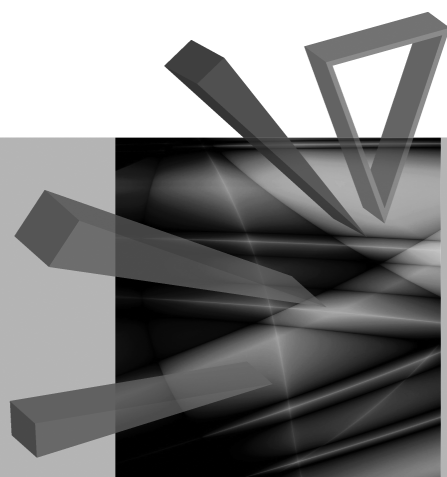


Prognozowanie w zarządzaniu firmą



Redaktorzy naukowi
Paweł Dittmann
Aleksandra Szpulak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2011

Senacka Komisja Wydawnicza

Zdzisław Pisz (przewodniczący),

Andrzej Bąk, Krzysztof Jajuga, Andrzej Matysiak, Waldemar Podgórski,

Mieczysław Przybyła, Aniela Styś, Stanisław Urban

Recenzenci

Włodzimierz Szkutnik, Jan Zawadzki

Redakcja wydawnicza

Barbara Majewska

Redakcja techniczna i korekta

Barbara Łopusiewicz

Skład i łamanie

Comp-rajt

Projekt okładki

Beata Dębska

Kopiowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2011

ISSN 1899-3192

ISBN 978-83-7695-141-6

Druk: Drukarnia TOTEM

Spis treści

Wstęp	7
Agnieszka Przybylska-Mazur: Optymalne zasady polityki pieniężnej w prognozowaniu wskaźnika inflacji	9
Alicja Wolny-Dominiak: Zmodyfikowana regresja Poissona dla danych ubezpieczeniowych z dużą liczbą zer	21
Andrzej Gajda: Doświadczenia i metody pozyskiwania danych eksperckich na potrzeby badań z wykorzystaniem metod foresight	30
Anna Gondek: Prognozy rozwoju gospodarczego Polski z użyciem metody analogii przestrzenno-czasowych	41
Bartosz Lawędziak: Sekurytyzacja papierów wartościowych opartych na hipotece odwrotnej	50
Filip Chybalski: Prakseologiczne aspekty prognozowania	59
Ireneusz Kuroпка, Paweł Lenczewski: Możliwość zastosowania modeli ekonometrycznych do prognozowania w przedsiębiorstwie Brenntag Polska	69
Jacek Szandula: Wyszukiwanie formacji w kursach giełdowych przy użyciu metod klasyfikacji danych	82
Joanna Perzyńska: Zastosowanie sztucznych sieci neuronowych do wyznaczania nieliniowych prognoz kombinowanych	94
Konstancja Poradowska, Tomasz Szkutnik, Mirosław Wójciak: Scenariusze rozwoju wybranych technologii oszczędności energii w życiu codziennym	102
Maciej Oesterreich: Wykorzystanie pakietu statystycznego R w prognozowaniu na podstawie danych w postaci szeregów czasowych z wahaniami sezonowymi	113
Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel: Algorytm zgodnego modelowania i prognozowania procesów ekonomicznych jako pakiet funkcji <i>Congruent Specification</i> programu Gretl	125
Marcin Błażejowski: Stacjonarność szeregów czasowych o wysokiej częstotliwości obserwowania – implementacja testu stacjonarności Dickeya w programie Gretl	137
Mirosław Wójciak: Wpływ czynników i zdarzeń kluczowych na rozwój nowych technologii – wybrane metody korygowania prognoz na przykładzie technologii energooszczędnych	149
Monika Dyduch: Grupowanie produktów strukturyzowanych	159
Piotr Bernat: Planowanie działalności przedsiębiorstwa wspomagane prognozowaniem	170

Roman Pawlukowicz: Informacje prognostyczne w rynkowych sposobach wyceny nieruchomości – identyfikacja i pozyskiwanie	182
Wojciech Zatoń: Uwarunkowania psychologiczne w prognozowaniu	189

