

Proceedings of the 6th Polish Symposium of Physics in Economy and Social Sciences (FENS2012), Gdańsk, Poland

Mining Associations on the Warsaw Stock Exchange

K. KARPIO, P. ŁUKASIEWICZ, A. ORŁOWSKI AND T. ZĄBKOWSKI

Department of Informatics, Warsaw University of Life Sciences (SGGW)

Nowoursynowska 159, 02-776 Warsaw, Poland

Identification of patterns in stock markets has been an important subject for many years. In the past, numerous techniques, both technical and econometric, were used to predict changes in stock markets, but dependences among all the companies listed on a stock market were considered in a limited extent. Numerous studies confirm that larger stocks items appear to influence smaller ones and that, on a global level, most of the world's stock markets are integrated. Therefore, this study implements the association rules using a data mining approach to explore the co-movement between stock items listed on the Warsaw Stock Exchange. We believe that in order to describe and to understand market's behavior, data mining techniques are more flexible in use than for instance pricing models based on a finance theory. The former seems to be more effective for explaining market behavior without making particular assumptions.

DOI: [10.12693/APhysPolA.123.553](https://doi.org/10.12693/APhysPolA.123.553)

PACS: 89.65.Gh, 07.05.Kf, 07.05.Rm

1. Introduction and aim of the study

Mining patterns in stock market data is one of the hottest issues of data mining due to its commercial applications and possible financial benefits. Finding patterns in stock data can help provide insight into market behavior, buying or selling habits and co-movements of stock shares. Numerous studies confirmed that there are typical patterns in such data. For instance, it was observed that larger stock items appear to lead smaller ones and that, on global level, most of the world's stock markets are integrated [1, 2]. In general, the research stream on stock market is focused on finding some characteristics such as anomalies, regularities, and predictive modeling [3]. The field of predictive modeling and forecasting can be characterized by data intensity, presence of a noise, non-stationarity, unstructured nature, high level of an uncertainty, and presence of hidden relationships in data. Additionally, many factors may influence the stock markets including political events, general economic conditions, and traders' expectations. Therefore, predicting price movements on stock market seems to be quite difficult.

Recently, data mining techniques have gained on importance [4, 5] despite negative connotations due to its far-reaching automation without insight into the nature of the problem. Due to technological development in information technology and software abilities gathering sufficient data on stocks traded by the hour and even by the minute have become much easier. This trend is another reason why data mining techniques have drawn the attention of investors interested in predicting stock prices. To improve the predictability of stock prices, previous studies utilized various methods, such as neural networks [6, 7] or support vectors [8]. Although these methods can increase predictability, it is difficult to ascertain how the results were determined and which factors have the most impact on stock prices. Furthermore, great expertise is required to use the result for investment purposes.

The analysis using the association rules allows people to understand and to implement the results very easily. Mining according to the association rules uncovers interesting correlation patterns among a large set of data items by showing the attributes that occur together frequently with a given quality measures. From a technical point of view, association rules are important classes of regularities that may exist in databases. Since first introduced in 1993 by Agrawal et al. [9], mining associations have received a great deal of attention in different applications such as market basket analysis and e-commerce, website patterns and clickstream analysis, survey data analysis or gene expression data analysis.

In this research, we implement the association rules to explore co-movements between stock items listed on Warsaw Stock Exchange. We believe that to describe and to understand market's behavior, data mining tools are more flexible in use than for instance the pricing models based on finance theory, because the former seems to be more effective for explaining market behavior without making particular assumptions.

The main research goal of the study is to apply the association rules to identify synchronous relationships in WSE stock market data based on return rates. To prove that the obtained rules are non-trivial and possess the ability to generalize we also test the association on random return rates data. The test procedure used in this work is typical for experiments in physics, in which we aim to prove the hypothesis by showing the evidence that no relationships are present on random data.

2. Data characteristics

In this article we analyzed the stock items listed on the main market of the Warsaw Stock Exchange (WSE). Data included return rates and trading volumes from 04-01-1999 to 23-02-2012, which corresponds to 3300 trading days. We limited stock items only to companies, excluding bonds, options and indices, bearing in mind that

over the years there was different number of companies appearing on the WSE market. Figure 1 shows the number of assets traded each day. The fluctuations are due to a number of companies which were not traded on certain day (volume equal to null).

