Forecasting Financial Stocks using Data Mining

California State University Fresno

Luna Christie Tjung (christieliem@csufresno.edu)

Craig School of Business, California State University

Fresno, CA 93740-8001

Ojoung Kwon (okwon@csufresno.edu)

Department of Information Systems & Decision Sciences

Craig School of Business, California State University

Fresno, CA 93740-8001

K.C. Tseng (kctseng@csufresno.edu)

Department of Finance

Craig School of Business, California State University

Fresno, CA 93740-8001

Jill Bradley (jbradley@csufresno.edu)

Department of Management

Craig School of Business, California State University

Fresno, CA 93740-8001

ABSTRACT

This study presents a Business Intelligence (BI) approach to forecast daily changes in seven financial stocks' prices. The purpose of our paper is to compare the performance of Ordinary Least Squares model and Neural Network model to see which model does a better job to predict the changes in the stock prices and identify critical predictors to forecast stock prices to increase forecasting accuracy for the professionals in the market.

The BI approach uses a financial data mining technique to assess the feasibility of financial forecasting compared to regression model using ordinary least squares estimation method. We used eight indicators such as macroeconomic indicators, microeconomic indicators, market indicators, market sentiment, institutional investor, politics indicators, business cycles, and calendar anomaly to predict changes in financial stock prices.

Keywords: Business Intelligence, Financial Forecasting, Investment Strategies, Data Mining, Forecasting Techniques, and Neural Networks

I. **Introduction**

Burstein and Holsapple (2008) state that 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. According to Han and Kamber (2006), data mining is extracting knowledge from large amounts of data.

Fama (1970) defined efficient markets hypothesis where the idea is a market in which prices provide accurate signals for resource allocation: that is, a market in which firms can make production-investment decisions, and investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at any time "fully reflect" all available information. We know that it is not easy to predict the stock prices but doable to the extent that we want to reduce the forecasting error by selecting better model. The purpose of our paper is to compare the performance of Ordinary Least Squares model and Neural Network model to see which model does better job to predict the changes in the stock prices.

OLS model has many advantages such as easy to use, to validate, and to generate best combination by using stepwise regression. However, OLS is a linear model that has relatively high forecasting error to forecasting a non linear environment in the stock markets. OLS model also can only trace one dependent variable at a time. On the other hand, neural network model has a high precision, improving prediction in nonlinear setting, and addressing problems with a great deal of complexity. So, we expect to see neural network does a better job compared to OLS model.

Prior research and common wisdom have suggested several factors that might be used in OLS or a Neural Network model to predict stock prices. Grudnitski and Osburn (1993) used general economic conditions and traders' expectations about what will happen in the market for these futures. Kahn (2006) stated that the sentiment indicator is the summation of all market expectation that is driven by emotions such as volatility index, put/call ratio, short interest, commercial activity, surveys, magazine over indicator, and many more. Tokic (2005) showed that political events like the war on terror, fiscal policy to lower taxes, and monetary policy to lower short-term interest resulted in the increase of the budget deficit. Nofsinger and Sias (1999) showed that there is a strong positive relation between annual changes in institutional ownership and returns over the herding interval across capitalizations.

In addition to determining which factors best predict changes in stock prices, we will also be comparing the two analytic strategies of OLS and Artificial Neural Networks. Hammad, Ali, and Hall (2009) showed that Artificial Neural Network (ANN) technique provides fast convergence, high precision and strong forecasting ability of real stock price. Traditional methods for stock price forecasting are based on statistical methods, intuition, or on experts' judgment. Traditional methods performance depends on the stability of the prices; as more political, economical, and psychological impact-factors get into the picture, the problem becomes non linear, and traditional methods need a more heuristic or nonlinear methods like ANN, Fuzzy logic, or Genetic Algorithms**.**

Along the same lines as Hammad et al., (2009), West, Brockett, and Golden (1997) concluded that the neural network offers superior predictive capabilities over traditional statistical methods in predicting consumer choice in nonlinear and linear settings. Neural networks can capture nonlinear relationships associated with the use of non compensatory decision rules. The study revealed that neural networks have great potential for improving model predictions in nonlinear decision contexts without sacrificing performance in linear decision contexts.

However, Neural Networks are not a panacea. For example, Yoon and Swales (1991) concluded that despite neural network's capability of addressing problems with a great deal of complexity, as the increase in the number of hidden units in Neural Network resulted in higher performance up to a certain point; additional hidden units beyond the point impaired the model's performance.

II. <u>Tools and Techniques of Stock Forecasting</u>

Moshiri and Cameron (2000) compared the most commonly used type of Artificial Neural Network (the Back-Propagation Networks (BPN) model) with six traditional econometrics models (three structural models and three time series models) in forecasting inflation. BPN models are static or feedforward-only (input vectors are fed through to output vectors, with no feedback to input vectors again); they are hetero-associative (the output vector may contain variables different from the input vector) and their learning is supervised (an input vector and a target output vector both are defined and the networks tend to learn the relationship between them through a specified learning rule). The three structural models include (1) the reduced-form inflation equation that follows from a fairly standard aggregate demand-aggregate supply model with adaptive expectations, (2) the inflation equation from Ray Fair's econometric forecasting model, and (3) a monetary model for forecasting inflation. The three time series models are (1) an ARIMA (Autoregressive integrated moving average) model is the single-variable model derived using Box-Jenkins (1978) methodology, (2) a Vector Autoregression or VAR model consider the joint behavior of several variables, and (3) a Bayesian Vector Autoregression or BVAR model is the combination of VAR model with prior information on the coefficients of the model and estimated using a mixed-estimation method (Theil, 1971). In one-period-ahead dynamic forecasting, the information contained in the econometric models is contained in the BPN and the BPN contains further information; the BPN models are superior for all four comparisons. Over a three-period forecast horizon the BPN models are superior in two comparisons (VAR and Structural) and inferior in two (ARIMA and BVAR). Over a twelve-period forecast horizon, the BPN models are superior in two comparisons (VAR and Structural) and equally good or bad in two (ARIMA and BVAR). Moshiri (2001) generalized that the BPN model has been able to outperform econometric models over longer forecast horizons.

There are many examples of the successful applications of data mining. DuMouchel (1999) used Bayesian data mining to work with large frequency tables, millions of cells, for FDA Spontaneous application. Giudici (2001) used Bayesian data mining for benchmarking and credit scoring in a highly dimensional complex datasets. Jeong, et.al. (2008) integrated data mining to a process design using the robust Bayesian approach.

III. Hypothesis

Fama (1970) defined efficient markets hypothesis where the ideal is a market in which prices provide accurate signals for resource allocation that is, a market in which firms can make production-investment decisions, and investors can choose among the securities that represent ownership of firms' activities under the assumption that security prices at any time "fully reflect" all available information. We know that it is not easy to predict the stock prices but doable to the extent that we want to reduce the forecasting error by selecting a better model. Because Neural Network can address problems with a great deal of complexity and improve its prediction in nonlinear setting, we expect to see that Neural Network does better job than OLS model.

H_0: OLS model better predicts stock prices than NN model

H_1: NN model better predicts stock prices than OLS model

IV. Methodology

In order to forecast the changes in financial stock prices, we used the daily changes in stock prices of seven financial stocks from September 1, 1998 to April 30, 2008. We used financial stocks because it is relatively volatile and more sensitive to economic news. Our participants and independent variables are the leading companies in financial industry found in table 2 and 3.

Predictors

We used eight indicators such as macroeconomic leading indicators (global market indices), microeconomic indicators (competitors), politics indicators (presidential election date and party), market indicators (USA index), institutional investor (BEN), and calendar anomaly as our independent variables to predict changes in daily financial stock prices. We also take into account the business cycle factors such as dot-com bubble in our forecasting horizon with dummy variables. We gathered our data through National Bureau of Economic Research (NBER), Yahoo Finance, Federal Reserve Bank, Market Vane (MV), NYSE, and FXStreet.