Summaries

Agnieszka Przybylska-Mazur: Optimal monetary policy rules in forecasting of inflation rate	20
Alicja Wolny-Dominiak: Zero-inflated Poisson Model for insurance data with a large number of zeros	29
Andrzej Gajda: Experience and methods of data collection from experts for research using foresight methods	40
Anna Gondek: Economic growth forecasts for Poland using the time-space analogy method	49
Bartosz Lawędziak: Securitization of survivor bonds based on the reverse mortgage	58
Filip Chybalski: Praxiological aspects of forecasting	68
Ireneusz Kuropka, Paweł Lenczewski: Econometric models usage feasibility in Brenntag Poland forecasting	81
Jacek Szandula: Searching for technical analysis formations in stock prices with the use of cluster analysis methods	93
Joanna Perzyńska: Application of artificial neural networks to build the nonlinear combined forecasts	101
Konstancja Poradowska, Tomasz Szkutnik, Mirosław Wójciak: The scenarios of development of selected technologies related to energy saving in everyday life	112
Maciej Oesterreich: The R application in forecasting unsystematic lacks in seasonal time series	124
Marcin Błażejowski, Paweł Kufel, Tadeusz Kufel: Congruent modelling and forecasting algorithm as function package Congruent Specification in GRETL	136
Marcin Błażejowski: Stationarity of high-frequency time series – implementation of Dickey’s stationarity test in GRETL	148
Mirosław Wójciak: The influence of key and events factors on the development of new technologies – selected methods of forecast correction on the example of energy-saving technologies	158
Monika Dyduch: Ranking of structured products	169
Piotr Bernat: Forecasting assisted business management planning	181
Roman Pawlukowicz: Prognostic data in market ways of property valuation – identification and acquisition	188
Wojciech Zatoń: Psychological aspects of forecasting	199

Jacek Szandula

Uniwersytet Ekonomiczny we Wrocławiu

**WYSZUKIWANIE FORMACJI
W KURSACH GIEŁDOWYCH
PRZY UŻYCIU METOD KLASYFIKACJI DANYCH**

Streszczenie: Artykuł stanowi próbę wprzęgnięcia metod klasyfikacji danych do rozpoznawania formacji giełdowych. Formacja jest tu rozumiana jako pewien powtarzalny wzór, którego pojawienie się oznacza określone konsekwencje. Możliwość tworzenia wzorów nie jest ograniczona jedynie do kursu instrumentu finansowego, ale dotyczy także wolumenu oraz innych wskaźników. W badaniu wykorzystano notowania giełdowe Agory SA. Uwzględniono szereg zmiennych opisujących zachowanie się kursu oraz wolumenu zarówno w krótkim, jak i dłuższym okresie. Do klasyfikacji danych – wyszukiwania wzorów – użyto metody Warda. Przeprowadzono także ocenę możliwości wykorzystania określonych formacji do celów praktycznych.

Słowa kluczowe: klasyfikacja danych, metoda Warda, analiza techniczna, formacja giełdowa.

1. Wstęp

W prognozowaniu kursów giełdowych najczęściej stosuje się analizę fundamentalną oraz analizę techniczną. W analizie fundamentalnej przyjmuje się, że wartość określonego instrumentu finansowego zależy od stanu obiektu gospodarczego, który ten instrument reprezentuje – przedsiębiorstwa przy kursie akcji, gospodarki przy kursie waluty. Analiza techniczna zakłada, że badanie istotnych czynników jest bezcelowe, gdyż wszystkie ważne informacje o obiekcie natychmiast są uwzględniane w bieżącej wycenie instrumentu finansowego. Oba te podejścia są często krytykowane jako mało skuteczne.

Ponadto analizie technicznej zarzuca się, że nie spełnia standardów metody naukowej, przeciwstawiając jej hipotezę rynku efektywnego [Fama 1970].

Mimo zastrzeżeń co do założeń analizy technicznej pozostaje ona dominującym podejściem przy prognozowaniu krótkookresowym kursów akcji. Jednym z jej podstawowych narzędzi jest analiza formacji, która przy formułowaniu prognozy wykorzystuje wykres cen instrumentu finansowego. Korzystający z formacji uzna-

ją, że pojawienie się określonego kształtu geometrycznego w przebiegu wykresu cen skutkuje określonym jego przebiegiem w nadchodzących okresach. Wyznaczenie prognozy opiera się najczęściej na ocenie wzrokowej przebiegu wykresu (np. [Edwards, Magee 2001; Pring 1998; Malkiel 2003; Jajuga 2007; Murphy 2008]) – co wprowadza sporą dozę dowolności w interpretacji analizowanego materiału – lub kształt wykresu analizowany jest z wykorzystaniem technik rozpoznawania obrazu (np. [Leigh, Purvis, Ragusa 2002; Leigh, Paz, Purvis 2002; Wang, Chan 2007; Liu, Kwong 2007]).

Celem artykułu jest weryfikacja przydatności statystycznych metod klasyfikacji danych do wyszukiwania formacji giełdowych. Przedstawiona zostanie propozycja redefinicji pojęcia formacji giełdowej oraz wynikające z tego konsekwencje.

