2004 ON WARSAW STOCK EXCHANGE VIA ZIPF ANALYSIS, SCATTER AND LAG PLOTS*

Wiktor Bachnik, Piotr Chomiuk, Szymon Fałtynowicz
Mariusz Gawin, Waldemar Gorajek, Jacek Kedzierski
Karol Kosk, Arkadiusz Kucharczyk, Piotr Leszczyński
Rafał Podsiadło

Students of Computer Science, Gdańsk University
Wita Stwosza 57, 80-952 Gdańsk, Poland

and Danuta Makowiec

Institute of Theoretical Physics and Astrophysics, Gdańsk University
Wita Stwosza 57, 80-952 Gdańsk, Poland

(Received April 19, 2005)

This paper presents the last year on Warsaw Stock Exchange (WSE) and world stock exchanges by graphical analysis: Scatter Plot, Zipf Analysis and Lag Plot of selected Polish (WIG, WIG20, WIG-BANKI, TECHWIG) and foreign (NIKKEI, DOW JONES Industrial Average) indices, and also selected companies listed on WSE. Zipf analysis proves that although, generally, holding securities was the best way to earn money in the last year, however, Zipf based strategy also could be profitable. Scatter Plots show no similarities between Polish and foreign indices, however, behaviour of Polish ones is similar. The volatility of indices and most companies was highest on Monday and lowest on Friday. Distribution of returns in continuous trading is neither Gaussian nor uniform.

PACS numbers: 05.40.–a, 05.45.Tp, 89.65.Gh

1. Introduction

Warsaw Stock Exchange have existed since the 16th of April 1991, but 2004 was one of the most prosperous years in its history. The stock exchange, along with a few other stocks in the world, acts rather abnormally, especially in comparison with the biggest world's stock exchanges like Wall Street or Tokyo SE. While many indices indicated high recession in the last four years

* Presented at the First Polish Symposium on Econo- and Sociophysics, Warsaw, Poland, November 19–20, 2004.

and only recently became more successive, WSE values of indices during this period, and especially last year, were still going up.

The most important facts and statistics describing changes of the Warsaw Stock Exchange in 2004 are as follows:

1) 2004 was the year of the highest session turnover of stocks. After last session on the 31st of December total turnover value for 2004 increased to 109.8 billion PLN (previous best 103.7 billion PLN in 2000) [1].

2) At the session on the 31st of December the value of WIG index reached 26 636.19 points and was the highest in history of WSE. WIG20 reached 1 960.573 points (best since the 4th of September 2000).

3) At the end of the year 151 (78%) among all listed companies ended with higher quotations than at the end of the previous year (record-holders are: Vistula which stocks price went up 508%, Tim — 425% and Wólczanka — 378%). Among companies which did not come up to investors' expectations was Prokom with 15% loss [2].

4) 2004 was the year of significant increase in capitalisation of the stock exchange. On the 31st of December capitalisation amounted to 214.3 billion PLN (only national companies). In relation to the last session of 2003 it meant a 53% increase.

5) During the year 36 new companies made their debuts on the WSE and the total number of listed companies increased from 203 to 230 (9 were de-listed). One of new companies was PKO BP which made its debut on the 10th of November. Value of turnover during trading session amounted to 3 272.7 million PLN and it was all-time high (previous reached on the 9th of February 2000 amounted to 1 110.8 million PLN).

6) Data for the first semester indicates that for the first time in the last 4 years, participation of individual investors in turnovers increased very significantly (from 29% in second semester of 2003 to 38% in first semester of 2004).

In this paper the last year on WSE and world stock exchanges will be presented with the use of some graphical analysis (Scatter Plot, Zipf analysis) of selected Polish and foreign indices and some companies listed on WSE with conclusions coming from graphs analysing. Time series data comes from web-sites `bossa.pl` [3] and `www.interia.pl` [4]. Zipf analysis gives answer to the question whether it is possible to predict return in next time step using the knowledge about changes in few previous steps. Scatter Plot shows volatility of values of the indices and companies stock prices in each weekday. It also enables one to find similarities in behaviour among indices and comparing securities prices with corresponding indices. Lag Plots are also presented for continuous quotations to qualify the distribution of data sets.

