



Instytut Badań Systemowych

Polskiej Akademii Nauk

Seria:

BADANIA SYSTEMOWE

TOM 65

Redaktor Naukowy

Prof. dr hab. inż. Jakub Gutenbaum

Rada Redakcyjna:

1. Prof. dr hab. inż. Olgierd Hryniewicz – *przewodniczący*
2. Prof. dr hab. inż. Jakub Gutenbaum – *redaktor naukowy*
3. Prof. dr hab. inż. Janusz Kacprzyk
4. Prof. dr hab. inż. Tadeusz Kaczorek
5. Prof. dr hab. inż. Roman Kulikowski
6. Doc. dr hab. inż. Marek Libura
7. Prof. dr hab. inż. Krzysztof Malinowski
8. Prof. dr hab. inż. Zbigniew Nahorski
9. Dr. hab. inż. Marek Niezgódka, prof. UW
10. Prof. dr hab. inż. Roman Słowiński
11. Doc. dr hab. inż. Jan Studziński
12. Prof. dr hab. inż. Stanisław Walukiewicz
13. Prof. dr hab. inż. Andrzej Weryński
14. Doc. dr hab. inż. Antoni Żochowski



Instytut Badań Systemowych
Polskiej Akademii Nauk

Antoni Wiliński

**GMDH – metoda grupowania argumentów
w zadaniach zautomatyzowanej predykcji
zachowań rynków finansowych**

Warszawa - Szczecin

Copyright © by Instytut Badań Systemowych PAN
Warszawa 2009

*Dr hab. inż. Antoni Wiliński – prof. nadzw.
Zachodniopomorski Uniwersytet Technologiczny w Szczecinie
Wydział Informatyki
Ul. Żołnierska 49, Szczecin 71-210
Tel. 091- 449 5660 ; fax 091-449 5540*

Autor jest profesorem Zachodniopomorskiego Uniwersytetu Technologicznego, poprzednio Politechniki Szczecińskiej. Habilitację uzyskał na Politechnice Warszawskiej w zakresie automatyki i robotyki. W obecnej kadencji jest dziekanem Wydziału Informatyki ZUT. Od lat zajmuje się problematyką inteligencji obliczeniowej i rozpoznawania wzorców w szeregach czasowych w celach predykcyjnych lub modelowania rzeczywistości. W Katedrze Systemów Multimedialnych, której jest kierownikiem, prowadzone są badania nad webowym systemem handlu automatycznego. Rozważane są rozmaite strategie analizy technicznej specyficzne dla decyzji podejmowanych z wysoką częstotliwością przez infoboty. Wśród tych strategii istotną rolę odgrywają omawiane tu algorytmy oparte na GMDH (Group Method of Data Handling). Praca przeznaczona jest raczej dla specjalistów będących zwolennikami analizy technicznej.

Recenzenci:

Prof. dr hab. inż. Zbigniew Banaszak
Doc. dr hab. inż. Maciej Krawczak

Redaktor techniczny:

Dr inż. Anna Samborska-Owczarek

Powyższej książki w całości lub części nie wolno powielać ani przekazywać w żaden sposób, nawet za pomocą nośników mechanicznych i elektronicznych (np. zapis magnetyczny), w tym też umieszczać ani rozpowszechniać w postaci cyfrowej zarówno w Internecie, jak i w sieciach lokalnych, bez uzyskania pisemnej zgody posiadacza praw autorskich.

ISBN 9788389475237
ISSN 0208-8029

To, co znane, nie jest jeszcze przez to, że jest znane, czymś poznanym
Fryderyk Hegel (1770-1831)

Wprowadzenie

Zasadniczym celem niniejszej monografii jest potwierdzenie tezy autora, że fuzja matematyki i technologii egzemplifikowana powiązaniem metod predykcji ze strategią inwestycyjną opartą na mechanizmach technologicznych platform brokerskich daje nową synergetyczną wartość. Wartości tej nie daje się uzyskać odrębnie. Przy niczym nieograniczonej fascynacji handlem¹ na platformach internetowych [10, 66, 81, 85, 86, 94], często dziś spotykanej wśród licznych, najczęściej młodych przedstawicieli przyszłego społeczeństwa informacyjnego, nie da się, bez odpowiedniej wiedzy, osiągnąć sukcesu. Podobnie sama wyłącznie teoretyczna wiedza, bez umiejętności jej zastosowania, pozostaje metafizyczną wartością, rzadko satysfakcjonująca twórcę. Trawestując Freddie Mercurę² można powiedzieć *The sale must go on* i coś energetycznego jest w tym aforyzmie. Jest w nim imperatyw wdrożenia odkrycia matematycznego, rozwiązania znalezionego w przestrzeni przeszłej. To zdarzenie rzadkie w świecie nauki. Z kolei szaleństwo nieprzygotowanego żywiołowego uczestnictwa w giełdowych³ i foreksowych⁴ transakcjach to zdarzenia częstotliwe, niejednokrotnie z odcieniem chorobliwego hazardu i najczęściej bez happy-endu.

¹ Przez handel na platformach internetowych rozumie się najczęściej częstotliwe zawieranie transakcji na odległość polegające na otwieraniu pozycji na rynkach udostępnianych przez daną platformę (najczęściej na rynkach walutowych). Takie zachowanie inwestora nazywane jest często, bez tłumaczenia, tradingiem.

² Freddie Mercury – słynny brytyjski piosenkarz ub. wieku, którego jednym z największych przebojów był utwór „The show must go on”.

³ W pracy rozpatrywane będą predykcje walorów (akcji) giełdowych i kursów walut.

⁴ Forex – to międzynarodowy rozproszony rynek walutowy udostępniany właśnie poprzez tysiące internetowych platform brokerskich.

Czego więc syntezą jest ostatecznie sukces na rynku⁵ finansowym? Jest zapewne sumą wiedzy (zastosowań matematyki, statystyki, udanych modeli predykcyjnych, znajomości realiów ekonomicznych), umiejętności (znajomości technologii brokerskich, przewidywania reakcji rynków na podstawie symptomów politycznych, ekonomicznych czy społecznych, stosowania oceny percepcyjnej zdarzeń) i możliwości ciągłego uczestnictwa w grze rynkowej (inaczej mówiąc wolnego czasu po podjęciu decyzji inwestycyjnej lub użycia systemu informatycznego do monitoringu rynku i podejmowania automatycznych decyzji wg wcześniej przetestowanej strategii).

W tym aspekcie cytowane dalej motto Kanta o znaczeniu matematyki w obiektywizacji reguł uogólniających opis rzeczywistości, tworzących model wiedzy o niej, jest wyrazem przekonań autora. Matematyki nie należy jednak w tym cytacie zawęzać do jej akademickiej definicji. Dziś Kantowi chodziłoby zapewne także o szersze rozumienie tego konceptu – uwzględniającego np. także dokonania metod sztucznej inteligencji, teorii chaosu czy mechaniki kwantowej. Szerzej, zgadzając się z Włodzisławem Duchem [15], należy raczej posługiwać się tu pojęciem inteligencji obliczeniowej, która jest pojęciem szerszym niż sztuczna inteligencja. Metoda staje się bowiem wartościowa poprzez obliczenia. Przed ich przeprowadzeniem często nie ma podstaw do uznania koncepcji za metodę.

Nie odzegnując się od niczego, autor, odwołując się do klasyfikacji postaw w nauce wg Andrzeja Wierzbickiego [124], skłonny jest zaliczyć prezentowane tu poglądy do skrajnie pragmatycznych, typowych dla przedstawicieli bardziej nauk technicznych niż nauk ścisłych.

Dzisiejsze postawy inwestorów wobec zadania predykcji na rynkach finansowych w najbardziej generalny sposób różnicuje podejście zaproponowane przez Eugena Famę [20], który wprowadzając kilkadziesiąt lat temu pojęcie rynku efektywnego podzielił inwestorów na zwolenników analizy fundamentalnej, analizy technicznej, zwolenników równoczesnego wykorzystywania metod z obu obszarów wiedzy oraz sceptyków przekonanych o nieprzewidywalności absolutnie niestacjonarnych zachowań rynków. Współcześnie, uważa się (np. Hia Jong Teoh [33], Goefrey Friesen[24], Bill Cai [8]), że klasyfikacja przekonań o możliwości przewidywania rynków winna przebiegać jeszcze nieco inaczej. Powinna przebiegać mianowicie wg podziału na trzy kategorie postaw -

⁵ Rynek to umowna nazwa dziedziny, w której obserwowane są zmiany ceny wybranego instrumentu finansowego np. kursu waluty, ceny akcji na giełdzie, ceny towaru itp.

że nic nie da się przewidzieć, że tylko analiza fundamentalna jest zasadna oraz, że tylko wyrafinowane metody współczesnej analizy technicznej (wykorzystujące sztuczne sieci neuronowe, rozpoznawania wzorców, algorytmy genetyczne itp.) mają sens.

Niewątpliwie do zwolenników tego ostatniego podejścia zaliczyć można większość noblistów w dziedzinie ekonomii (tych, związanych z ekonometrią, ekonomią i inżynierią finansową) ostatnich kilkunastu lat takich jak Myron Scholes, Robert Merton, Robert Engle, Harry Markowitz, Joseph Stiglitz, William Sharpe. Wszyscy oni uważali, że jest sens tworzenia złożonych modeli odwzorowania gospodarczej rzeczywistości i pomimo świadomie popełnianych błędów (na poziomie uproszczeń w założeniach koniecznych do zbudowania modelu) przyczynili się do istotnego poszerzenia wiedzy matematycznej o modelowaniu rynków.

Częste są także opinie, że prawdopodobnie najlepsze wyniki predykcyjne na poziomie zapewniającym ich praktyczną użyteczność inwestycyjną uzyskuje się poprzez kojarzenie elementów analizy technicznej i fundamentalnej. To oczywiście pewna odmiana przytoczonego wyżej poglądu o celowości poszukiwania skutecznego rozwiązania w gąszczu współczesnej inteligencji obliczeniowej. Metametodą jest tu hybrydyzacja metod o różnym charakterze i różnych technik wykonania decyzji rekomendowanych przez te modele. Bardzo dynamicznie rozwijają się ostatnio w rozpatrywanym obszarze badania chińskich analityków np. Sheng-Hsun Hsu [107] i Zhang Yudong [133] łącząc modelowanie za pomocą sztucznych sieci neuronowych z teorią zbiorów przybliżonych [88, 89] uważają, że to najbardziej perspektywiczny kierunek opanowywania niestacjonarnych procesów, a Hia Jong Teoh [33] przedstawia metodę łączącą sieci neuronowe z GMDH (Group Method of Data Handling). S. Dehuri i S. Cho [14] łączą techniki regresyjne z optymalizacją rojem cząstek. Optymalizacja rojem jest najczęściej zabiegiem pomocniczym dla określania najlepszych parametrów w wielowymiarowej przestrzeni predyktorów (indykatorów, zmiennych objaśniających, składowych portfela itp.).

Interesującym aspektem metod stosowanych dla predykcji rynków jest także umowny podział na dwie duże grupy metod (niezależnie od wykorzystywania danych o naturze fundamentalnej lub technicznej). Do pierwszej grupy można byłoby zaliczyć wszelkie metody oparte na idei pozyskiwania pewnych parametrów z przeszłości pozwalających na zbudowanie modelu wykorzystującego te parametry (czasami tylko chwilowo) i umożliwiających niejako kontynuację modelu w przyszłości. Metod tych nie należy jednak zaliczać do parametrycznych wg taksonomii stosowanej w statystyce – nie są to metody oparte na założeniu zna-

jomości rozkładu statystycznego zmiennej losowej. Do tej grupy z pewnością należy rozpatrywana tu metoda GMDH i wszelkie inne metody oparte na regresji, korelacji, średnich kroczących i ich pochodnych, modelach rodziny GARCH, algorytmach genetycznych czy sieciach neuronowych.

Druga duża rodzina to metody oparte na poszukiwaniu wzorców. Podstawowa idea przyświecająca zwolennikom tej grupy metod jest hipoteza, że jeżeli zmiany w szeregu czasowym wielokrotnie obserwowane w przeszłości następowały wg określonego powtarzalnego schematu, to prawdopodobne jest powtórzenie tego schematu także wówczas, gdy zaczyna być obserwowany jego początek. Charakter wzorców może być bardzo zróżnicowany. Wzorcami są powszechnie stosowane w analizie percepcyjnej rynków, wynikające z teorii fal Elliotta formacje graficzne (takie jak np. formacja głowy i ramion, podwójnego szczytu, spodka, diamentu itp.). Wzorcami mogą być określone sekwencje świec japońskich, może to być także zbiór jednocześnie spełnionych wielu warunków związanych z indykatorami czy parametrami rozkładu statystycznego.

Większość uznanych traderów⁶ takich jak Larry Williams, Jack Schwager, Joe Kruttsinger czy Thomas DeMark, opisani w pracy Joe Kruttsingera [66], posługiwała się i posługuje technikami, które dziś można nazwać poszukiwaniem wzorców. Większość z opisywanych technik (spośród tych ujawnianych) to stosunkowo proste zabiegi mające jednak wspólną cechę. Są rezultatem ciężkiej wieloletniej pracy, tysięcy obserwacji i zauważenia trudnych do logicznego objaśnienia, jednak powtarzalnych sekwencji zdarzeń. Przez stosunkowo proste wzorce należy rozumieć takie, które daje się zaobserwować lub obliczyć w czasie rzeczywistym bez budowania złożonych modeli prognostycznych, zawsze jednak, jak twierdzi większość spośród tych wspomnianych zwycięzców rynku, podstawowym narzędziem wspomagającym ich decyzje jest ustawicznie pracujący komputer.