The macroeconomic indicators include the 18 major global stock indices. The microeconomic indicators include the competitors and companies from different industries. There are 213 of them. The daily market indicators include changes in price and volume of S&P500, Dow Jones Industrial, Dow Jones Utility, and Dow Jones Transportation. The sentiment indicators include the Volatility Index (VIX) and CBOE OEX Implied Volatility (VXO).

The politics indicators include the election and party. We denote non-election date and Democratic Party with 0. The calendar anomalies include daily, weekly, monthly, and pre holiday calendar. The daily calendar includes Monday, Tuesday, Wednesday, Thursday, and Friday. The weekly calendar includes week one to week 5. The monthly calendar includes January to December. We denote Tuesday, week 2, and September with dummy variable zero in the OLS model. On the other hand, we included them in the NN model.

The business cycle includes the recession from technological crash and current bear market with dummy variable one.

OLS Model

We used SPSS to perform stepwise regression to create a unique regression model for each company.

NN Model

We have been going through many trials and errors before we get the desirable results. First, we used BrainMaker software to create the NN model for 4 companies (C, GS, JPM, and MS), but it has the major drawback that we cannot include more than 20 variables into our model. So, we search other software that can assist us to include more variables. Also, we do not know what the best network architecture is for each dependent variable.

We found the Alyuda NeuroIntelligence that helped us to handle many more data and find the best network architecture. So, with this benefit, we increase more independent variables from 100 to 267 variables and dependent variables from 4 to 7 variables (T. Rowe Price Group, Plum Creek Timber, HCP, Host Hotels & Resorts, Public Storage, Boston Properties, and Simon Property Group). The major differences are we included all sub-industries in the market to increase our performance accuracy and took out the market vane indicators on commodity market because its lack of the correlation with our dependent variables.

We ran the neural network with Alyuda NueroIntelligence to create a NN model. We did data manipulation by using the changes or the first derivative of our independent variables except for the dummy variables. Also, we normalized the data because we have negative to positive numbers.

We followed seven-step neural network design process to build up the network. We used the Alyuda NeuroIntelligence to perform data analysis, data preprocessing, network design, training, testing, and query.

We used hyperbolic tangent method to design the network. We used Batch Back Propagation model with stopping training condition of MSE of 1% to find the best network during the network training. Hyperbolic tangent is a sigmoid curve and is calculated using the following formula: $F(x)= (e^{x}-e^{-x})/(e^{x}+e^{-x})$. “Back propagation algorithm is the most popular algorithm for training of multi-layer perceptrons and is often used by researchers and practitioners. The main drawbacks of back propagation are: slow convergence, need to tune up the learning rate and momentum parameters, and high probability of getting caught in local minima.”(Alyuda NeuroIntelligence Manual, 2010)

Also, we used overtraining control such as retain and restore best network and add 10% jitter to inputs, weights randomization method such as Gaussian distribution of network inputs, and retrains network 2 times with the lowest training error to train the network. By retaining and restoring best network, we can prevent over-training such as memorizing data instead of generalizing and encoding data relationships and thus reduce the network error. As a result, the validation errors rise while training errors may still reduce in the training graph. By adding jitter, we not only can prevent over-training but also allow the network to escape local minima during training (major drawback from the batch-back propagation) by adding 10% random noise to each input variable during training. By randomizing the weights, we avoid sigmoid saturation problems that cause slow training. We used Gaussian distribution because it is characterized by a continuous, symmetrical, bell-shaped curve.

Data - Model Building Data Set and Performance Testing Data Set for OLS

We used 2431 data do build the OLS model by running stepwise regression. By doing that, we reduced our independent variables range between 31 and 61 variables. We tested the model by using a randomly selected 152 data to test the forecasting accuracy of the OLS with NN methods.

Data – Training Data Set, Validation Data Set, and Testing Data Set

Unlike the OLS model, NN model used all independent variables. There are three sets of data used in the neural network model such as training set, validation set, and testing set. The training set is used to train the neural network and adjust network weights. The validation set is used to tune network parameters other than weights, to calculate generalization loss and retain the best network. The testing set is used to test how well the neural network performs on new data after the network is trained. We used training and validation data to train the network and come up with a model. Finally, we used testing data to test the forecasting errors between the actual and predicted values. Out of 2431 data, we have 152 testing data. The remaining is equally distributed among the training and validation data.

Analysis

We measured our success by testing the accuracy the NN with OLS model in term of the significant % error of the mean and standard error. After we were getting the results, our mean for NN is low (2.47% to 19.68%) but our standard deviation is too high (218.73% to 584.26%). The same happened to the OLS model with mean of 7.29% to 167.43% and standard deviation of 160.33% to 962.01%. Then, we realized that our % error has both positive and negative numbers because we are using the first derivative for all our variables except dummy variables. So, we took the absolute value of the error percentage of all variables. Even after we took the absolute value of the error percentage, our mean (127% to 206%) and standard deviation (174% to 532.8%) for NN model are still so high. For the OLS model, we got mean of 104%-381% and standard deviation of 127% to 849%.

So, we want to normalize the data to create better network training by adding 0.1 to the absolute value of the minimum value in each variable to avoid minus sign from the rounding down. After we normalized the data, we have both lower mean (2.13% to 3.27%) and standard deviation (1.78% to 3.39%) for NN model. We have similar result for the OLS model with the mean ranges from 4%-32% and standard deviation ranges from 2.46%-8%. Finally, we conducted a paired-t test to measure the performance between two models by using the % NN error and |% OLS error|.

V. Results

The Adj R^2 ranges from 0.523 to 0.766 with DW from 2.039 to 2.209.

FINANCIAL INDUSTRY	COMPANIES	OLS ADJ R^2	DW
1 Asset Management	T. ROWE PRICE GROUP INC. [TROW]	0.766	2.085
2 REIT - Diversified	PLUM CREEK TIMBER CO. INC. [PCL]	0.758	2.128
3 REIT - Healthcare Facilities	HCP INC. [HCP]	0.597	2.171
4 REIT - Hotel/Motel	HOST HOTELS & RESORTS INC. [HST]	0.695	2.153
5 REIT - Industrial	PUBLIC STORAGE [PSA]	0.818	2.085
6 REIT - Office	BOSTON PROPERTIES INC. [BXP]	0.811	2.039
7 REIT - Retail	SIMON PROPERTY GROUP INC. [SPG]	0.523	2.209

After normalized the data, we reduced our error significantly to < 3.28% for NN and < 33% for OLS model.

NORMALIZED DATA %ERROR

FINANCIAL INDUSTRY	COMPANIES	NN Mean (Stdev)	OLS |Mean| (Stdev)
Asset Management	T. ROWE PRICE GROUP INC. [TROW]	3.18% (2.94%)	32% (8%)
REIT - Diversified	PLUM CREEK TIMBER CO. INC. [PCL]	2.13% (1.78%)	4% (3%)
REIT - Healthcare Facilities	HCP INC. [HCP]	2.66% (3.25%)	25% (5%)
REIT - Hotel/Motel	HOST HOTELS & RESORTS INC. [HST]	3.22% (2.84%)	6% (3%)
REIT - Industrial	PUBLIC STORAGE [PSA]	2.82% (3.39%)	5% (3%)
REIT - Office	BOSTON PROPERTIES INC. [BXP]	2.54% (2.33%)	5.43% (2.46%)
REIT - Retail	SIMON PROPERTY GROUP INC. [SPG]	3.27% (3.3%)	6% (4%)

The Adj R^2 ranges from 0.523 to 0.766 with DW from 2.039 to 2.209.

