DE GRUYTER OPEN

Folia Oeconomica Stetinensia
DOI: 10.2478/foli-2014-0101

WYDZIAŁ NAUK EKONOMICZNYCH I ZARZĄDZANIA
UNIWERSYTET SZCZECIŃSKI

FORECASTING CHANGES IN STOCK PRICES ON THE BASIS OF PATTERNS IDENTIFIED WITH THE USE OF DATA CLASSIFICATION METHODS

Jacek Szanduła, Ph.D.

Wroclaw University of Economics
Department of Economic Forecasts and Analyses
Komandorska 118/120, Wrocław, Poland
e-mail: Jacek.Szandula@ue.wroc.pl

Received 18 November 2013, Accepted 1 July 2014

Abstract

The paper develops the concept of harnessing data classification methods to recognize patterns in stock prices. The author defines a formation as a pattern vector describing the financial instrument. Elements of such a vector can be related to the stock price as well as sales volume and other characteristics of the financial instrument. The study uses data concerning selected companies listed on the stock exchange in New York. It takes into account a number of variables that describe the behavior of prices and volume, both in the short and long term. Partitioning around medoids method has been used for data classification (for pattern recognition). An evaluation of the possibility of using certain formations for practical purposes has also been presented.

Keywords: data classification, cluster analysis, partitioning around medoids, technical analysis, price patterns.

JEL classification: C1, G17.

Introduction

Fundamental and technical analyses are the two main approaches to forecasting the stock market. Fundamental analysis involves identifying the state of the economy or company – represented by a financial instrument – considering both its environment and development prospects. It maintains that markets may underestimate the value of a security in the short run but that in the long run the ‘proper’ price will be reached. Profits can be made by purchasing the mispriced security and then waiting for the market to recognize its ‘mistake’ and correct the price of the security[1]. In technical analysis it is assumed that all the important information about the object are immediately reflected in the current valuation of the financial instrument, and therefore the study of important factors inside the object and its surroundings is pointless. Both of these approaches are often criticized as ineffective. Moreover, it is alleged that technical analysis does not meet the requirements of the scientific method, and the efficient market hypothesis is set against it[2].

Despite the significant formal shortcomings, technical analysis is frequently used, if not the dominant, tool in predicting short-term stock prices. Among its basic techniques one should mention price pattern analysis, which uses a price chart of a financial instrument in the formulation of price forecasts. In this case, the occurrence of a particular geometric shape in the price chart gives the signal that stock prices will be formed in a characteristic way in oncoming periods. Pattern recognition is usually based on a visual evaluation of the graph[3]. Such an assessment gives as a result a subjective interpretation of the analyzed material, what entails a considerable degree of discretion. The identification of the pattern may also be carried out using image recognition techniques[4].

Szanduła presents the concept of redefining the idea of the price pattern[5]. The pattern is interpreted as a pattern-vector consisting of different variables describing the financial instrument and not the graphical shape. This article aims to show that such an understanding of the pattern allow to effectively predict stock prices. It presents a way to use such patterns for making investment decisions in the stock market. The usefulness of this approach for forecasting stock prices changes is examined.

To achieve the goal, the idea of the price-patterns is described. The article presents a data classification method used in the further empirical research as well as the method to evaluate a classification quality. Then the procedure describing how to verify the usefulness of the proposed method of forecasting stock prices is explained. The last part of the article is an empirical study carried out on the basis of selected companies listed on the NYSE.

1. Methods

1.1. The idea of price patterns

Fundamental principles of technical analysis[6]:

- market action discounts influence of all other factors,
- prices move in trends,
- history tends to repeat itself,

determine the way in which analysts proceed while preparing the stock price forecast. If the price of the financial instrument discounts the impact of all the potential factors, it is pointless to concern oneself about them – the analysis of the price changes should be just adequate. In addition, if from time to time the market behaves according to a certain pattern, identification of such a pattern can be sufficient for making correct forecasts.

