.

32. Tjung, L. C., Kwon, O., & Tseng, K. C. (2012). Comparison study on neural network and ordinary least squares model to stocks' prices forecasting. *Financial Management, 9*(1), 32-54. https://doi.org/10.18374/JIFE-19-1.2

33. Tjung, L.C., Kwon, O., Tseng, K. C., & Bradley-Geist, J. (2010). Forecasting financial stocks using data mining. *Global Economy and Finance Journal, 3*(2), 13-26. https://doi.org/10.2139/ssrn.1566268

34. Wang, M., Qiu, C., & Kong, D. (2011). Corporate social responsibility, investor behaviors, and stock market returns: Evidence from a natural experiment in China. *Journal of Business Ethics, 101*(1), 127-141. https://doi.org/10.1007/s10551-010-0713-9

35. Wei, J., Huang, J., & Hui, P. (2013). An agent-based model of stock markets incorporating momentum investors. *Physica A-Statistical Mechanics and Its Applications, 392*(12), 2728-2735. https://doi.org/10.1016/j.physa.2013.02.011

36. Yao, J., Ma, C., & He, W. (2014). Investor herding behavior of Chinese stock market. *International Review of Economics & Finance, 29,* 12-29. https://doi.org/10.1016/j.iref.2013.03.002

37. Yao, Y., & Yueh, L. (2009). Law, finance, and economic growth in China: an introduction. *World Development [HW. Wilson – SSA], 37*(4), 753. https://doi.org/10.1016/j.worlddev.2008.07.009

38. Zhang, H., Wei, J., & Huang, J. (2014). Scaling and predictability in stock markets: A comparative study. *PloS One, 9*(3), 91707-91704. https://doi.org/10.1371/journal.pone.0091707

39. Zhang, X., Zhang, Y., Wang, S., Yao, Fang, Y. B., & Yu, P. S. (2018). Improving stock market prediction via heterogeneous information fusion. *Knowledge-Based Systems, 143,* 236-247. https://doi.org/10.1016/j.knosys.2017.12.025

APPENDIX A

Table A1. Summary tables of analysis by industries and by individual companies from A-share of the Shanghai Stock Exchange

Basic material

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev	Aver.	Max	Stand. Dev	Aver.	Max	Stand. Dev
600019	1.00	1.00	3.28	12.98	2.09	1.79	14.96	2.12	2.46	13.20	2.00
600028	1.00	0.00	1.56	13.51	1.79	1.29	14.18	1.86	1.74	12.27	1.75
600075	0.90	0.90	1.42	15.06	1.71	1.40	16.07	1.69	1.40	15.85	1.69
600078	0.04	0.00	1.58	12.42	1.72	1.62	11.81	1.72	3.79	15.72	2.14
600091	0.41	0.61	2.04	13.38	2.11	2.04	13.50	2.13	2.03	13.31	2.13
600111	0.03	0.63	0.75	9.16	1.11	0.77	8.82	1.11	0.75	8.86	1.11
600117	0.51	0.75	1.35	12.02	1.73	1.35	13.07	1.81	1.34	12.16	1.79
600123	1.00	1.00	3.05	23.97	3.66	2.85	22.80	3.60	2.81	22.20	3.59
600139	0.03	0.61	0.82	8.28	1.04	0.84	9.04	1.06	0.81	8.86	1.07
600165	0.00	0.00	1.00	16.21	1.41	1.05	16.48	1.45	1.05	16.09	1.44
600172	0.87	0.70	1.14	9.19	1.30	1.12	9.82	1.35	1.13	9.44	1.35
600176	0.17	0.56	1.12	12.84	1.57	1.14	13.40	1.59	1.12	13.51	1.58
600179	1.00	1.00	6.31	23.18	2.30	1.32	17.04	1.75	1.30	16.73	1.76
600188	0.10	0.24	1.00	8.22	1.13	1.02	7.95	1.15	1.01	8.20	1.13
600219	1.00	1.00	2.40	26.07	3.23	2.27	27.09	3.25	2.24	26.30	3.24
600225	1.00	1.00	2.38	11.82	1.92	1.75	13.08	1.82	1.73	13.05	1.79
600256	0.62	0.80	0.79	5.43	0.97	0.78	5.54	0.97	0.78	5.33	0.97
600281	0.31	0.48	0.79	5.43	0.97	0.78	5.54	0.97	0.78	5.33	0.97
600282	0.16	0.51	1.34	14.93	1.95	1.36	15.97	1.92	1.34	16.38	1.94