Interesującym podejściem jest kojarzenie analizy technicznej z fundamentalną. Do tej grupy można na przykład zaliczyć techniki prognostyczne oparte na obserwacji zdarzeń gospodarczych np. ogłaszanych o stałych porach dnia czy miesiąca informacjach o nastrojach lub

⁶ Trader w powyższym kontekście to zawodowo trudniący się prognozowaniem rynków inwestor, często mentor dla innych graczy. Wielcy amerykańscy traderzy to ludzie sukcesu, którzy wbrew rozpowszechnionym poglądom o nieprzewidywalności rynków dowiedli, że stosowane przez nich techniki, rzadko ujawniane, są skuteczne.

wskaźnikach rozwoju gospodarczego, stopach procentowych ustalanych przez banki centralne itp. Posługiwanie się tymi technikami w sposób automatyczny umożliwia serwis fxengines.com [142]. Do tej grupy metod można także zaliczyć opisywaną w niniejszej pracy fuzję GMDH z wynikami obserwacji inwestorów uczestniczących w handlu na wybranym rynku.

W Polsce problematyką GMDH bezpośrednio zainteresowało się niewielkie grono naukowców i powstało niewiele publikacji, z których najważniejsze to prace Witolda Pedrycza [2, 90] i Józefa Korbicza [62, 68]. Spośród innych autorów zainteresowanych metodą należy wymienić nazwiska Zbigniewa Banaszaka, Marcina Mrugalskiego, Janusza Kusia. Powstała także jedna (prawdopodobnie) praca doktorska wykorzystująca wprost metodę GMDH napisana przez Justynę Patalas pod kierunkiem Tadeusza Krupy.

Tytułowa metoda grupowania argumentów (atrybutów) to opracowane przez ukraińskiego uczonego Aleksieja Ivakhnenkę [34-51] podejście do interpolacji i predykcji w formie samoorganizującego się opisu rzeczywistości (modelu). W podejściu tym kojarzona jest wspomniana wiedza i umiejętności. Wiedza to zdolność do opracowania najlepszego w danych okolicznościach modelu (np. predykcyjnego), umiejętność to zdolność do zastosowania sugestii wynikających z tego modelu do podjęcia decyzji (np. inwestycyjnej, z wykorzystaniem platformy brokerskiej). Najstarsze z tych prac powstały już w latach siedemdziesiątych ub. wieku [37, 39, 47, 50].

Aleksiej Ivakhnenko [36, 38] i jego następcy Tetsuo Sawaragi [101], Hema Madala [73], Johann Mueller i Frank Lemke [40, 41, 46, 70] rozwinęli kilka algorytmów mieszczących się w obrębie tej metody.

Metoda w części dotyczącej modelu predykcyjnego polega na równoczesnym zastosowaniu dwóch zasad – zbudowania najlepszego regresyjnego modelu według kryterium opartego na informacji nieużywanej do budowy modelu, oraz drugiej zasady – ograniczenia złożoności struktury modelu regresyjnego do najmniejszej akceptowalnej przez badacza (zgodnie z zasadą brzytwy Ockhama). Istotą metody, od której pochodzi jej nazwa – grupowanie argumentów lub atrybutów, jest uwzględnianie w modelu regresyjnym tylko kilku spośród wielu argumentów – zmiennych wejściowych/zmiennych objaśniających. Oznacza to kolejne rozpatrywanie niejednokrotnie tysięcy modeli regresyjnych o różnych konfiguracjach tworzących je danych wejściowych i wyborze tylko najlepszych z nich. W tym sensie metoda jest samoorganizująca się. Na każdym kroku szeregu czasowego najlepszy w danej chwili model regresyjny ma nie tylko różne współczynniki liczbowe przy poszcze-

gólnych wyrazach równania (w ogólnym przypadku nieliniowego), ale także różną strukturę tego równania w sensie wartości stopnia wykładników wyrazów równania – modelu. Wybór kilku (trafniej, wielu) najlepszych, a nie jednego najlepszego modelu, na danym kroku złożoności algorytmu, Ivaknenko uzasadnia koniecznością pozostawienia pewnej swobody w procesie decyzyjnym. Nawiązując do prac Gabora Ivakhnenko [42] powtarza argumentację selekcyonerów np. roślin czy zwierząt, którzy pozostawiają zawsze kilka-kilkanaście najlepszych par czy osobników. W ten sposób uzyskuje się niejako szerszą bazę odwzorowania rzeczywistości. Nie wszystkie najlepsze cechy mogą być przekazane przez pojedynczy egzemplarz (przez najlepszy model rzeczywistości). Ta chwilowa dominacja może być przypadkowa i w wielokrotnie powtarzanej predykcji może się nie sprawdzić.

Z punktu widzenia stopnia naszej determinacji w procesie poznania rzeczywistości metoda ta jest zapewne specyficznym konsensem pomiędzy filozoficznym ewolucjonizmem, wyrażonym równaniem regresji, a kreacjonizmem, wyrażonym poprzez odrzucenie przekonania o skuteczności prognozowania na podstawie przeszłości. Autorski kreacjonizm jest tu wyrażony poprzez zwątpienie w możliwość precyzyjnej prognozy. Sceptycyzm ten wyrażany jest wielokrotnie – myślą przewodnią tej pracy jest przecież odrzucenie wiary w gaussowski rozkład ceny walorów względem średniej (np. rozdział 3) i ciągłe stosowanie rozbudowanej przestrzeni zmiennych decyzyjnych związanych z technologiami platform brokerskich. Sceptycyzm ten wyraża się także w rozpatrywaniu wyłącznie prognoz krótko- i bardzo krótkoterminowych, całkowicie wynikających z matematycznej czy statystycznej wiedzy pozyskanej z przeszłości, ignorującej „chiromancje” ekonomiczne czy polityczne. Postawa taka rzutuje zresztą na przyszłe zastosowania metody – jako automatycznie realizowanego scalpingu⁷.

Metoda z pewnością należy do grupy wielu współczesnych metod opartych na myśleniu indukcyjnym takich jak sieci neuronowe, RBF (*radial basis functions*), algorytmy ewolucyjne, teoria zbiorów przybliżonych. W aspekcie rozwiązywanych zadań predykcyjnych i decyzyjnych metoda łączy podejścia charakterystyczne dla ekstrakcji wiedzy i data mining'u (eksploracji danych) z rozpoznawaniem wzorców i klasyfikacją. Ta część metody z pewnością nawiązuje do postawy przeświadczenia o możliwości ekstrakcji wiedzy z bazy danych. Ta postawa

⁷ Scalping – to specyficzna dla foreksowych platform brokerskich technika inwestowania polegająca na częstotliwym otwieraniu i zamykaniu pozycji, co kilka-kilkanaście minut, a nawet częściej.

występuje wyraźnie u protoplasty metody – prof. Ivaknenki, który pisał [39, 49], cytując Gabora, że świat realny jest bardzo złożony, bardziej, niż to się powszechnie uważa, nie oznacza to jednak, że należy zrezygnować z możliwości rozpoznania tej złożoności. Mało tego, bardzo często złożony model jest niespodziewanie precyzyjny i nie przeuczony⁸.

Zastosowanie indukcji w poszukiwaniu reguły rządzącej rzeczywistością, czyli wyprowadzenie nieznanego do tej pory prawa na podstawie obserwacji, zaczęło być równie ważne jak dedukcyjny sposób myślenia dopiero w epoce informacyjnej i towarzyszącej jej burzliwemu wzrostowi możliwości obliczeniowych maszyn. Pomimo tej technologicznej rewolucji autor konstatuje ciągle duży dyskomfort w zderzeniu potrzeb obliczeniowych rozpatrywanej tytułowej metody GMDH (Group Method of Data Handling)⁹ i możliwości aktualnych komputerów. Dodatkową obliczeniową komplikacją podejścia proponowanego przez Aleksieja Ivakhnenkę i naśladowców – Hema Madalę [73], Johanna Mullera [70], jest rozszerzenie regresyjno-neuronowej asocjacji metody GMDH o kryterium zewnętrzne (oparte na informacji spoza rozpatrywanego zbioru danych uczących) związane z uwarunkowaniami technologicznymi. Te uwarunkowania technologiczne to możliwość uwzględnienia ograniczeń brokerskich.

Istotną wnoszoną w niniejszej pracy innowacją jest zastosowanie ograniczeń w strategii inwestycyjnej związanych bezpośrednio z technologiami stosowanymi na platformach internetowych. Innowacja ta, to specyficzna przestrzeń $SL^{10} \times TP^{11} \times Tdop$ (poziom akceptowalnej straty/poziom satysfakcjonującego zysku/ustalony maksymalny okres

⁸ To twierdzenie o krok od teorii chaosu, której Ivaknenko nie przywołuje.

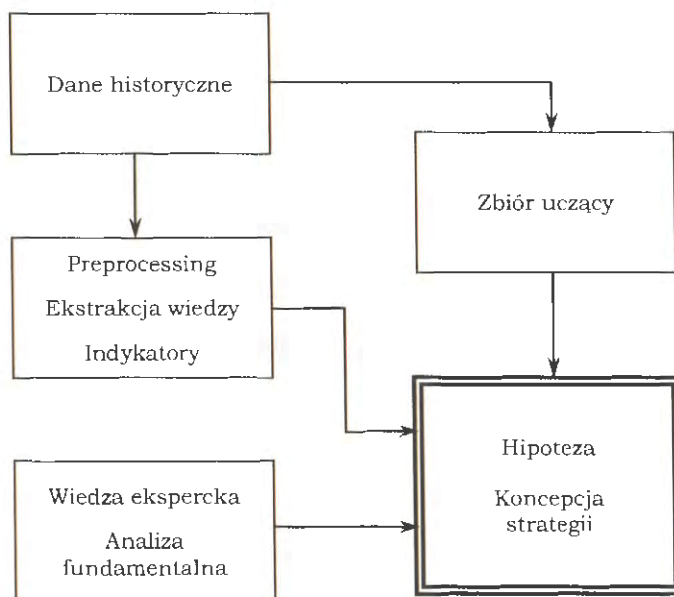
⁹ Group Method of Data Handling – angielskie tłumaczenie metody opracowanej przez Ivakhnenkę w oryginale zwane *Metod Gruppovogo Uceta Argumetov* (ros.). W języku polskim nie ma dotąd jednoznacznego przyjętego powszechnie tłumaczenia. Właściwe wydaje się być – Metoda Grupowania Atrybutów czy Metoda Grupowego Uwzględniania Argumentów.

¹⁰ SL – Stop Loss – stosowany na internetowych platformach brokerskich mechanizm półautomatycznego ograniczania starty. Polega on na ręcznym ustawieniu dopuszczalnej przez inwestora ponoszonej przez niego straty, w przypadku zmian rynku w kierunku niezgodnym z oczekiwaniami inwestora i automatycznym wykonaniu tego ograniczenia przez platformę, jeżeli rynek rzeczywiście zmieni się na niekorzyść gracza.

¹¹ TP – *Take Profit* – inny mechanizm półautomatycznego handlu na platformie brokerskiej. Jeżeli rynek zmieni się zgodnie z oczekiwaniami inwestora, to mechanizm TP dyskontuje zysk. To wyraz obawy inwestora przed utratą chwilowych korzyści, które już osiągnął.

otwarcia pozycji), charakterystyczna dla współczesnych platform internetowych w handlu (tradingu) automatycznym. Tu właśnie technologia miesza się z modelem teoretycznym – modelem wiedzy o obserwowanej rzeczywistości. Rzeczywistością tą dla inwestora jest szereg czasowy danych obserwowanych na rynkach finansowych. W pracy, w dalszej jej części, przedstawione zostaną przykłady zastosowania metody GMDH (dokładniej, niektórych, opracowanych na jej podstawie autorskich algorytmów) na różnych rynkach finansowych. Podstawą do oceny skuteczności algorytmu będzie najczęściej kryterium Sharpe’a (lub Sortino) [106, 109, 140] uwzględniające jednocześnie zysk i ryzyko (średnią stopy zwrotu i jej wariancję).

Ogólna koncepcja kojarzenia dwóch systemów – generowania koncepcji predykcyjnej i systemu podejmowania decyzji w warunkach konkretnej platformy technologicznej przedstawiona jest na dwu schematach.



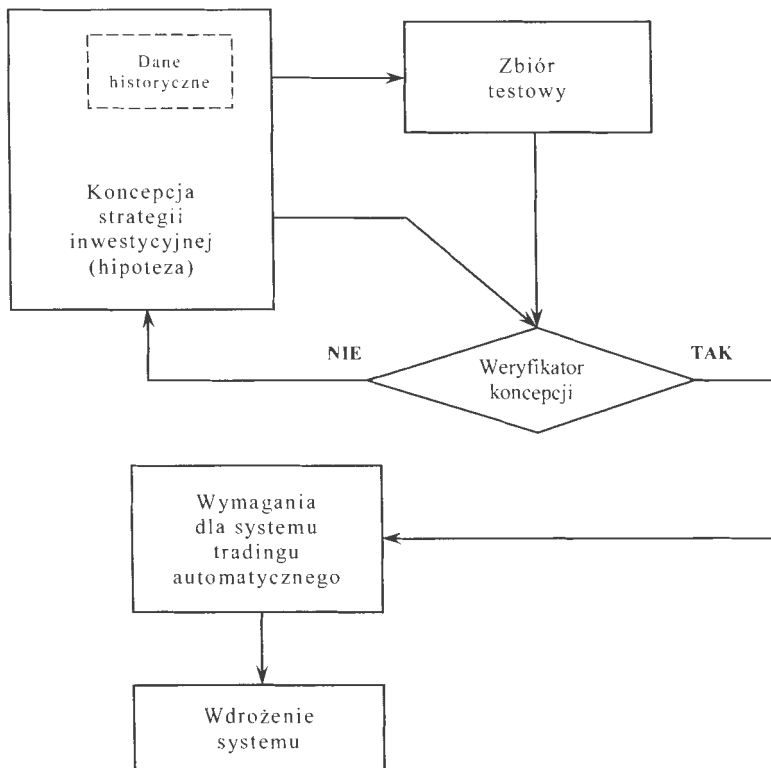
Rys. 1. Schemat blokowy procesu powstawania hipotez badawczych dotyczących koncepcji strategii inwestycyjnych.