2. Metodologia

2.1. Idea formacji giełdowych

Idea wykorzystania formacji giełdowych wiąże się z koniecznością zaakceptowania podstawowych założeń analizy technicznej:

1. Cena instrumentu finansowego odzwierciedla wpływ wszystkich czynników mogących wpływać na cenę.

2. Na rynku istnieją powtarzające się co pewien czas wzory zachowań.

Istotą i celem analizy technicznej jest umiejętność poprawnego zidentyfikowania takiego wzorca.

Mimo że analiza formacji należy do podstawowych narzędzi analizy technicznej, literatura przedmiotu z reguły pomija definicję tego pojęcia, pozostawiając jego znaczenie definicji słownikowej. Jajuga określa formację jako graficzny wzorek, według którego kształtował się kurs akcji w przeszłości [Jajuga 2007, s. 206]. Uzupełniając tę myśl, można dodać, że klasycznie rozumiana formacja jest pewnym układem, formą geometryczną, wzorem przebiegu wykresu cen instrumentu finansowego.

Wadę takiego rozumienia formacji stanowi ograniczanie się do analizy graficznej wykresu cen instrumentu finansowego. Tymczasem powtarzające się wzory zachowań rynku mogą dotyczyć także na przykład wolumenu czy zmienności. Analiza formacji nie powinna z góry odrzucać możliwości wykorzystania innych cech opisujących instrument finansowy do identyfikacji wzoru. W związku z tym autor proponuje rozszerzyć pojęcie formacji do modelu wzorca wektora cech opisującego instrument finansowy.

Aby analizę formacji można było uznać za przydatną do praktycznych zastosowań, muszą być spełnione następujące warunki:

1. Musi istnieć możliwość zidentyfikowania formacji – modele wzorców poszczególnych formacji powinny być unikalne, aby w sytuacji wystąpienia w szere-

gu statystycznym podobnego zestawu cech zakwalifikowanie tegoż zestawu do konkretnej formacji było jednoznaczne.

2. Pojawienie się określonej formacji powinno oznaczać, że po niej nastąpi określony skutek (np. wzrost kursu). Nie oznacza to, że ma tu pojawić się związek przyczynowo-skutkowy, wystarczy powiązanie symptomatyczne.

W spełnieniu pierwszego warunku powinny pomóc metody klasyfikacji i dyskryminacji danych. Wykorzystanie metod klasyfikacji danych ma za zadanie wyodrębnić jednorodnych (podobnych) grup wektorów cech instrumentu finansowego. Na podstawie każdej z grup ustalony zostanie model wzorca formacji. Po ustaleniu modeli wzorców formacji przyporządkowanie poszczególnych sytuacji empirycznych do konkretnych formacji będzie możliwe dzięki metodom dyskryminacyjnym.

Kluczowe znaczenie dla jakości przeprowadzonej analizy ma wybór cech diagnostycznych – zmiennych. Zmienne użyte w badaniu powinny reprezentować możliwie szerokie spektrum informacji związanych z badanym instrumentem finansowym. Wśród nich można rozważyć np.:

- a) bieżący poziom kursu,
- b) zmienność kursu na ostatniej sesji,
- c) zmienność kursu na poprzednich sesjach,
- d) względną zmianę kursu w określonych przedziałach czasowych,
- e) liczbę dni od lokalnych maksimów i minimów,
- f) trend kursu w określonych przedziałach czasowych,
- g) obroty na ostatniej sesji,
- h) średnie obroty w ciągu poprzednich sesji.

2.2. Klasyfikacja danych

Istnieje wiele metod taksonomicznych, umożliwiających wyodrębnić grupy obiektów podobnych ze względu na wskazane cechy¹. Niezależnie od rozpatrywanej metody konieczność zrównania zakresów zmienności zmiennych narzuca przeprowadzenie procedury normalizacyjnej, na przykład standaryzacji:

$$x'_{i,j} = \frac{x_{i,j} - \bar{x}_j}{s_j}, \quad (1)$$

gdzie: $x'_{i,j}$ – standaryzowana wartość zmiennej X_j w i -tym momencie,

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

n – liczba obserwacji,

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

¹ Przegląd metod taksonomicznych można znaleźć np. w [Grabiński 1992; Pocięcha i in. 1988].

- m – liczba zmiennych,
 $x_{i,j}$ – wartość zmiennej X_j w i -tym momencie,
 \bar{x}_j – średnia wartość zmiennej X_j :

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

s_j – odchylenie standardowe zmiennej X_j :

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

Podobieństwo zmiennych na ogół określane jest poprzez ich wzajemną odległość². Im odległość między zmiennymi mniejsza, tym większe podobieństwo. W kolejnym kroku należy wyznaczyć macierz odległości:

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

której elementy są odległościami euklidesowymi³ między wektorami standaryzowanych wartości poszczególnych obserwacji:

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

Dysponując macierzą odległości, można przystąpić do klasyfikacji obserwacji. Autor proponuje postępowanie dwuetapowe. W pierwszym kroku obserwacje dzielone będą z wykorzystaniem metody Warda [1963], a wstępnie uzyskany podział na zbiory zostanie następnie zoptymalizowany metodą środków ciężkości [Beale 1969; Sparks 1973; Sparks 1975].