Table A1 (cont.). Summary tables of analysis by industries and by individual companies from A-share of the Shanghai Stock Exchange

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev	Aver.	Max	Stand. Dev	Aver.	Max	Stand. Dev
600291	0.00	0.00	1.53	18.39	2.16	1.60	16.71	2.14	1.73	16.43	2.08
600293	0.01	0.08	1.31	12.11	1.75	1.35	12.61	1.74	1.33	12.39	1.75
600301	0.54	0.00	1.74	16.39	2.07	1.73	16.97	2.07	14.82	33.91	3.11
600307	0.34	0.75	1.32	13.86	1.97	1.33	13.64	1.96	1.31	14.01	1.97
600311	0.11	0.56	1.05	9.06	1.31	1.07	9.10	1.31	1.05	9.15	1.30
600318	0.95	0.99	1.18	11.29	1.48	1.14	10.94	1.50	1.13	12.10	1.52
600319	0.00	0.24	1.66	17.76	1.95	1.85	14.74	1.97	1.67	17.19	1.99
600331	0.01	0.25	1.58	14.95	2.09	1.64	15.00	2.11	1.60	15.34	2.09
600333	0.32	0.40	1.43	8.83	1.48	1.43	9.54	1.48	1.43	9.31	1.49
600339	0.01	0.14	1.21	9.69	1.27	1.25	9.60	1.32	1.23	10.05	1.31
600367	0.00	0.04	1.10	8.31	1.20	1.14	8.04	1.24	1.12	8.29	1.22
600378	0.39	0.27	1.28	10.88	1.40	1.29	9.74	1.39	1.29	10.44	1.42
600381	0.97	0.98	1.86	26.27	2.72	1.80	24.38	2.62	1.80	24.78	2.63
600390	0.96	0.93	1.02	11.21	1.26	0.98	10.64	1.26	0.99	11.03	1.27
600395	1.00	1.00	0.76	5.45	0.82	0.69	5.19	0.84	0.69	5.19	0.85
600399	0.31	0.83	0.73	6.73	0.97	0.74	6.98	0.96	0.73	6.85	0.96

Finance

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600000	0.12	0.22	1.78	14.81	2.52	1.80	16.09	2.54	1.79	16.29	2.53
600007	0.44	0.90	1.77	17.01	2.27	1.78	16.95	2.26	1.75	16.86	2.24
600016	0.17	0.41	1.99	15.88	2.33	2.02	14.79	2.33	1.99	14.97	2.32
600052	0.41	0.77	2.59	23.92	3.42	2.60	23.77	3.39	2.57	23.23	3.38
600061	0.95	0.00	1.49	16.54	2.02	1.46	16.59	2.04	3.72	20.75	2.22
600109	0.72	0.90	0.75	8.04	1.13	0.74	8.43	1.16	0.74	8.38	1.16
600162	0.07	0.74	3.06	28.45	3.66	3.11	29.92	3.72	3.05	27.84	3.67
600193	0.46	0.29	1.39	11.61	1.63	1.39	11.90	1.63	1.40	11.56	1.64
600215	0.23	0.85	1.53	13.23	1.82	1.54	13.32	1.82	1.51	12.93	1.80
600239	0.00	0.53	1.11	12.73	1.81	1.18	12.41	1.81	1.11	12.39	1.79
600240	0.13	0.10	1.43	16.08	1.91	1.46	14.96	1.90	1.46	15.27	1.90
600275	0.06	0.33	1.88	21.38	2.44	1.92	20.81	2.44	1.89	20.31	2.44
600322	0.34	0.69	1.16	9.38	1.40	1.17	10.15	1.43	1.16	9.50	1.41