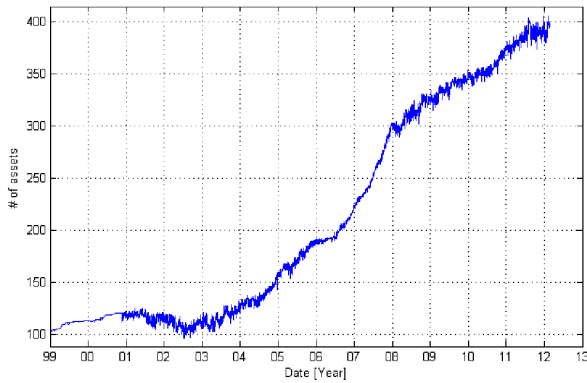


Fig. 1. The number of companies traded each day.

The total number of companies traded each day increased from slightly above 100 up to about 390 on average. Every asset was traded from 22 up to 3300 days (Table I). The companies traded less days are the companies that have just appeared on the WSE. No conclusions can be drawn regarding their size. There are big companies among relatively small ones. On the other hand, the companies traded longer are the companies which emerged on the WSE in 90'. These are mostly the biggest companies listed on WSE.

TABLE I

The companies with the greatest and the smallest numbers of trading days.

Name	# of days	Name	# of days
KGHM	3301	BOWIM	22
MOSTALEXP	3301	SOPHARMA	22
MOSTALZAB	3301	ESTAR	23
PEKAO	3301	EUROHOLD	25
STALEXP	3301	GETBANK	25
SYGNITY	3301	OPERATERR	49
TPSA	3301	ZAMET	92
RAFAKO	3300	PGODLEW	95
INGBSK	3300	TOYA	99
ALMA	3300	NOVAKBM	99
ASSECOPOL	3300		
BBIDEVNFI	3300		
BUDIMEX	3300		
DEBICA	3300		

With the mean number of assets traded daily of 217 the total number of prices analyzed was 698 897. The log-mean distribution of daily return rates is presented in Fig. 2. It follows the well known log-normal like shape with the exception of fat tails.

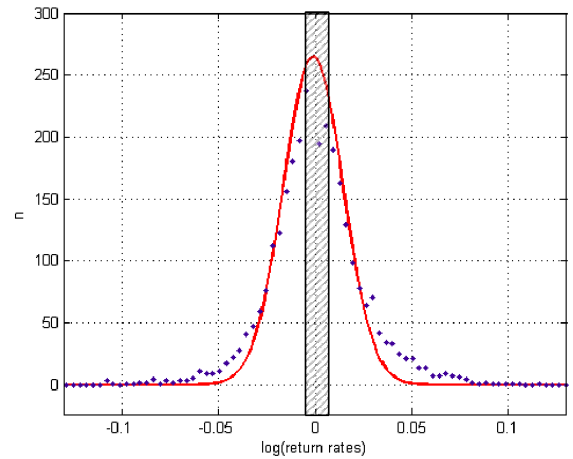


Fig. 2. The log-mean distribution of the return rates with the range of -0.008 to 0.008 selected (see the following paragraph for explanation). The curve represents normal shape.

For the purpose of analysis, special attention has been made to define positive or negative return rates. Prices and thus return rates fluctuate over time. Moreover, a distribution of return rates is always concentrated around zero. The return rates within the range of $[-0.008; 0.008]$ are called a tie and we are not interested in it. So, we remove the tie from our data records. The interval of $[-0.008, 0.008]$ is proposed based on transaction costs. Typical transaction costs for buying or selling stock items via internet for an individual investor amounts to about 0.4% of the transaction value, hence the 0.8% would cover at least these costs. Moreover distributions of the return rates are concentrated around zero so the applied cut has also the purpose to exclude the return rates, which are so small so they do not have practical application.