Paired Samples Test

		Paired Differences					t	df	Sig. 2-tailed
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper			
Pair 1	TROWnn - TROWols	-0.2929	0.08307	0.00674	-0.3063	-0.2796	-43.48	151	0
Pair 2	PCLnn - PCLols	-0.0181	0.03225	0.00262	-0.0233	-0.013	-6.933	151	0
Pair 3	HCPnn - HCPols	-0.2236	0.04953	0.00402	-0.2315	-0.2157	-55.66	151	0
Pair 4	HSTnn - HSTols	-0.0295	0.03996	0.00324	-0.0359	-0.0231	-9.111	151	0
Pair 5	PSAnn - PSAols	-0.0173	0.02853	0.00231	-0.0219	-0.0127	-7.47	151	0
Pair 6	BXPnn - BXPols	-0.029	0.034	0.003	-0.034	-0.024	-10.59	151	0
Pair 7	SPGnn - SPGols	-0.0273	0.04385	0.00356	-0.0344	-0.0203	-7.685	151	0

Based on the result from paired t-test, we reject the Ho that OLS better predicts stock prices. The negative sign in the t-statistics shows that NN has lower errors compared to the OLS model.

VI. Discussion

The stock market is made of investors' risk and return characteristics, perception and expectation about stocks, and how they interpret the news. Each investor is reacting to the market at various times, looking at different relevant information, and making different conclusions at different times. With this assumption, Fama hypothesis is not strong.

Investors are made up from noise traders who do not maximize utility but instead trade exogenously, smart-money trader or utility-maximizing risk averse, and contrary opinion trader. Investors can over-react, under-react, or normally react to certain economic news. We also do not know how big and long are the impact of economic news to the stock prices. With these uncertainties, we need a model that can work best with the shocks. Our finding shows that NN and OLS model are not capable to accommodate market shocks.

The OLS model is easy to use and validate. It also works fast. However, it is a linear model with a relatively higher error to forecast non-linear environment in the stock market. Also, it only traces one dependent variable at a time.

The NN model is complex and requires more efforts to train the network over and over again to find the best model. Some critical factors to create the best model such as the network architecture (number of layers and neurons) and design (logistic/ hyperbolic tangent/ linear), training algorithms, and stop training conditions (number of iterations). Although we can choose low MSE, it does not guarantee that is the best model because the network might be over-trained causing it not learning but memorizing. Also, the Alyuda NeuroIntelligence used different set of data each time we run the network. The software can tell us what is the best network architecture but not the training part. We cannot be certain about if the model is the best or not when it comes to train the network. With these uncertainties, it is hard to measure the performance of the neural network. It takes more time to train and learn how to use the neural network.

Our results show that NN does a better job than OLS model. Also, our paper shows a significant contribution to the financial forecasting where we can see how one industry affect the others.

VII. Conclusion and future direction

We have limited data resources to find the institutional investors' holding in particular company and for how long. Also, we missed out other important variables such as war and news. The news we mean here is the words spoken by credible investors and politicians in the market such as Warrant Buffett and senators/president/the Fed. One of our research limitations is we are only comparing two methods while there are many other models that requires our attention. For the future direction, we recommend to include more techniques to find the best model for financial forecasting purpose. Also, there are many other learning algorithms in the NN that we have not explored yet.

To improve the accuracy of forecasting variables, we suggest researchers to conduct find the percentage of noise traders who do not maximize utility but instead trade exogenously, smart-money trader or utility-maximizing risk averse, and contrary opinion trader in the market. Profiling the investors is important to understand how they behave.

We learned that data normalization can make a huge difference to the result.

References

Alyuda NeuroIntelligence. (2010). Alyuda NeuroIntelligence Manual. (http://www.alyuda.com/neural-networks-software.htm).

Brainmaker, (2010). California Scientific. (http://www.calsci.com/BrainMaker.html)

Burstein, F., Holsapple, C. (2008). Handbook on Decision Support System 2. Springer Berlinn Heidelbert, page 175-193.

DuMouchel, W. (1999). Bayesian Data Mining in Large Frequency Tables, With an Application to the FDA Spontaneous. *American Statistician*. Volume 53, page 177.

Fama, E. (1970). Efficient Capital Market: A review of Theory and Empirical Work. *Journal of Finance*, Volume 25, page 383-417.

Giudici, P. (2001). Bayesian data mining, with application to benchmarking and credit scoring. *Applied Stochastic Models in Business & Industry*. Volume 17, page69-81.

Grudnitski, G., Osburn, L. (1993). Forecasting S&P and Gold Futures Prices: An Application of Neural Network. *Journal of Futures Markets*, Volume 13, page 631-643.

Hammad, A.A.B., Ali, S.M.A., Hall, E.L. (2009) Forecasting the Jordanian Stock Price using Artificial Neural Network.

Han, J., Kamber, M. (2006). Data Mining: Concepts and Techniques. 2nd Edition. Morgan Kaufmann, page 5.

Jeong, H., Song, S., Shin, S., Cho, B. (2008). Integrating Data Mining to a Process Design Using the Robust Bayesian Approach. *International Journal of Reliability, Quality & Safety Engineering*. Volume 15, page 441-464.

Kahn, M. (2006). Technical Analysis Plain and Simple: Charting the Markets in Your Language. Financial Times Prentice Hall Books.

Moshiri, S., Cameron, N. (2000). Neural network versus econometric models in forecasting inflation. *Journal of Forecasting*, Volume 19, page 201-217.

Nofsinger, J., Sias, R. (1999). Herding and Feedback Trading by Institutional and Individual Investors. *Journal of Finance*, Volume 54, page 2263-2295.

Tokic, D. (2005). Explaining US stock market returns from 1980 to 2005. *Journal of Asset Management*, Volume 6, page 418-432.

West, P., Brockett, P., Golden, L. (1997). A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice. *Marketing Science.* Volume 16, page 370-391.

Yoon, Y., Swales, G. (1991) Predicting Stock Price Performance: A Neural Network Approach.