2. Zipf analysis

Zipf analysis is useful to study changes of stock prices or values of stock indices [5]. In this work some financial data series were used and translated into sequence of characters. Each subsequent change in price

$$r(t) = P(t) - P(t-1)$$

was replaced by a following character: 'v' when $r(t) < 0$, '^' when $r(t) > 0$, '–' when $r(t) = 0$. From this sequence of characters all subsequences made of n successive characters were taken and words were created. Then number of occurrences for each word w was measured and divided by a total number of words to obtain probability for each one of fixed length n, $p(w)$. Probabilities for single character words $p(\text{'v'})$, $p(\text{'}\wedge\text{'})$ and $p(\text{'–'})$ were calculated.

In the next step normalisation of results was performed following procedure described in [5]. The last step was sorting words by normalised values.

Graphs in Fig. 1 present results of Zipf analysis of some stock indices at successive closing times. Studied words are made of 5 characters.

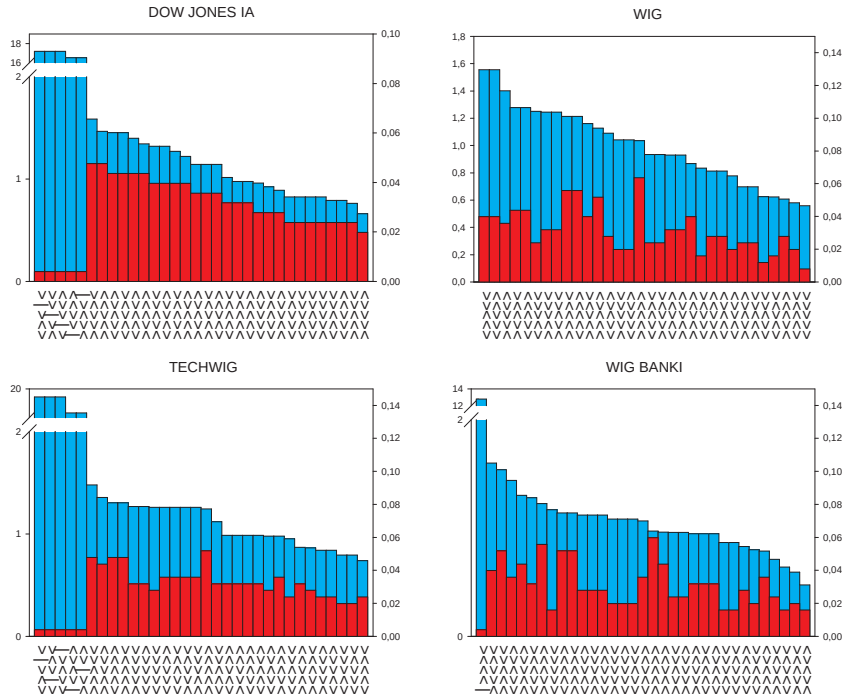

Fig. 1. Zipf analysis of WIG, TECHWIG, WIG BANKI and DOW JONES IA. Words under each graph should be read from top to bottom. The grey bars (cyan colour online) represent normalised values (given on the left axis) for each word and the black bars (red colour online) — probabilities (values on the right axis).

Graphs in Fig. 2 and 3 show Zipf analysis of some stock companies (prices at successive closing times). The length of word is decreased to 4 due to obtained words diversity.