Rys. 1 przedstawia schemat poszukiwania koncepcji strategii predykcyjnej. Rodzi się ona w przestrzeni danych historycznych w wyniku ustawicznej penetracji różnymi (z założenia heurystycznymi)

Wprowadzenie

algorytmami tej przestrzeni, stosowaniu wyrafinowanej filtracji, uznanych i nowych indyktorów wysyłających sygnały do otwarcia lub zamknięcia pozycji. Do tych zabiegów, należących do analizy technicznej, dodać można ekspercką wiedzę o przyszłości rozpatrywanego rynku, wynikającą z przewidywanych skutków zdarzeń politycznych, gospodarczych czy społecznych. Symboliczną sumę tej wiedzy (model wiedzy) przedstawia rys. 1. Jej rezultatem jest pewna strategia, z którą w danej chwili utożsamia się inwestor w swym przekonaniu o jej skuteczności.

Gdy jest już o jej wartości przekonany, to warto podjąć decyzję o jej wdrożeniu. Schemat tego postępowania ilustruje rys. 2.



Rys. 2. Schemat blokowy procesu opracowywania wymagań dla systemu tradingu automatycznego.

Inwestor (gracz giełdowy, użytkownik platformy foreksowej) rozważa decyzje o testowaniu strategii. Racjonalne i logiczne jest, by dokonać tego na nowych, nieuwzględnianych wcześniej danych. To ty-

powy zabieg zwany cross-walidacją lub sprawdzeniem krzyżowym. Może on polegać na sekwencyjnym potraktowaniu przestrzeni danych historycznych (najpierw uczenie, potem testowanie) lub wykonywaniu testów co pewien czas w miarę przemieszczania się wzdłuż szeregu czasowego.

Jeżeli weryfikacja strategii jest pomyślna, to może nastąpić faza wdrożenia systemu z uwzględnieniem wymagań stawianych przez daną platformę inwestycyjną. Są to najczęściej wymagania natury technologicznej, np. uwzględnienie API¹² serwera, jeżeli jest udostępniane, uwzględnienie ograniczeń brokera, np. dopuszczalnej częstotliwości tradingu, dopuszczalnej odległości pomiędzy otwarciem pozycji a zleceniem zamknięcia, wielkości prowizji zależnej od zainwestowanej kwoty (najczęściej nieliniowej) itp.

Ograniczenia te odniesione do konkretnej rozpatrywanej strategii, o obiecujących wynikach uzyskanych w fazie testowania, mogą uczynić ją całkowicie nieprzydatną w praktyce. To właśnie istota rozpatrywanego tu zderzenia teoretycznej sprawności strategii z twardą rzeczywistością.

Twórca metody Group Method of Data Handling – prof. Ivakhnenko, początkowo nie rozpatrywał możliwości jej zastosowania na rynkach finansowych. Rozważał częściej przypadki predykcji ekonomicznej, ekologicznej, czy w systemach technicznych [37, 39, 48, 51]. Zrobił to dopiero pod koniec ubiegłego wieku wspólnie z niemieckimi badaczami Muellerem i Lemke [40, 41]. W tym czasie burzliwie rozwijały się statystyczne metody predykcyjne oparte najczęściej na ułomnych założeniach, np. o normalnym rozkładzie stopy zwrotu lub stacjonarności rynku. Pomimo kilku nagród Nobla dotyczących tego obszaru aktywności człowieka (Markowitz, Sharp, Merton, Scholes, McFadden, Engle) modele budowane przez te znakomitości nie doprowadziły, na szczęście, do ostatecznego zdefiniowania wiarygodnego modelu predykcyjnego. Nie doprowadza do niego, także na szczęście, ta praca. Doprowadza do lepszego zrozumienia aforyzmu Hegla: „To, co znane, nie jest jeszcze przez to, że jest znane, czymś poznanym”.

W odróżnieniu od wielu prac ekonometrycznych praca koncentruje się na pragmatycznym powiązaniu predykcji z możliwościami jej

¹² API – Application Programming Interface – interfejs programowania aplikacji, system umożliwiający, w danym rozumieniu, komunikację pomiędzy serwerem, na którym umieszczona jest aplikacja brokerska a klientem, który dzięki API ma możliwość oprogramowania swoich aplikacji w sposób umożliwiający komunikację z serwerem.

Wprowadzenie

internetowej (automatycznej) realizacji. Jak wspomniano, wymaga to uwzględnienia szeregu niespodziewanych ograniczeń natury technologicznej. By udowodnić ten związek w pracy przedstawiono wyniki wielu eksperymentów w przestrzeniach danych historycznych różnych instrumentów finansowych. Wnioski są interesujące – zwłaszcza w wymiarze praktycznym.

Praca składa się z 6 rozdziałów. Pierwsze dwa poświęcone są szczegółowemu badaniu różnych odmian algorytmu kombinatorycznego zbudowanych na koncepcji algorytmu liniowego rozpatrywanego przez Ivakhnenkę [42, 50]. Zmodyfikowane algorytmy GMDH są doprowadzone do takiego stopnia szczegółowości, który umożliwia ich implementację w dowolnie wybranym środowisku programowym. Autor nie publikuje własnych skryptów opracowanych w środowisku Matlab z uwagi na ich rozmiary, być może także ze względu na niedoskonałość kodowania, ale także mając prawo do ochrony tych utworów jako dodatkowej wartości. Skrypty programowe towarzyszące tej pracy są niejednokrotnie bardzo rozbudowane i są często rezultatem wielu miesięcy intensywnych eksperymentów. W dodatku do pracy przedstawiony jest fragment najbardziej złożonego algorytmu GMDH pozwalający Czytelnikowi na opracowanie własnego programu symulacyjnego.

W rozdziale trzecim zbadane zostały wszelkie możliwe fragmenty wielowymiarowej przestrzeni predyktorów, które użyte były w pierwszych rozdziałach do badania skuteczności algorytmów GMDH. Trzy ostatnie rozdziały to prezentacja wyników rozmaitych badań z zastosowaniem tytułowej metody. Badania przeprowadzane są w różnych przestrzeniach danych, także wstępnie przetworzonych i w zaskakująco różnych, odmiennych stylach handlu. Ich wspólnym mianownikiem jest zastosowanie krocząco zmienianej regresji opartej na GMDH – ogólnie podsumowując z wyjątkowo dobrymi rezultatami.

Praca jest przygotowaniem do handlu automatycznego. Wyników takich eksperymentów, z wykorzystaniem systemu informatycznego umożliwiającego handel automatyczny, autor tu nie zamieszcza. Wszystkie symulacje wykonane są na różnych danych historycznych ze starannym odcięciem danych uczących od danych testowych. Celem zasadniczym, jest przygotowanie strategii do zastosowania w kompletnych systemach internetowych – zawierających moduły pozyskiwania bieżących danych, moduły dostępu do zasobów historycznych i moduły realizacji handlu wg przetestowanych i adaptowanych na bieżąco strategii inwestycyjnych.

4

GMDH w predykcji wielorakiej rynków finansowych

Doświadczenie stworzyło sztukę, brak zaś doświadczenia – przypadek

Arystoteles (384-322 p.n.e)

Wielu, także często wybitnych, teoretyków analizy technicznej rynków finansowych niejednokrotnie przez całe życie zajmuje się doskonaleniem metod prognostycznych tak, że po pewnym czasie celem ostatecznym badań staje się jakość predykcji w szeregach czasowych. Tymczasem w istocie rzeczy doskonalenie metod prognostycznych jest najczęściej jedynie etapem pośrednim do osiągnięcia celu głównego – zbudowania lepszego narzędzia spekulacji. Idzie więc nie o to, by precyzyjnie prognozować, lecz by umieć dobrą metodę prognostyczną zastosować w strategii inwestycyjnej, umieć zbudować dobry system tradingowy. Predykcja wieloraka (wielokrotna) jest przykładem takiego podejścia do pośredniej roli dokładności prognozowania. Dokładności, która przestaje być celem samym w sobie a staje się podstawą do zbudowania dobrej strategii, najlepiej automatycznego tradingu. Predykcja wieloraka to umowne określenie swobody wyboru zmiennej objaśnianej spośród wielu innych zmiennych stanowiących rezultat pewnych obserwacji (pomiarów). Przez analogię do pojęcia regresji wielorakiej [83, 103, 115, 119] zastosowana tu zostanie zasada możliwości wyrażania jednej zmiennej za pomocą innych. Zazwyczaj jest tak, że z góry wiadomo, co jest wynikiem jakiegoś eksperymentu, a co (jaki zbiór zmien-

nych) wymaga obserwacji dla ustalenia statystycznego związku pomiędzy wejściami do obserwowanego systemu i obserwowanym wynikiem. Np. poziom wody w rzece zależy zapewne od m.in. sumy opadów (nie odwrotnie), liczba strażaków gaszących pożar zależy od wielkości pożaru (nie odwrotnie). Ten związek przyczynowo-skutkowy nie zawsze jest tak oczywisty, często jest zupełnie nieoczywisty. Np. czy wartość pary walutowej EUR/USD zależy od ceny ropy, czy odwrotnie?

Na rynkach finansowych występuje interesująca sytuacja w aspekcie, który można byłoby nazwać aksjologicznym. Z punktu widzenia poszukiwania jakiejś wartości (przydatności, korzyści) w zbiorze danych nie będzie miało znaczenia, która zmienna jest w istocie skutkiem, a która przyczyną. Można nawet wyobrazić sobie okresową przemienność tych ról. Może tak być, gdy perturbacje ekonomiczne występują wyraziście na jednym rynku – „ważniejszym” – i wpływają na inny, a po jakimś czasie duża zmienność na rynku przed chwilą traktowanym jako rynek „skutku” może stać się „przyczyną” dla rynku pozornie „ważniejszego”.

Jeżeli więc celem nadrzędnym inwestora interesującego się rynkiem papierów wartościowych jest spekulacja, to zrozumiałą będzie zamiar zbudowania modelu umożliwiającego zarabianie w jakikolwiek sposób na wiedzy wyekstrahowanej z danych. Zamiar pełnego zrozumienia ekonofizycznej istoty (fundamentalnej) wszystkich procesów zachodzących na rynkach będzie wówczas oczywistym etapem pośrednim do osiągnięcia celu końcowego – zarabiania pieniędzy [7, 91].

Można sobie oczywiście wyobrazić, że ten cel pośredni – model zachowania się rynków – może mieć inne bardzo istotne np. społeczne czy militarne znaczenie, nie jest jednak ono przedmiotem rozważań w niniejszej pracy.

4.1. Charakterystyka danych

Stosując się do powyższego podejścia wybrano $M=12$ zmiennych jako podstawową macierz danych, na której przeprowadzono eksperymenty. Na te 12 zmiennych złożyły się najważniejsze, zdaniem autora, papiery wartościowe współczesnego świata – najważniejsze indeksy giełdowe, pary walutowe, ceny ropy i złota.

Wybór tych a nie innych zmiennych ma tu znaczenie symboliczne. Są to po prostu najważniejsze papiery wartościowe, a celem będzie wykazanie, czy zbiór oparty na nich danych pozwoli na skuteczną spekulację. Wracając do wyrażonej we wprowadzeniu retoryki należa-

łoby powiedzieć, że zbiór papierów wartościowych, w którym poszukuje się możliwości zysku, nie powinien być budowany według takiego kryterium jak „medialna wielkość”. Racjonalny inwestor nigdy nie buduje portfela posługując się takimi pozarynkowymi kryteriami. Korzystniejsze więc byłoby poszukiwanie zmiennych (papierów wartościowych) maksymalizujących prawdopodobieństwo udanej spekulacji – maksymalizujących potencjalną stopę zwrotu z inwestycji.

Wśród tych zmiennych (papierów wartościowych) znalazł się indeks WIG20, niebędący oczywiście jednym z najważniejszych na świecie. Został tu jednak ułożony z oczywistych powodów subiektywnych – ciekawości badacza.

Przed wykorzystaniem wspomnianych zmiennych sprawdzono współczynnik korelacji liniowej Pearsona wszystkich par wytypowanych zmiennych. Zakładano, że rozpatrywane zmienne będą skorelowane ze sobą, zwłaszcza zmienne objaśniające z wytypowaną zmienną wyjściową. Zakładano także, że każda z rozpatrywanych zmiennych będzie posiadała odpowiednio dużą własną zmienność (wariancję), gdyż inaczej iluzoryczny byłby jej wpływ na pozostałe zmienne.

Współczynnik korelacji liniowej badano za pomocą formuły:

$$r_{ij} = \frac{\text{cov}(x_i, x_j)}{s^2(x_i) \cdot s^2(x_j)} \quad (45)$$

gdzie:

$\text{cov}(x_i, x_j)$ – kowariancja wybranej pary zmiennych,

$s^2(x_i)$ – wariancja zmiennej x_i , $x_i \in \mathbf{X}$ ($i=1, 2, \dots, M$).

Zmienność własną badano jako:

$$v_i = \frac{s(x_i)}{\bar{x}_i} \quad (46)$$

gdzie:

$s(x_i)$ – odchylenie standardowe zmiennej x_i ,

\bar{x}_i – średnia arytmetyczna zmiennej x_i .

Rozpatrywany zbiór 12 zmiennych to kolejno:

1. Indeks giełdy we Frankfurcie DAX.
2. Para walutowa EUR/USD (kurs euro wyrażony w dolarach amerykańskich).
3. Indeks giełdy w Londynie FTSE 100.

