A TEMPORAL PATTERN APPROACH FOR PREDICTING WEEKLY FINANCIAL TIME SERIES

DAVID H. DIGGS
Department of Electrical and Computer Engineering
Marquette University
P.O. Box 1881,
Milwaukee, WI 53201-1881, USA
Email: david.diggs@marquette.edu
Ph: 414.288.6046

RICHARD J. POVINELLI
Department of Electrical and Computer Engineering
Marquette University
P.O. Box 1881,
Milwaukee, WI 53201-1881, USA
Email: richard.povinelli@marquette.edu
Ph: 414.288.7088

ABSTRACT

Discovering patterns and relationships in the stock market has been widely researched for many years. The goal of this work is to find hidden patterns within stock market price time series that may be exploited to yield greater than expected returns. A data mining approach provides the framework for this research. The data set is composed of weekly financial data for the stocks in two major stock indexes. Experiments are conducted using a technique designed to discover patterns within the data. Results show that these methods can outperform the market in longer time ranges with bull market conditions. Results include consideration of transaction costs.

INTRODUCTION

Data mining is the process of discovering hidden patterns in data. Due to the large size of databases, importance of information stored, and valuable information obtained, finding hidden patterns in data has become increasingly significant. The stock market provides an area in which large volumes of data are created and stored on a daily basis, and hence an ideal dataset for applying data mining techniques.

Statistical analysis has been widely used for many years to make predictions on the future values of a security price and study its behavior over time. Times series such as the stock market are often seen as non-stationary which present challenges in predicting future values. The focus of this research is in analyzing and predicting weekly financial time series. This work will show the advantage of using a weekly trading strategy, which is an extension of the daily trading strategy, to overcome the transaction cost associated with trading.

The proposed method is a data mining approach that uses time-delay embedding and temporal patterns to characterize events. The method is designed to analyze non-stationary time series and provides the basis for this work. The paper is broken into five sections, which describe the goal of this work, overview of the Time Series Data Mining method, financial applications, results, and research conclusions.

PROBLEM STATMENT

Previously, we have applied our Time Series Data Mining (TSDM) approach to making one-step daily price predictions (Povinelli, 2000). We have shown that a simple trading strategy based on these predictions can yield greater returns than expected by the efficient market hypothesis. However, these returns are adversely affected by transaction

costs. Hence, the goal of the current research is to develop an approach that increases returns and overcomes transaction costs. This is achieved by studying weekly stock price time series and making one-step weekly predictions.

Li and Tsang (2002) previously researched stock prediction using genetic programming in order to do financial forecasting. Results showed they were able to trade with better accuracy on long positions in indexes and individual stocks. Sauer (1994) also proposed a method similar to TSDM, in which time-delay embedding with interpolation and weighted regression were used to make time series predictions.

TIME SERIES DATA MINING

The proposed approach is based on TSDM (Povinelli, 2000). This method discovers hidden temporal structures predictive of sharp movements in price, using a time-delay embedding process that reconstructs the time series into a phase space that is topologically equivalent to the original system under certain assumptions. These assumptions are discussed in (Sauer et al., 1991). The TSDM technique was developed to make one-step predictions (Povinelli, 2000; Povinelli and Feng, 1998). The approach is used to make a prediction, which determines the buying and selling of stocks in a given time period. Figure 1 below shows two possible examples of the types patterns for which the TSDM method searches. The illustrated patterns are predictive of sharp increases in a time series.
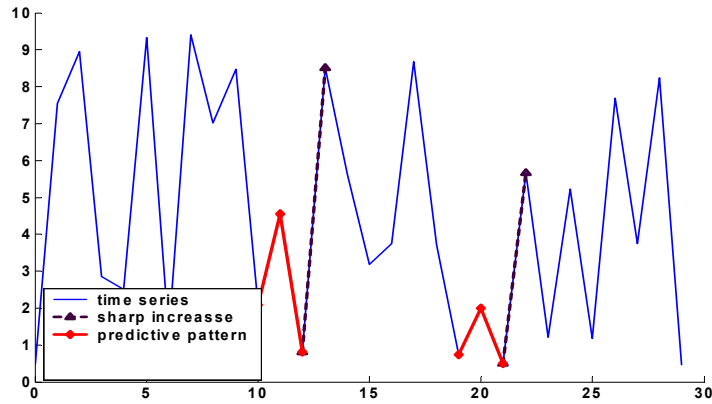

Figure 1 – Time series example

The patterns are determined by the three previous points to make a prediction for the next time step. The method involves clustering patterns, which are used to detect these sharp increases in the stock price. To find these temporal patterns the time series is embedded into a reconstructed phase space with a time delay of one and a dimension of three (Sauer et al., 1991). Once the data is embedded, temporal structures are located using a genetic algorithm search. Clusters are made of points within a fixed distance of the temporal structures. A percent change function

$$g(t) = (x_{t+1} - x_t)/x_t \tag{1}$$

determines the value given to the prediction made from the clustering using the temporal structures. This value is the percent change in the security price for the next week. The

temporal structures are next ordered by how well each predicts the stock price movements. A ranking function is defined as the average value within a temporal structure, and it is used to order the structures for optimization. The optimization is a search to find the best temporal structures and is done with a simple genetic algorithm (sGA) that finds fitness value parameters that maximize the ranking function $f(P)$. The sGA uses a combination of Monte Carlo search for population initialization with roulette selection and locus crossover to find $P*$, and a criteria of fitness values halt the genetic algorithm. See reference (Povinelli, 2000) for more details on the algorithm.