Metoda Warda należy do aglomeracyjnych metod klasyfikacji danych. Punktem wyjścia jest zbiór n jednoobiektowych skupień. Są one sukcesywnie łączone w skupienia wyższego rzędu, aż do uzyskania jednego skupienia obejmującego

² Inne sposoby pomiaru podobieństwa zmiennych można znaleźć np. w [Cieślak, Jasiński 1979; Strahl 1990; Szandula 2002].

³ Odległości euklidesowe można w razie potrzeby zastąpić innymi miarami odległości, np. Hamminga, Czebyszewa czy Mahalanobisa.

wszystkie obiekty. W pierwszym kroku wyszukiwane są pary skupień p i q najmniej odległe od siebie [Ward, Hook 1963]:

$$d_{pq} = \min_{i,j} \{d_{i,j}\} \quad i < j; j = 2, \dots, n. \quad (6)$$

Następnie skupienia p i q łączone są w jedno nowe skupienie p , a skupienie q usuwa się. Dla nowo uzyskanego skupienia p wyznacza się odległości od wszystkich pozostałych skupień według wzoru:

$$d'_{pr} = \frac{(n_p + n_r)d_{pr} + (n_q + n_r)d_{qr} - n_r d_{pq}}{n_p + n_q + n_r}, \quad (7)$$

gdzie: d'_{pr} – odległość nowego skupienia od skupienia r ,

d_{pr} – odległość pierwotnego skupienia p od skupienia r ,

d_{qr} – odległość pierwotnego skupienia q od skupienia r ,

d_{pq} – wzajemna odległość pierwotnych skupień p i q ,

n_p – liczebność pierwotnego skupienia p ,

n_q – liczebność pierwotnego skupienia q ,

n_r – liczebność skupienia r .

Na początku kolejnej iteracji dla wszystkich r przyjmuje się, że

$$d_{pr} = d'_{pr}. \quad (8)$$

Powyższe kroki powtarzane są, aż do uzyskania jednego skupienia.

Jednym z najważniejszych etapów klasyfikacji danych jest ustalenie ostatecznej liczby klas (skupień). Spośród wielu prac podejmujących ten problem (zob. np. [Hardy 1996; Herbin i in. 2001; Cheong, Lee 2008]) żadna nie daje jednoznacznej wskazówki co do właściwego podejścia. Dopuszczalnym rozwiązaniem jest arbitralne ustalenie liczby klas.

Po ustaleniu k początkowych skupień można przystąpić do optymalizacji podziału. W przypadku optymalizacji metodą środków ciężkości należy określić maksymalną liczbę iteracji. Dla każdego ze skupień wyznaczany jest środek ciężkości:

$$\bar{\mathbf{c}}_l = \frac{1}{n_l} \sum_{i=1}^{n_l} \mathbf{x}_{l_i}, \quad (9)$$

gdzie: $\bar{\mathbf{c}}_l$ – środek ciężkości l -tej klasy, $l = 1, 2, \dots, k$,

n_l – liczebność l -tej klasy,

\mathbf{x}_{l_i} – i -ty wektor obserwacji należący do l -tej klasy:

$$\mathbf{x}_{l_i} = \begin{pmatrix} x'_{l_i,1} \\ x'_{l_i,2} \\ \vdots \\ x'_{l_i,m} \end{pmatrix}_{m \times 1}. \quad (10)$$

Środki ciężkości stają się jądrami koncentracji. Optymalizacja polega na przyporządkowaniu poszczególnych wektorów obserwacji do najbliższego jądra koncentracji. W ten sposób powstaje nowy podział na k skupień. W następnej iteracji ponownie wyznacza się jądra koncentracji i dokonuje optymalizacji. Kroki powtarza się, dopóki nowy podział nie przynosi zmian lub nie zostanie osiągnięta maksymalna liczba iteracji.