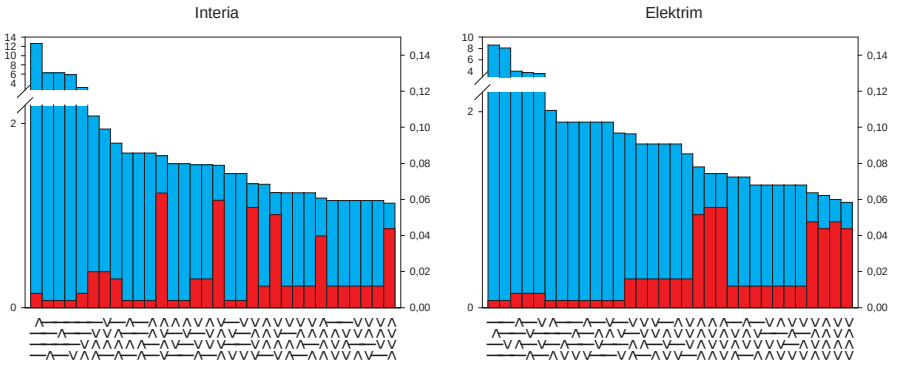

Fig. 2. Zipf analysis of stock companies: Interia, Elektrim.

If the results of Zipf analysis were known at the beginning of the year, it would have been possible to make profit using the following strategy:

1. Pairs of words which differ at the last character, and the probability that one word is significantly greater than other one, are selected. Words chosen for simulation are available at [8]. In this research the difference in probabilities is fixed to at least 75% for companies and at least 66% for indices.

2. If a sequence of $n-1$ characters occurs then, depending on the last character in word, the stock should be bought (for the '$\wedge$' character) or sold (for 'v' character) and on the next day the opposite transaction should be done.

3. Words ending with '–' character are ignored, because application of this strategy does not, mathematically, offer any profits.

Results of this simulation are shown in Table I. Assuming opening capital of 1000, first column shows how much could be earned by applying Zipf strategy. The second column shows how much could be earned by buying stock in the beginning of year and selling at the end of year. Last column shows how many words were chosen to this simulation. The first observation is that using of Zipf is profitable for every company and indices. It can be observed that picking up more words brings larger profits. Despite this fact keeping shares of most companies through the year was more profitable because of general positive trend on the market. It should be noted that