APPLICATION TO FINACIAL TIME SERIES

The stock market is a platform for millions of investors to interact through the buying and selling of securities on various equity markets, such as the NASDAQ, AMEX, and NYSE. The goal is to use the TSDM method to achieve small and unexpected returns that are greater than transaction costs associated with trading. The data is in the form of weekly stock price data comprised of the Dow Jones 30 and NASDAQ 100 component stocks.

The method is applied to the stocks of a given market index. Weekly buy or do nothing signals are generated by the TSDM method for each stock in the index. The TSDM returns are compared the index as a benchmark. Transaction costs are also considered to show the adverse effects of actually making these weekly trades. Transaction costs are calculated based on the number of stocks in the weekly portfolio selected by the TSDM model. A transaction cost of $20 is used for each buy and sell transaction.

METHOD RESULTS

Four experiments with various time ranges were run on each data set. Training the data was done using the TSDM method to find predictive structures. Testing for each experiment used a weekly trading strategy for buying and selling a set of stocks that are output from our model. Each week a new set is traded based on the model predictions. The time ranges for each data set were current year (1/01/2003-5/01/2003), previous year (1/01/2002-1/01/2003), previous 5 years (1/01/1998-1/01/2003), and previous 10 years (1/01/1993-1/01/2003). The training period for the four experiments was 25 weeks because previous test experiments with this training period gave the best results. The genetic algorithm parameters were the same for each experiment, consisting of population size 30, and fitness convergence value of 0.9, which is used as a stopping criterion.

The total portfolio values are set to $10,000 and $100,000 to calculate transaction costs and returns for the model. The weekly geometric average rate of return is calculated for the model and both indexes for each experiment, shown below in Table 1. The average rate of return for the model is based on the returns calculated from each stock in the portfolio, which are bought and sold equally on a weekly basis. The table also shows the adjusted return with transaction costs and the average transaction cost per week. Figures 2 through 5 plot the previous year and previous five years of TSDM model weekly returns and adjusted model returns based on a $10,000 portfolio value. The figures plot model and adjusted returns for both the Dow Jones 30 and NASDAQ 100 data sets along with a comparison against the index benchmarks.

Table – 1 Weekly Returns Comparison

Rate of Return	TSDM Model	TSDM Model With Transaction Costs ($10,000)	TSDM Model With Transaction Costs ($100,000)	Index
Dow Jones 30 ytd	-0.119 %	-0.126 %	-0.120 %	-0.000768 %
Dow Jones 30 1 Year	-0.0098 %	-0.017 %	-0.011 %	-0.000127 %
Dow Jones 30 5 Years	0.730 %	0.720 %	0.729 %	0.0000616 %
Dow Jones 30 10 Years	1.440 %	1.429 %	1.439 %	0.00000324 %
NASDAQ 100 ytd	-0.172 %	-0.193 %	-0.174 %	0.00586 %
NASDAQ 100 1 Year	-0.017 %	-0.042 %	-0.019 %	-0.00887 %
NASDAQ 100 5 Years	1.018 %	0.989 %	1.015 %	-0.00011 %
NASDAQ 100 10 Years	1.873 %	1.845 %	1.871 %	0.00201 %
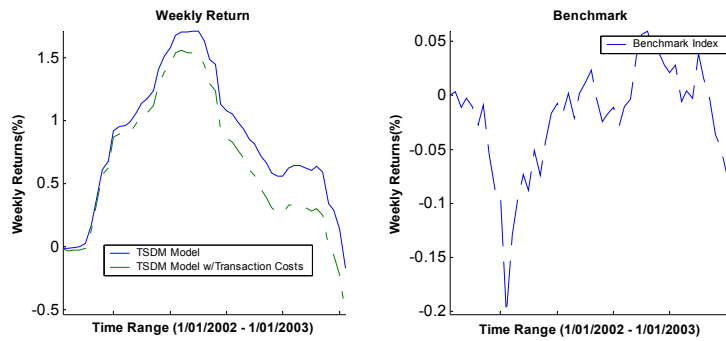
Figure 2 - Dow Jones 30 Experiment 2
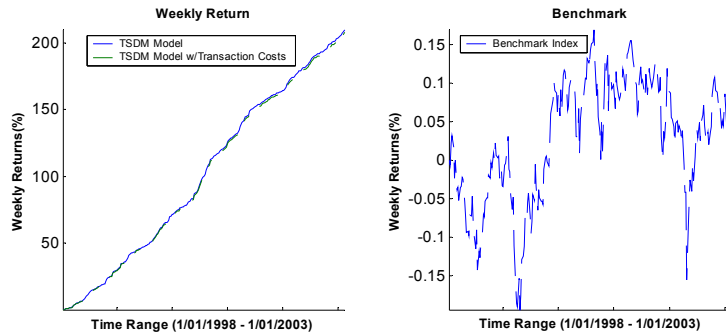
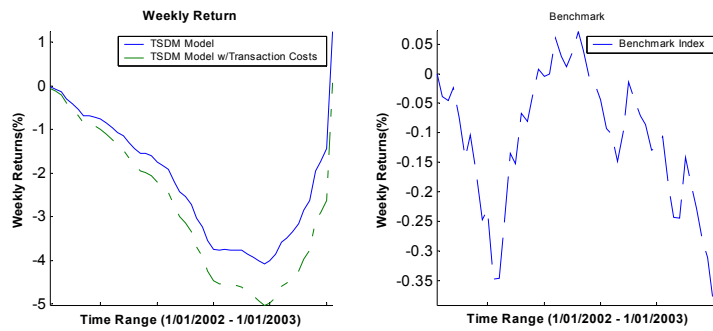**Figure - 3 Dow Jones 30 Experiment 3**
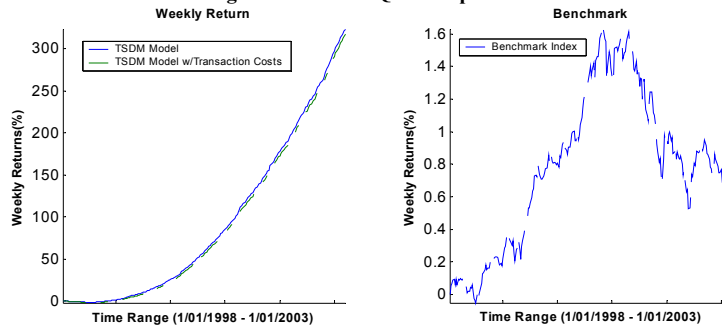
Figure - 4 NASDAQ 100 Experiment 2

