Computer Science Journal of Moldova, vol.**19**, no.2(56), 2011

Prediction Models of Financial Markets Based on Multiregression Algorithms

Antoni Wiliński

Abstract

The paper presents the results of simulations performed for predictive goals for the main Polish index named WIG20, using the historical quotes on several connected financial time series. The data (monthly and daily tested) used to predict WIG20 are such series as economical supply of money, level of unemployment, inflation and lagged series of the main index. In order to reach prediction goal, the author's algorithms were used. These algorithms are the hybrid of two methods – simple rules and multiregression prediction. The results reveal some interesting features of regression models, indicating the prospect of further applications of the method, especially in Internet area. The main hypothesis is that markets have a short term memory which allows to create different strategies.

Keywords: multivariate regression, simple rules, investigation strategy, prediction, trading systems, algotrading.

1 Introduction

Investment strategy based on so called simple rules and regression prediction has a long story, as well in theory, as in real trade on financial markets [1, 2, 11]. Today, most of scientists, especially those whose opinions agree with Eugene Fama's theory of efficient markets [3], claim that looking for a successful investment strategy based on regression is a task with no chances for success. The author of this paper limits the search space only to technical analysis, which, according to Fama's theory followers, makes this task totally hopeless.

However, there is a fundamental difference between efficiency of prediction and investment strategy goals [4]. The aim of prediction is to predict a change (sometimes only the direction of a change), and the aim of investment strategy is to achieve an investment success. These are entirely different categories. Of course, an efficient prognosis can be the natural base for a simple and quite successful strategy. Also, we can say that an inefficient prognosis does not eliminate chances for building a good, often very complex, strategy in which a prognostic model is only a small component of the whole model of making an investment decision. In the paper just this approach is considered. For some different sets of data and some regression models the results of prediction are used as condition for next steps of algorithms.

The term multiple prediction refers to such a prediction process, where the dependent variable is chosen in a special way from the set of all variables observed. Usually it is main goal of an investigation – the main index, the profit or the returns from market. Analogically to the concept of multiple regression [2, 5], the principle of expressing one variable in terms of the others will be applied [5]. Typically it is predetermined what is the outcome of a given experiment (in this case – the main Polish index WIG20) and what (which variables set) needs to be observed in order to explore the statistical relationship between the input and the output of the observed system [1, 6].

WIG20 is the most important Polish financial instrument - it is a price index covering 20 largest and most liquid Warsaw Stock Exchange companies. It is analyzed by many authors from academic field, especially in the context of predicting its changes [4, 6, 7].

The term simple rules in investigation strategy is a set of special parameters which are used to prepare and perform an invest decision [8, 9, 10]. Usually these are parameters connected to technological restriction of specific broker platform e.g. *Stop Loss, Take Profit, spread* and others.

On financial markets, when focusing on searching for a certain value (usefulness, profit) in a given data set, it is irrelevant which of the independent variables are actually used for modelling the outcome. In our case the main Polish index was defined as a result of prediction and

some series were chosen to build models allowing for gaining a profit in any possible way.

The paper is organized as follows. In the first part a data used in investigation will be presented. Next, the base algorithm and three different strategies based on different input sets will be described. The profits for all models will be compared.

2 Data characteristics

There were prepared three sets of data for the presented simulation models. The first one was extracted from typical long term time series.

For verifying the strategy mentioned above, two associated time series were chosen: a time series of unemployment rate in Poland and a time series of money supply in Poland. Data for both time series were gathered with the same (monthly) frequency as in case of WIG20 and concerned the same 12-years period [4]. Hence, the number of terms in all three time series was the same. Additionally, the matrix of input data was enlarged by introducing two more time series, created by shifting (lagging) the original time series of WIG20 one and two terms (months) back. Hence, the size of the input matrix was equal to $(i_k - 2) \times 4$, where i_k – number of terms in the original matrix ($i_k = 145$). Let us denote the final matrix of input data as X_{ij}. Each row of this matrix is composed of four terms $X_i = [X_{i1}X_{i2}X_{i3}X_{i4}]$, where:

- x_{i1}, $i = 1, 2, ...i_k - 2$ – time series created by shifting the original time series of WIG20 two terms (months) back;
- x_{i2}, $i = 2, 3...i_k - 1$ – time series created by shifting the original time series of WIG20 one term back;
- x_{i3}, $i = 3, 4...i_k$ – time series of unemployment rate in Poland;
- x_{i4}, $i = 3, 4, ...i_k$ – time series of money supply in Poland;
- y_i, $i = 3, 4, ...i_k$ – time series of output variable WIG20, composed of values x_{i1}, $i = 3, 4, ...i_k$.

The given set of attributes describing the output variable y_i (WIG20) was a result of a few circumstances. Firstly, data describing these attributes were easily available for the author [4] and it was appealing to check whether they can be the source of regression rules suitable for prediction. Secondly, intuitively, these attributes seemed to be responsible for changes in the analyzed index. The intuition often misleads the economists but it is always worth to try new ideas, especially in terms of optimization of some parameters describing time series such as: value of shifting, levels of barriers used for automatic closing of losing positions, structure of regression models etc. The author analyzed also other matrixes of input data, however they were not so interesting (in terms of the overall profits) as the matrix described in the paper [4].

The second set of data was built on quite another basis. It was a different period under consideration, also monthly tested, but with a wider spectrum of factors. In the second case some new economical series were added. In this set were used: full candlestick data of WIG20 (named OHLC – Open, High, Low, Close), also volume of transaction, and new series of economical factors such as: level of national growing, interest rate of national currency, inflation, and also as previously - level of unemployment and supply of money. Some of the series were modified with one or two steps shifted back (lagged series).

For the first data set the horizon of investigated period was about 12 years (1995-2007). The second data set covered monthly sampled date of period 2003-2011. For the third set it was about 4 years up to now.

Quite another structure was built for the third experiment. It was a set created from one day candlesticks of WIG20. In this case also lagged series and volume of day trading were used.

3 Base algorithm

The base algorithm is a hybrid of two different approaches to trading strategies. The first component is connected to so called simple trading rules [4, 8, 10]. The parameters which were used by the author were –

Stop Loss barrier (SL), barrier b (after crossing of which the new action was performed), number n candlesticks back which was considered in process to determine a predict action.

The general description of the algorithm in some steps:

1. The initial values of three main parameters SL, n, b were set.

2. For the set of M independent variables and for n steps back a multiregression model based on Penrose pseudoinversion [4, 11] was created.

3. As a result of the model (step 2) a one step prediction was determined.

4. Based on the prediction and parameter b a strategy was performed:

   - when $\hat{x}_{i+1} \leq x_i + b \wedge \hat{x}_{i+1} > x_i$, a short position is opened;
   - when $\hat{x}_{i+1} > x_i + b$, a long position is opened;
   - when $\hat{x}_{i+1} \leq x_i - b$, a short position is opened;
   - when $\hat{x}_{i+1} > x_i - b \wedge \hat{x}_{i+1} < x_i$, a long position is opened;
     for $i = m + 1, ..., i_k$

5. Inside the next step (candlestick) the crossing of the values *High* and *Low* was tested.

6. If the test from step 5 was positive then the result for the current step was equal to SL.

7. The result (return) after one step was compared with all previous ones. If the result was the best, then the new optimal parameters were fixed.

8. If loops for parameters from step 1 were not finished we went to step 1.

9. The optimal parameters SL, n, b were used to perform the next step. It was the step made without any information from future, so it was the step in future. The return after this step was the main result of the algorithm.

4 Experiments

Basing on the algorithm described above and the three sets of data, three different simulation experiments were considered.

The first one was completed with the first set of data and with constant values of the three main parameters (step 8 of the main algorithm was not repeated).