Jakość klasyfikacji zależy od homogeniczności i heterogeniczności skupień. Syntetycznym miernikiem umożliwiającym ocenę poprawności klasyfikacji dla ustalonej liczby klas jest tzw. szerokość sylwetki (*silhouette width*) [Rousseeuw 1987]:

$$\bar{s} = \frac{1}{n} \sum_{l=1}^k \sum_{i=1}^{n_l} s_{l_i}, \quad \bar{s} \in [-1, 1], \quad (11)$$

gdzie \bar{s} – szerokość sylwetki, przy czym:

$$s_{l_i} = \begin{cases} \frac{b_{l_i} - a_{l_i}}{\max(a_{l_i}, b_{l_i})}, & \text{dla } n_l > 1 \\ 0, & \text{dla } n_l = 1 \end{cases}, \quad (12)$$

gdzie: a_{l_i} – średnia odległość wektora \mathbf{x}_{l_i} od pozostałych wektorów obserwacji w l -tej klasie,

$$b_{l_i} = \min_{j \neq l} (b_{l_i, j}), \quad (13)$$

$b_{l_i, j}$ – średnia odległość wektora \mathbf{x}_{l_i} od wszystkich wektorów obserwacji w j -tej klasie.

Im szerokość sylwetki \bar{s} jest większa (bliższa jedności), tym lepsza jakość klasyfikacji.

2.3. Weryfikacja przydatności formacji giełdowych

Wyszukiwanie i określanie formacji będzie miało sens, jeżeli choćby część z nich ułatwi podjęcie decyzji co do zachowań inwestycyjnych – kupna lub sprzedaży. Oznacza to, że formacja powinna poprzedzać określone zachowanie się kursu in-

strumentu finansowego. Autor proponuje zweryfikować, czy poszczególne klasy dają sygnały co do ewentualnego kupna lub sprzedaży w następujący sposób:

1. Ocena istotności zmiany kursu po upływie wybranej długości okresu dla każdej klasy przeprowadzona na podstawie danych użytych do klasyfikacji (zbioru uczącego). Ocena ta polega na przyjęciu lub odrzuceniu hipotezy o równości średniej zmiany kursu w danej klasie do średniej zmiany kursu w badanej populacji. Hipoteza zerowa i alternatywna przyjmuje postać:

$$\begin{aligned} H_0: \bar{y}_{t,l} &= \bar{y}_t \\ H_1: \bar{y}_{t,l} &\neq \bar{y}_t \end{aligned} \quad (14)$$

gdzie: \bar{y}_t – średnia zmiana kursu w badanej populacji po t sesjach,
 $\bar{y}_{t,l}$ – średnia zmiana kursu w l -tej klasie po t sesjach.

Statystyka testowa, przy założeniu normalności rozkładu badanej zmiennej, przyjmuje postać (zob. np. [Ostasiewicz i in. 2006]):

$$z = \frac{\bar{y}_{t,l} - \bar{y}_t}{s_{t,l}} \sqrt{n_l}, \quad (15)$$

gdzie $s_{t,l}$ – odchylenie standardowe kursu w l -tej klasie po t sesjach.

Statystyka z dla dużych n_l ma rozkład asymptotycznie zbieżny do standardowego rozkładu normalny. Dla

$$|z| \leq z^*, \quad (16)$$

nie ma podstaw do odrzucenia hipotezy H_0 . W przypadku gdy

$$|z| > z^*, \quad (17)$$

hipotezę H_0 należy odrzucić na rzecz H_1 . Wartość krytyczną z^* odczytuje się ze standardowego rozkładu normalnego dla przyjętego poziomu istotności α . Alternatywnie zamiast wyznaczać wartość krytyczną z^* można wyznaczyć p -value:

$$p\text{-value} = 2 \cdot P\{\phi > |z|\}, \quad (18)$$

przy czym $\phi \sim N(0,1)$ – jest zmienną o standardowym rozkładzie normalnym.

Im mniejsza jest wartość p -value, tym mniejsze prawdopodobieństwo popełnienia błędu statystycznego pierwszego rodzaju, czyli odrzucenia hipotezy H_0 wówczas, gdy jest ona prawdziwa.

2. Przyporządkowanie nowych danych do odpowiednich klas, tj. przeprowadzenie dyskryminacji zbioru testowego. Przed przystąpieniem do dyskryminacji zbioru testowego zmienne należy poddać standaryzacji. Aby nie wprowadzać niepotrzebnych zniekształceń w obrazie danych, autor proponuje przeprowadzać stan-

daryzację wartości zbioru testowego parametrami (średnią i odchyleniem standardowym) uzyskanymi dla zbioru uczącego. Dyskryminację można przeprowadzić, wykorzystując środki ciężkości poszczególnych klas. Nowy wektor obserwacji przyporządkowany zostanie do klasy, której środek ciężkości jest mu najbliższy.

3. Ocena istotności odchyleń od średniej zmian kursu i obrotu po upływie wybranej długości okresu dla każdej klasy przeprowadzona na podstawie zbioru testowego. Przebieg postępowania jest taki jak w punkcie 1, ale z tą różnicą, że populację stanowi teraz zbiór testowy.