This idea is often employed in the price pattern analysis. The definition of the term 'price pattern' is usually omitted in the literature, leaving its meaning to the dictionary definition. Murphy understands price patterns as pictures or formations that appear on price charts, that can be classified into different categories, and that have predictive value[7]. This definition demonstrates that price patterns are a kind of geometric forms or shapes. Limiting the analysis only to the shape of a price chart of the financial instrument is a significant disadvantage. The pattern's repetition of the market behavior could also concern such variables as daily trade volume, price volatility and others. Szanduła notes that there is no reason to reject in advance the possibility of using other features to identify the pattern of the market behavior[8]. It is therefore proposed to extend the concept of the pattern to the model of vector of variables describing the financial instrument.

The purpose of the pattern analysis is its practical use to calculate forecasts. A necessary condition to recognize the pattern analysis as useful is the ability to identify unique patterns. On the basis of the information set (vector data) collected for the given period there must be a way to clearly determine whether the vector is to be assigned to one or the other pattern. In addition, the emergence of a specific pattern should be a signal that at a specified time horizon one can expect specific changes in the stock price. It means that the pattern should precede the increase or decrease of the stock price. There is no need to find causal relationship between the occurrence of the pattern and the continuation of price behavior – symptomatic linking would be good enough.

The task of determining homogeneous groups of feature vectors of a financial instrument can be completed using cluster analysis methods. The model pattern is established on the basis of vectors classified into one class. The assignment of the new (current) vector to the correct pattern is a problem that can be solved with data discrimination methods.

What is essential for the quality of the analysis is the choice of diagnostic features – variables. The variables used in the study should represent the widest possible range of information relating to the financial instrument. Among them one can consider:

- the current price level,
- price volatility during last session (or trading day),
- price volatility in the previous sessions,
- relative change in price at specified intervals,
- the number of days from local maxima and minima,
- trend slope at specified intervals,
- turnover during last session,
- average turnover during previous sessions,
- stock market indicators and oscillators, such as RSI, MACD, PPO and others.

1.2. Data classification

There are many taxonomic methods enabling the separation of similar objects according to the specific characteristics[9]. Most methods of data classification require the range of variation of the variables describing the analyzed objects to be similar. That means that it is necessary to choose the normalization procedure – such as standardization:

$$x'_{i,j} = \frac{x_{i,j} - \bar{x}}{s_j} \tag{1}$$

where:

$x'_{i,j}$ – standardized value of the variable X_j in i^{th} time point,

$i = 1, 2, \ldots, n$,

n – number of observations,

$j = 1, 2, \ldots, m$,

m – number of variables,

$x_{i,j}$ – value of the variable X_j in i^{th} time point,

$\bar{x}_j$ – average of X_j:

$$\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{i,j} \tag{2}$$

s_j – standard deviation of X_j:

$$s_j = \frac{1}{n-1}\sum_{i=1}^{n}(x_{i,j} - \bar{x}_j)^2 \tag{3}$$

The similarity of objects is determined by the distance between them. The smaller the distance between objects, the greater the similarity. So the next step is to determine the distance matrix:

$$\mathbf{D} = \begin{pmatrix} 0 & d_{1,2} & \cdots & d_{1,n} \\ d_{2,1} & 0 & \cdots & d_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n,1} & d_{n,2} & \cdots & 0 \end{pmatrix}_{n \times n} \tag{4}$$

whose elements are the Euclidean[10] distances between vectors of standardized values describing two objects:

$$d_{i,k} = \sqrt{\sum_{j=1}^{m}(x'_{i,j} - x'_{k,j})^2} \tag{5}$$

where: $d_{i,k}$ – distance between objects i and k.

With the distance matrix one can proceed to the object classification. The author has no preference in choosing a method of classification. For practical reasons, the presented empirical research is based on partitioning around medoids (PAM)[11]. This method proved to be the fastest in the classification of a large data set.

PAM belongs to partitioning methods and requires the establishment of the number of classes (clusters) from the very beginning. Among the many works coping with the problem of the optimal number of clusters[12] none gives a clear guidance about the correct approach. Thus the arbitrary choice of number of clusters is an acceptable solution.