4. Para walutowa GBP/USD (kurs funta brytyjskiego wyrażony w dolarach amerykańskich).
5. Złoto (cena jednej uncji troy złota wyrażona w dolarach amerykańskich).
6. Indeks giełdy w Hong Kongu HAI (*Hang Seng Index*).
7. Indeks giełdy w Tokio Nikkei 225.
8. Ropa naftowa (*Brent Crude Oil*).
9. Warszawski indeks giełdowy WIG 20.
10. Amerykański indeks giełdowy DJIA (*Dow Jones Industrial Average*).
11. Amerykański indeks giełdowy S&P 500.
12. Amerykański indeks giełdowy NASDAQ-100.

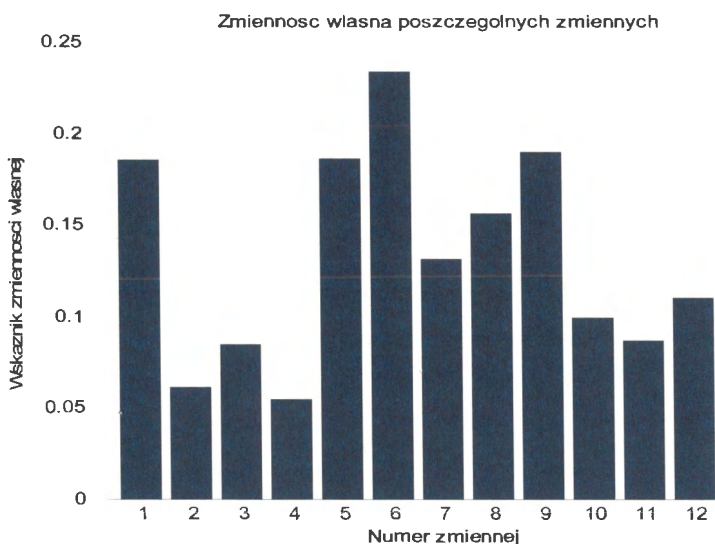
Przygotowano $N=626$ wierszy macierzy danych X . Każda kolumna zawierała jednodobowe wartości Close świecy danego papieru wartościowego. Przykład stosowanego formatu i dokładności tych zmiennych przedstawiono w tabeli 12.

Tab. 12. Fragment macierzy danych, którą zastosowano do predykcji wielokrotnej

	DAX	EUR/USD	FTSE 100	GBP/USD	GOLD	HAI	NIKKEI 225	OIL	WIG 20	DJIA	S&P 500	NASDAQ
1	4350	1,2920	4917	1,8741	425,70	13412	11600	53,10	1970	10404	1163,0	1462
2	4346	1,2914	4906	1,8787	425,90	13426	11566	51,70	1980	10539	1179,0	1490
3	4349	1,2962	4897	1,8904	427,90	13517	11669	54,07	1998	10521	1181,0	1484
4	4365	1,2905	4912	1,8816	425,65	13491	11724	55,43	1989	10388	1168,0	1466
5	4340	1,2853	4899	1,8762	424,05	13513	11668	56,20	1985	10417	1174,0	1476
...
626	7479	1,4648	5954	1,9616	882,00	24410	13481	88,64	2964	12513	1377,8	1875

Na wykresie (rys. 72) przedstawiono wyniki badania zmienności własnej, obliczonej zgodnie z (46). Pożądaną wysoką zmienność rozpatrywanych czynników zaobserwowano u większości zmiennych. Jedy-
nym wyjątkiem z wartością v mniejszą od 0.1 są tu zmienne o indeksach 2, 3 i 4 (to kolejno para walutowa EUR/USD, londyński indeks FTSE 100, para walutowa GBP/USD).

Do możliwości prognozy tych trzech czynników należałoby podejść ostrożniej, o ile zbudowane modele dawałyby interesujące rezultaty. Pozostałe wyniki wstępnych testów nie wykluczają ewentualności pomyślniej prognozy.



Rys. 72. Badania zmienności własnej czynników rozpatrywanych w macierzy danych – papierów wartościowych.

4.2. Zastosowane algorytmy kombinatoryczne GMDH

Zgodnie z zamiarem autorów, do prognozy wielorakiej wybranej grupy czynników (papierów wartościowych) zastosowano proste stonkowo algorytmy kombinatoryczne z rodziny algorytmów należących do metody GMDH.

Istotą metody grupowania argumentów jest zasada samoorganizacji modelu opisującego rzeczywistość. Przez tę samoorganizację rozumie się najczęściej zdolność systemu odpowiedzialnego za automatyczne dążenie do celu do adaptacyjnego określania struktury modelu

najlepiej (w sensie pewnego kryterium) opisującego obserwowany proces.

W rozpatrywanym przypadku celem systemu jest predykcja zdarzeń na rynkach finansowych – to cel bezpośredni, bliższy. Celem dalszym, pośrednim, osiąganym poprzez realizację celu bliższego, jest zysk osiągany w pewnym systemie tradingowym za pomocą strategii zbudowanej na metodzie predykcji zdarzeń.

Przez model rozumie się pewien wielomian (jakikolwiek, prosty lub złożony z innych elementów podobnie jak (5)) szeregu Gabora-Kolmogorowa [50, 73]:

$$y = a_0 + \sum_{i=1}^M a_i x_i + \sum_{i=1}^M \sum_{j=1}^M a_{ij} x_i x_j + \sum_{i=1}^M \sum_{j=1}^M \sum_{k=1}^M a_{ijk} x_i x_j x_k + \dots \quad (47)$$

gdzie:

y – zmienna wyjściowa (objaśniana);

x_i – i -ta zmienna wejściowa (objaśniająca);

a_i, a_{ij}, a_{ijk} – współczynniki wielomianu, estymowane w procesie uczenia modelu.

Jest on modelem wówczas, gdy przedstawiająca go zmienna \hat{y} jest odnoszona do rzeczywiście zaobserwowanych czy pomierzonych wartości zmiennej objaśnianej y .

Szereg ten (47) jest oczywiście nieskończony i heurystyczne podejście badacza sprowadzi się do takiej selekcji modeli cząstkowych (opisywanych przez niektóre tylko wyrazy szeregu Gabora), by rozwiązać problem predykcji w rozsądnym czasie (maszynowym). Ten czas ma zresztą także inny aspekt praktyczny. Czas obliczeń może mieć bezpośredni związek z systemem tradingowym. Jeżeli byłby zbyt długi dla pewnego wybranego wariantu strategii (np. systemu zbudowanego na bardzo krótkich świecach), to cała metoda byłaby nieprzydatna, gdyż nie spełniałaby warunku jej realizacji w czasie rzeczywistym.

Sytuacja taka to dowód na to, że technologia (tu oferta platformy brokerskiej) ogranicza ideę matematyczną, o czym pisze np. Jajuga [52].

Tak więc, jeżeli będzie ustalona w ten sposób dopuszczalna złożoność modelu jako pewnej instancji szeregu Gabora, to można przystąpić do analizy dokładności tych modeli i ich przydatności predykcyjnej.

Uwaga o ograniczeniu złożoności przez rozsądny czas obliczeń nie będzie raczej dotyczyła algorytmu zbudowanego dla omawianych tu danych. Rozpatrywane są bowiem świece jednodniowe, co oznacza, że

w ewentualnym systemie tradingowym byłyby teoretycznie doba na wykonanie obliczeń i podjęcie decyzji.

Należy jednak uzmysłowić sobie, że złożoność modelu wybranego z szeregu Gabora może być tak znaczna, że i ten czas (tu - doba, wspomniano o świecach jednodobowych) nie będzie wystarczający w sposób oczywisty. Rozważania takie mają jednak sens tylko wtedy, gdy dowiedzie się istnienia związku pomiędzy rosnącą złożonością modelu a jego przydatnością predykcyjną. Twórca metody Ivakhnenko [50], jak wcześniej wspomniano, zauważył, że występuje optimum złożoności modelu, powyżej którego zaszumienie modelu pogarsza precyzję odtwarzania (lub prognozowania) rzeczywistości.

Uwzględniając powyższe zbadano dwie klasy modeli (47) – liniowe i nieliniowe. Modele liniowe miały postać:

$$\hat{Y}_l = [1 \quad x_i \quad x_j \quad x_k \quad x_l] \cdot A_l^T \quad (48)$$

lub inaczej:

$$\hat{Y}_l = X_w \cdot A_l^T \quad (49)$$

gdzie X_w to macierz danych utworzona z wybranych (tu czterech) wektorów (kolumn) macierzy wszystkich danych X . Do tych czterech wybranych wektorów nie mogą wchodzić dane z pierwszej kolumny, gdyż reprezentuje ona zmienną obserwowaną (umownie). W związku z tym:

$$X_w = [1 \quad x_i \quad x_j \quad x_k \quad x_l] \quad (50)$$

Każda zmienna x_m , $m \in \{i, j, k, l\}$ to wektor kolumnowy o długości L . To liczba obserwacji wzięta pod uwagę przy obliczaniu współczynników liniowego równania regresji opisującego zmienną obserwowaną y .

$\mathbf{1}$ to kolumna jedynek rozszerzająca podmacierz danych X_w , po to, by równanie regresji liniowej posiadało wyeksponowany tzw. wyraz wolny, nie związany z żadnym czynnikiem x_m .

A_l^T to wektor (transponowany) współczynników liniowego równania regresji, zawierający także wyraz wolny a_0 .

W ostatecznej więc postaci liniowy model (wybrany z szeregu Gabora) będzie miał postać:

$$\hat{y} = a_0 + a_1 x_i + a_2 x_j + a_3 x_k + a_4 x_l \quad (51)$$

Współczynniki A_l pozwalające na utworzenie powyższego modelu są otrzymywane poprzez zastosowanie mechanizmu regresji na da-

nych X_w z rzeczywiście obserwowanymi wynikami wybranej zmiennej y w taki sposób, że:

$$A_l = X_w \setminus Y \quad (52)$$

Ponieważ w równaniu (52) macierz X_w nie jest macierzą kwadratową (ma wymiary $L \times 5$), to nie można tego równania rozwiązać stosując inwersję macierzy X_w . Równanie to jest rozwiązywane z zastosowaniem tzw. pseudoinwersji Moore'a-Penrose'a [4, 74], która daje przybliżone rozwiązanie i spośród wielu możliwych wskazuje najlepsze w sensie metody najmniejszych kwadratów. Przez najlepsze rozumie się tu najmniejszą długość wektora $\|Y - XA\|^2$, czyli pewnego kryterium:

$$C_R = \|Y - \hat{Y}\|^2 \rightarrow \min \quad (53)$$

Zarówno model \hat{Y} jak i oceniające go kryterium C_R może być wynikiem jednej z wielu prób (kombinacji) takiego doboru kolumn ze zbioru $\{x_i, x_j, x_k, x_l\}$, które wyznaczają lepsze i gorsze modele \hat{Y}^k , $k=1,2,\dots,K$, gdzie k to numer kolejnej symulacji, kolejnego k -go rozpatrywanego szeregu czasowego.

Niech p oznacza kolejny numer próby wykonywanej z k -tym szeregiem czasowym, a rozsądek badacza wyznacza liczbę dopuszczalnych badanych modeli jako P . Dla każdego k -tego szeregu poszukiwany jest więc taki model \hat{Y}_{opt}^k , dla którego indeks p (w sensie kolejno rozpatrywanej kombinacji) jest taki, że:

$$p_{opt} = \arg \min_p C_R^p \quad (54)$$

gdzie:

$$C_R^p = \frac{1}{L} \sum \|Y - \hat{Y}^p\|^2 \text{ dla } p=1,2,\dots,P. \quad (55)$$

P jest właśnie tą liczbą, tym ograniczeniem heurystycznym, które daje nadzieję na znalezienie dobrego rozwiązania w rozsądnym czasie. Jej wyznaczenie jest najczęściej syntezą doświadczenia, intuicji, rozsądku i oczekiwań badacza. Wymienione czynniki z trudem (jeżeli w ogóle) dają się opisać formalnie. Podkreślają z pewnością indukcyjny charakter rozważanego algorytmu.

W przypadku rozpatrywanych teraz modeli liniowych liczba P może być wyznaczana po prostu jako liczba kombinacji $P = C_4^M$ i dla

rozpatrywanych tu danych ($M=11$) nie jest liczbą uniemożliwiającą rozważenie wszystkich kombinacji.

Inaczej będzie w przypadku modeli nieliniowych. Nieliniowy model opisujący zmienną obserwowaną ma według autorskiej koncepcji postać:

$$\hat{Y}_n = [1 \quad x_i \quad x_j \quad x_k x_l \quad x_m x_n] \cdot A_n^T \quad (56)$$

lub inaczej:

$$\hat{Y}_n = X_w^n \cdot A_n^T \quad (57)$$

Gdzie X_w^n jest teraz macierzą – podmacierzą nowej macierzy danych, rozszerzonej o kolumny uzyskane z wymnożenia kolumn części liniowej (pierwotnych danych). X_w^n jest więc podmacierzą 4 kolumn plus kolumna jedynek, uzyskaną z konkatenacji poziomej:

$$X^n = \text{conh}(X + X_2^n) \quad (58)$$

Gdzie X_2^n jest macierzą rozszerzającą X o kolumny zawierające iloczyny drugiego stopnia macierzy X , takie że:

$$y_m = x_k x_l \quad (59)$$

dla $\forall k, l \in \{i, j\}, i, j = 1, 2, \dots, M$, także dla $i=j$.

Pozostałe rozważania dotyczące pseudoinwersji i uzyskiwania wektora współczynników równania regresji A_n są takie same, jak dla modelu liniowego.