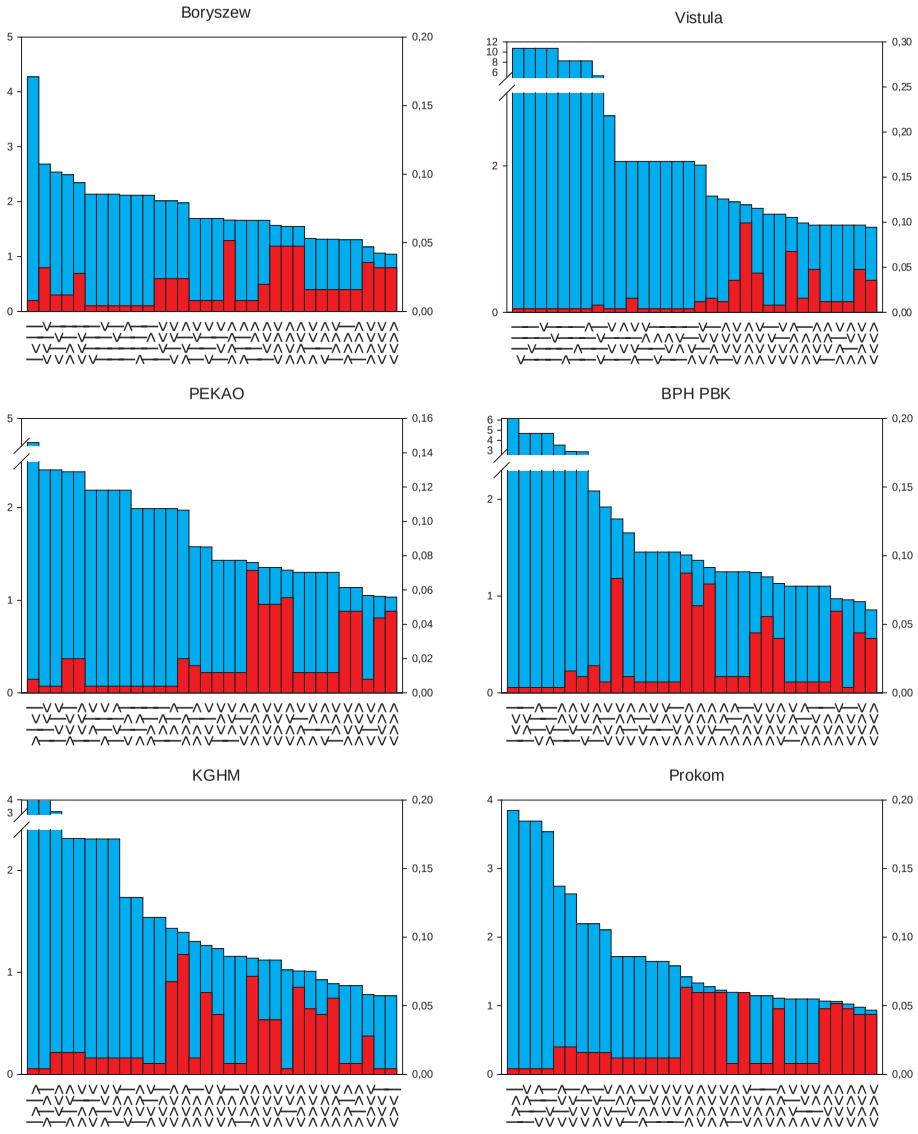

Fig. 3. Zipf analysis of stock companies: Boryszew, Vistula, PEKAO, BPH PBK, KGHM, Prokom.

using studied strategy it was possible to profit also on Prokom's shares, which recorded general loss. This proves that Zipf analysis is independent of general trend on the market. Taking these factors into consideration, the Zipf analysis calculation should be changed accordingly. We should assign '–' to interval represented by ±doubled commission.

TABLE I

Results of Zipf simulation.

	Using Zipf	Normal	Number of words
Boryszew	1339.3	2610.3	5
BPH PBK	1182.6	1404.9	3
Elektrim	1341.1	1990.5	3
Interia	1189.2	1250.0	2
KGHM	1099.5	1178.5	2
PEKAO	1229.1	1243.2	4
Prokom	1085.4	835.2	2
Vistula	1025.0	6078.4	1
DOW JONES	1120.7	1031.6	4
TECHWIG	1113.1	1146.6	1
WIG BANKI	1234.6	1322.9	3
WIG	1219.3	1250.6	4

3. Scatter analysis

Scatter plots show the relationship between two variables by displaying data points on a two-dimensional graph. The variable that might be related to the response is plotted on the X axis, and the response variable is plotted on the Y axis. Scatter plots can provide answer to the following question: Are variables X and Y related?

In this analysis some financial data series were taken (prices at successive closing times) grouped by weekday. They are presented on scatter plots, where on X axis are days from Monday to Friday and on Y axis are placed normalised changes of stock prices (R).

To normalise data the formula was used

$$R(t) = \frac{\log(P(t)) - \log(P(t-1))}{S}\,, \tag{1}$$

where S is standard deviation for $\{P(t)\}$.

Generally, increases are more frequent on Fridays and drops in value occur more often on Wednesdays when companies are considered. The largest fluctuations of prices occur on Fridays and the smallest are observed on Mondays. The best day for selling and buying shares should be individually determined for each company.