The algorithm of PAM is as follows:

1. Establish number of clusters k.
2. Randomly select k of the n observations as the medoids.
3. Associate each observation to the closest medoid.
4. For each medoid m and each data point o associated to m swap m and o and compute the total cost of the configuration (that is, the total distance of o to all the data points associated to m):

$$c(o,m) = \sum_{j \in M} d_{o,j} \tag{6}$$

where: $c(o,m)$ – total cost of configuration,

M – set of indices describing objects assigned to m.

After that select the medoid o with the lowest cost of the configuration.

5. Repeat steps 3 and 4 until there is no change in the assignments.

The correctness of assigning objects to different classes can be evaluated on the basis of homogeneity and heterogeneity of the clusters. A silhouette width is a coefficient designed to judge the quality of the classification for a fixed number of classes[13]:

$$SC = \frac{1}{n} \sum_{l=1}^{k} \sum_{i=1}^{n_l} s_{l_i} \tag{7}$$

where: $SC \in [-1, 1]$ – silhouette coefficient, while:

$$s_{l_i} = \begin{cases} \dfrac{b_{l_i} - a_{l_i}}{\max(b_{l_i}, a_{l_i})}, & \text{for } n_l > 1 \\ 0, & \text{for } n_l = 1 \end{cases} \tag{8}$$

a_{l_i} – average distance of i^{th} object belonging to l^{th} cluster to other objects from the same l^{th} cluster:

$$a_{l_i} = \frac{1}{\#L - 1} \sum_{j \in L} d_{i,j} \tag{9}$$

L – set of indices describing objects belonging to l^{th} cluster,

$\#L$ – number of objects into l^{th} cluster,

$$b_{l_i} = \min_{k \neq l}(b_{l_i,k}) \tag{10}$$

$b_{l_i,k}$ – average distance of i^{th} object belonging to l^{th} cluster to all objects from k^{th} cluster:

$$b_{l_i,k} = \frac{1}{\#K} \sum_{j \in K} d_{i,j} \tag{11}$$

K – set of indices describing objects belonging to k^{th} cluster,

$\#K$ – number of objects into k^{th} cluster.

The greater the SC (closer to 1), the better the quality of classification.

1.3. Verification of suitability of patterns

The main objective of the identification of a price pattern is its practical use in making investment decisions – to buy or to sell shares. Therefore, the emergence of a particular pattern should indicate that a stock price should behave in a predictable way in the oncoming periods. The author proposes to verify whether clusters – patterns give signals as to the possible shares purchase or sale in the following steps:

1. Verification of the significance of price changes after the chosen length of time for each cluster on the basis of the data used for classification (training set). This assessment is the result of the acceptance or rejection of the null hypothesis of the equality of the average price change in the cluster and in the population. The null and the alternative hypotheses take the form:

$$\begin{aligned} H_0 &: \quad \bar{x}_{t,l} = \mu_t \\ H_1 &: \quad \bar{x}_{t,l} \neq \mu_t \end{aligned} \tag{12}$$

where:

μ_t – average price change after t sessions,

$\bar{x}_{t,l}$ – average price change after t sessions in the l^{th} cluster.

The test statistic takes the form[14]:

$$z = \frac{\bar{x}_{t,l} - \mu_t}{s_{t,l}} \sqrt{n_l}\,, \tag{13}$$

where $s_{t,l}$ – standard deviation of the price change after t sessions in the l^{th} cluster.

On the basis of the central limit theorem the test statistic z for large n_l is asymptotically converging to the standard normal distribution.

2. Assignment of new data to the appropriate clusters, i.e. carrying out a discrimination of the test set. Before the test set is discriminated, variables must be standardized. In order not to introduce unnecessary distortions, the author proposes to carry out the standardization of the test set with the use of parameters (mean and standard deviation) obtained for the training set. Discrimination may be performed using the centers of gravity of each cluster. In such manner, new observation is assigned to the cluster whose center of gravity is the nearest to this observation.

3. Confirmation or rejection of the relevance of price changes indicated in point 1 on the basis of new data – the test set. Again, the hypothesis about equality of the average price change in the cluster and in the population is verified. The procedure is analogous to the one

from point 1, with the difference that now the population forms the test set. When, for the given cluster, the test set confirms the properties of the training set – for example, after some period in both sets there is a significant increase in the price – it can be said that on the basis of that cluster, it is possible to find a pattern to predict the stock price change.