Wybór takiej, a nie innej liczby wyrazów w modelu liniowym (48) i nieliniowym (56) jest rezultatem arbitralnie podjętej decyzji. Jest to po prostu świadomy (jednak nie przypadkowy) wybór takiej postaci modeli. Zbudowanie modeli na części dostępnych danych zapewnia pewną równowagę pomiędzy dokładnością modelu w części „uczącej”, tam gdzie model poprzez regresję „uczy się” współczynników, a częścią „testową”, gdzie następuje predykcja. Wzięcie do modelu większej liczby wyrazów np. 5, 6 czy więcej spowoduje lepsze opisanie przeszłości i niekoniecznie lepszą predykcję. Zazwyczaj mówi się wówczas o „przeuczeniu” modelu. Wstępne autorskie testy spowodowały, że można było wcześniej napisać, że dobór modelu nie jest przypadkowy.

W rozbudowanych systemach GMDH, tu nieomawianych, często tworzy się, np. w formie lokalnej sieci neuronowej, moduł rozpoznający optymalną w danym kroku strukturę modelu opisującego zmienna obja-

śnianą. W rozpatrywanym tu przypadku oznaczałoby to automatyczne rozpoznawanie liczby wyrazów w modelu (48) lub (56). W dalszej części opracowania, z uwagi na jego objętość, przedstawione będą wyniki badań modelu liniowego.

4.3. Liniowy model regresji w predykcji wielorakiej

Opisana na wstępie macierz obserwowanych danych o najważniejszych papierach wartościowych zawiera $M=12$ kolumn i kilkaset wierszy. Postawiona teza, że być może są takie zmienne (kolumny), które dadzą się dobrze opisać za pomocą pozostałych (w modelu liniowym za pomocą wybranych najlepszych czterech) wymaga sprawdzenia i statystycznego potwierdzenia.

Niech Y będzie tożsamy z x_s , gdzie x_s jest s -tą kolumną w macierzy X , $s=1,2,\dots, M$, i będzie ustawiony na miejscu pierwszej kolumny, a kolumna pierwsza zajmie miejsce s . Tak zmodyfikowana macierz danych X_s stanowi wyjściowy materiał badawczy do weryfikacji jakości modelu predykcyjnego. To pewna konwencja organizacji obliczeń – zawsze na miejscu pierwszej kolumny w macierzy danych będzie znajdowała się zmienna pełniąca funkcję zmiennej objaśnianej w danym wariancie obliczeń.

$$\hat{Y}_s = f(x_1, x_2, \dots, x_i, \dots, x_M), i \neq s, \text{ dla } s = 1, 2, \dots, M \quad (60)$$

Powstaje w ten sposób $M=12$ systemów predykcyjnych, z których każdy obliczany jest według algorytmu opisanego poprzez (50) do (55), czyli każdy najlepszy model w systemie wybierany jest według kryterium (53) spośród wielu modeli charakterystycznych dla danej struktury przestrzeni danych X_s .

W kryterium (53-54) pozornie nie widać jego predykcyjnego charakteru, o ile nie wyjaśni się wzajemnego przesunięcia wprzód o horyzont predykcyjny h_p kolumny pierwszej (zmiennej obserwowanej) w stosunku do kolumn pozostałych.

Jest to zabieg charakterystyczny dla całej idei GMDH stosowanej w celu predykcyjnym.

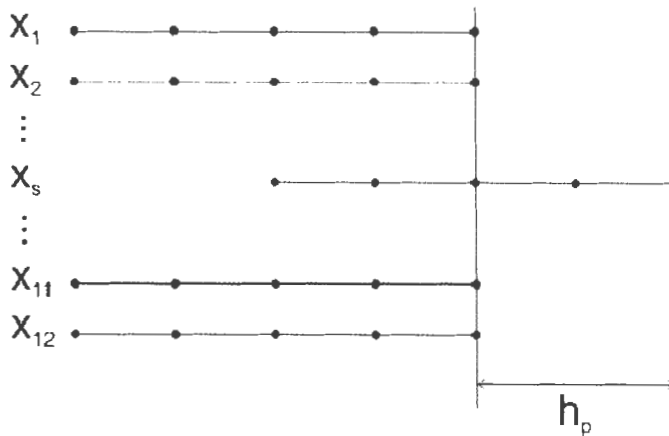
Polega on na wykonaniu regresji i obliczeniu współczynników dla sytuacji jak na rys. 73a, a następnie obliczeniu kryterium jakości predykcji w sytuacji jak na rys. 73b.

Na rysunku symbolicznie zaznaczono wybraną zmienną x_s jako ułożoną wewnątrz macierzy danych, przed jej przemieszczeniem na

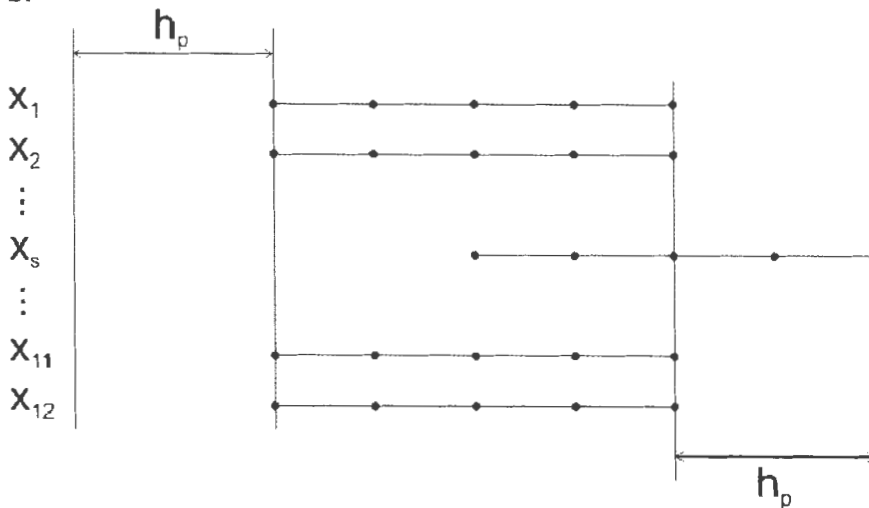
pozycję pierwszej kolumny (macierz danych na rysunku przedstawiona jest po jej „położeniu” – kolumny są tu poziomo).

Na rys. 73b widać więc predykcyjny charakter obliczeń dla zmiennej x_s (w poprzedniej ogólnej notacji posiadającej symbol zmiennej obserwowanej y). Zmienna ta wyprzedza bowiem chwilę bieżącą o horyzont predykcyjny h_p .

a:



b:



Rys. 73. Przesunięcie $M-1$ (z wyjątkiem x_s) szeregów czasowych o h_p wstecz (a) oraz przesunięcie ich do chwili bieżącej t_0 dla obliczenia predykcji zmiennej x_s .

Podsumowując, dla jednego systemu danych (dla wybranego indeksu s zmiennej obserwowanej) należy kolejno wykonać następujące kroki:

1. Utworzyć chwilową macierz danych X_s , wybierając z obserwowanych szeregów czasowych x_i , $i=1,2,\dots,12$, $i \neq s$ wektory o długości L_w , rozpoczynające się L_w+h_p przed chwilą bieżącą, a kończące się h_p przed chwilą bieżącą.
2. Utworzyć wektor zmiennej obserwowanej x_s o długości L_w kończący się w chwili bieżącej.
3. Obliczyć współczynniki równania regresji liniowej według (52); tu $A_l = X_s \setminus x_s$.
4. Utworzyć nową chwilową macierz danych X_s^o z wykorzystaniem wszystkich dostępnych danych pomiarowych, czy obserwowanych, czyli macierz złożoną z wektorów długości L_w , ale kończących się w chwili bieżącej (oczywiście do tej macierzy nie wchodzi wybrana zmienna x_s).
5. Obliczyć model predykcyjny x_s przesunięty o h_p do przodu według (49) czyli tu $\hat{x}_s = X_s^o \cdot A_l$.
6. Obliczyć kryterium rozbieżności modelu \hat{x}_s i rzeczywistości (po upływie odpowiedniego czasu lub po pobraniu informacji z danych historycznych). Kryterium to może dotyczyć rozbieżności obserwowanej tylko na odcinku predykcyjnym (lepiej) lub na całej długości wektorów x_s i \hat{x}_s .

Powyższy, złożony z 6 kroków algorytm należy powtarzać dla każdego $s=1, 2, \dots, 12$ po to, by osiągnąć cel postawiony na wstępie niniejszego opracowania – wybrać zmienną (lub kilka), która jest dobrze opisywana przez pozostałe.

Ekspertyzy dla modelu liniowego przeprowadzono w różnych miejscach szeregu czasowego, dla różnych wartości L_w (np. od 10 do 100), dla różnych wartości horyzontu predykcyjnego h_p (np. od 1 do 10). W modelu liniowym brano pod uwagę kombinacje 4 spośród 11 zmiennych (dwunastą zmienną z macierzy danych była zmienna obserwowana „wyprowadzana” poza macierz danych wejściowych).

Dla $L_w=40$ i $h_p=2$ uzyskano wyniki przedstawione na rys. 74 i 75. W badaniach odniesiono się także do praktycznej przydatności prognoz w możliwym do implementacji systemie tradingowym – opisanym nieco dalej.

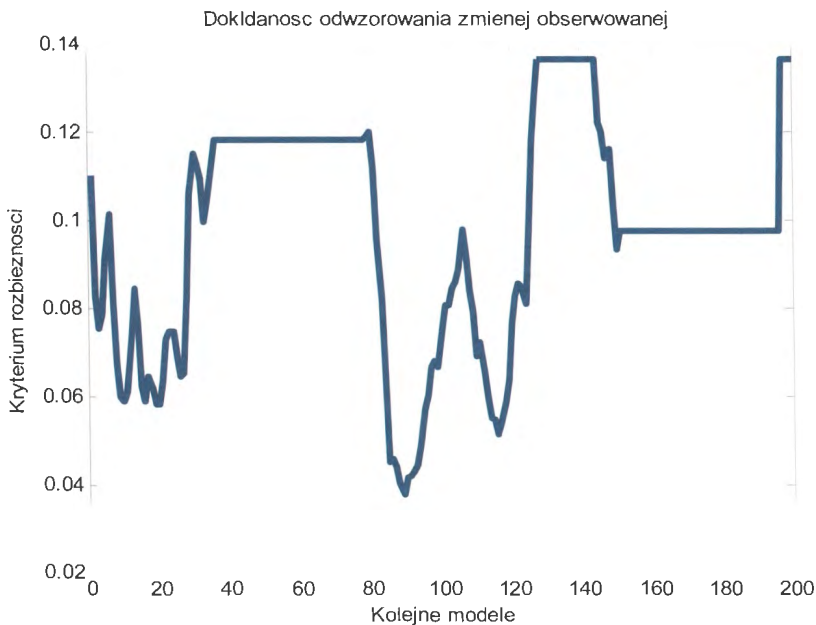
Indeks $s=9$ oznacza wybór indeksu WIG20 jako zmiennej obserwowanej. Rezultaty dotyczą więc początkowo wyłącznie tej zmiennej.

Na rys. 74 przedstawiono wyniki symulacji dokładności szacowanej wg kryterium (53) określającej rozbieżność pomiędzy najlepszym modelem a rzeczywistym przebiegiem zmiennej obserwowanej na całym odcinku.

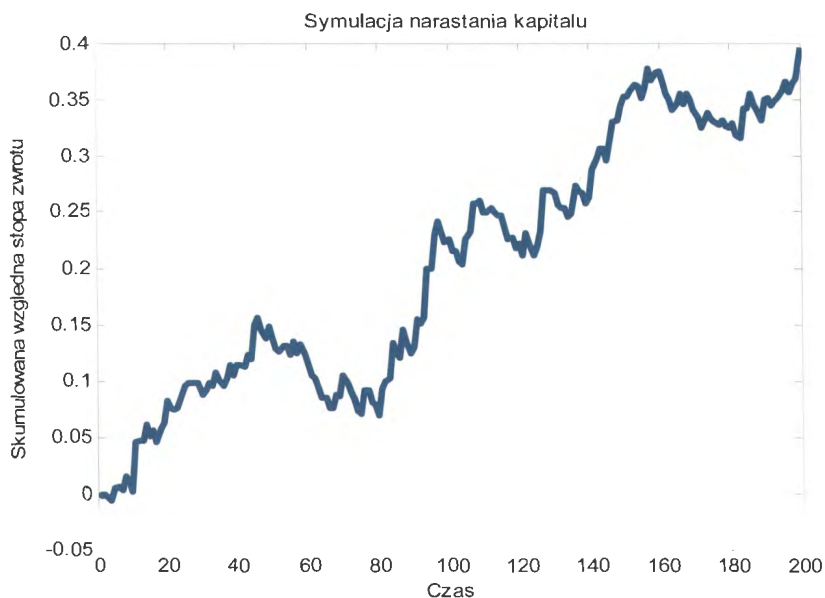
Badania wykonano dla $s=9$, czyli dla pojedynczej wybranej zmiennej obserwowanej, którą był WIG20. Wykonano je dla 200 kolejnych szeregów czasowych. Kryterium dotyczy rozbieżności pomiędzy zmiennymi (\hat{y} i y) skonwertowanymi do wartości względnych – po normalizacji takiej, że $\max(y)=1$, a $\min(y)=0$.

Wyniki przedstawione na rys. 74 wskazują na dość dobrą dokładność odtwarzania. Ivakhnenko [50] uważa, że dokładność rzędu 10% jest bardzo dobrym wynikiem, zwiastującym sukcesy prognostyczne.

Wykorzystując tę wstępną pozytywną ocenę zbudowano koncepcję systemu tradingowego i przeprowadzono symulację narastania kapitału. Wynik tej symulacji przedstawia rys. 75.



Rys. 74. Symulacja dokładności odwzorowania szeregów czasowych dla kolejnych 200 przypadków.



Rys. 75. Symulacja narastania kapitału w systemie tradingowym opartym na predykcji krótkoterminowej indeksu WIG20.