The results we can see in Fig. 1. The values of the return are determined in normalized WIG20, so the final result equal 1,43 is a mean profit compared to 143% of maximum of WIG20 during about 12 years. Calmar ratio for the cumulated profit showed in Fig. 1 is 5,38.

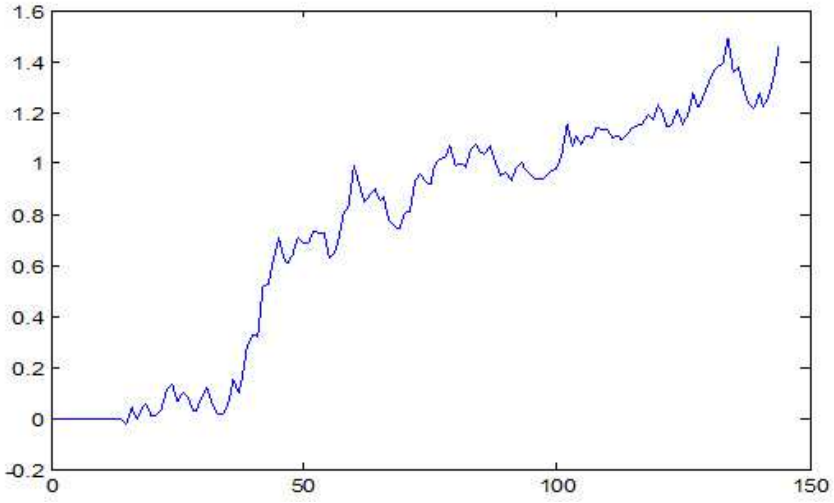

Figure 1. Cumulated profit for the first strategy. 130 investigation decisions in monthly sampled time series of WIG20.

The indicator means the final result of the profit divided by maximum drawdown.

The second experiment was performed with the second set of data and with a modification of the main algorithm. After each 9^{th} step in the main loop of the above algorithm an investment decision based on a prediction rule (step 4th) was made. The result of a cumulated profit is shown in Fig. 2. Calmar ratio for this simulation is rather bad – here the result of 1.95 was obtained.

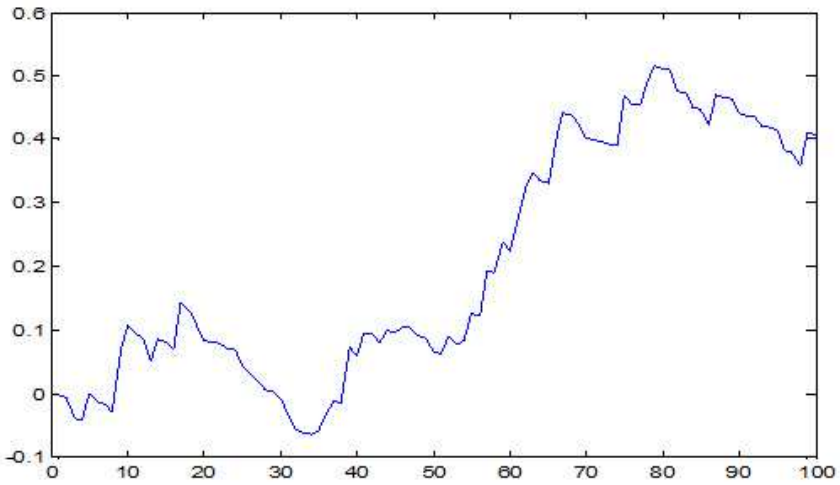

Figure 2. Cumulated profit for experiment with the second set of data (monthly tested economical factors and candlesticks of WIG20).

The third experiment was performed for quite another data. There were used only day tested candlesticks of WIG20 with lagged series, as described above. First, the simulation was made for 100 days only. The result of cumulative profit is shown in Fig. 3 and it is not too interesting. The Calmar ratio is 1,54 for this experiment and is the worst. But for 1000 steps (days) the result is excellent – Fig. 4.

For the horizon of 1000 days we achieve Calmar ratio very high, equal to 24.21.