VIII. FIGURES

FIGURE 1. TROW- Batch Back Propagation: TESTING

Model Architecture: 267-40-1 No. of Iterations: 751

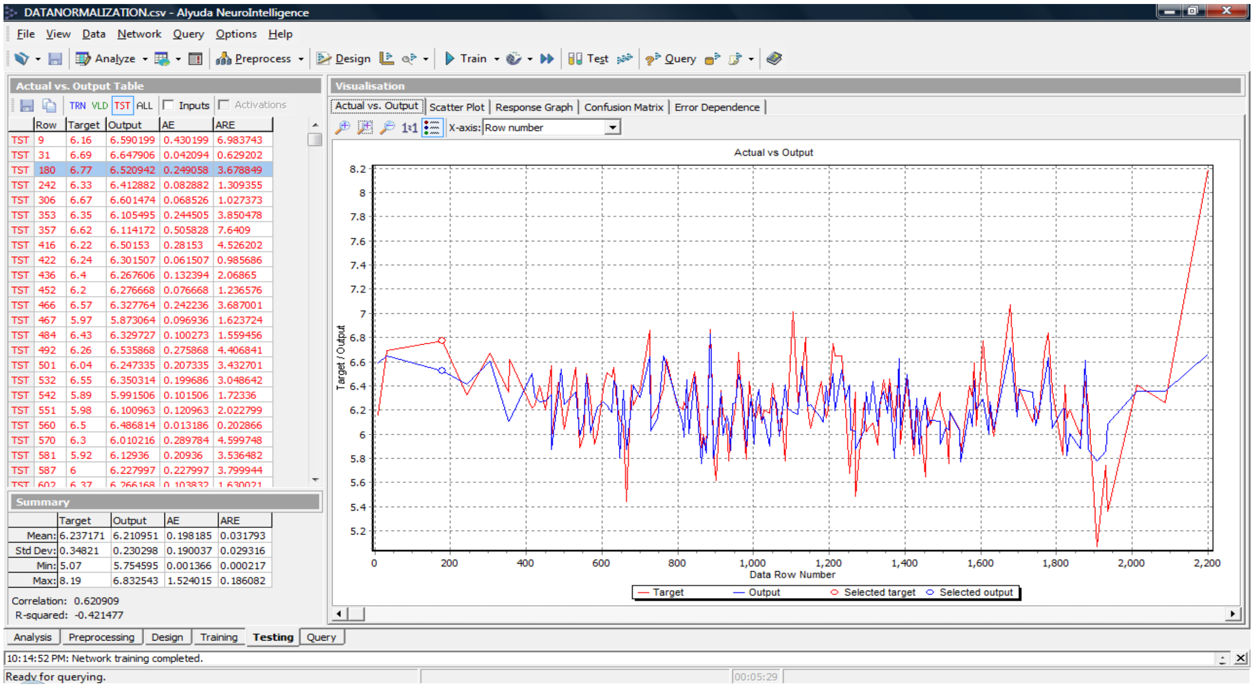


FIGURE 2. PCL- Batch Back Propagation: TESTING

Model Architecture: 267-40-1 No. of Iterations: 1001

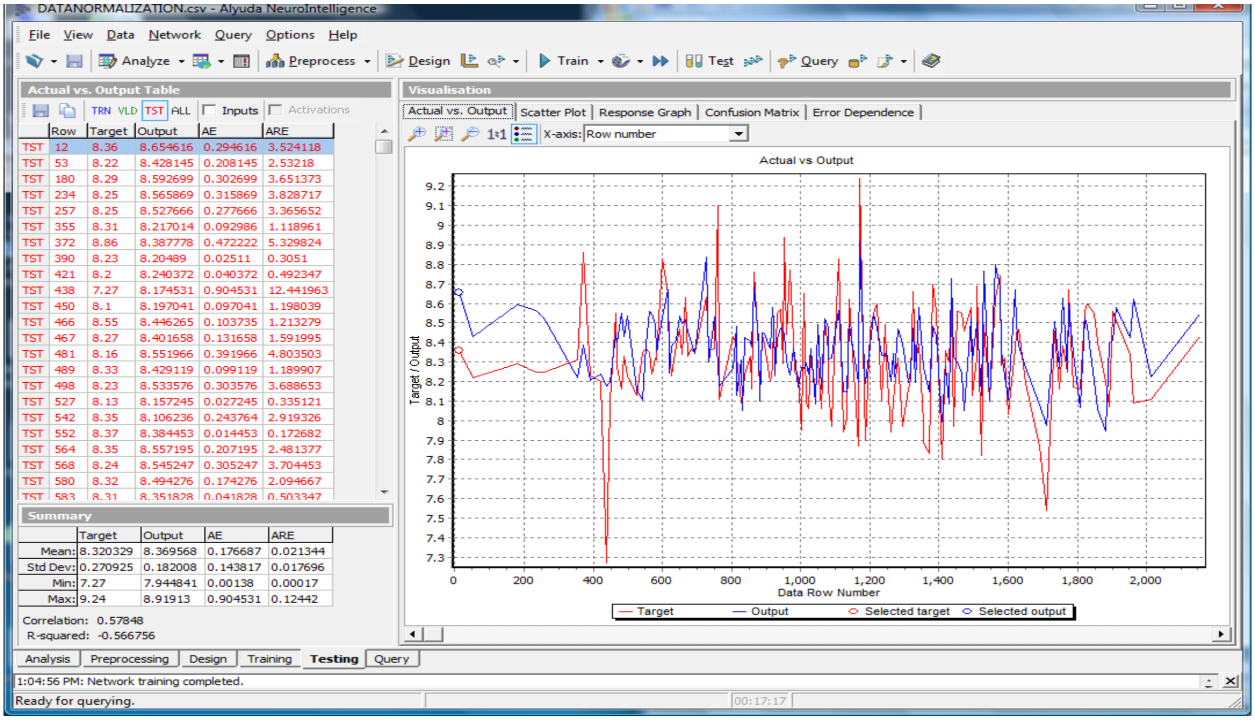


FIGURE 3. HCP- Batch Back Propagation: TESTING

Model Architecture: 267-40-1 No. of Iterations: 1001