Wyniki te są na pierwszy rzut oka całkowicie satysfakcjonujące. Przy zauważalnej wariancji wartość średnia zwrotu jest wyraźnie dodatnia, co skutkuje dość systematycznym wzrostem kapitału. Dla wykresu przedstawionego na rys. 75 średnia wartość stopy zwrotu w każdym z 200 kroków wyniosła $m_k = 0.0020$ przy odchyleniu standardowym $s_k = 0.0107$.

Niech:

$$r_k = m_k / s_k \quad (61)$$

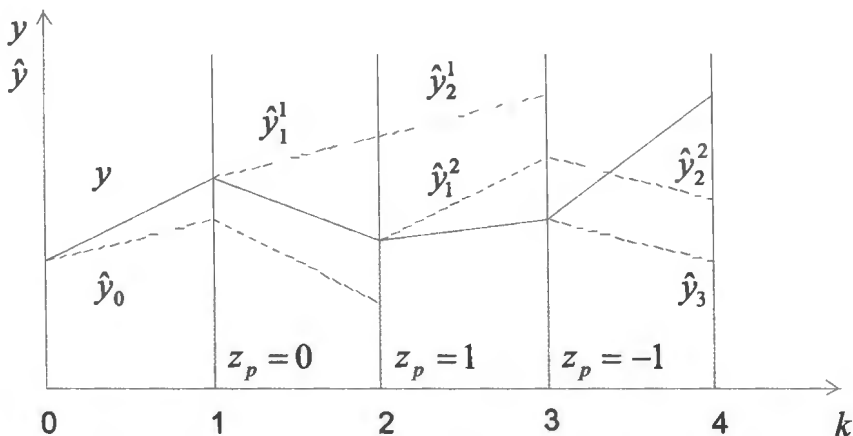
oznacza znormalizowaną stopę zwrotu (zmodyfikowany wskaźnik Sharpe'a [141]) na jeden krok w przestrzeni symulacji (tu na dobę). Niech wskaźnik ten będzie kryterium średniej jakości strategii inwestycyjnej realizowanej poprzez tworzenie szeregu czasowego stóp zwrotu odnoszonego do miary zmienności (odchylenia standardowego). Podejście takie umożliwi porównanie pomiędzy sobą wszelkich rozpatrywanych zmiennych w roli zmiennych objaśnianych.

Dla pierwszego rozpatrywanego przypadku WIG20 wskaźnik ten wynosi $r_k = 0.19$. Stopa zwrotu rejestrowana co dobę i dodawana do dotychczasowego wyniku (skumulowana) obliczana jest w przestrzeni względnych wartości – tu jako ułamek maksymalnej wartości indeksu

WIG20 odnotowanej w dostępnych i wykorzystywanych tu danych. Tak więc obserwowany przyrost rzędu kilku dziesiątych (maksymalnie ok. 0.4) to przyrost rzędu 40% (1600 jednostek indeksu WIG20) od wartości maksymalnej WIG20, która wynosiła dla badanego okresu 3933. Wynik ten uzyskany został dla $k=200$ odcinków czasowych (tu 200 dni). W dziedzinie rzeczywistych zmian tego indeksu średni przyrost dobowy dla w wyników względnych to ok. 8 jednostek WIG20 przy standardowym odchyleniu rzędu 42 jednostek.

Strategia inwestycyjna, która doprowadziła do takiego przebiegu skumulowanej wartości obserwowanego indeksu opierała się na kilku założeniach. Po pierwsze – wprowadzono powszechnie stosowany w platformach brokerskich mechanizm ograniczania straty SL (Stop Loss). Przyjęto, że dobowe ograniczenie SL będzie równe 1% wartości maksymalnej indeksu, czyli ok. 40 jednostek WIG20. Następnym warunkiem postawionym w strategii był sposób przewidywania kierunku zmian zmiennej obserwowanej (w górę lub w dół) oraz potwierdzenie tego kierunku przez poprzednią prognozę.

Sytuację tę przedstawiono na rys. 76. Na rysunku tym dla każdego początku symulacji $k=0,1,2...K$ (przeprowadzono $K=200$ takich symulacji) dokonywano prognozy dwa kroki wprzód y_1^k i y_2^k . Przez y_1^k rozumiano prognozę wartości y po jednym odcinku prognostycznym zaczynającym się od k -go kroku symulacji. Rzeczywiste wartości zmiennej obserwowanej oznaczano $y_1, y_2, \dots, y_k, \dots$



Rys. 76. Przebiegi rzeczywiste (y) i modelowane \hat{y} zmiennej obserwowanej. Odcinek \hat{y}_1^k oznacza pierwszy odcinek horyzontu prognostycznego, a \hat{y}_2^k – drugi.

W praktycznych implementacjach systemów handlowych bardziej istotna jest trafność przewidywania kierunku zmian niż trafność oceny rzeczywistych wartości stopy zwrotu. Jeżeli bowiem poprawnie byłby przewidziany kierunek zmian, to przy wielokrotnie wyższym rzeczywistym odchyleniu (ale w prognozowanym kierunku) wielokrotnie wyższa będzie wygrana.

W aspekcie zarządzania ryzykiem inwestycyjnym, taka pomyłka nie prowadzi więc do frustracji. W przypadku pomyłki co do znaku zmian obserwowanej zmiennej (np. otwarcie pozycji *long* zamiast *short*) wynik będzie wyłącznie stratą. Z tych powodów stratę tę próbuje się ograniczyć wspomnianym mechanizmem *Stop Loss*.

W sensie matematycznym zgodność prognozy co do kierunku zmiany zmiennej obserwowanej oznacza badanie pierwszych pochodnych y i \hat{y} na odcinku prognostycznym Δt , niech będzie to $\frac{dy}{dt}$ i $\frac{d\hat{y}}{dt}$ lub

w układzie dyskretnym $\frac{y_2^k - y_1^k}{\Delta t}$ i $\frac{\hat{y}_2^k - \hat{y}_1^k}{\Delta t}$.

Niech pewna zmienna $z(y)$ określa zgodność znaku (*sign*) zmiennej prognozowanej i rzeczywistej.

$$z(y^k) = \begin{cases} 1 & \text{gdy } \text{sign}\left(\frac{y_2^k - y_1^k}{\Delta t}\right) = \text{sign}\left(\frac{\hat{y}_2^k - \hat{y}_1^k}{\Delta t}\right) \\ 0 & \text{w przeciwnym przypadku} \end{cases} \quad (62)$$

Zmienna przyjmuje więc wartość jeden wtedy, gdy prognozowany kierunek zmian wyrażony poprzez modelowaną stopę zwrotu $(\hat{y}_2^k - \hat{y}_1^k)$ pokrywa się z kierunkiem rzeczywistych zmian.

W systemie tradingowym dodatnia wartość zwrotu $(\hat{y}_2^k - \hat{y}_1^k)$ sugeruje otwarcie pozycji długiej - zmienna leżąca w szeregu czasowym później \hat{y}_2^k jest większa niż wcześniejsza \hat{y}_1^k .

Jeżeli później (w warunkach rzeczywistych lub w danych historycznych) wystąpi także dodatnia różnica $(y_2^k - y_1^k)$ to znaki obu różnic będą jednakowe (tu dodatnie) a $z(y^k)$ będzie równa jeden. Wystąpi zgodność prognozy z rzeczywistością i wygrana w systemie tradingowym.

Podobnie dla sytuacji, gdy obie, modelowana i rzeczywista stopa zwrotu będą ujemne. Oznaczać to będzie otwarcie pozycji krótkiej zakończone sukcesem. Prawdopodobieństwo wystąpienia $z(y^k)=1$, dla

$k=1,2,\dots$ może być miarą skuteczności prognozy i jej praktycznej przydatności w systemie handlu instrumentem finansowym.

W rozważaniach przedstawionych na rys. 76 rozpatrywano także inną zmienną opisującą jakość prognozy – $z_p(y^k)$. Zmienna ta zawiera w sobie element potwierdzenia prognozy bieżącej przez prognozę poprzedniego kroku.

$$z_p(y^k) = \begin{cases} 1 & \text{gdy} & \text{sign}(y_2^k - y_1^k) = \text{sign}(\hat{y}_2^k - \hat{y}_1^k) \wedge \\ & & \text{sign}(y_2^k - y_1^k) = \text{sign}(\hat{y}_3^{k-1} - \hat{y}_2^{k-1}) \\ 0 & \text{gdy} & \text{sign}(y_2^k - y_1^k) \neq \text{sign}(\hat{y}_3^{k-1} - \hat{y}_2^{k-1}) \\ -1 & \text{gdy} & \text{sign}(y_2^k - y_1^k) \neq \text{sign}(\hat{y}_2^k - \hat{y}_1^k) \wedge \\ & & \text{sign}(y_2^k - y_1^k) = \text{sign}(\hat{y}_3^{k-1} - \hat{y}_2^{k-1}) \end{cases} \quad (63)$$

Zastosowanie reguł (63) do przykładowych przebiegów rzeczywistych (y) i modelowanych (\hat{y}) zmian przedstawiono na rys. 76. W dolnej części rysunku przedstawiono wartości z_p ustalone według (64).

Rozważane są tam kolejno trzy różne przypadki:

- gdy drugi odcinek predykcyjny rozpoczynającej się w y_0 prognozy sugeruje pozycję krótką, a pierwszy odcinek prognozy rozpoczynającej się w y_1 – pozycję długą, to $z_p=0$, gdyż jest rozbieżność prognoz; system tradingowy nie wykonuje żadnej akcji;
- gdy drugi odcinek predykcyjny prognozy rozpoczynającej się w y_1 wskazuje na pozycję długą, a także potwierdza to pierwszy odcinek prognozy rozpoczynającej się w y_2 , to przy zgodności z rzeczywistym przebiegiem wystąpi $z_p=1$; system tradingowy otworzy pozycję długą i odniesiony zostanie sukces;
- gdy drugi odcinek predykcyjny modelu rozpoczynającego się w y_2 wskazuje na pozycję krótką i potwierdzony jest przez pierwszy odcinek modelu z y_3 , a rzeczywisty przebieg jest przeciwny, to wystąpi sytuacja $z_p = -1$. System tradingowy otworzy pozycję krótką i wystąpi porażka. Wielkość straty jest w tym systemie ograniczana przez mechanizm SL.

Dla tak zdefiniowanych funkcji $z_p(y^k)$ i $z(y^k)$ przeprowadzono badania w tej samej przestrzeni danych, dla której przeprowadzono symulację przedstawioną na rys. 74 i 75.

Dla rozpatrywanych 200 szeregów czasowych, każdy o długości $L_w=40$ uzyskano dla zmiennej z według (17) wynik $\sum_k z(y^k) = 108$ (tj. 54% trafnych prognoz), a dla zmiennej z_p według (18) wynik $\sum_k z_p(y^k) = 14$.

Oznacza to, że uzyskano przewagę zwycięstw nad porażkami wynoszącą 14 zdarzeń (na 200 badanych świec, spośród których podjęto 185 decyzji o otwarciu pozycji). Przy zastosowaniu mechanizmu Stop Loss uzyskano narastanie kapitału jak na rys. 75. Symulacje te wykonano przy założeniu, że jedyną interesującą zmienną obserwowaną jest indeks WIG20.

Kolejnym i najistotniejszym, w zgodzie z tytułem rozdziału, rozpatrywanym tu zadaniem było sprawdzenie przydatności prognostycznej różnych zmiennych obserwowanych (nie tylko wspomnianego WIG20), wybieranych kolejno z tej samej macierzy danych. Zadanie to jest określone jako wspomniana predykcja wieloraka. Stawia się tu bowiem hipotezę, że w pewnej skończonej przestrzeni danych pomiarowych mogą być lepsze i gorsze zmienne (kolumny w macierzy danych) w aspekcie ich wykorzystania jako pewnego obiektu opisywanego za pomocą pozostałych danych.

Badania te przeprowadzono dla opisanych wcześniej 12 zmiennych charakteryzujących wybrane papiery wartościowe.

Stosując te same zasady obliczania predykcji potwierdzonej i ten sam sposób obliczania skumulowanej wartości stopy zwrotu co poprzednio dla pojedynczego przypadku (WIG20), na rys. 77 przedstawiono równocześnie narastanie kapitału wszystkich zmiennych. Badania te, jak poprzednio, wykonano dla $L_w=40$ i $h_p=2$, $n=12$ oraz przy założeniu liniowości modeli – każdy z modeli miał postać (48) lub (51).

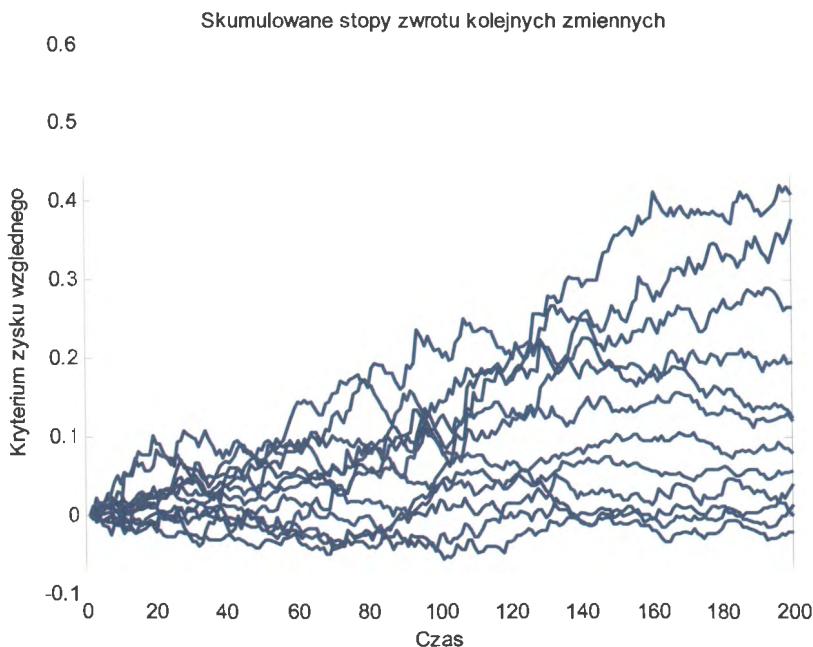
Tak jak poprzednio dla badań pojedynczej wyróżnionej zmiennej (WIG20) zastosowano tu mechanizm SL (Stop Loss) zawsze jednakowy dla każdej rozpatrywanej zmiennej obserwowanej ustalony na poziomie 1% wartości maksymalnej tej zmiennej.