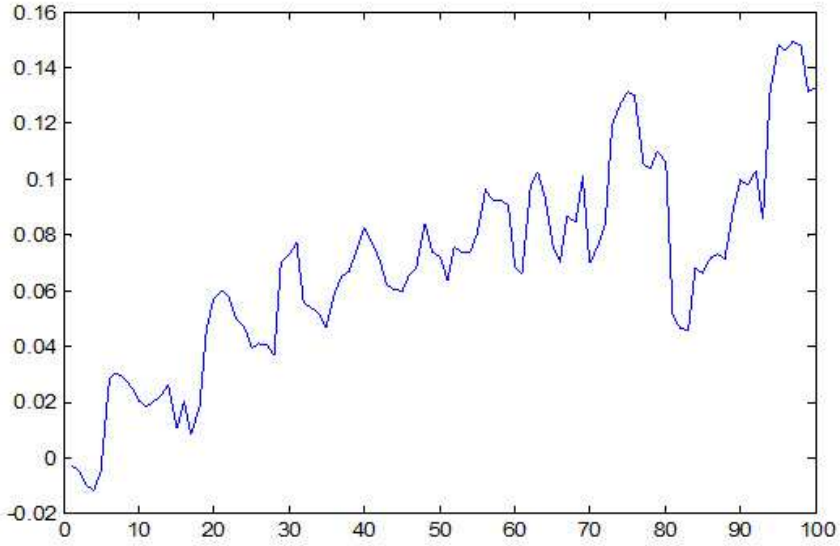

Figure 3. Cumulated profit for experiment with the third set of data performed for 100 daily steps.

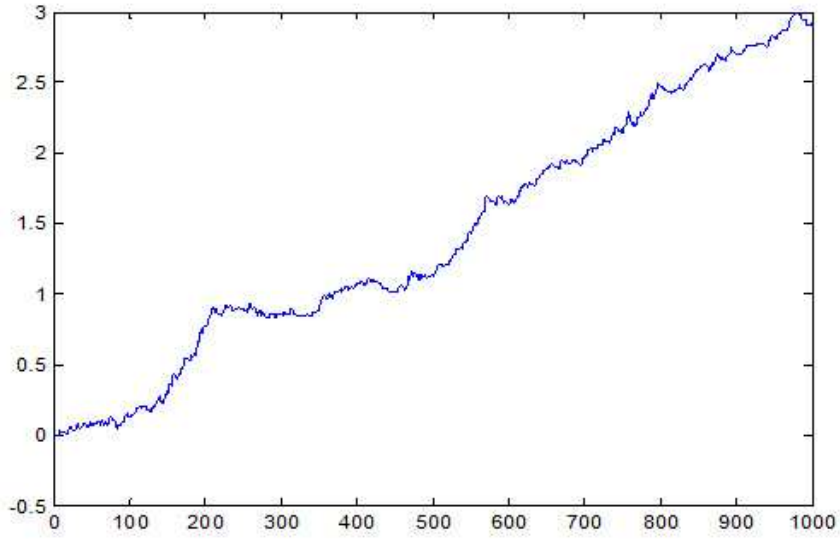

Figure 4. Cumulated profit for experiment with the third set of data performed for 1000 daily steps.

5 Conclusions

The above presented results of some simulations of author's algorithm were performed in Matlab environment and were implemented by the author. The main conclusion is that they are rather very optimistic results and they are in opposition to the efficient market hypothesis. As we can see it is the reason to test if markets have a kind of short term memory. All experiments were performed under conditions when no future information was available and every next decision was made with new adaptive parameters.

In comparison with typical Buy and Hold strategy, the presented results are without any doubts much better. We can estimate the results of Buy and Hold strategy by looking in Fig. 5.