Figure - 5 NASDAQ 100 Experiment 3

CONCLUSIONS

Our data mining approach combined with a weekly trading strategy is used to overcome trading cost and compared against market index used as benchmarks. The benchmarks used in comparing results were two major stock market indexes, the Dow Jones 30 and NASDAQ 100. The model produced positive returns in the five and ten year experiments when transaction costs were both taken into account and ignored. When transaction costs were calculated, the results from the NASDAQ experiments were more affected due to larger portfolio sizes than the Dow Jones experiments. The different time ranges enabled us to see how are model reacted to the changes in the market setting. The first two experiments were set directly in the recent bear market, which has taken both major indexes down significantly over the past two years. Our model was unable to do better than the market due to a lack of diversity in the weekly portfolios. In the five and ten year time ranges our model was able to overcome transaction costs and outperform the market, showing a strong predictive stock selection process. In comparison to the Dow Jones experiments, the NASDAQ produced better results due to the larger selection of stocks to choose from to create better weekly portfolios With this trading strategy, a larger portfolio value with the ability to purchase more shares reduces the adverse impact of trading costs on the returns. As shown in the Table 1 the difference between the model return and the adjusted return decreases with a higher portfolio value.

This work leads itself into an optimization problem for future work. The optimization is to be done in the form of portfolio management that will optimize the weekly portfolio predictions made using the TSDM method. The future directions of this

work will include the use of Markowitz Portfolio Theory and the Capitol Asset Pricing Model (CAPM) (Reilly and Brown, 1997) to find optimal portfolios while meeting specific goals such as maximizing return and minimizing risk. This addition focuses on increasing the returns of the TSDM method over time by altering the weights of the weekly portfolios. In addition to portfolio management techniques the transaction cost model can be modified to include the bid-ask spread.

ACKNOWLEDGEMENTS

REFERENCES

Li, J., and Tsang, E. P. K., 2002, "Eddie for Financial Forecasting," Genetic Algorithms and Genetic Programming in Computational Finance,, S.-H. Chen, ed., Kluwer Academic Publishers, Norwell, Massachusetts, 161 -174.

Povinelli, R. J., 2000, "Identifying Temporal Patterns for Characterization and Prediction of Financial Time Series Events," *Temporal, Spatial and Spatio-Temporal Data Mining: First International Workshop; TSDM2000*, Lyon, France, 46-61.

Povinelli, R. J., and Feng, X., 1998, "Temporal Pattern Identification of Time Series Data Using Pattern Wavelets and Genetic Algorithms," *Artificial Neural Networks in Engineering*, St. Louis, Missouri, 691-696.

Reilly, F. K., and Brown, K. C., 1997, *Investment Analysis and Portfolio Management*, Dryden Press, Fort Worth, Texas.

Sauer, T., 1994, "Time Series Prediction Using Delay Coordinate Embedding," Time Series Prediction: Forecasting the Future and Understanding the Past, A. S. Weigend and N. A. Gershenfeld, eds., Addison-Wesley, 175-194.

Sauer, T., Yorke, J. A., and Casdagli, M., 1991, "Embedology," *Journal of Statistical Physics*, Vol. 65, 579-616.