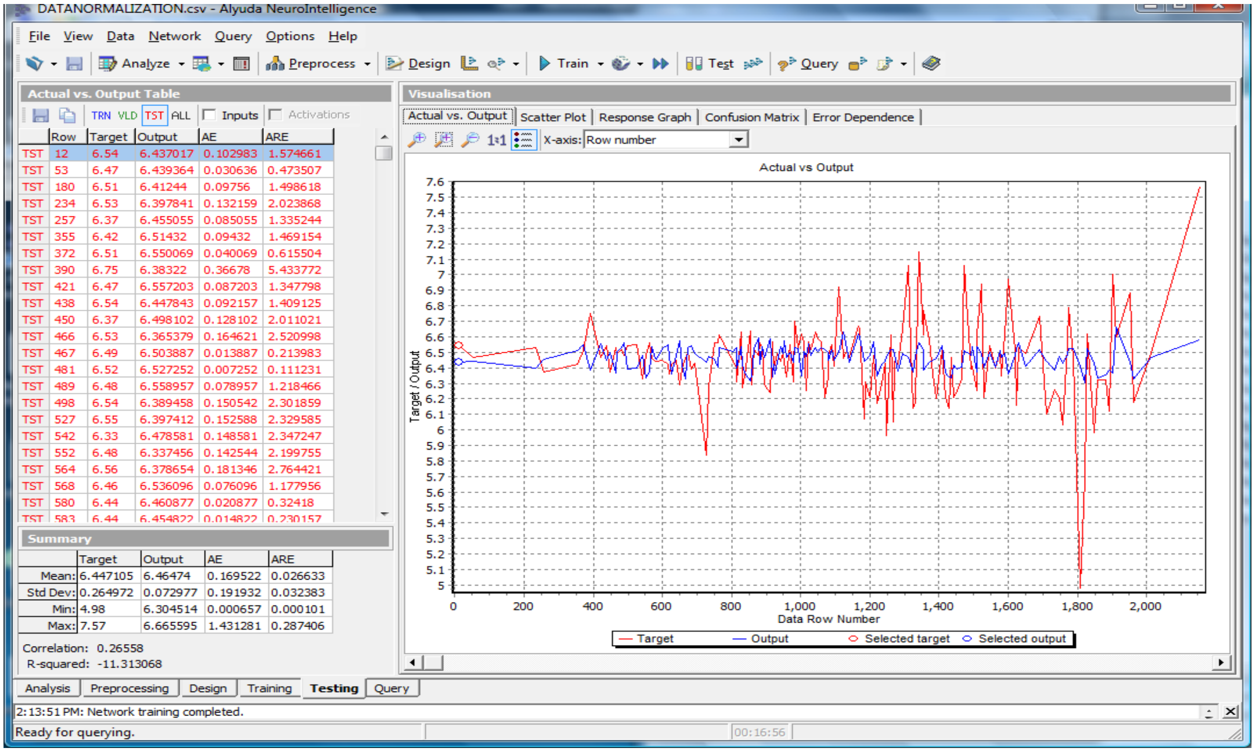


FIGURE 4. HST- Batch Back Propagation: TESTING

Model Architecture: 267-40-1 No. of Iterations: 1001

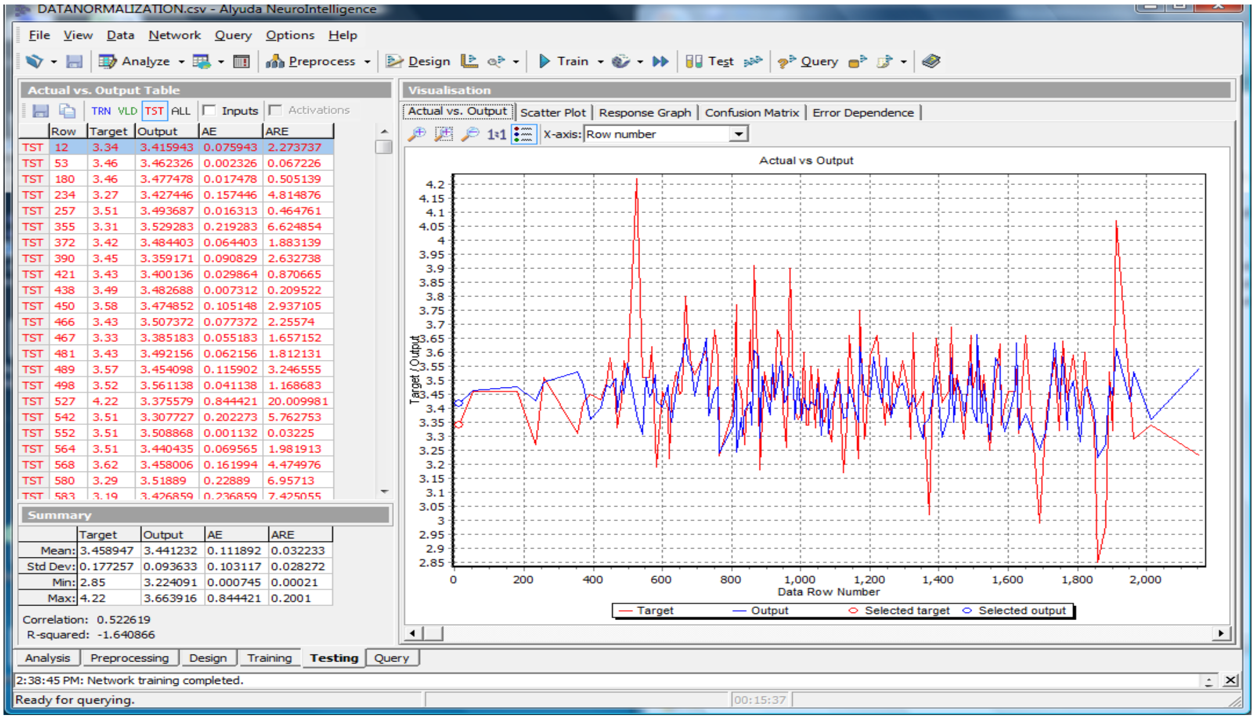


FIGURE 5. PSA- Batch Back Propagation: TESTING

Model Architecture: 267-40-1 No. of Iterations: 1001

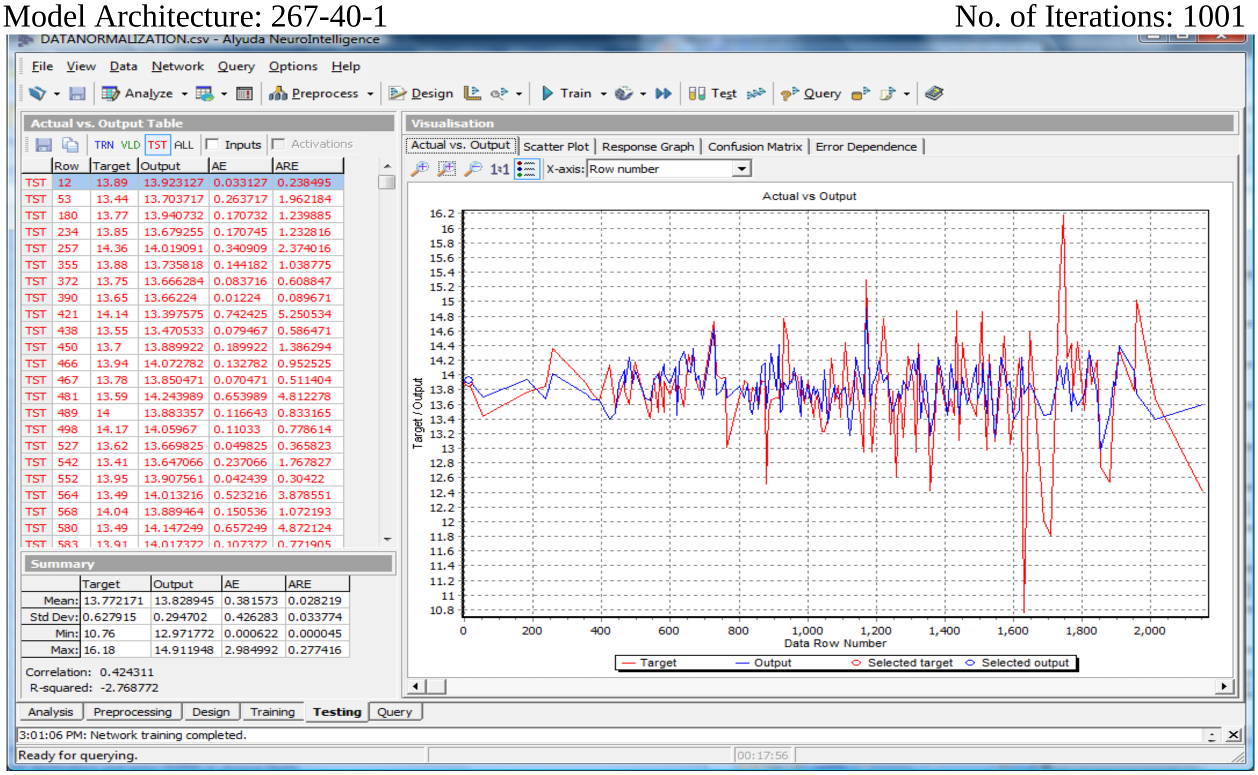


FIGURE 6. BXP- Batch Back Propagation: TESTING
Model Architecture: 267-40-1 No. of Iterations: 1001