3. Mining correlations with association rules

The association rules algorithms are used mainly to determine the relationships between items or features that occur frequently in databases. For instance, in market basket analysis context, if people who purchase item A also purchase item B, then we can say that there is a relationship between item A and item B, and such information can be useful for decision making. Therefore, the aim of applying the association rules algorithm is to find interesting relationships by analyzing data and use them to support a decision making process.

The association rules are defined as follows [9]. We take $I = (i_1, i_2, \dots, i_m)$ as the item set and D as a set of transactions in a database in which each transaction T represents an item set such that $T \subseteq I$. That is, each item set T is a non-empty sub-item set of I . The association rules are an implication of the form $A \rightarrow B$, where $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$. The rule $A \rightarrow B$ holds in transaction set D according to three popular measures: support, confidence and lift.

Support, denoted as $\text{Sup}(A, D)$, represents the rate of transactions in D containing the item set A . Support is used to evaluate the statistical importance of D , and the higher its value, the more important transaction set D is. Therefore, the rule $A \rightarrow B$ has support $\text{Sup}(A \cup B, D)$ that represents the rate of transactions in D containing $A \cup B$.

The other measure of rule $A \rightarrow B$, called confidence and denoted as $\text{Conf}(A \rightarrow B)$, represents the rate of transactions in D that contain A and also contain B . That is

$$\text{Conf}(A \rightarrow B) = \frac{\text{Sup}(A \cap B)}{\text{Sup}(A, D)}. \quad (1)$$

In this case, $\text{Conf}(A \rightarrow B)$ denotes that if the transaction includes A , the chance that the transaction also contains B is relatively high. Confidence is an estimate of $P(B|A)$, that is the probability of observing B given A . It ranges from 0 to 1 and may be expressed as percentage. For the rules with the same confidence level, the rule with the highest support is preferred. The rationale is that the estimate for confidence is more reliable.

The confidence alone may not be sufficient to assess the descriptive power of a rule. Rules with high confidence may occur by chance. Therefore, we also investigated the reliability of support and confidence (see Sect. 3.4). Such spurious rules can be detected by determining whether the antecedent item and the consequent item are statistically independent [10]. This inspired a number of measures for association rule interest. One of them is lift, defined as:

$$\text{Lift}(A \rightarrow B) = \frac{\text{Conf}(A \cap B)}{\text{Sup}(B, D)}. \quad (2)$$

Lift measures how far from independence are items A and B . It ranges within $[0, +\infty]$ and it is symmetric with respect to antecedent and consequent items. Values close to 1 imply that A and B are independent and the rule is not interesting. Values far higher than 1 indicate that the occurrence of A provides information about B . Lift is a measure of co-occurrence and does not inform about implication of the items.

Given a set of transactions D , the problem of association rules discovery is to generate all transaction rules that have certain level of minimum support (called MinSup) and confidence (called MinConf). To explore association rules we used Apriori algorithm [9] because it is simple and easy to implement. Therefore, the problem of mining association rules was decomposed into two steps [11]. The first step was to detect a large item set whose support is greater than MinSup and the second step was needed to generate association rules using the large item set. Such rules must satisfy two conditions: $\text{Sup}(A \cup B) \geq \text{MinSup}$ and $\text{Conf}(A \rightarrow B) \geq \text{MinConf}$. For the purpose of our analyses we set the value of 2% of the largest single item frequency for MinSup and 20% for MinConf.

In our research, the aim of the analysis was to find interesting associations among the transactional stock mar-

ket data presented in Sect. 2. More specifically, we aim to perform the following analyses:

1. Mining association rules among the all stock items on the main market that obtained day to day positive rate of return higher than 0.008;
2. Link analysis for WIG20 companies;
3. Mining the upstream and downstream relations among the all stock items on the main market that obtained day to day rate of return higher or lower than specified interval $[-0.008, 0.008]$;
4. Evaluating the statistical significance of the support and confidence.

The source data included stock items, trading dates and closing prices. Therefore, in order to mine the association rules the input data were prepared to indicate positive rate of return on day by day basis. In the following part of work, to measure upstream and downstream relationships we transformed the data to have the status of going up or down for all the stock items.