4. Ocena skutków ekonomicznych podjętych decyzji. W przypadku występowania statystycznie istotnych zmian poziomu kursu w zbiorach uczącym i testowym można przystąpić do weryfikacji przydatności strategii inwestycyjnej wykorzystującej formacje giełdowe.

3. Badanie empiryczne

W badaniu wykorzystane zostały dane dotyczące spółki Agora SA, notowanej na Giełdzie Papierów Wartościowych w Warszawie. Dane obejmują okres od 17.11.2000 do 12.08.2010 r. Uwzględniono zmienne:

- a) określające zmienność kursu:
- stosunek kursu zamknięcia do kursu otwarcia,
 - stosunek kursu maksymalnego do kursu minimalnego,
 - stosunek kursu zamknięcia do kursu minimalnego,
 - stosunek ruchomego 10-sesyjnego odchylenia standardowego kursu zamknięcia do 10-sesyjnej średniej ruchomej kursu zamknięcia (10-sesyjny ruchomy współczynnik zmienności kursu zamknięcia):

$$v_{t,k} = \frac{s_{t,k}}{\bar{x}_{t,k}}, k = 10, \quad (19)$$

gdzie: $\bar{x}_{t,k}$ – średnia ruchoma k -sesyjna na t -tą sesję,

$s_{t,k}$ – k -sesyjne odchylenie standardowe na t -tą sesję;

- b) określające kształtowanie się kursu w ostatnim czasie:
- liczba sesji od najwyższego kursu zamknięcia w ciągu ostatnich 20 dni,
 - liczba sesji od najwyższego kursu zamknięcia w ciągu ostatnich 10 dni,
 - liczba sesji od najwyższego kursu zamknięcia w ciągu ostatnich 5 dni,
 - liczba sesji od najniższego kursu zamknięcia w ciągu ostatnich 20 dni,
 - liczba sesji od najniższego kursu zamknięcia w ciągu ostatnich 10 dni,
 - liczba sesji od najniższego kursu zamknięcia w ciągu ostatnich 5 dni,
 - zmiana kursu zamknięcia z sesji na sesję,
 - zmiana kursu zamknięcia w ciągu 5 sesji,
 - zmiana kursu zamknięcia w ciągu 10 sesji,

- c) związane z obrotem:
- bieżący obrót,
 - średni obrót z 10 sesji.

Obserwacje poszczególnych zmiennych zostały podzielone na dwa rozłączne chronologicznie zbiory: uczący o licznosci 1500 obserwacji i testowy składający się z 934 obserwacji. Dla zbioru uczącego dokonano standaryzacji, klasyfikacji (wzory 1-8). Liczba klas ustalona została arbitralnie na $k = 25$. Mimo dokonania optymalizacji klasyfikacji (wzory 9-10), jej jakość należy określić jako słabą ze względu na niską wartość szerokości sylwetki (wzór 11) – poniżej 0,2. Obliczenia przeprowadzone zostały z wykorzystaniem programu Taksonomia numeryczna autorstwa K. Kolendy, będącego dodatkiem do pracy [Kolenda 2006].

Niska jakość klasyfikacji nie jest wystarczającym powodem, by zakończyć wyszukiwanie formacji. Szerokość sylwetki określa średnią wartość dla wszystkich klas. W przypadku analizy formacji nie jest konieczne ani nawet oczekiwane, aby podczas każdej sesji (dla każdego wektora obserwacji) generowany był sygnał kupna lub sprzedaży. Wystarczy, aby część wyodrębnionych klas była homogeniczna i generowała określony sygnał. Dlatego w kolejnym kroku zbadana została istotność średniej względnej zmiany kursu po jednej, pięciu i dziesięciu sesjach.

W przypadku większości klas średnie zmiany kursu w analizowanych odstępach czasowych nie różniły się istotnie od średniej zmiany kursu dla całej zbiorowości. Jednak w przypadku 10 klas zaobserwowane zostały zmiany, które można uznać za istotne. Środki ciężkości tych klas stanowiąc będą modele wzorców formacji. Zestawienie otrzymanych wyników prezentuje tabela 1. Pogrubioną czcionką zaznaczone zostały zmiany kursów, dla których p -value nie przekracza 0,05. Wartości te należy uznać za istotnie różne od średniej wyznaczonej dla całego okresu. W związku z tym wektor zakwalifikowany do klasy/formacji 12 przynosił (poprzedzał) spadek kursu akcji średnio 2,7-procentowy po pięciu sesjach i 3,5-procentowy po 10 sesjach. Z kolei wektory zakwalifikowane do klasy 23 dawały średnio 0,9-procentowy wzrost kursu po 5 sesjach i 2,5-procentowy po 10.