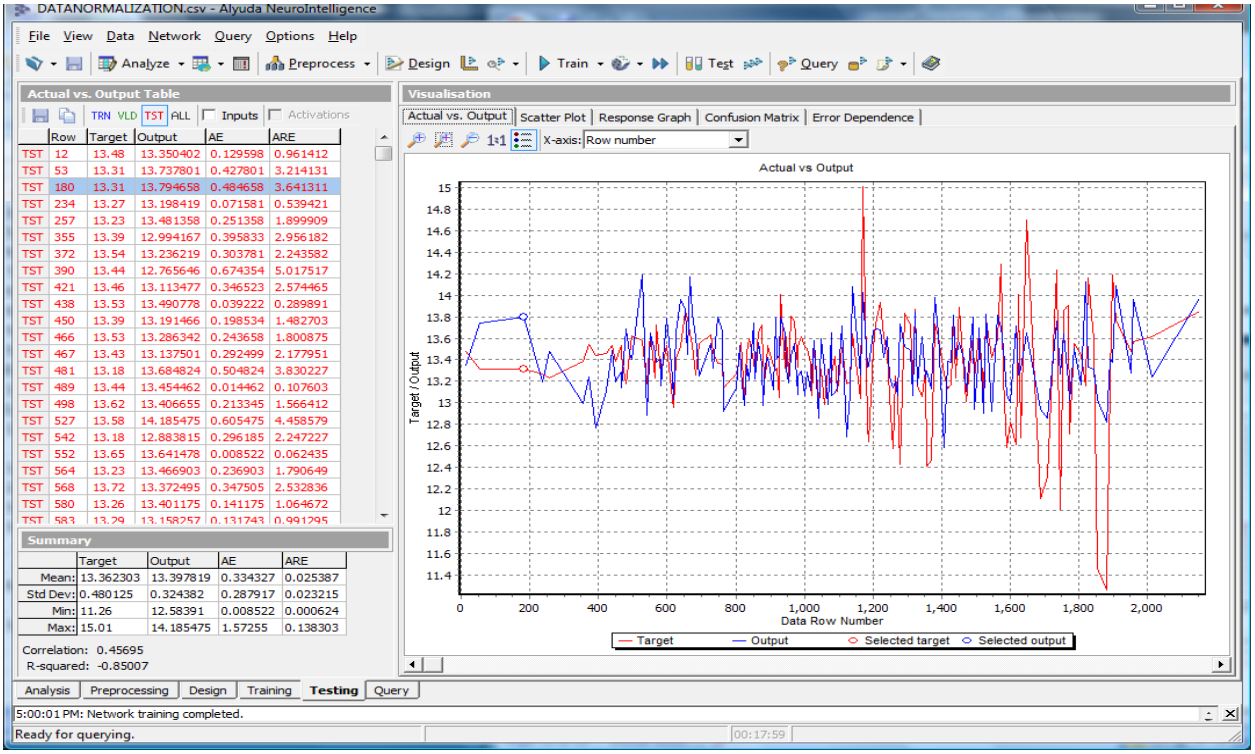
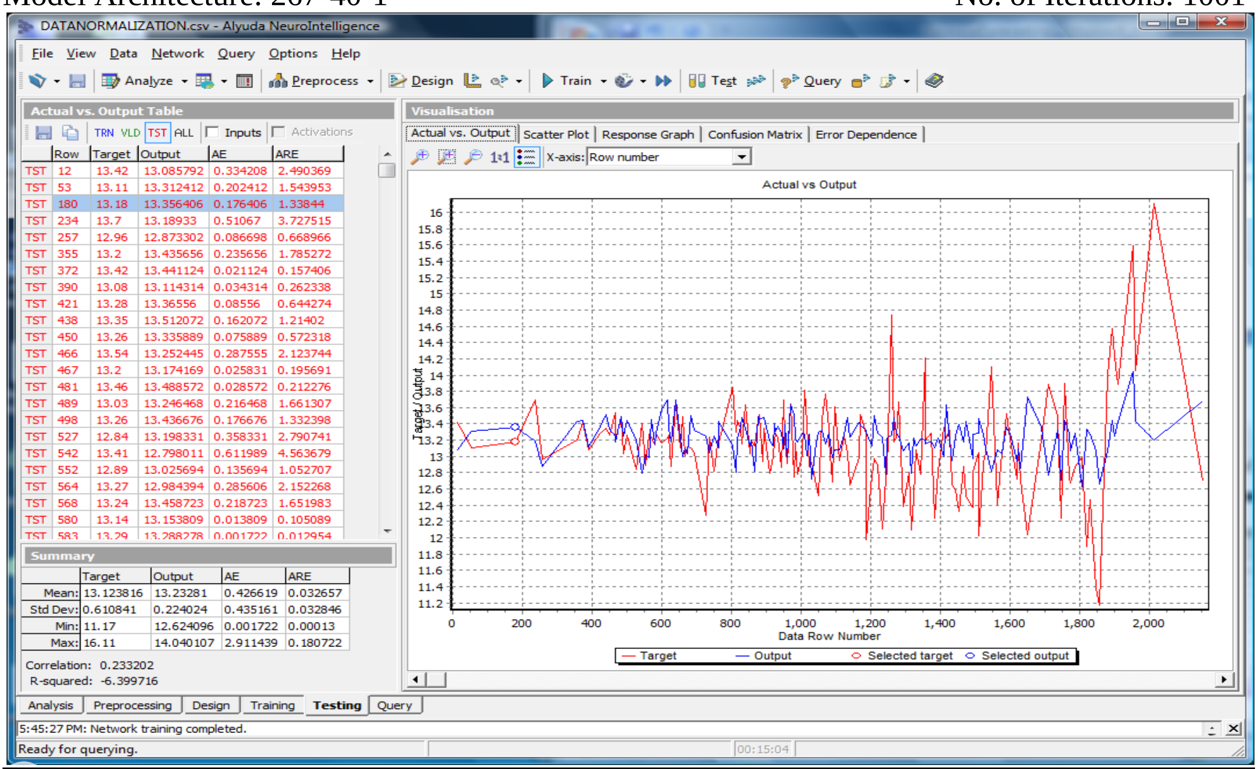


FIGURE 7. SPG- Batch Back Propagation: TESTING
Model Architecture: 267-40-1 No. of Iterations: 1001



IX. TABLES

TABLE 1. Macroeconomic indicators (World Indexes)
Source: YahooFinance

AEX	Amsterdam Price
CAC	Paris Price
DAX	German Index
FTSE	FTSE Index Price
SMI	Swiss Market Index
STI	Straits Times Index Singapore
IPC	Mexico Index
HSI	Hang Seng Index
BSE	Bombay Stock Exchange Sensex
BVSP	Bovespa-Brazillian Index
ATX	Vienna Stock Exchange
MERV	Merval Buenos Aires Index
KLSE	FTSE Bursa Malaysia Klci
TSEC	Taiwan Weighted Index
KOSPI	Kospi Composite Index
N225	Nikkei 225
JKSE	Jakarta Stock Exchange Index
TA	Tel Aviv 100

TABLE 2. Microeconomic indicators
Source: YahooFinance

BASIC MATERIALS	**COMPANIES**
1 Agricultural Chemicals	POTASH CP SASKATCHEWAN [POT]
2 Aluminum	ALCOA INC [AA]
3 Chemicals - Major Diversified	DOW CHEMICAL [DOW]
4 Copper	FREEPORT MCMORAN [FCX]
5 Gold	BARRICK GOLD [ABX]
6 Independent Oil & Gas	OCCIDENTAL PETROLEUM [OXY]
7 Industrial Metals & Minerals	BHP BILLITON [BHP]
8 Major Integrated Oil & Gas	EXXON MOBIL [XOM]
9 Nonmetallic Mineral Mining	HARRY WINSTON DIAMOND [HWD]
10 Oil & Gas Drilling & Exploration	TRANSOCEAN [RIG]
11 Oil & Gas Equipment & Services	SCHLUMBERGER [SLB]
12 Oil & Gas Pipelines	KINDER MORGAN ENERGY PARTNERS [KMP]
13 Oil & Gas Refining & Marketing	IMPERIAL OIL [IMO]
14 Silver	COEUR D' ALENE MINES COPR [CDE]
15 Specialty Chemicals	LUBRIZOL CORP [LZ]

16Steel & Iron	RIO TINTO PLC [RTP]
17 Synthetics	PRAXAIR INC [PX]
CONGLOMERATES	
18 Conglomerates	GENERAL ELECTRIC [GE]
CONSUMER GOODS	
19 Appliances	WHIRLPOOL CORP [WHR]
20 Auto Manufacturers - Major	HONDA MOTOR CO. LTD [HMC]
21 Auto Parts	JOHNSON CONTROLS INC [JCI]
22 Beverages - Brewers	FORMENTO ECONOMICO MEXICANO [FMX]
23 Beverages - Soft Drinks	THE COCA-COLA CO. [KO]
24 Beverages - Wineries & Distillers	DIAGEO PLC [DEO]
25 Business Equipment	XEROX CORP. [XRX]
26 Cigarettes	BRITISH AMERICAN TOBACCO PCL [BTI]
27 Cleaning Products	ECOLAB INC [ECL]
28 Confectioners	CADBURY PLC [CBY]
29 Dairy Products	LIFEWAY FOODS INC [LWAY]
30 Electronic Equipment	SONY CORPORATION [SNE]
31 Farm Products	ARCHER-DANIELS-MIDLAND [ADM]
32 Food - Major Diversified	HJ HEINZ CO. [HNZ]
33 Home Furnishings & Fixtures	FORTUNE BRANDS INC [FO]
34 Housewares & Accessories	NEWELL RUBBERMAID INC [NWL]
35 Meat Products	HORMEL FOODS CORP. [HRL]
36 Office Supplies	ENNIS INC. [EBF]