For convenience, we defined UP and DOWN flags to indicate positive (greater than 0.008) and negative (less than -0.008) return rates, respectively.

3.1. Mining association rules among the stocks with positive rate of return

Based on the source table with transactions data we prepared the dataset for the association analysis, see Fig. 3.

DATE	STOCK_NAME	RATE_OF_RETURN
04JAN1999	BUDIMEX	1.0875
04JAN1999	BUDOPOL	0.9637
04JAN1999	BYTOM	1.0222
04JAN1999	BZWBK	1.0415
04JAN1999	CALATRAVA	-1
04JAN1999	CDRED	1.0623
04JAN1999	CEGNBR	-1
04JAN1999	COLIAN	1.0959

Items removed because of the tie.

→

DATE	STOCK_NAME
04JAN1999	BUDIMEX
04JAN1999	BUDOPOL
04JAN1999	BYTOM
04JAN1999	BZWBK
04JAN1999	CDRED
04JAN1999	COLIAN

Fig. 3. Illustration of data preparation for the analysis.

After completing the rule discovery the resultant rules ordered by confidence have been summarized in Table II. It was found that the biggest companies (mainly from WIG20) formed the rules with the highest confidence level. Based on that we can generalize that there is high probability of observing in a basket some companies from WIG20 given the presence of other companies from WIG20. The association rules with biggest confidence are between biggest banks and oil as well as metallurgy companies: PKOBP, PEKAO, BZWBK, LOTOS, PKNORLEN, KGHM. It can indicate the inter-sector financials relations between companies as well as intra-sector common market conditions. Noteworthy the presence of the LCCORP which is the group of developer companies implementing investments in the largest cities. The rule $\text{LCCORP} \Rightarrow \text{PEKAO}$ may indicate the

PEKAO as the source of a potential credit. In fact, PEKAO Pioneer Investments TFI (subsidiary company of PEKAO) is the owner of more than 5% of the LC-CORP shares. Presence of such rule opens way to further analysis because such rule is probably not permanent and reflects rather short time business relation. For instance there could be hardly to imagine creation of such rule in the recent years because of the bad condition of developer sector of economy. The prices of apartments and commercials are going down. The banks are not interested neither in opening new credit lines nor keeping current financing. The other potential credit/investment relation could be present in the case of observed banks and other companies. However, those relations are of longer term indicating a coming far further besides simple credit of separate investments. The association rule for LCCORP \Rightarrow PEKAO is present for slightly more than 6% trading days, whereas supports of the rules: PKOBP \Rightarrow KGHM (11.09%), PKOBP \Rightarrow PKNORLEN (10.75%), PEKAO \Rightarrow KGHM (20.15%) are up to three times bigger.

TABLE II
Top 10 rules with the highest confidence.

Conf	Support	Lift	Rule
64.25	11.54	1.87	PKOBP \Rightarrow PEKAO
63.88	10.88	2.02	LOTOS \Rightarrow PKNORLEN
62.39	19.75	1.65	PKNORLEN \Rightarrow KGHM
62.10	10.57	1.65	LOTOS \Rightarrow KGHM
61.72	11.09	1.64	PKOBP \Rightarrow KGHM
61.25	6.51	1.62	LCCORP \Rightarrow KGHM
59.87	10.75	1.89	PKOBP \Rightarrow PKNORLEN
58.69	20.15	1.55	PEKAO \Rightarrow KGHM
57.84	18.54	1.69	BZWBK \Rightarrow PEKAO
57.83	6.15	1.69	LCCORP \Rightarrow PEKAO

The closer look at the rules ordered by lift (Table III) enabled us to see that there are many rules with the very high lift indicating strong associations among particular stocks.

TABLE III
Top 10 rules with the highest lift.