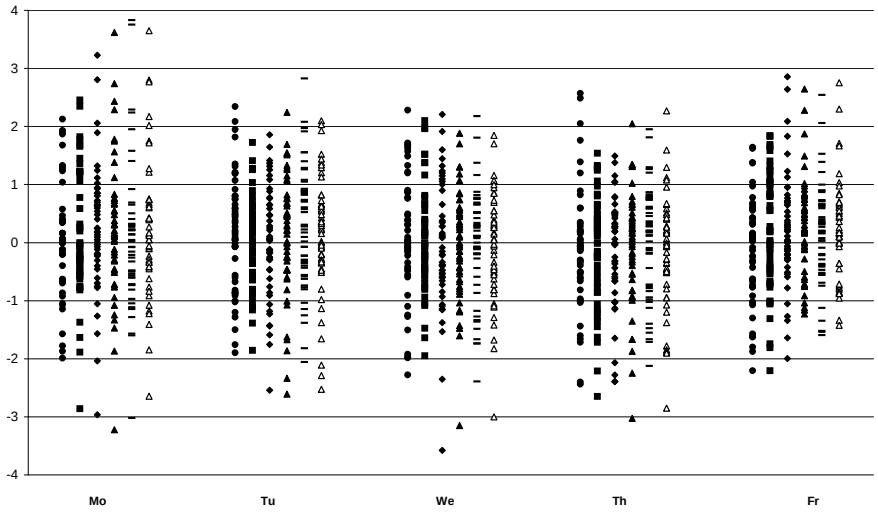

Fig. 4. Scatter analysis of stock indices, from left for each day: DOW JONES, NIKKEI, TECHWIG, WIG, WIG BANKI, WIG20.

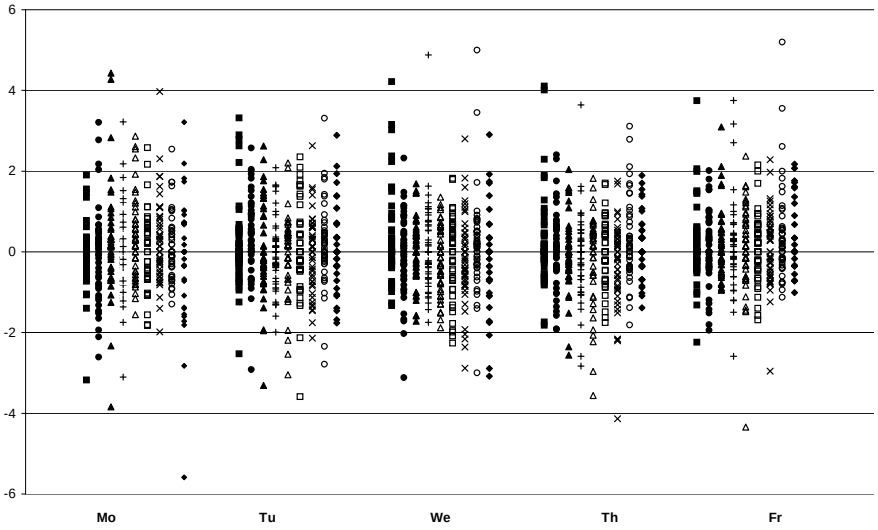

Fig. 5. Scatter analysis of companies, from left for each day: Boryszew, BPHPBK, Elektrim, Interia, KGHM, PEKAO, Prokom, Vistula, Zywiec.

Value of Polish indices increases mostly on Mondays, Tuesdays and Fridays. On Wednesdays and Thursday decreases dominate. The smallest fluctuations are observed on Fridays and the largest occur on Mondays. No relations can be found between Polish and foreign indices.

TABLE II

Standard deviation for scatter analysis (indices).

	DJ IA	NIKKEI	TECHWIG	WIG	WIG BANKI	WIG20	Average
Mo	0.989	1.288	1.090	1.206	1.248	1.191	1.184
Tu	0.983	0.777	0.934	1.024	1.025	1.026	1.002
We	1.049	0.916	1.061	0.914	0.918	0.920	0.953
Th	1.055	0.965	0.936	0.982	0.940	0.960	0.954
Fr	0.928	1.017	0.950	0.835	0.849	0.850	0.871
Average	1.001	0.992	0.994	0.992	0.996	0.989	

TABLE III

Average for scatter analysis (indices).