Z rys. 77 wynika wyraźnie, że jest grupa zmiennych, które użyte jako zmienne objaśniane nie rokują, a w zasadzie wykluczają możliwość wykorzystania ich jako obiektu obserwowanego w praktycznym celu prognostycznym. To co najmniej 8 zmiennych, których wartość skumulowanego zysku oscyluje wokół zera, a w dwóch przypadkach jest ujemna. Na tym tle wyróżniają się zmienne, które dają dość wyraźny wynik dodatni. Wyniki końcowe dla badanych 200 przypadków przedstawiono na rys. 78.

Dwie spośród nich (na rys. 78, na którym przedstawiono końcowe wartości skumulowanego wyniku po 200 krokach) osiągają wartości skumulowanego zysku rzędu 40% wartości maksymalnej zmiennej (są to zmienne o indeksach 8 (Oil) i 9 (WIG20)). Dwie inne dają także wyraźny dodatni wynik rzędu 30% wartości maksymalnej (to zmienna nr 7 (Nikkei 225) i 5 (Gold)).

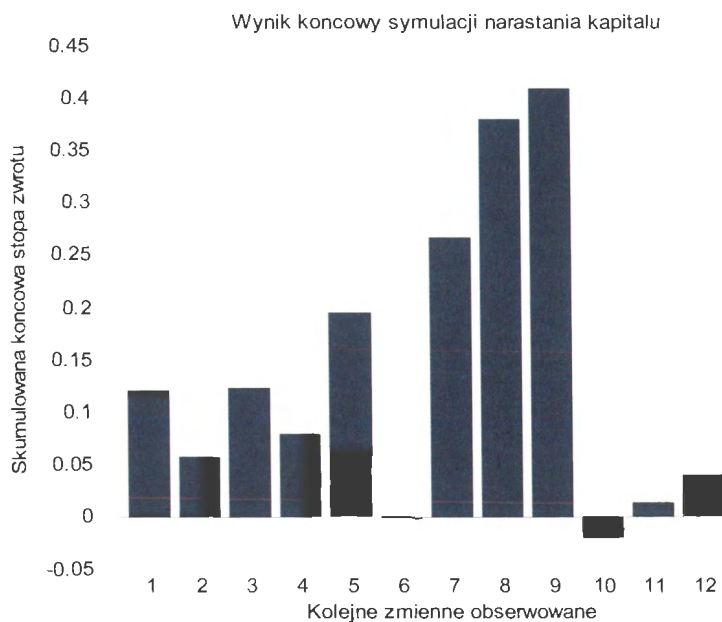
W zasadzie na totalną porażkę skazane są wszystkie pary walutowe i indeksy amerykańskie. To w zasadzie logiczny rezultat – to one raczej kształtują światowy rynek niż go „słuchają”.

To, że w rozpatrywanej tu predykcji wielorakiej wyraźnie wyróżniają się jako zmienne obserwowane te papiery wartościowe, których miejsce w łańcuchu przyczynowo-skutkowym jest zgodne z intuicją i logiką (często dwiema sprzecznymi kategoriami filozoficznymi), świadczy na korzyść rozpatrywanego tu modelu opisującego nieprzewidywalny rynek (rynki).



Rys. 77. Skumulowane stopy zwrotu dla kolejno rozpatrywanych 200 szeregów czasowych dla kolejno rozpatrywanych 12 zmiennych obserwowanych.

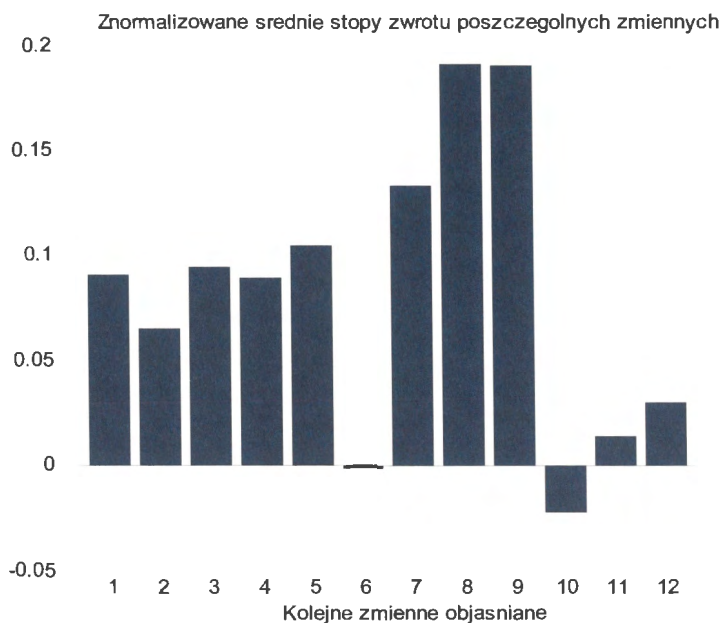
Rys. 78 nie przedstawia wyników predykcji wielokrotnej na tle zmienności wyrażanej formułą (61). W związku z tym dla każdej zmiennej objaśnianej kolejno obliczono średnią dobową stopę zwrotu i podzielono ją przez średnie odchylenia standardowe tych stóp. Wyniki przedstawiono na rys. 79.



Rys. 78. Końcowe rezultaty symulacji skumulowanej stopy zwrotu kolejnych zmiennych traktowanych jako zmienne objaśniane przez pozostałe.

Na rys. 79 widać porównywalną atrakcyjność (niezawodność) inwestycyjną dwóch zmiennych – o numerach 8 i 9 (*Oil* i *WIG20*) oraz wyraźnie niższą kilku innych (o numerach 2, 3, 4, 5 i 7). Do zdecydowanego odrzucenia kwalifikują się zmienne o indeksach 6, 10, 11 i 12. Te trzy ostatnie to indeksy amerykańskie, nr 6 – to indeks Hang Seng. Potwierdza to spostrzeżenie o „objaśniającym” a nie „objaśnianym” charakterze najważniejszych indeksów giełdowych.

Powyższy model predykcji wielorakiej sporządzony został dla danych, które pozostawały względem siebie w pozycji pozyskanej w sposób naturalny, a więc nie przesunięte. Tymczasem Ivakhnenko rozpatrując w pełni rozwinięty model Gabora-Kolmogorowa sugeruje, by rozpatrywać w modelach regresji także przesunięte względem siebie wektory danych opisujących.



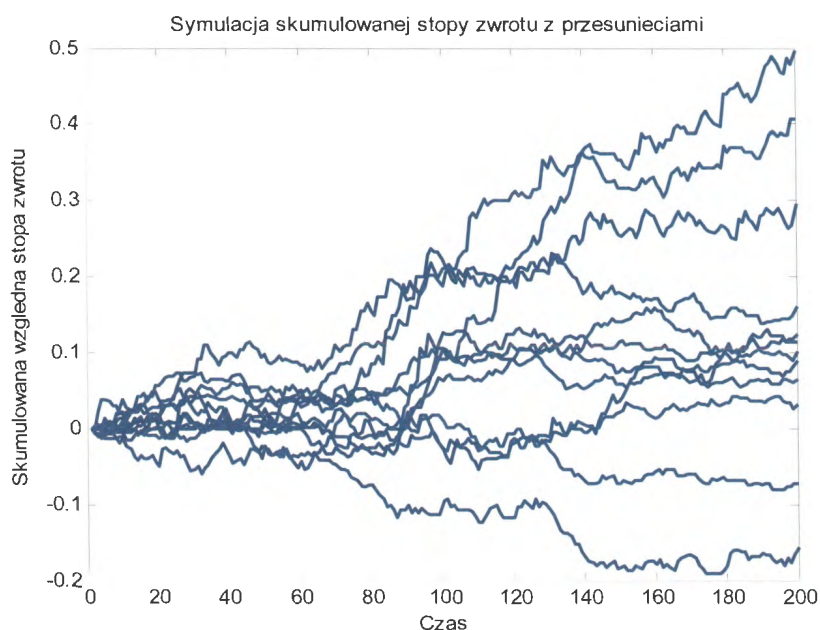
Rys. 79. Względna niezawodność poszczególnych zmiennych objaśnianych wyrażona poprzez równoczesne uwzględnienia położenia (średniej stopy zwrotu) i ryzyka (średniego odchylenia standardowego).

W kolejnych rozpatrywanych tu badaniach dokonano losowego przemieszczenia wektorów zmiennych wejściowych względem chwili bieżącej wstecz. Opisać można tę procedurę za pomocą następujących kolejno wykonywanych operacji:

1. Spośród określonych tu 12 zmiennych wybrano jedną jako zmienną obserwowaną, opisywaną za pomocą pozostałych 11 zmiennych.
2. Dla każdej z pozostałych 11 zmiennych wybrano wektor o rozpatrywanej tu długości L_w przesunięty wstecz od chwili bieżącej o losową wartość od 0 (bez przesunięcia) do $Mlag$ – maksymalna wartość przesunięcia (w przedstawionych dalej wynikach badań przyjęto $Mlag=12$).
3. Dla wylosowanych przemieszczeń wektorów zmiennych opisujących wybraną zmienną opisywaną poszukiwano optymalnej struktury modelu wielomianowego (tu liniowego) według zasad wcześniej opisanych.

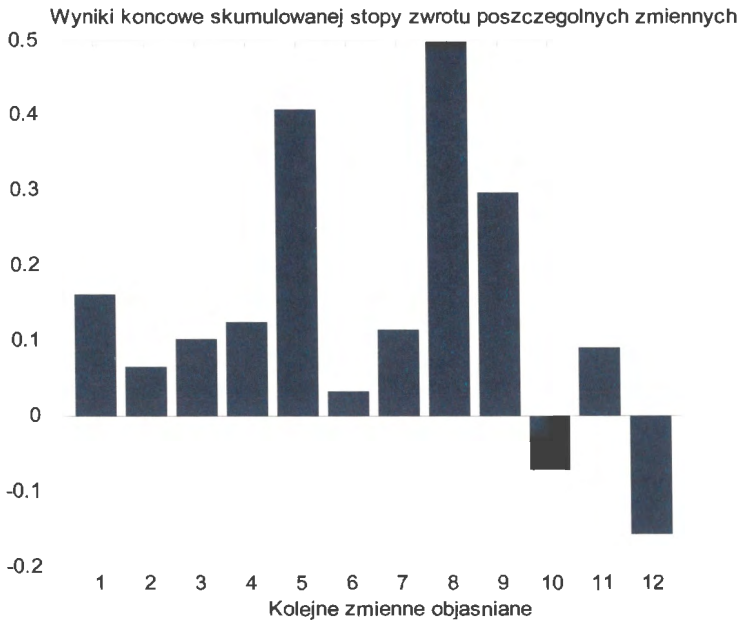
4. Po znalezieniu modelu według punktu 3 zmieniano numer zmiennej opisywanej (punkt 1) i powtarzano powyższe kroki.

Badania tak zbudowanego algorytmu przedstawiono na kolejnych rysunkach 80, 81 i 82. Na rys. 80 przedstawiono wyniki symulacji zorganizowanej podobnie jak przedstawiona na rys. 77 z tą różnicą, że dla każdej kolejno rozpatrywanej zmiennej obserwowanej, pozostałe zmienne przestrzeni danych (kolumny) były przesuwane losowo wstecz od chwili bieżącej.



Rys. 80. Skumulowane stopy zwrotu 12 rozpatrywanych zmiennych opisywanych każdorazowo przez pozostałe 11 zmiennych z przesunięciami wstecz od chwili bieżącej.

Wyniki te generalnie nie wykazują istotnych zmian (w szczególności poprawy) w stosunku do wcześniej badanego przypadku modeli bez przesunięcia. Pojawia się jednak szczególny jeden przypadek – to zmienna *Oil* – cena ropy naftowej (numer 8). W przypadku symulacji tej zmiennej nastąpiła wyraźna poprawa wyniku końcowego. Dominację tej zmiennej przedstawia rys. 81.