Conf	Support	Lift	Rule
42.08	2.33	6.28	BOGDANKA \Rightarrow MIRBUD
34.84	2.33	6.28	MIRBUD \Rightarrow BOGDANKA
40.51	2.39	5.81	IPOPEMA \Rightarrow EKOEXPORT
34.35	2.39	5.81	EKOEXPORT \Rightarrow IPOPEMA
34.43	1.91	5.68	BOGDANKA \Rightarrow ENEA
31.50	1.91	5.68	ENEA \Rightarrow BOGDANKA
39.20	2.09	5.63	APLISENS \Rightarrow EKOEXPORT
30.00	2.09	5.63	EKOEXPORT \Rightarrow APLISENS
38.89	2.12	5.58	PATENTUS \Rightarrow EKOEXPORT
30.43	2.12	5.58	EKOEXPORT \Rightarrow PATENTUS

The rules apply to the companies of a small size. Namely, those rules are for instance BOGDANKA and MIRBUD, IPOPEMA and EKOEXPORT, BOG-

DANKA and ENEA, APLISENS and EKOEXPORT, PATENTUS and EKOEXPORT. Some of these relations may have explanation: BOGDANKA is supplier of coal for ENEA's power plants; IPOPEMA "buy" recommendations for EKOEXPORT in 2011; but majority of rules are not linked through commercial or ownership relationships.

3.2. Link analysis for WIG20 companies

In the research we also used so called *link analysis*, which is mathematical and graphical analysis for highlighting the linkages between items of interest. In our case, we use it to provide web-like graphs indicating the strength of connections between the stocks from WIG20 with the positive rate of return, see Fig. 4. As the measure of the general strength we used a normalized count of all connections for the stock, ranging from 0 to 1. It was observed that the strongest relations were formed by KGHM, PEKAO, BRE, BORYSZEW, TPSA, PKNORLEN, and ASSECOPOL. These stocks may be considered as the leaders driving the others when the upward trend is present.

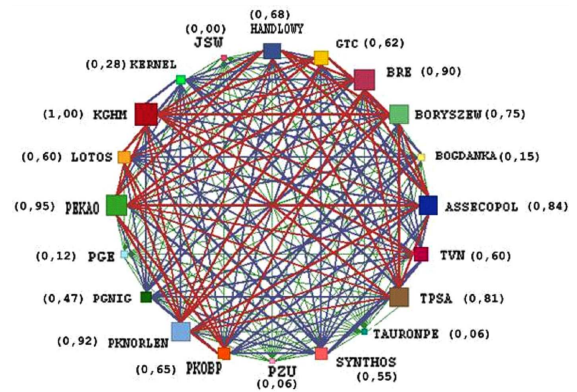


Fig. 4. Link analysis between the stocks from WIG20 with positive rate of return. The strength of the relationships for the stocks is presented in brackets.

3.3. Mining the upstream and downstream relations among the stocks

To analyze the upstream and downstream relations among the stocks we prepared data set according to the scheme in Fig. 5. For this purpose the source data were transformed to indicate the status of going up or down of the stock prices in the respective trading days.

After completing the rule discovery the resultant rules were ordered by confidence (Table IV). As previously, it was found that the biggest companies (mainly from WIG20) formed the rules with the highest confidence level. What was interesting and most noticeable that majority of the rules with the high confidence were those with DOWN direction. There are no mixed rules (UP-DOWN, DOWN-UP) among rules with highest confidence. That means one company "gives" its trend to the

DATE	STOCK_NAME	RATE_OF_RETURN
04JAN1999	BUDIMEX	1.0875
04JAN1999	BUDOPOL	0.9697
04JAN1999	BYTOM	1.0222
04JAN1999	BZWBK	1.0415
04JAN1999	CALATRAYA	-1
04JAN1999	CDRED	1.0623
04JAN1999	EDGNGR	-1
04JAN1999	COLJAN	1.0959

DATE	STOCK_NAME
04JAN1999	BUDIMEX UP
04JAN1999	BUDOPOL DOWN
04JAN1999	BYTOM UP
04JAN1999	BZWBK UP
04JAN1999	CDRED UP
04JAN1999	COLJAN UP

Fig. 5. Illustration of data preparation for the analysis.

Top 10 rules with the highest confidence.