	DJ IA	NIKKEI	TECHWIG	WIG	WIG BANKI	WIG20	Average
Mo	0.051	0.036	0.238	0.229	0.061	0.242	0.192
Tu	0.166	−0.052	0.018	0.137	0.248	0.155	0.140
We	0.106	0.066	−0.065	−0.044	−0.021	−0.099	−0.057
Th	−0.121	−0.121	−0.158	−0.093	0.037	−0.156	−0.092
Fr	−0.137	0.177	0.210	0.254	0.175	0.211	0.213
Average	0.013	0.021	0.049	0.097	0.100	0.070	

TABLE IV

Standard deviation for scatter analysis (companies).

	Borysz.	Bphpbk	Elektr.	Interia	Kghm	Pekao	Prokom	Vist.	Zywiec	Aver.
Mo	0.783	1.168	1.334	1.039	0.998	1.304	1.030	0.790	1.328	1.086
Tu	1.045	0.965	1.100	0.863	0.972	1.084	0.921	0.960	0.916	0.981
We	1.105	0.970	0.780	1.033	0.796	0.972	1.067	1.116	1.105	0.994
Th	1.099	0.956	0.886	1.032	1.104	0.787	1.032	0.981	0.725	0.956
Fr	0.917	0.890	0.820	1.043	1.055	0.793	0.873	1.117	0.763	0.919
Aver.	0.990	0.990	0.984	1.002	0.985	0.988	0.984	0.993	0.968	

TABLE V

Average for scatter analysis (companies).

	Borysz.	Bphpbk	Elektr.	Interia	Kghm	Pekao	Prokom	Vist.	Zywiec	Av.
Mo	−0.075	−0.102	0.211	0.027	0.352	0.152	0.252	0.104	−0.091	0.092
Tu	0.158	0.386	−0.001	−0.059	−0.029	0.040	−0.030	0.166	0.087	0.080
We	0.321	−0.065	−0.006	0.049	−0.205	−0.059	−0.210	0.166	−0.174	−0.020
Th	0.239	0.062	−0.045	−0.084	−0.108	−0.005	−0.306	0.280	0.128	0.018
Fr	0.163	0.129	0.175	0.178	0.108	0.111	0.122	0.409	0.224	0.180
Av.	0.161	0.082	0.067	0.022	0.023	0.048	−0.035	0.225	0.035	

4. Continuous trading

We have collected continuous trading data for two major Polish indices: WIG and TECHWIG. Data from WIG was collected from the 8th until the 17th of November 2004. New index values are recorded every minute. Data from TECHWIG was collected from the 4th until the 12th of November 2004. New index values are recorded every 30 seconds.

4.1. Scatter plots

Similarly to Zipf analysis above, we examined scatter plots of continuous trading data for WIG and TECHWIG (Table VI).

TABLE VI

Standard deviation and average for scatter plot for continuous quotation.

	Standard deviation		Average	
Hours	TECHWIG	WIG	TECHWIG	WIG
10–11	1.296	0.767	0.060	0.291
11–12	1.043	0.873	0.032	−0.054
12–13	0.999	1.001	−0.018	−0.152
13–14	0.737	0.930	−0.002	−0.098
14–15	0.838	1.025	0.023	−0.065
15–16	1.122	1.165	0.021	−0.125
16–16:15	0.617	1.247	−0.053	−0.087
Average	0.950	1.001	0.009	−0.041

WIG index usually increases in the mornings and decreases throughout the rest of the day. The smallest fluctuations are at the beginning and the largest at the end of the trading session. TECHWIG index increases mostly in the mornings when its fluctuations are high.

4.2. Lag plots

Lag plot represents pairs of subsequent values on a 2D plane. If an array of n values $x_1, \ldots, x_n$ is given, the coordinates of lag plot are $(x_1,x_2), (x_2,x_3), \ldots (x_{n-1}, x_n)$. In our case, we used index value returns normalised (1).