Tabela 1. Podsumowanie wyników zbioru uczącego

Numer klasy		3	4	7	11	12	17	18	20	23	25
Średnia zmiana kursu po 1 sesji	[%]	-0,4	-0,1	-0,9	0,7	-0,8	-0,1	-0,1	-0,3	-0,1	-0,3
	p -value	0,282	0,949	0,003	0,002	0,056	0,748	0,813	0,539	0,847	0,640
Średnia zmiana kursu po 5 sesjach	[%]	-1,8	1,0	-2,0	0,1	-2,7	-1,7	-1,7	-0,9	0,9	-2,2
	p -value	0,044	0,051	0,002	0,497	0,000	0,006	0,165	0,561	0,043	0,022
Średnia zmiana kursu po 10 sesjach	[%]	-0,9	2,9	-1,8	-0,7	-3,5	-4,0	-3,0	-4,8	2,5	-3,2
	p -value	0,588	0,000	0,178	0,600	0,002	0,002	0,026	0,008	0,000	0,015

Źródło: opracowanie własne.

Na podstawie uzyskanych wyników można stwierdzić, że część formacji lepiej nadaje się do prognozowania krótkookresowego – na następną sesję – a część w przypadku dłuższych okresów – na 10 sesji naprzód. Spośród 15 sygnałów 11 zapowiada spadki, a tylko 4 wzrosty. Możliwe więc, że formacje zapowiadające spadki są łatwiejsze do zidentyfikowania.

Podsumowanie wyników zbioru testowego prezentuje tabela 2. Zbiór testowy nie potwierdził wyników uzyskanych dla zbioru uczącego. Spośród 10 klas i 15 istotnych wskazań w zbiorze uczącym jedynie jeden przypadek w zbiorze testowym był istotny na poziomie 0,05 – klasa 23, prognoza na 10 sesji naprzód. Dwa przypadki (szary kolor w tabeli 2) można by dodatkowo zaakceptować przy odpowiednim zwiększeniu poziomu istotności. W przypadku klasy 3 i prognozy pięciosesyjnej średnia zmiana kursu była istotna, lecz jej kierunek był inny – w zbiorze uczącym nastąpił spadek, natomiast w zbiorze testowym wzrost kursu (szare tło komórki w tabeli 2).

Tabela 2. Podsumowanie wyników zbioru testowego

Numer klasy		3	4	7	11	12	17	18	20	23	25
Średnia zmiana kursu po 1 sesji	[%]	0,5	0,4	-0,6	0,7	-0,5	0,1	-0,6	0,5	-0,2	1,1
	<i>p</i> -value	0,456	0,532	0,228	0,102	0,214	0,882	0,077	0,181	0,529	0,653
Średnia zmiana kursu po 5 sesjach	[%]	2,2	2,3	0,1	0,0	-0,3	0,5	-0,3	2,1	0,0	-0,7
	<i>p</i> -value	0,002	0,266	0,964	0,971	0,786	0,718	0,784	0,077	0,999	0,903
Średnia zmiana kursu po 10 sesjach	[%]	1,3	6,1	-3,3	3,4	-1,0	0,4	0,6	0,5	2,4	1,5
	<i>p</i> -value	0,251	0,065	0,011	0,007	0,524	0,752	0,748	0,731	0,019	0,803

Źródło: opracowanie własne.

4. Wnioski

Próba testowa nie potwierdziła predyktywnych własności poszczególnych formacji, co z punktu praktycznego wykorzystania proponowanego podejścia czyni go bezużytecznym. W związku z tym autor odstąpił od badania oceny skutków ekonomicznych decyzji podjętych na podstawie analizy formacji.

Przykład empiryczny nie potwierdził zasadności koncepcji wykorzystania metod klasyfikacji danych do wyszukiwania formacji giełdowych w ich rozszerzonym rozumieniu. Zdaniem autora nie oznacza to jednak fiaska samej koncepcji. Na uwagę zasługują bowiem choćby własności zbioru uczącego. Gdyby wszystkie klasy były zbiorami przypadkowych wektorów, nie powinny dawać żadnych sygnałów co do dalszego kształtowania się kursu. Na brak potwierdzenia własności zbioru uczącego w zbiorze testowym może składać się szereg przyczyn. Najważniejszą przesłanką niepowodzenia wydaje się niska jakość klasyfikacji zbioru uczącego, która ma wpływ na późniejszą poprawność dyskryminacji zbioru testowego. Poprawę jakości klasyfikacji zbioru uczącego można osiągnąć poprzez zmianę w do-

borze zmiennych diagnostycznych, a także zmianę metody klasyfikacji. Zasadne wydaje się także przeprowadzenie analizy z wyłączeniem obserwacji, których dyskryminacja może być wątpliwa – leżących daleko od wzorca modelu formacji.