TABLE IV

Conf	Support	Lift	Rule
65.72	11.97	1.97	PKOBP DOWN \Rightarrow PEKAO DOWN
65.33	2.97	1.84	PGE DOWN \Rightarrow MIDAS DOWN
64.25	11.54	1.87	PKOBP UP \Rightarrow PEKAO UP
63.88	10.88	2.02	LOTOS UP \Rightarrow PKNORLEN UP
63.76	5.76	3.83	AZOTYTARNOW DOWN \Rightarrow PETROLINV DOWN
63.41	6.30	3.81	CENTROZAP DOWN \Rightarrow PETROLINV DOWN
63.25	10.85	2.00	LOTOS DOWN \Rightarrow PKNORLEN DOWN
63.10	6.79	1.74	ARMATURA DOWN \Rightarrow SKOTAN DOWN
62.88	8.72	3.77	LCCORP DOWN \Rightarrow PETROLINV DOWN
62.85	9.48	1.73	HERKULES DOWN \Rightarrow SKOTAN DOWN

Top 10 rules with the highest lift.

TABLE V

Conf	Support	Lift	Rule
40.94	2.12	6.50	BOGDANKA DOWN \Rightarrow IPOPEMA DOWN
33.65	2.12	6.50	IPOPEMA DOWN \Rightarrow BOGDANKA DOWN
45.75	2.12	6.32	FERRO DOWN \Rightarrow MIRBUD DOWN
29.29	2.12	6.32	MIRBUD DOWN \Rightarrow FERRO DOWN
42.08	2.33	6.28	BOGDANKA UP \Rightarrow MIRBUD UP
34.84	2.33	6.28	MIRBUD UP \Rightarrow BOGDANKA UP
56.00	2.54	6.18	PGE DOWN \Rightarrow CHEMOS DOWN
44.72	2.18	6.18	KOV DOWN \Rightarrow MIRBUD DOWN
30.13	2.18	6.18	MIRBUD DOWN \Rightarrow KOV DOWN
28.09	2.54	6.18	CHEMOS DOWN \Rightarrow PGE DOWN

Taking into account the highest lift (Table V) we could observe that there is strong association among particular stocks but this is not necessarily due to some commercial or ownership relationships between the stocks.

The biggest lift values exceed 6. That means: if the first company in the pair has the indicated change of the price, then the probability of occurrence of indicated price change of the second company is more than 6 times greater than in the case of uncorrelated behavior. The majority of the rules are of the DOWN \Rightarrow DOWN type. As previously, all the rules have rather small support, which means those rules, with high lift are strong but occur seldom (2–3% of trading days).

3.4. Reference distribution and statistical reliability of parameters

In order to quantitatively check the strength of the observed associations we used the comparison of the real

data to the reference one. The reference data have been built by assigning to each company random daily return rates from the set of return rates for all companies traded on the same day. In this way we do not distinguish companies. The only differences between companies are due to different periods they were traded. Some companies were introduced to WSE during boom and were traded mostly during days when prices were going up. The other companies have less optimistic history because of global trend present during their appearance on the stock market. Our reference data takes into account global trend each day and any traded company has the same chance appearing in the associated pair. The only differences are due to statistical fluctuations.

In case of support, we can observe that it has similar values for real and reference data up to the value of 11%. This means that those values could have random nature. However, beyond the value of 11% we observe system-

atically increasing differences. The real data show values of support up to about 22% while for the reference data it does not exceed 15%, see Fig. 6. That means that the pairs of companies with the biggest values of support, e.g.: PEKAO and KGHM, PKNORLEN and KGHM, BRE and KGHM, BZWBK and PEKAO, BRE and PEKAO form some patterns on real data but not necessarily on random data.

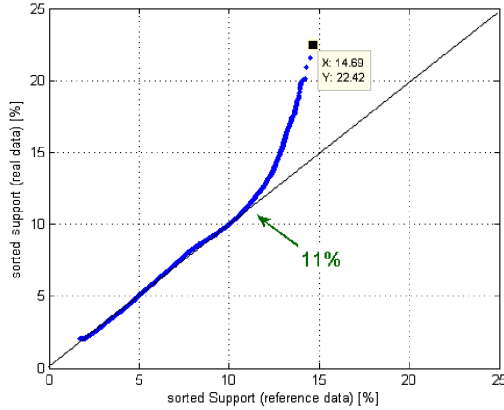


Fig. 6. Support for association rules (X — reference data, Y — real data).