37 Packaging & Containers	OWENS-ILLINOIS [OI]
38 Paper & Paper Products	INTERNATIONAL PAPER CO. [IP]
39 Personal Products	PROCTER & GAMBLE CO. [PG]
40 Photographic Equipment & Supplies	EASTMAN KODAK [EK]
41 Processed & Packaged Goods	PEPSICO INC. [PEP]
42 Recreational Goods, Other	FOSSIL INC. [FOSL]
43 Recreational Vehicles	HARLEY-DAVIDSON INC. [HOG]
44 Rubber & Plastics	GOODYEAR TIRE & RUBBER CO. [GT]
45 Sporting Goods	CALLAWAY GOLF CO. [ELY]
46 Textile - Apparel Clothing	VF CORP. [VFC]
47 Textile - Apparel Footwear & Accessories	NIKE INC. [NKE]
48 Tobacco Products, Other	UNIVERSAL CORP. [UVV]
49 Toys & Games	MATTEL INC. [MAT]
50 Trucks & Other Vehicles	PACCAR INC. [PCAR]
FINANCIAL	
51 Accident & Health Insurance	AFLAC INC. [AFL]
52 Asset Management	T. ROWE PRICE GROUP INC. [TROW]
53 Closed-End Fund - Debt	ALLIANCE BERNSTEIN INCOME FUND INC. [ACG]
54 Closed-End Fund - Equity	DNP SELECT INCOME FUND INC. [DNP]
55 Closed-End Fund - Foreign	ABERDEEN ASIA-PACIFIC INCOME FUND INC. [FAX]
56 Credit Services	AMERICAN EXPRESS CO. [AXP]

57 Diversified Investments	MORGAN STANLEY [MS]
58 Foreign Money Center Banks	WESTPAC BANKING CORP [WBK]
59 Foreign Regional Banks	BANCOLOMBIA S.A. [CIB]
60 Insurance Brokers	MARSH & MCLENNAN [MMC]
61 Investment Brokerage - National	CHARLES SCHWAB CORP. [SCHW]
62 Investment Brokerage - Regional	JEFFERIES GROUP INC. [JEF]
63 Life Insurance	AXA [AXA]
64 Money Center Banks	JPMORGAN CHASE & CO. [JPM]
65 Mortgage Investment	ANALLY CAPITAL MANAGEMENT [NLY]
66 Property & Casualty Insurance	BERKSHIRE HATHAWAY [BRK-A]
67 Property Management	ICAHN ENTERPRISES, L.P. [IEP]
68 REIT - Diversified	PLUM CREEK TIMBER CO. INC. [PCL]
69 REIT - Healthcare Facilities	HCP INC. [HCP]
70 REIT - Hotel/Motel	HOST HOTELS & RESORTS INC. [HST]
71 REIT - Industrial	PUBLIC STORAGE [PSA]
72 REIT - Office	BOSTON PROPERTIES INC. [BXP]
73 REIT - Residential	EQUITY RESIDENTIAL [EQR]
74 REIT - Retail	SIMON PROPERTY GROUP INC. [SPG]
75 Real Estate Development	THE ST. JOE COMPANY [JOE]
76 Regional - Mid-Atlantic Banks	BB & T CORP. [BBT]
77 Regional - Midwest Banks	US BANCORP [USB]

78 Regional - Northeast Banks	STATE STREET CORP. [STT]
79 Regional - Pacific Banks	BANK OF HAWAII CORP. [BOH]
80 Regional - Southeast Banks	REGIONS FINANCIAL CORP. [RF]
81 Regional - Southwest Banks	COMMERCE BANCSHARES INC. [CBSH]
82 Savings & Loans	PEOPLE’S UNITED FINANCIAL INC. [PBCT]
83 Surety & Title Insurance	FIRST AMERICAN CORP. [FAF]
HEALTHCARE	
84 Biotechnology	AMGEN INC. [AMGN]
85 Diagnostic Substances	IDEXX LABORATORIES INC. [IDXX]
86 Drug Delivery	ELAN CORP. [ELN]
87 Drug Manufacturers - Major	JOHNSON & JOHNSON [JNJ]
88 Drug Manufacturers - Other	TEVA PHARMACEUTICAL INDUSTRIES LTD [TEVA]
89 Drug Related Products	PERRIGO CO. [PRGO]
90 Drugs - Generic	MYLAN INC. [MYL]
91 Health Care Plans	UNITEDHEALTH GROUP INC. [UNH]
92 Home Health Care	LINCARE HOLDINGS INC. [LNCR]
93 Hospitals	TENET HEALTHCARE CORP. [THC]
94 Long-Term Care Facilities	EMERITUS CORP. [ESC]
95 Medical Appliances & Equipment	MEDTRONIC INC. [MDT]
96 Medical Instruments & Supplies	BAXTER INTERNATIONAL INC. [BAX]
97 Medical Laboratories	QUEST DIAGNOSTICS

& Research	INC. [DGX]
98 Medical Practitioners	TRANSCEND SERVICES INC. [TRCR]
99 Specialized Health Services	DAVITA INC. [DVA]
INDUSTRIAL GOODS	
100 Aerospace/Defense - Major Diversified	BOEING CO. [BA]
101 Aerospace/Defense Products & Services	HONEYWELL INTERNATIONAL INC. [HON]
102 Cement	CRH PLC[CRH]
103 Diversified Machinery	ILLINOIS TOOL WORKS INC. [ITW]
104 Farm & Construction Machinery	CATERPILLAR INC. [CT]
105 General Building Materials	VULCAN MATERIALS CO. [VMC]
106 General Contractors	EMCOR GROUP INC. [EME]
107 Heavy Construction	MCDERMOTT INTERNATIONAL INC. [MDR]
108 Industrial Electrical Equipment	EATON CORPORATION [ETN]
109 Industrial Equipment & Components	EMERSON ELECTRIC CO. [EMR]
110 Lumber, Wood Production	WEYERHAEUSER CO. [WY]
111 Machine Tools & Accessories	STANLEY WORKS [SWK]
112 Manufactured Housing	SKYLINE CORP [SKY]
113 Metal Fabrication	PRECISION CASTPARTS CORP. [PCP]
114 Pollution & Treatment Controls	DONALDSON COMPANY INC. [DCI]
115 Residential Construction	NVR INC. [NVR]
116 Small Tools &	THE BLACK &
Accessories	DECKER CORP. [BDK]
117 Textile Industrial	MOHAWK INDUSTRIES INC. [MHK]
118 Waste Management	WASTE MANAGEMENT INC. [WM]
SERVICES	
119 Advertising Agencies	OMNICOM GROUP INC. [OMC]
120 Air Delivery & Freight Services	FEDEX CORP. [FDX]
121 Air Services, Other	BRISTOW GROUP INC. [BRS]
122 Apparel Stores	GAP INC. [GPS]
123 Auto Dealerships	CARMAX INC. [KMX]
124 Auto Parts Stores	AUTOZONE INC. [AZO]
125 Auto Parts Wholesale	GENUINE PARTS CO. [GPC]
126 Basic Materials Wholesale	AM CASTLE & CO. [CAS]
127 Broadcasting - Radio	SIRIUS XM RADIO INC. [SIRI]
128 Broadcasting - TV	ROGERS COMMUNICATIONS INC. [RCI]
129 Business Services	IRON MOUNTAIN INC. [IRM]
130 CATV Systems	COMCAST CORP. [CMCSA]
131 Catalog & Mail Order Houses	AMAZON.COM INC. [AMZN]
132 Computers Wholesale	INGRAM MICRO INC. [IM]
133 Consumer Services	MONRO MUFFLER BRAKE INC. [MNRO]
134 Department Stores	THE TJX COMPANIES INC. [TJX]
135 Discount, Variety Stores	WAL-MART STORES INC. [WMT]