4. The evaluation of the economic result of using the signals given by the found patterns. If in step 3 significant price changes in training and test sets are confirmed, practical verification of the suitability of the investment strategy using these patterns may be performed. This evaluation should be carried out on data not previously included in the training or test sets – and therefore requires a second, independent test set.

2. Empirical research

The survey covers following companies: Alcoa, Boeing, Caterpillar, Coca Cola, Disney, General Electric and IBM in the period from 2 January1962 to 24 August 2012. The choice of companies is narrowed to few companies belonging to the DJI index with a relative long history. The data are adjusted for dividends and splits. Vector of features include following variables:

a) connected with volatility:
   - the ratio of the closing to the opening price,
   - the ratio of the maximum to the minimum price,
   - the ratio of the closing price to the maximum price,
   - the ratio of a 10-session closing price moving standard deviation to the 10-session closing price moving average (10-session moving coefficient of variation):

$$v_{t,k} = \frac{s_{t,k}}{\bar{x}_{t,k}}, \ \mathrm{k} = 10 \tag{14}$$

   where:

   $\bar{x}_{t,k}$ – k-session moving average for t^{th} session,

   $s_{t,k}$ – k-session moving standard deviation for t^{th} session;

b) connected with the price trend behavior in the recent time:
   - number of sessions from the highest closing price during last 20 sessions,
   - number of sessions from the highest closing price during last 10 sessions,
   - number of sessions from the highest closing price during last 5 sessions,
   - number of sessions from the lowest closing price during last 20 sessions,
   - number of sessions from the lowest closing price during last 10 sessions,
   - number of sessions from the lowest closing price during last 5 sessions,

– percentage change in the closing price per session,
– percentage change in the closing price over 5 sessions,
– percentage change in the closing price over 10 sessions;

c) connected with the volume:
– the ratio of the current volume to the average volume of the last 300 trading sessions,
– the ratio of the average volume of 10 sessions for the average volume of the last 300 sessions;

d) stock exchange indicators:
– *RSI* – relative strength index:

$$RSI_n = \begin{cases} 100 - \dfrac{100}{1 - \dfrac{EMA_n(U,27)}{EMA_n(D,27)}} & \text{for } EMA_n(D,27) \neq 0 \\ 100 & \text{for } EMA_n(D,27) = 0 \end{cases} \tag{15}$$

where:

$$U_t = \max(close_t - close_{t-1}; 0) \tag{16}$$

$$D_t = \max(close_{t-1} - close_t; 0) \tag{17}$$

$$EMA_t(X,k) = \begin{cases} X_1 & \text{for } t = 1 \\ \dfrac{2}{k+1} X_t + \left(1 - \dfrac{2}{k+1}\right) EMA_t(X,k) & \text{for } t > 1 \end{cases} \tag{18}$$

– *PPO* – percentage price oscillator:

$$PPO_n = \frac{EMA_n(close, 12)}{EMA_n(close, 26)} - 1 \tag{19}$$

– signal *PPO*:

$$signal\ PPO_n = EMA_n(PPO, 9) \tag{20}$$

The first 299 observations for each variable are used only to calculate the starting values (the average volume of 300 sessions also include the present value), and then are put away. The next set of 3,000 observations forms the training set, followed by a test set of 3,000 observations. The rest of the observations is left for the second test set – to be used for the evaluation of the economic result of the investment made in accordance to the signals generated by the training set and confirmed by the first test set. The number of classes is set arbitrarily at k = 25. The classification is carried out using the method of PAM. Calculations are performed with the use of the program of 'Taksonomia numeryczna' by Kolenda, which is an add-on to the work[15]. Table 1 shows the *SC* coefficients obtained after classification for each company. The quality of the classification should be regarded as very low – *SC* values are below 0.2.

Table 1. Silhouette widths after classification of training sets

Company	Alcoa	Boeing	Caterpillar	Coca Cola	Disney	GE	IBM
Silhouette width *SC*	–0.07	0.08	0.10	0.12	0.11	0.10	0.10

Source: own calculations.