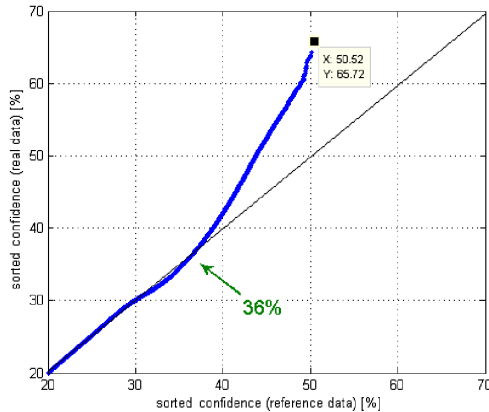


Fig. 7. Confidence for association rules (X — reference data, Y — real data).

The similar behavior was observed in case of confidence. Here, we can observe slightly lower values for real data in the range of $\text{Conf} \in [30\%; 36\%]$. The differences are not too big and do not exceed 2%. However, when confidence exceeds 36%, then again we can observe increasing values for real data in comparison to reference one, see Fig. 7. The differences are becoming bigger and bigger up to the value of 60% as compared to 50% for reference data. Then the confidence for real data increases sharply up to 66%. This indicates that the rules with high confidence and high support form strong relations which are not accidental.

Based on these experiments, we can conclude that the rules, which exceeded the threshold of 11% for support

and 36% for confidence possess the ability to generalize. They reveal the real relationships among stock items which did not arise from statistical coincidences.

We can also narrow the rules to these having very high confidence and high support. Table VI contains rules with confidence $> 55\%$ and support $> 17\%$. Those rules have very high values of both parameters, highly exceeding values found in the reference data.

TABLE VI

Rules with confidence and support exceeding values in reference data.

Conf	Support	Lift	Rule
58.69	20.15	1.55	PEKAO \Rightarrow KGHM
62.39	19.75	1.65	PKNORLEN \Rightarrow KGHM
55.57	18.75	1.47	BRE \Rightarrow KGHM
57.84	18.54	1.69	BZWBK \Rightarrow PEKAO
55.83	17.84	1.48	TPSA \Rightarrow KGHM
55.58	17.81	1.47	BZWBK \Rightarrow KGHM
55.50	17.57	1.62	PKNORLEN \Rightarrow PEKAO

Again we observe similar rules between three banks PEKAO, BRE, BZWBK and KGHM. The other group consists of the rules between banks and between other companies: BZWBK \Rightarrow PEKAO, PKNORLEN \Rightarrow KGHM, TPSA \Rightarrow KGHM. There is also rule between PKNORLEN and PEKAO stating that increase of prices of PKNORLEN is associated with the price increase of PEKAO.

4. Results and concluding remarks

In this paper we analyzed association rules in the WSE. In opposite to the traditional approach based on the measurement of linear correlations we adopt data mining technique. That allowed us to study all companies in the analyzed period of data at the same time. The association rules with Apriori algorithm allowed us to discover companies whose returns rates influence return rates of other companies. The analysis was aimed to the discovery of companies' pairs, on the other words, to study bilateral associations. We calculated values for commonly used measures: support, confidence and lift distinguishing association rules appearing most frequently. The analysis was performed in four steps: (i) analysis of the association rules between positive return rates; (ii) link analysis for WIG20 companies with positive return rates; (iii) analysis of the association rules between combinations of positive and negative return rates; (iv) evaluation of statistical significance of measured parameters on reference data.