136 Drug Stores	CVS CAREMARK CORP. [CVS]
137 Drugs Wholesale	MCKESSON CORP. [MCK]
138 Education & Training Services	DEVRY INC. [DV]
139 Electronics Stores	BEST BUY CO. INC. [BBY]
140 Electronics Wholesale	AVNET INC. [AVT]
141 Entertainment - Diversified	WALT DISNEY CO. [DIS]
142 Food Wholesale	SYSCO CORP. [SYY]
143 Gaming Activities	BALLY TECHNOLOGIES INC. [BYI]
144 General Entertainment	CARNIVAL CORP. [CCL]
145 Grocery Stores	KROGER CO. [KR]
146 Home Furnishing Stores	WILLIAMS-SONOMA INC. [WSM]
147 Home Improvement Stores	THE HOME DEPOT INC. [HD]
148 Industrial Equipment Wholesale	W.W. GRAINGER INC. [GWW]
149 Jewelry Stores	TIFFANY & CO. [TIF]
150 Lodging	STARWOOD HOTELS & RESORTS WORLDWIDE INC. [HOT]
151 Major Airlines	AMR CORP. [AMR]
152 Management Services	EXPRESS SCRIPTS INC. [ESRX]
153 Marketing Services	VALASSIS COMMUNICATIONS INC. [VCI]
154 Medical Equipment Wholesale	HENRY SCHEIN INC. [HSIC]
155 Movie Production, Theaters	MARVEL ENTERTAINMENT INC. [MVL]
156 Music & Video Stores	BLOCKBUSTER INC. [BBI]

157 Personal Services	H&R BLOCK INC. [HRB]
158 Publishing - Books	THE MCGRAW-HILL CO. INC. [MHP]
159 Publishing - Newspapers	WASHINGOTN POST CO. [WPO]
160 Publishing - Periodicals	MEREDITH CORP. [MDP]
161 Railroads	BURLINGTON NORTHERN SANTA FE CORP. [BNI]
162 Regional Airlines	SOUTHWEST AIRLINES CO. [LUV]
163 Rental & Leasing Services	RYDER SYSTEM INC. [R]
164 Research Services	PAREXEL INTL CORP. [PRXL]
165 Resorts & Casinos	MGM MIRAGE [MGM]
166 Restaurants	MCDONALD'S CORP. [MCD]
167 Security & Protection Services	GEO GROUP INC. [GEO]
168 Shipping	TIDEWATER INC. [TDW]
169 Specialty Eateries	STARBUCKS CORP. [SBUX]
170 Specialty Retail, Other	STAPLES INC. [SPLS]
171 Sporting Activities	SPEEDWAY MOTORSPORTS INC. [TRK]
172 Sporting Goods Stores	HIBBETT SPORTS INC. [HIBB]
173 Staffing & Outsourcing Services	PAYCHEX INC. [PAYX]
174 Technical Services	JACOBS ENGINEERING GROUP INC. [JEC]
175 Trucking	JB HUNT TRANSPORT SERVICES INC. [JBHT]
176 Wholesale, Other	VINA CONCHA Y TORO S.A. [VCO]
TECHNOLOGY	

177 Application Software	MICROSOFT CORP. [MSFT]
178 Business Software & Services	AUTOMATIC DATA PROCESSING INC. [ADP]
179 Communication Equipment	NOKIA CORP. [NOK]
180 Computer Based Systems	ADAPTEC INC. [ADPT]
181 Computer Peripherals	LEXMARK INTERNATIONAL INC. [LXK]
182 Data Storage Devices	EMC CORP. [EMC]
183 Diversified Communication Services	TELECOM ARGENTINA S A [TEO]
184 Diversified Computer Systems	INTERNATIONAL BUSINESS MACHINES CORP. [IBM]
185 Diversified Electronics	KYOCERA CORP. [KYO]
186 Healthcare Information Services	CERNER CORP. [CERN]
187 Information & Delivery Services	DUN & BRADSTREET CORP. [DNB]
188 Information Technology Services	COMPUTER SCIENCES CORPORATION [CSC]
189 Internet Information Providers	YAHOO! INC. [YHOO]
190 Internet Service Providers	EASYLINK SERVICES INTERNATIONAL CORP. [ESIC]
191 Internet Software & Services	CGI GROUP INC. [GIB]
192 Long Distance Carriers	TELEFONOS DE MEXICO, S.A.B. DE C.V. [TMX]
193 Multimedia & Graphics Software	ACTIVISION BLIZZARD INC. [ATVI]
194 Networking &	CISCO SYSTEMS INC.
Communication 19Devices	[CSCO]
195 Personal Computers	APPLE INC. [AAPL]
196 Printed Circuit Boards	FLEXTRONICS INTERNATIONAL LTD. [FLEX]
197 Processing Systems & Products	POLYCOM INC. [PLCM]
198 Scientific & Technical Instruments	THERMO FISHER SCIENTIFIC INC. [TMO]
199 Security Software & Services	SYMANTEC CORP. [SYMC]
200 Semiconductor - Broad Line	INTEL CORP. [INTC]
201 Semiconductor - Integrated Circuits	QUALCOMM INC. [QCOM]
202 Semiconductor - Specialized	XILINX INC. [XLNX]
203 Semiconductor Equipment & Materials	APPLIED MATERIALS INC. [AMAT]
204 Semiconductor- Memory Chips	MICRON TECHNOLOGY INC. [MU]
205 Technical & System Software	AUTODESK INC. [ADSK]
206 Telecom Services - Domestic	AT&T INC. [T]
207 Telecom Services - Foreign	NIPPON TELEGRAPH & TELEPHONE CORP. [NTT]
208 Wireless Communications	CHINA MOBILE LIMITED [CHL]
UTILITIES	
209 Diversified Utilities	EXELON CORP. [EXC]
210 Electric Utilities	SOUTHERN COMPANY [SO]
211 Foreign Utilities	ENERSIS S.A. [ENI]
212 Gas Utilities	TRANSCANADA CORP. [TRP]
213 Water Utilities	AQUA AMERICA INC. [WTR]

TABLE 3. Market Indicators
Source: YahooFinance

S&P	S&P 500's price changes
DJI	Dow Jones Industrial's price changes
DJT	Dow Jones Transportation's price changes
DJU	Dow Jones Utility's price changes

TABLE 4. Market Sentiment Indicators
Source: YahooFinance

VIX	CBOE Volatility Index changes
VXO	CBOE OEX Volatility Index

TABLE 5. Institutional Investor

BEN	FRANKLIN RESOURCES INC.

TABLE 6. Politics Indicators

Election	Presidential Election day	White House
Party	Party: Republican or Democratic	Wikipedia

TABLE 7. Business Cycles

Tech crash	Technological Crash 3/24/2000-10/9/2002
Current bear	Current bear 10/9/2007-11/16/2009

TABLE 8. Calendar anomalies

Mon	Monday
Wed	Wednesday
Thurs	Thursday
Fri	Friday
W1	First week
W3	Third week
W4	Fourth week
W5	Fifth week
Jan	January
Feb	February

Mar	March
Apr	April
May	May
June	June
July	July
August	August
Oct	October
Nov	November
Dec	December
pHoliday	Pre holiday