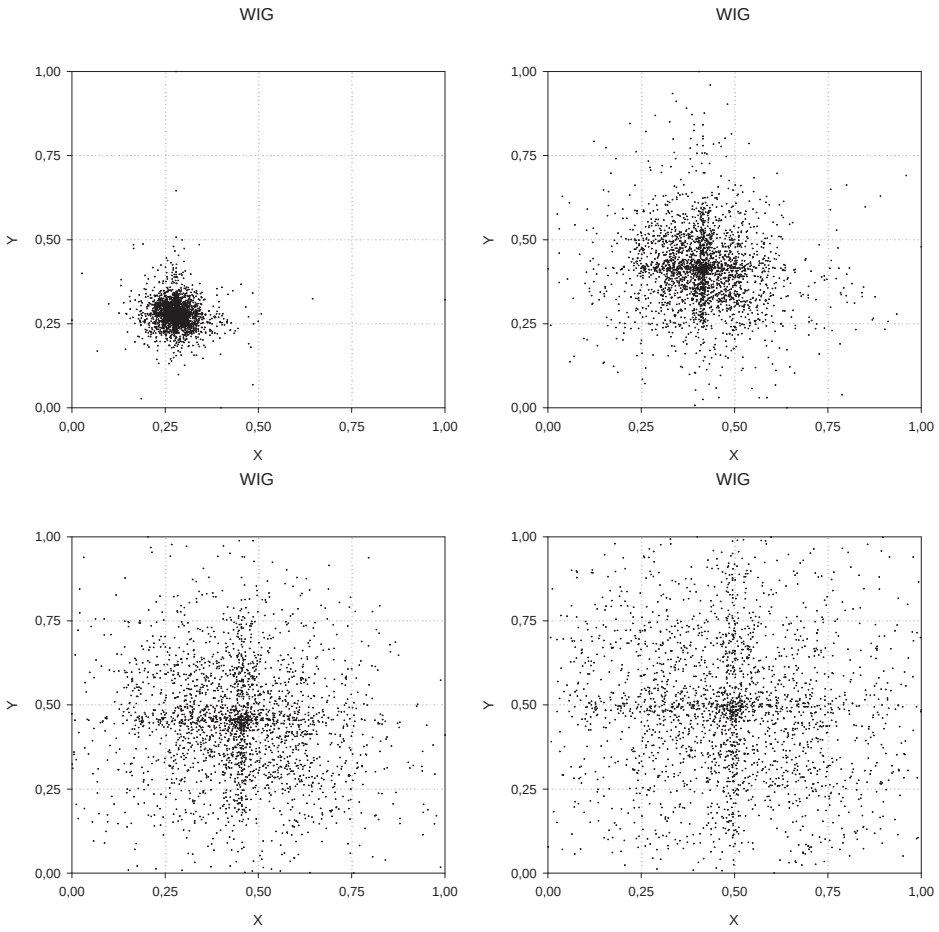

Fig. 6. Lag plots for WIG index. Upper-left panel shows plot with all the values, upper-right has 10 minimal and 10 maximal values removed, lower-left 50 extremes removed, lower-left 100 extremes removed.

Figs. 6 and 7 present lag plots for WIG and TECHWIG indices. First panel represents the whole data. Points on upper-left panel are located in left bottom quarter, which is caused by the presence of a point with large value. Therefore, we subsequently remove extreme events. Finally, after removing 100 extreme values, WIG dotplot has the increasing density towards the center of down-right panel. This makes the plot similar to the plot of Gaussian noise. The TECHWIG data provides almost uniform distribution, however points are concentrated on horizontal and vertical lines.