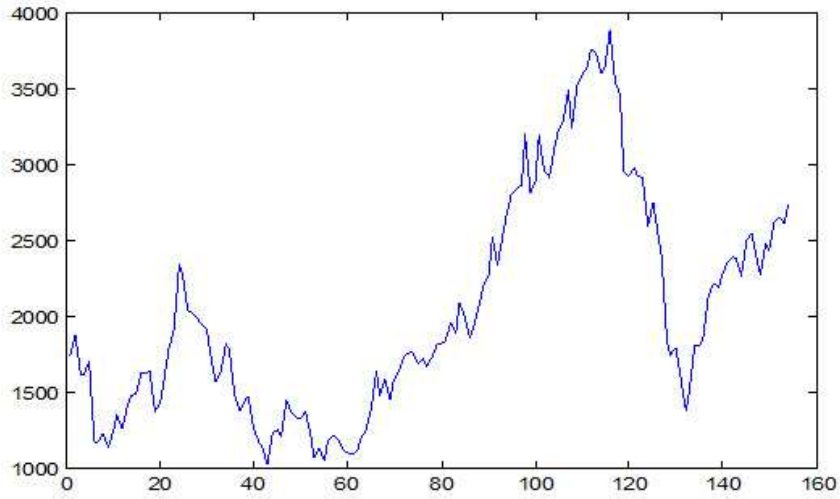

Figure 5. Main index of Polish stock exchange WIG20 in about 12 years horizon.

The presented results should be considered as a chance for next generation automated trading systems. All of them need very often adaptation to fast changes on markets – in this case it was the main Polish index. Of course, regarding to main principles of computational

intelligence, we can and should try to complete experiments like these with another markets. Before calculation we can not resign from the chance.

References

[1] N.R. Draper, H. Smith. *Applied regression analysis.* Wiley. New York, 1981.

[2] C.M. Queen, J.Q. Smith. *Multiregression Dynamic Models.* Journal of Royal Statistical Society, vol. 55, no 4 (1993), pp. 849–870.

[3] E. Fama. *Efficient capital markets.* Journal of Financial Economics, vol. 11, (1991), pp. 1575–1617.

[4] A. Wiliński, J. Rejer. *The Management of Prediction Method in the System of Investment Decision Making.* Management and Production Engineering Review, vol. 1 no.2 (2010), pp. 10–20.

[5] A. Wiliński, P. Czapiewski. *GMDH-based Multiple Prediction of Financial Markets.* Polish Journal of Environmental Studies, vol. 17, no 3b (2008).

[6] D. Żebrowska–Suchodolska, D. Witkowska. *Momentum and winner-loser strategies: Evidence for the Warsaw Stock Exchange.* Polish Journal of Environmental Studies, vol.18, no 5b (2009), pp. 220–227.

[7] K. Kompa, A. Matuszewska–Janica. *Examination of Warsaw Stock Exchange indexes gehaviour: Applications of rolling windows variance ratio test.* Polish Journal of Environmental Studies, vol.17, no 3b (2008), pp. 150–154.

[8] B. LeBaron. *Technical trading rules and regime shifts in foreign exchange intervention.* Journal of International Economics, vol. 49 (1999), pp. 125–143.

[9] B.M. Cai, C.X. Cai, K. Keasey. *Market efficiency and returns to simple technical trading rules: further evidence form US, UK, Asian and Chinese Stock Markets.* Asia-Pacific Financial Markets, Springer, (2005), pp. 45–60.

[10] W. Brock, J. Lakonishok, B. LeBaron. *Simple technical trading rules and stochastic properties of stock returns.* Journal of Finance, vol. 47 (1992), pp. 1731–1764.

[11] A.G. Ivakhnenko, J.P. Zaicenko, V.D. Dimitrov. *Priniatije resenij na osnove samoorganizacji.* Sovetskoe Radio, Moskwa, 1976.

Antoni Wiliński

Received May 28, 2011

West Pomeranian University of Technology of Szczecin,
Faculty of Computer Science and Information Technology
ul. Żonierska 52, 71-210 Szczecin, Poland
Phone: (+4891) 449-56-60
E–mail: *awilinski@wi.zut.edu.pl*