Conglomerates

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600149	0.09	0.70	2.70	33.63	3.71	2.74	31.68	3.66	2.69	31.24	3.68
600200	0.72	0.94	1.42	16.75	1.79	1.41	16.27	1.74	1.39	16.98	1.76
600212	0.87	0.98	1.95	18.17	2.46	1.93	18.14	2.45	1.91	17.79	2.45
600260	0.98	0.98	2.70	24.26	3.25	2.61	25.59	3.24	2.61	24.64	3.24

Consumer goods

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600006	1.00	1.00	2.08	20.03	2.37	1.84	16.88	2.24	1.81	16.76	2.24
600073	0.99	1.00	1.91	12.35	1.93	1.83	11.58	1.84	1.81	11.78	1.85
600074	0.05	0.37	1.69	15.41	2.33	1.73	15.68	2.25	1.70	15.95	2.26
600084	0.94	0.60	1.57	12.38	1.65	1.52	9.46	1.40	1.57	11.29	1.50
600086	0.22	0.09	0.88	8.73	1.19	0.89	9.41	1.16	0.90	9.60	1.16
600090	1.00	1.00	1.61	14.28	1.87	1.51	11.86	1.79	1.47	11.48	1.79
600093	0.68	1.00	2.40	24.33	3.10	2.39	25.06	3.07	2.32	25.70	3.11
600097	0.15	0.81	1.46	10.40	1.60	1.48	10.49	1.58	1.44	10.80	1.59
600103	0.00	0.00	1.72	14.48	2.06	1.83	14.06	2.09	1.82	13.76	2.08
600108	0.52	0.65	1.70	17.49	2.29	1.70	18.01	2.24	1.69	16.90	2.23
600127	0.11	0.44	1.58	11.48	1.90	1.61	11.62	1.89	1.58	12.24	1.88
600132	1.00	1.00	0.91	11.61	1.12	0.82	11.70	1.14	0.81	11.74	1.14
600156	0.11	0.12	1.81	18.48	2.19	1.85	17.41	2.23	1.84	16.82	2.21
600166	0.23	0.40	0.86	7.08	1.04	0.87	6.88	1.02	0.87	7.44	1.02
600177	0.51	0.73	1.85	16.36	2.42	1.85	16.84	2.42	1.83	16.19	2.41
600182	0.58	0.90	1.73	11.57	1.97	1.72	12.84	1.97	1.69	12.84	2.00
600186	0.04	0.44	1.48	12.42	1.72	1.52	10.60	1.71	1.48	10.81	1.70
600191	0.05	0.47	2.09	19.78	2.69	2.13	22.59	2.73	2.09	22.29	2.74
600197	0.05	0.13	1.51	8.52	1.49	1.55	9.04	1.52	1.54	9.30	1.53
600202	0.86	0.97	2.01	16.60	2.35	1.97	17.38	2.38	1.95	16.67	2.37
600220	0.69	0.59	2.18	19.40	2.59	2.16	21.66	2.55	2.17	21.51	2.56
600232	0.14	0.49	1.64	13.63	1.71	1.67	12.40	1.68	1.64	13.48	1.67
600233	0.73	0.95	1.34	13.43	1.65	1.32	10.86	1.58	1.30	10.94	1.59
600257	0.02	0.00	2.84	18.90	2.95	2.92	20.10	2.92	9.81	22.41	3.47
600261	1.00	1.00	1.00	11.37	1.02	0.90	11.08	1.03	0.88	10.90	1.02
600265	0.48	1.00	1.78	12.12	2.05	1.78	13.02	2.09	13.00	26.46	3.08
600298	0.09	0.59	1.18	10.53	1.46	1.20	10.17	1.42	1.17	9.86	1.43
600300	0.01	0.01	2.12	20.15	2.64	2.20	20.28	2.64	2.20	19.84	2.64
600303	0.00	0.04	0.90	7.74	1.14	0.95	8.08	1.15	0.93	8.19	1.15
600308	0.00	0.00	1.68	18.42	2.26	2.11	19.64	2.44	1.97	19.40	2.33
600337	0.06	0.00	1.74	15.99	2.12	1.79	16.93	2.16	2.34	14.41	2.07
600356	0.07	0.15	1.37	14.14	1.69	1.39	13.69	1.66	1.38	14.43	1.68
600365	0.14	0.52	1.70	17.31	2.03	1.72	16.88	2.00	1.70	16.68	2.01
600400	0.11	0.37	1.53	12.57	1.77	1.56	14.24	1.81	1.54	14.30	1.80
600419	0.00	0.00	1.77	24.37	2.55	1.86	23.66	2.67	1.86	23.52	2.64