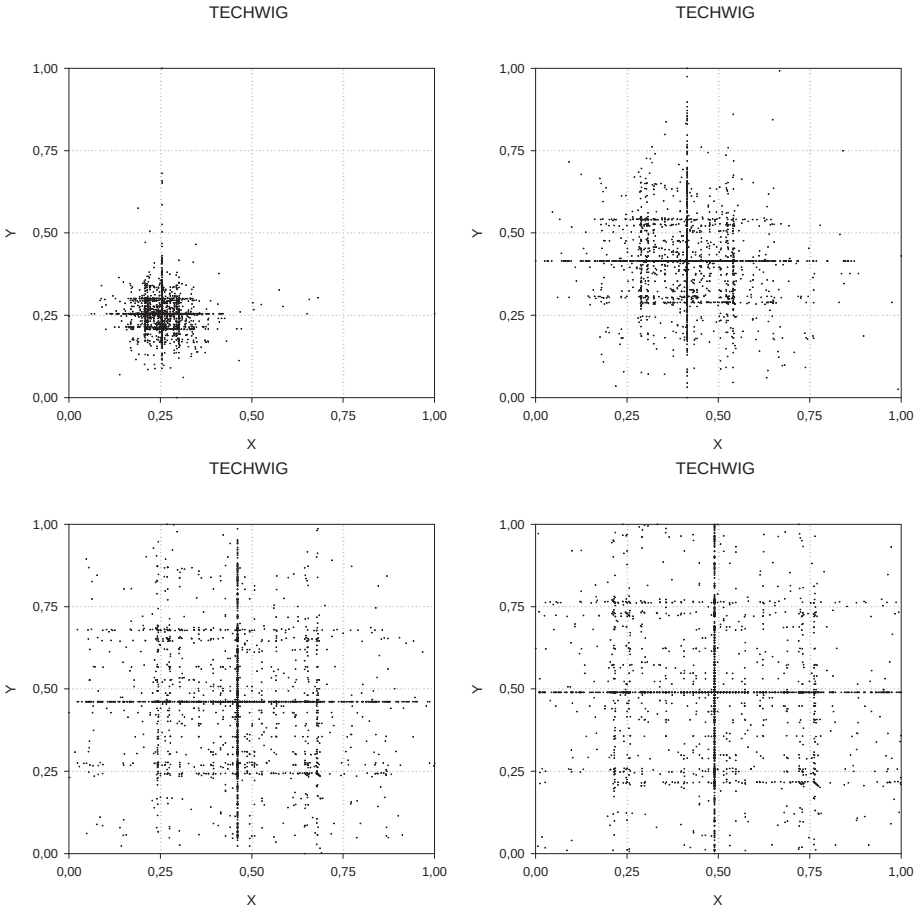

Fig. 7. Lag plots for TECHWIG index. Extreme values are removed as in Fig. 6.

5. Concluding remarks

In this paper we have examined selected indices and stock companies from Warsaw Stock Exchange in 2004, using graphical methods of Zipf and scatter analysis. Zipf analysis shows probabilities of price change sequences and thus allows us to react to these changes in order to maximise profits. We have presented a strategy that can be profitable, even when share prices are generally decreasing. Scatter analysis for daily quotations reflects stock and indices value changes in each weekday (in each hour for continuous trading). Research showed that it is possible to determine best moments to buy and sell shares. Additionally, we have made lag plots for two major Polish indices; WIG and TECHWIG. The results were very different. Plot for WIG index indicated that changes were of Gaussian noise type, while those in TECHWIG were more uniform.

W.B., P.L. and R.P. would like to thank organisers of the FENS meeting for their hospitality.

REFERENCES

[1] http://www.gpw.com.pl/zrodla/podsumowanie2004.pdf, in Polish.

[2] http://biznes.interia.pl/news?inf=577636, in Polish.

[3] Source data for daily quotations: http://bossa.pl/pub/ciagle/mstock/mstcgl.zip.

[4] Source data for continuous quotations: http://mojeinwestycje.interia.pl/gie/notgpw/notc/c_akcje.

[5] http://www.unifr.ch/cgi-bin/econophysics/etraitad.pl?number=6&art=feb-mar-99/art-nature-correlations-financial.html

[6] N. Vandewalle, M. Ausloos, *Physica A* **268**, 240 (1999).

[7] NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/eda/section3/scatterp.htm.

[8] Table of chosen words for Zipf simulation, http://iftia9.univ.gda.pl/network/WSE2004/zipf-words.pdf.