Low quality of classification is not sufficient to cancel pattern searching. A silhouette width reflects the average quality classification for all clusters. For the pattern analysis it is not necessary or even expected that during each session (for each vector of observations) buy or sell signal is generated. It is enough that some separate cluster are homogeneous and generates a specific signal. Thus only Alcoa company is excluded from the further research due to the negative value of *SC*.

For most clusters the average change rate in the analyzed intervals did not differ significantly from the average change rate for the whole population. However, based on the training set one could also find clusters in which the average changes are significantly different from zero. The level of significance is set to $\alpha = 0.05$. The summary of the results for the training set in the case of Boeing is presented in the Table 2. Price changes for which the p-value is less than 0.05 are marked in bold. These values should be considered as significantly different from the average determined for all observations. Therefore, the vector of observations assigned to the cluster number 5 brings (precedes) an increase in the stock price on average of 1.44% after five sessions and 2.26% after 10 sessions.

Table 2. Results of the Boeing training set (fragment)

Cluster number		5	8	9	11	19	24	25
Number of observations in the cluster		142	121	105	84	126	125	247
Average price change after 1 session	%	0.22	–0.05	0.12	**–0.90**	–0.01	0.40	0.26
	p-value	0.293	0.722	0.699	**0.000**	0.810	0.085	0.050
Average price change after 5 sessions	%	**1.44**	**–1.18**	**1.08**	–0.76	**–0.84**	**1.55**	**0.92**
	p-value	**0.002**	**0.023**	**0.027**	0.181	**0.012**	**0.015**	**0.009**
Average price change after 10 sessions	%	**2.26**	–1.56	**1.61**	–0.73	**–1.60**	1.61	**1.04**
	p-value	**0.000**	0.070	**0.036**	0.387	**0.006**	0.111	0.020

Source: own calculations.

Basing on the obtained results it can be concluded that some patterns are more suitable for short-term forecasting – for the next session – while others are so for longer periods. For the Boeing company, on the basis of the training set, there are 7 signals indicating an increase and 4 indicating a decrease in shares price.

The summary of the results of a test set of Boeing is presented in the Table 3. The test set in most cases does not confirm the results obtained for the training set. Only in the case of the cluster number 9 for 10 sessions ahead, both in the training and the test sets, a significant increase in the price is observed. It should be noted that for a stock trader what is more important than the number of signals is their credibility. In this case, on the basis of observations belonging to the cluster No. 9, there can be determined a pattern preceding growth after 10 sessions. This pattern is established by the center of gravity of the cluster.

Table 3. Results of the Boeing test set (fragment)

Cluster number		5	8	9	11	19	24	25
Number of observations in the cluster		274	71	98	68	125	171	267
Average price change after 1 session	%	0.11	–0.20	0.54	0.00	0.23	–0.11	0.06
	p-value	0.945	0.186	0.081	0.658	0.539	0.143	0.611
Average price change after 5 sessions	%	0.46	0.81	1.33	0.13	–0.06	0.66	0.67
	p-value	0.667	0.673	0.170	0.478	0.148	0.834	0.717
Average price change after 10 sessions	%	0.83	–0.76	**3.30**	–0.48	–0.56	1.06	0.85
	p-value	0.699	0.293	**0.000**	0.226	0.106	0.440	0.697

Source: own calculations.

Table 4. Number of patterns fund for the investigated companies

Company	Number of patterns generating signals for:		
	1 session ahead	5 sessions ahead	10 sessions ahead
Boeing	–	–	1
Caterpillar	1	–	1
Coca Cola	1	–	–
Disney	2	–	–
GE	–	–	–
IBM	–	1	1

Source: own calculations.

The survey of other companies shows that only for GE case it is impossible to identify any pattern that would generate a reliable signal to buy or sell. Summary of discovered patterns for the investigated companies is shown in the Table 4.

The final stage of the study is to evaluate the return on investment which is the implementation of the strategy based on the signals generated by the identified patterns. This evaluation is conducted on the basis of the data from the second test set. Investment assumptions are as follows:

– for each of the 5 companies the initial investment amount is $ 100,
– both long or short positions are allowed,
– there are no transaction costs,
– after the signal is received, the total amount is invested. After the investment period the position is completely closed.