Healthcare

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600055	0.09	0.20	1.66	22.95	2.09	1.69	22.23	2.11	1.68	22.86	2.13
600062	0.03	0.03	1.25	8.72	1.45	1.29	9.07	1.46	1.28	8.81	1.45
600079	0.03	0.33	1.10	8.90	1.22	1.13	8.41	1.24	1.10	8.04	1.24
600080	0.48	0.46	1.35	9.47	1.26	1.35	10.47	1.26	1.35	10.27	1.27
600085	0.93	1.00	1.06	12.42	1.26	1.03	12.06	1.25	1.00	12.23	1.25
600195	0.25	0.52	1.25	12.81	1.73	1.26	12.38	1.71	1.25	11.80	1.71
600201	0.89	0.93	1.15	11.36	1.32	1.13	11.59	1.34	1.13	11.30	1.33
600211	0.01	0.11	1.94	19.77	2.39	2.00	19.26	2.35	1.97	19.70	2.39
600216	0.54	0.97	0.93	6.16	1.09	0.93	6.36	1.10	0.91	6.04	1.10
600267	0.37	0.88	0.72	5.65	0.78	0.72	5.35	0.79	0.70	5.38	0.79
600297	0.31	0.68	1.71	12.49	2.02	1.73	12.72	2.01	1.71	12.51	1.99
600380	0.10	0.45	1.58	16.35	2.11	1.61	16.77	2.15	1.58	15.72	2.12
600385	0.97	0.99	2.20	20.62	2.64	2.11	23.11	2.70	2.08	21.91	2.71
600466	0.54	0.53	0.85	10.80	1.11	0.84	11.76	1.14	0.85	11.50	1.13
600488	0.73	0.97	1.57	15.39	1.85	1.56	16.83	1.92	1.53	16.82	1.90

Industrial goods

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600089	0.61	0.77	1.00	10.47	1.42	0.99	10.65	1.42	0.99	10.57	1.43
600150	1.00	1.00	1.46	18.58	2.26	1.21	17.55	2.24	1.20	17.93	2.26
600169	1.00	1.00	1.11	10.76	1.60	1.04	11.89	1.59	1.04	11.92	1.60
600243	0.01	0.32	1.66	13.92	1.96	1.72	14.01	1.98	1.67	14.06	1.95
600290	1.00	1.00	1.51	13.12	1.75	1.31	12.86	1.80	1.29	13.03	1.82
600302	0.42	0.67	1.51	14.06	1.57	1.52	14.01	1.61	1.51	13.54	1.59
600320	0.00	0.00	1.79	19.36	2.42	1.96	20.02	2.37	1.96	19.62	2.37
600335	0.88	0.93	1.44	16.16	1.91	1.42	15.38	1.89	1.42	16.10	1.89
600388	0.99	0.99	1.29	12.62	1.60	1.25	12.65	1.61	1.26	12.75	1.59