The results showed that the association rules allowed finding the strongest relationships that occur in the data, which in turn may help investors or those interested to learn about stock market to gather and easily present the knowledge of underlying patterns on the market. In particular, the results can be summarized as follows:

- We observed significant association rules among positive return rates for some of the biggest banks, oil and metallurgy companies. Typical confidence was about 60% while support was ranging from 6% to more than 20%. Some of these associations, like LCCORP and PEKAO could be present due to the credit relationship and real estate market condition.
- The rules regarding banks and other companies imply that positive return rate of the banks are accompanied by the positive return rate for the companies.
- Rules with the highest values of lift are formed by relatively small companies. For those rules lift exceeded 6 but support was very small, of about 2%. That means the rules are strong but occur rather seldom. One explanation is that the rules are related to short-time relationships or due to the external random conditions.
- Almost all rules with the highest lift are two-directional. In other words stock price increase on one company is accompanied by price increase for the other company and vice-versa, with similar confidence level.
- Some of the rules with high lift could be explained by business relations between companies, e.g. BOGDANKA as supplier of coal for ENEA's power plants.
- The rules with the highest confidence between combinations of positive and negative return rates are mostly DOWN-DOWN or UP-UP. This appears to be as one stock "gives" its trend to the other within a rule. The majority are DOWN-DOWN rules among them, so the rules seem to be stronger during market declines. However, for PKOBP and PEKAO as well as for LOTOS and PKNORLEN we observe two-directional rules. That could be due to the common market conditions or external influences.
- Taking into account the highest lift values within the upstream and downstream relations we can conclude that, as previously, the rules are utilizing relatively small companies. Some of them have been observed for rules with positive return rates. However, the values of support are very small, thus the rules occur very seldom.
- The comparison of the real data association rules with reference data with no non-statistical relations between the stocks showed the differences for high values of confidence and support. Noticeable increase in values of the parameters for real data as compared to the reference one are starting from 11% for support and 36% for confidence. The values of support amounted to about 20% for real data and only up to about 13% for reference ones. Similarly, the big differences were observed for largest

values of support. For the reference data its values amounted up to 50% while for real data — up to 66%.

This means that the relationships (rules) discovered, especially these with high lift and high support are present in data and indicate strong relations which are not accidental.

- The comparison on reference data showed also that rules having up to 11% for support and up to 36% for confidence were present both in real data and reference data.

Association rules seem to be appropriate technique for effective exploration of underlying patterns in huge amount of data. This technique is often used in the financial markets in the context of investment processes automation. In order to avoid the technical aspects related to the system architecture and database requirements which in case of automation are important, in this paper we rather focused on knowledge extraction from the financial data. Nevertheless, the issue concerning processes automation is an interesting proposal for future research.

Acknowledgments

The work was funded by the National Science Centre based on decision number DEC-2011/03/B/HS4/05092.

References

- [1] C. Wu, Y. Su, *Int. Rev. Econom. Fin.* **7**, 63 (1998).
- [2] S.H. Na, S.Y. Sohn, *Exp. Syst. Appl.* **38**, 9046 (2010).
- [3] S.H. Chen, P.P. Wang, *Computational Intelligence in Economics and Finance (Advanced Information Processing)*, Springer, Berlin 2006.
- [4] M. Gavrilov, D. Anguelov, P. Indyk, R. Motwani, in: *Proc. Sixth Int. Conf. on Knowledge Discovery and Data Mining*, Eds. R. Ramakrishnan, S. Stolfo, R. Bayardo, I. Parsa, ACM, New York 2000, p. 487.
- [5] B. Kovalerchuk, E. Vityaev, *Data Mining in Finance: Advances in Relational and Hybrid Methods*, Kluwer Academic, New York 2000.
- [6] E.M. Azoff, *Neural Network Time Series Forecasting of Financial Markets*, Wiley, New York 1994.
- [7] J.W. Hall, in: *Adaptive Selection of US Stocks with Neural Nets*, Ed. G.J. Deboeck, Wiley, New York 1994, p. 45.
- [8] S. Chen, M.T. Leung, H. Daouk, *Comput. Operat. Res.* **30**, 901 (2003).
- [9] R. Agrawal, T. Imielinski, A. Swami, in: *Proc. 1993 ACM SIGMOD Int. Conf. on Management of Data*, Eds. P. Buneman, S. Jajodia, ACM Press, Washington 1993, p. 207.
- [10] P.J. Azevedo, A.M. Jorge, *Lect. Notes Comput. Sci.* **4701**, 510 (2007).
- [11] R. Agrawal, J. Shafer, *IEEE Trans. Knowledge Data Eng.* **8**, 962 (1996).