Table 5. The results of investment using signals generated by the patterns

Company	Boeing	Caterpillar	Coca Cola	Disney	IBM
Final value of investment ($)	176.31	210.38	92.06	60.54	153.82
TRR (%)	76.3	110.4	–7.9	–39.5	53.8
NoID	600	1911	550	1449	1240
$\overline{YRR}*$ (%)	26.9	10.3	–3.7	–8.4	9.1

Source: own calculations.

Table 5 summarizes the results obtained for the investigated companies. The investment in 3 companies has given a profit while in 2 it has been a loss. The capital investment of the initial $ 500 has brought in total $ 693. The theoretical average annual rate of return for each company is calculated based on the number of trading days in which the position was open and on the assumption that during a year there are 252 trading days:

$$\overline{YRR}* = (1+TRR)^{\frac{252}{NoID}} - 1 \tag{21}$$

where: $\overline{YRR}*$ – theoretical average yearly rate of return, *TRR* – total rate of return, *NoID* – number of investment days (days with an open position). The theoretical average annual rate of return is an approximate measure that shows what would be the average annual rate of return if there were no breaks in investment – if we could invest at the given efficiency constantly during the whole period.

As a benchmark to the proposed method, signals generated by percentage price oscillator are used[16]. A buying signal is generated when *PPOn* (equation 19) is negative and crosses *signal PPOn* (equation 20) from underneath. A selling signal is generated when *PPOn* is positive and crosses *signal PPOn* from above. Table 6 presents the results obtained using percentage price oscillator. In this case initial investment of $ 600 produced $ 1625 as a result.

Table 6. The results of investment using signals generated by percentage price oscillator (PPO)

Company	Alcoa	Boeing	Caterpillar	Coca Cola	Disney	IBM
Final value of investment ($)	71.2	106.9	456.1	307.7	168.4	514.4
TRR (%)	–28.8	6.9	356.1	207.7	68.4	414.4
NoID	2773	2773	2726	2626	2713	2773
$\overline{YRR}*$ (%)	–3.0	0.6	15.1	11.4	5.0	16.0

Source: own calculations.

Conclusions

In the shown example the result of the activity in accordance to the above procedure is the profit of a total rate of return of 38.6%. Still, the result should be approached with caution since the investments in two out of five companies have produced losses. Moreover, the benchmark method beats it with the total rate of return at 170.8%. It should also be noted that the analysis was conducted without transaction costs, which are always incurred in economic reality.

The poor results does not mean, however, that the procedure is useless and should be abandoned. However it definitely requires improvements. The main problem to be solved seems to be enhancing the classification quality of a training set. Key issues are the selection of variables to the vector of features describing a financial instrument, the choice of classification method, and the final number of clusters. It may also be relevant to make an analysis without observations for which a cluster assignment is not certain – observations that are located far from a model pattern.

Notes

1 See for example Murphy (1999), p. 5.

2 Fama (1970).

3 E.g. Murphy (1999); Edwards, Magee (2001), Pring (1998); Malkiel (2003); Jajuga (2007).

4 For example: Leigh, Purvis, Ragusa (2002); Leigh, Paz, Purvis (2002); Wang, Chan (2007); Liu, Kwong (2007).

5 Szanduła (2011).

6 See for example Murphy (1999), p. 2.

7 Ibidem, p. 100.

8 Szanduła (2011).

9 The review of taxonomic methods can be found for example in Rencher (2002); Tan, Steinbach, Kumar (2006); Grabiński (1992); Pociecha et al. (1988).

10 Euclidean distances can be also replaced with other measures of distance, see Hamming, Chebyshev or Mahalanobis.

11 Kaufman, Rousseeuw (1987).

[12] See for example Hardy (1996); Herbin, Bonnet, Vautrot (2001); Cheong, Lee (2008).

[13] Rousseeuw (1987).

[14] See for example Aczel, Sounderpandian (2008); Ostasiewicz, Rusnak, Siedlecka (2006).

[15] Kolenda (2006).

[16] Appel (2005).

References

Aczel, A.D. & Sounderpandian, J. (2008). *Complete Business Statistics*. Irwin Professional.