Services

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600009	0.81	0.87	1.58	14.00	1.85	1.55	11.75	1.84	1.55	12.07	1.84
600054	1.00	1.00	1.24	8.40	1.25	0.98	7.12	1.06	0.97	6.74	1.06
600115	0.97	1.00	1.16	11.08	1.39	1.11	9.95	1.43	1.08	10.27	1.43
600122	0.94	0.98	0.80	8.28	1.00	0.77	8.52	1.01	0.76	8.58	1.00
600125	1.00	1.00	0.91	9.12	0.94	0.87	8.50	0.92	0.86	8.57	0.92
600136	0.99	1.00	1.60	15.29	1.88	1.53	14.41	1.85	1.52	14.88	1.87
600138	1.00	0.00	1.98	10.72	2.03	1.42	9.25	1.46	4.91	14.18	2.02
600190	1.00	0.99	1.54	15.45	1.67	1.48	14.15	1.65	1.48	14.16	1.66
600203	0.96	0.97	1.44	11.60	1.54	1.37	12.00	1.48	1.37	12.67	1.51
600221	1.00	1.00	1.10	8.21	1.19	0.91	8.15	1.08	0.90	8.15	1.07
600242	0.99	0.98	0.95	6.30	0.93	0.90	7.05	0.96	0.90	7.34	0.97
600250	1.00	1.00	1.47	16.27	1.70	1.40	16.40	1.67	1.38	16.77	1.69
600258	0.26	0.22	1.07	9.29	1.21	1.09	11.40	1.27	1.09	10.88	1.28
600270	1.00	1.00	2.15	20.70	2.27	1.71	10.25	1.80	1.71	10.48	1.80
600278	0.97	0.97	1.28	12.20	1.48	1.23	13.38	1.47	1.23	13.85	1.48
600279	1.00	1.00	1.20	12.82	1.45	1.12	12.82	1.42	1.11	12.27	1.42
600306	1.00	1.00	0.90	6.93	0.99	0.72	5.64	0.78	0.72	5.51	0.79
600361	0.14	0.10	1.75	22.47	2.60	1.79	26.18	2.67	1.79	25.25	2.68
600382	1.00	1.00	1.30	12.65	1.52	1.16	12.36	1.43	1.16	12.49	1.44

Utilities

Company code at SSE	t-tests (p-value, alpha = .05)		Neural Network			Regression_All			Regression_Stepwise		
	RegAll vs NN	RegStep vs NN	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.	Aver.	Max	Stand. Dev.
600101	1.00	1.00	0.97	7.15	1.00	0.80	7.15	0.97	0.80	7.19	0.97
600168	1.00	1.00	1.23	7.74	1.33	1.19	8.58	1.32	1.18	8.49	1.31
600207	1.00	1.00	1.52	14.55	1.73	1.21	12.76	1.46	1.20	12.61	1.46
600283	0.98	0.86	1.21	8.34	1.41	1.17	8.66	1.42	1.19	8.98	1.43
600292	1.00	1.00	1.21	8.68	1.36	1.14	8.29	1.34	1.13	8.24	1.33
600310	0.79	0.87	0.85	9.49	1.29	0.84	10.08	1.29	0.83	10.12	1.30
600396	0.97	0.97	1.63	17.50	1.98	1.58	13.80	1.89	1.58	13.79	1.90

APPENDIX B

Economic, market, and institutional investors indicators

DIA	Dow Jones Industrial Average
AEX	Amsterdam Price
ATX	Vienna Stock Exchange
BEN	Franklin Resources, Inc.
BVSP	Bovespa-Brazillian Index
FCHI(CAC40)	Cotation Assistée en Continu
GDAXI(DAX)	Deutsche Boerse AG German Stock Index
FTSE	FTSE Index Price
JKSE	Jakarta Stock Exchange Index
KLSE	FTSE Bursa Malaysia Klci
KS11(KOSPI)	Kospi Composite Index
MERV	Minimum efficiency reporting value
N225	Nikkei 225
BSESN	Bombay Stock Exchange Sensex
SSMI(SMI)	Swiss Market Index
STI	Straits Times Index Singapore
TWII(TSEC)	Taiwan Weighted Index
VXO	Volatility Index
VIX	Cboe Index
S&P500	S&P 500