Literatura

- Beale E.M.L., *Euclidean cluster analysis*, „Bulletin of the International Statistical Institute” 1969, vol. 43, no. 2, s. 92-94.
- Cheong My-Young, Lee Hakbae, *Determining the number of clusters in cluster analysis*, „Journal of the Korean Statistical Society” 2008, vol. 37, issue 2, s. 135-143.
- Cieślak M., Jasiński R., *Miara podobieństwa funkcji*, „Przegląd Statystyczny” 1979, nr 3/4.
- Edwards R.D., Magee J., *Technical Analysis of Stock Trends*, 8th ed., St. Lucie Press, Boca Raton, London – New York – Washington D.C. 2001.
- Fama E.F., *Efficient capital markets: A review of empirical work*, „Journal of Finance” 1970, vol. 25, no. 2, s. 383-417.
- Grabiński T., *Metody taksonometrii*, Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 1992.
- Hardy A., *On the number of clusters*, „Computational Statistics & Data Analysis” 1996, vol. 23, issue 1, s. 83-96.
- Herbin M., Bonnet N., Vautrot P., *Estimation of the number of clusters and influence zones*, „Pattern Recognition Letters” 2001, vol. 22, issue 14, s. 1557-1568.
- Jajuga K., *Podstawy inwestowania na giełdzie papierów wartościowych*, Giełda Papierów Wartościowych w Warszawie, Warszawa 2007.
- Kolenda M., *Taksonomia numeryczna. Klasyfikacja, porządkowanie i analiza obiektów wielocechowych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006.
- Leigh W., Paz N., Purvis R., *Market timing: A test of a charting heuristic*, „Economics Letters” 2002, vol. 77, s. 55-63.
- Leigh W., Purvis R., Ragusa J.M., *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” 2002, vol. 32, issue 4, s. 361-377.
- Liu J.N.K., Kwong R.W.M., *Automatic extraction and identification of chart patterns towards financial forecast*, „Applied Soft Computing” 2007, vol. 7, s. 1197-1208.
- Malkiel B.G., *Błądząc po Wall Street. Dlaczego nie można wygrać z rynkiem*, WIG-Press, Warszawa 2003.
- Murphy J.J., *Analiza techniczna rynków finansowych*, WIG-Press, Warszawa 2008.
- Ostasiewicz S., Rusnak Z., Siedlecka U., *Statystyka: elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006.
- Pociecha J., Podolec B., Sokołowski A., Zając K., *Metody taksonomiczne w badaniach społeczno-ekonomicznych*, PWN, Warszawa 1988.
- Pring M.J., *Podstawy analizy technicznej*, WIG-Press, Warszawa 1998.
- Rousseeuw P.J., *Silhouettes: A graphical aid to the interpretation and validation of cluster analysis*, „Journal of Computational and Applied Mathematics” 1987, vol. 20, s. 53-65.
- Sparks D.N., *Algorithm AS 58: Euclidean cluster analysis*, „Journal of the Royal Statistical Society” 1973, Series C (Applied Statistics), vol. 22, no. 1, s. 126-130.
- Sparks D.N., *Remark AS R13: A remark on algorithm AS 58: Euclidean cluster analysis*, „Journal of the Royal Statistical Society” 1975, Series C (Applied Statistics), vol. 24, no. 1, s. 160-161.
- Strahl D., *Metody programowania rozwoju społeczno-gospodarczego*, PWE, Warszawa 1990.

- Szandula J., *Metody analogowe w prognozowaniu społeczno-gospodarczym*, Akademia Ekonomiczna we Wrocławiu, Wydział Gospodarki Narodowej, praca doktorska, 2002.
- Wang Jar-Long, Chan Shu-Hui, *Stock market trading rule discovery using pattern recognition and technical analysis*, „Expert Systems with Applications” 2007, vol. 33, issue 2, s. 304-315.
- Ward J.H., *Hierarchical grouping to optimize an objective function*, „Journal of American Statistical Association” 1963, vol. 58, no. 301, s. 236-244.
- Ward J.H., Hook M.E., *Application of an hierarchical grouping procedure to a problem of grouping profiles*, „Educational and Psychological Measurement” 1963, vol. 23, s. 69-81.

SEARCHING FOR TECHNICAL ANALYSIS FORMATIONS IN STOCK PRICES WITH THE USE OF CLUSTER ANALYSIS METHODS

Summary: The article attempts to employ the methods of data classification to identify stock formations. Formation is understood as a repetitive pattern which appearance means certain consequences. The ability to form patterns is not limited only to the price of a financial instrument, but it is also applicable to volume and other indicators. The study uses stock prices of Agora SA. A number of variables describing the behaviour of price and volume of stock in both the short and long term are taken into account. For the data classification – pattern searching – Ward’s method is used. The paper also conducts an evaluation of the possible use of certain formations for practical purposes.