Cheong, M.-Y. & Lee, H. (2008). Determining the number of clusters in cluster analysis. *Journal of the Korean Statistical Society*, 37 (2), 135–143. DOI: 10.1016/j.jkss.2007.10.004.

Appel, G. (2005). *Technical Analysis: Power Tools for Active Investors*. Financial Times Prentice Hall, Upper Saddle River.

Edwards, R.D. & Magee J. (2001). *Technical Analysis of Stock Trends*. 8th edition, London–New York–Washington D.C.: St. Lucie Press, Boca Raton.

Fama, E.F. (1970). Efficient capital markets: A review of empirical work. *Journal of Finance*, 25 (2), 383–417. DOI: 10.1111/j.1540-6261.1970.tb00518.x.

Grabiński, T. (1992). *Metody taksonometrii*. Kraków: Wydawnictwo Akademii Ekonomicznej w Krakowie.

Hardy, A. (1996). On the number of clusters. *Computational Statistics & Data Analysis*, 23 (1), 83–96. DOI: 10.1016/S0167-9473(96)00022-9.

Herbin, M., Bonnet, N. & Vautrot, P. (2001). Estimation of the number of clusters and influence zones. *Pattern Recognition Letters*, 22 (14), 1557–1568. DOI: 10.1016/S0167-8655(01)00103-9.

Jajuga, K. (2007). *Podstawy inwestowania na giełdzie papierów wartościowych*. Warszawa: Giełda Papierów Wartościowych w Warszawie.

Kaufman, L. & Rousseeuw, P.J. (1987). *Clustering by means of Medoids*. In: Ed. Y. Dodge, *Statistical Data Analysis Based on the L1–Norm and Related Methods* (pp. 405–416). North-Holland.

Kolenda, M. (2006). *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.

Leigh, W., Paz, N. & Purvis, R. (2002). Market timing: A test of a charting heuristic. *Economics Letters*, 77, 55–63. DOI: 10.1016/S0165-1765(02)00110-6.

Leigh, W., Purvis, R. & Ragusa, J.M. (2002). Forecasting the NYSE composite index with technical analysis, pattern recognizer, neural network, and genetic algorithm: A case study in romantic decision support. *Decision Support Systems*, 32 (4), 361–377. DOI: 10.1016/S0167-9236(01)00121-X.

Liu, J.N.K. & Kwong, R.W.M. (2007). Automatic extraction and identification of chart patterns towards financial forecast. *Applied Soft Computing*, 7, 1197–1208. DOI: 10.1016/j.asoc.2006.01.007.

Malkiel, B.G. (2003). *Błądząc po Wall Street. Dlaczego nie można wygrać z rynkiem*. Warszawa: WIG-Press.

Murphy, J.J., (1999). *Technical Analysis of the Financial Markets: A Comprehensive Guide to Trading Methods and Applications*. Paramus, NJ: New York Institute of Finance.

Ostasiewicz, S., Rusnak, Z. & Siedlecka, U. (2006). *Statystyka: elementy teorii i zadania*. Wrocław: Wydawnictwo Akademii Ekonomicznej we Wrocławiu.

Pociecha, J., Podolec, B., Sokołowski, A. & Zając, K. (1988). *Metody taksonomiczne w badaniach społeczno-ekonomicznych*. Warszawa: Państwowe Wydawnictwo Naukowe.

Pring, M.J. (1998). *Podstawy analizy technicznej*. Warszawa: WIG-Press.

Rencher, A.C. (2002). *Methods of Multivariate Analysis*. New York: John Wiley & Sons. DOI: 10.1002/0471271357.

Rousseeuw, P.J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. DOI: 10.1016/0377-0427(87)90125-7.

Szanduła, J. (2011). Wyszukiwanie formacji w kursach giełdowych przy użyciu metod klasyfikacji danych. *Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, 185, 82–93.

Tan, P., Steinbach, M. & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Addison Wesley.

Wang, J.-L. & Chan, S.-H. (2007). Stock market trading rule discovery using pattern recognition and technical analysis. *Expert Systems with Applications*, 33 (2), 304–315. DOI: 10.1016/j.eswa.2006.05.002.