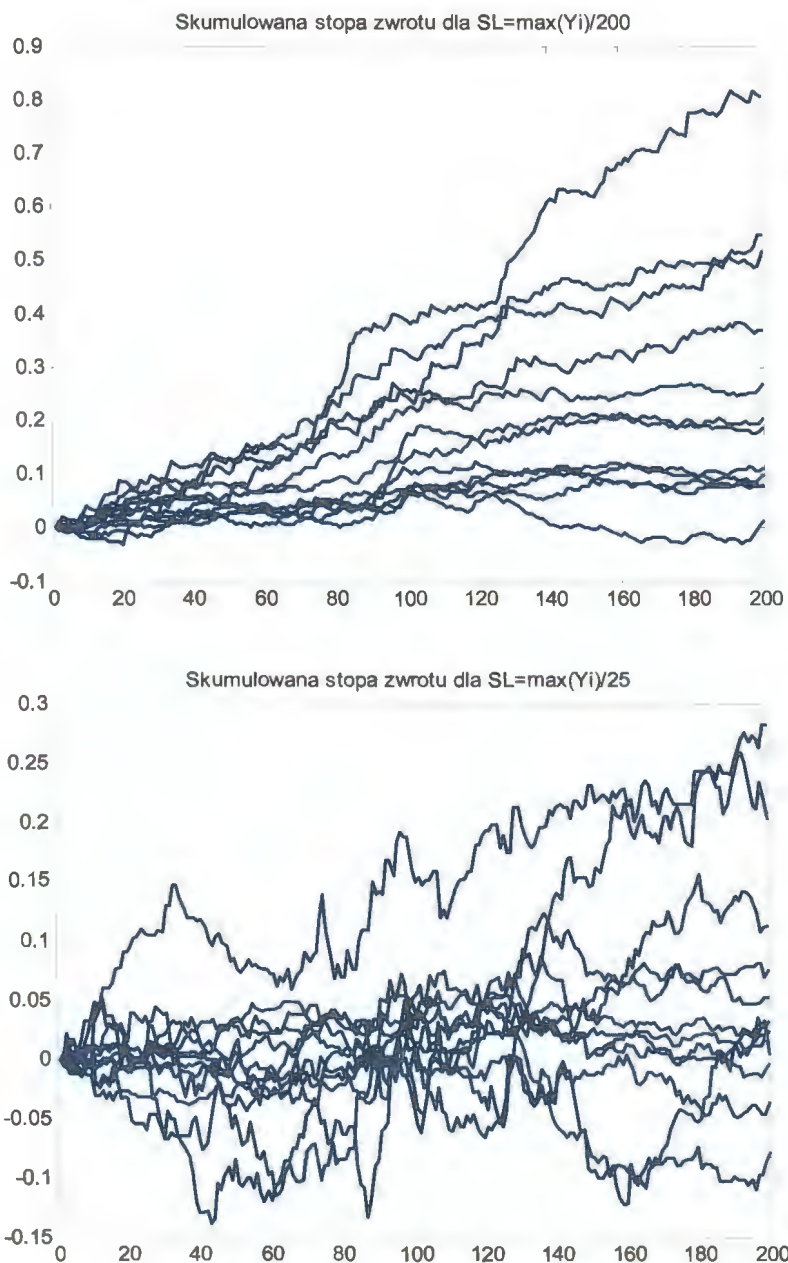


Rys. 81. Wyniki końcowe skumulowanej stopy zwrotu poszczególnych zmiennych objaśnianych uzyskane w warunkach modelowania zmiennej objaśnianej za pomocą przesuniętych w czasie wektorów zmiennych objaśniających.

Należy zaznaczyć, że pomimo prostoty rozpatrywanego tu modelu liniowego wprowadzenie przesunięć znacznie już komplikuje model i wydłuża czas obliczeń. Z tego powodu prezentowane tu wyniki są rezultatem symulacji dla przypadkowych pojedynczych przesunięć wektorów zmiennych objaśniających. Dlatego nawet dość przypadkowo uzyskany dobry wynik predykcji (z drugiej jednak strony to rezultat selekcji spośród wielu modeli) ze zmienną *Oil* jako wynikową może tu być inspiracją dla przeglądania wszelkich innych kombinacji przesunięć.

Na rysunku 81 przeprowadzono symulację 200 kolejnych szeregów czasowych stosując wspomniany wcześniej mechanizm *Stop Loss* dla wartości $SL = \max(Y_i)/100$. Oznacza to, że każda kolejno rozpatrywana zmienna objaśniana y_i miała ustaloną wartość maksymalną w przedziale rozpatrywanych zmiennych historycznych i 1% tej wartości był przyjęty jako SL. Zasada ta powodowała, że wszystkie rozpatrywane zmienne umieszczone były w znormalizowanej przestrzeni względnych wartości. Przy tej okazji sprawdzono, czy wartość SL istotnie wpływa na skuteczność selekcji zmiennych predestynowanych do roli bycia opisywaną przez inne w predykcji wielorakiej.

Rozdział 4. GMDH w predykcji rynków finansowych



Rys. 82. Skumulowane stopy zwrotu 12 rozpatrywanych zmiennych dla dwóch różnych parametrów Stop Loss symulacji. Na lewym rysunku SL jest $1/200$ wartości maksymalnej rozpatrywanej zmiennej a na prawym $1/25$ tej wartości.

Wyniki tych badań przedstawiono na rys. 82. W przeprowadzonych symulacjach zbadano wpływ dwóch wartości SL na końcowe wyniki predykcji wielokrotnej. Z lewej strony przedstawiono rezultaty dla $SL = \max(Y_i)/200$, a z prawej dla $SL = \max(Y_i)/25$.

Widoczne jest znaczne pogorszenie wyników na rysunku z lewej – najlepszy wynik po 200 krokach przy mianowniku 200 jest rzędu 0.7, a przy mianowniku 25 jest ok. dwukrotnie gorszy. Pomimo tego ten wynik (ok. 0.35 uzyskany jest dla zmiennej *Gold*) jest dość obiecujący przy wyjątkowo dużej wartości SL (tu 4% od wartości maksymalnej). Oznacza to, że dla tej zmiennej prognozy są tak udane, że praktycznie nie jest potrzebne stosowanie mechanizmu *Stop Loss*. Dla pozostałych zmiennych obserwuje się jednak już dużą zależność (znaczące pogorszenie skumulowanej stopy zwrotu) przy zwiększeniu SL. Rezultat tych badań jest dość oczywisty – zmniejszenie SL oznacza zmniejszenie strat wynikających z błędnej decyzji inwestycyjnej (otwarcia pozycji w niewłaściwym kierunku), ale jednocześnie rośnie ryzyko zamknięcia pozycji przez przypadkowe ekstremalne odchylenia wewnątrz każdego kolejnego okresu (wartość *High* lub *Low* kolejnej świecy – w zależności od kierunku otwarcia pozycji).

Modele regresji w formie wielomianów liniowych pomimo wszelkich wskazywanych wad wydają się jednak być interesującym narzędziem w strategiach inwestycyjnych a dowodem na to mogą być przedstawione w tym rozdziale wykresy narastania kapitału.

4.4. Nieliniowe modele predykcji wielorakiej

Po modelach nieliniowych, takich jak zawierający pięć wyrazów, opisany formułą (12):

$$\hat{Y}_n = \left[1 \quad x_i \quad x_j \quad x_k x_l \quad x_m x_n \right] \cdot A_n^T$$

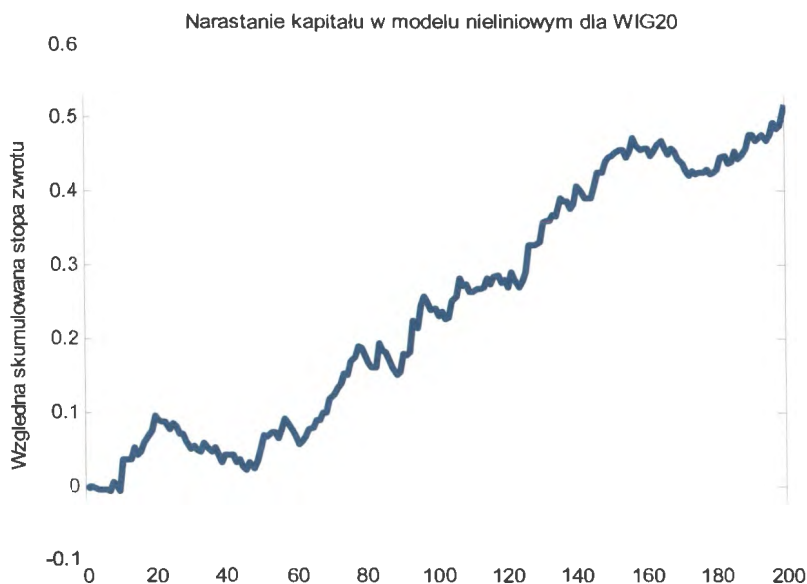
spodziewać się należy poprawienia wyników uzyskanych poprzez symulacje rynków finansowych za pomocą modeli liniowych.

Wynika to z oczywistego pozornie powodu – skoro są one bardziej złożone i wymagają więcej wysiłku (czasu maszynowego) do ich zastosowania, to powinny dać lepsze wyniki. Po uważnej lekturze pierwszych rozdziałów niniejszej pracy nie jest to jednak aż tak oczywiste. Wraz ze wzrostem złożoności modeli nie musi następować poprawa jakości prognozy. Związek ten nie jest oczywisty, gdyż nigdy a priori nie wiadomo, jaka jest statystyczna struktura danych i jak złożone wielo-

miany, najczęściej przez chwilę, ją opisują. Także w aspekcie prognostycznym.

Nie oznacza to w żadnym stopniu gotowości do zaniechania poszukiwań w kierunku podwyższania złożoności – wszak rzeczywistość jest prawdopodobnie bardziej złożona niż się wydaje [91].

Pierwsze symulacje wykonano dla 200 kolejnych szeregów czasowych z WIG20 jako zmienną objaśnianą przez pozostałe 11 zmiennych i ich kombinacje drugiego stopnia (iloczyny dwóch różnych czynników i kwadraty tych czynników). Jak w modelach liniowych przyjęto $L_w=40$ i $h_p=2$ (długość wektorów zmiennych branych pod uwagę w obliczaniu współczynników równania regresji; długość horyzontu predykcyjnego). Narastanie kapitału przy zastosowaniu mechanizmu *Stop Loss* $SL=\max(\text{WIG20})/100$ przedstawiono na rys. 83.



Rys. 83. Symulacja narastania kapitału w strategii inwestycyjnej zbudowanej na nieliniowym modelu dla WIG20.

Na wykresie tym najwyższą skumulowaną wartość stóp zwrotu po 200 krokach uzyskano na poziomie 51.76% wartości maksymalnej. W jednostkach bezwzględnych indeksu WIG20 oznaczało to 2035 jednostek w ciągu 200 dni. Jest to suma sukcesów predykcyjnych w obu kierunkach – zarówno z pozycji długich jak i krótkich. Średnio daje to

dobowy wynik dodatni rzędu 0.26% przy odchyleniu standardowym 1.1%. Stad znormalizowana stopa zwrotu na jeden krok wynosi 0.2364. To dość dobry wynik na tle dotychczasowych (liniowych) rozważań. Dobry także w sensie intuicyjnym – dobre raczej wrażenie sprawia wykres 13 z wyraźnie rosnącą skumulowana stopa zwrotu przy relatywnie niewielkiej wariancji. Szczegółowa analiza ryzyka potwierdzi zapewne tę pozytywną intuicyjną ocenę.

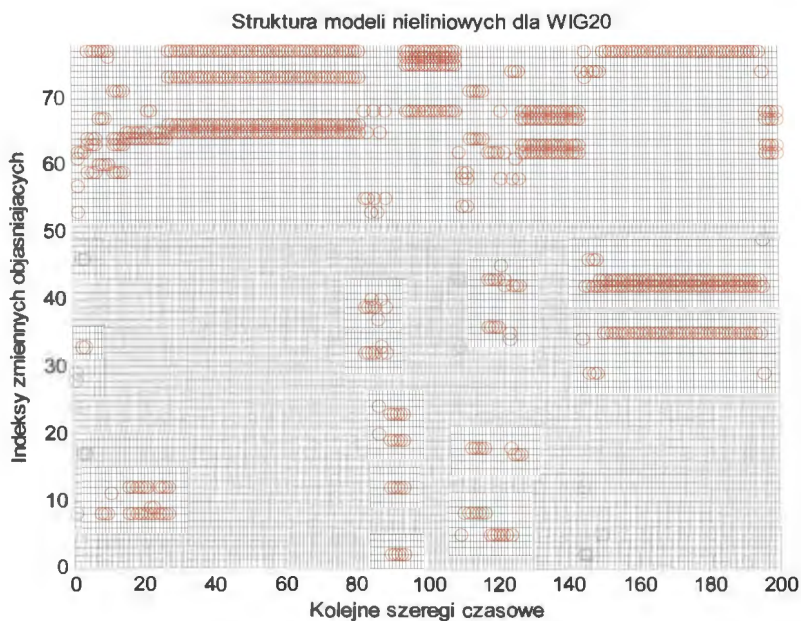
Interesujące jest w związku z tym, skąd biorą się wyraźnie lepsze wyniki symulacji modelu nieliniowego (rys. 83) od wyników modelu liniowego (rys. 75), gdzie znormalizowana stopa zwrotu była rzędu 0.19. Odpowiedz na to pytanie znaleźć można analizując diagram na rys. 84. Na wykresie tym każda pionowa linia reprezentuje pewną chwilę czasową, w której zainicjowano proces symulacji – poszukiwania nieliniowego modelu najlepiej opisującego w sensie regresyjnym przeszłość (wektor zmiennej objaśnianej o długości $L_w = 40$) i na tej pionowej zaznaczone są zawsze cztery indeksy zmiennych tworzących z zgodzie z (56) strukturę modelu nieliniowego:

$$\hat{Y}_n = \left[1 \quad x_i \quad x_j \quad x_k x_l \quad x_m x_n \right] \cdot A_n^T$$

W formule (56) zakładano, że dwa pierwsze wyrazy wielomianu będą miały charakter liniowy a dwa kolejne kwadratowy. W wyniku rozwinięcia tej koncepcji rozpatrywano bardziej ogólny algorytm generowania modeli, w którym na każdej pozycji wielomianu mógł pojawić się zarówno człon liniowy jak i nieliniowy. Rezultat takiego właśnie podejścia przedstawia rys. 44. Pierwszych 12 dolnych poziomów reprezentuje liniowa postać wyrazów wielomianu – modelu, wszystkie powyżej (aż do 78) to iloczyny drugiego stopnia lub w szczególności kwadraty zmiennych objaśniających.

Z rys. 84 wynika, że większość modeli, najlepszych w sensie aproksymacji przeszłości i jakości predykcji, tworzona jest z wykorzystaniem wyrazów nieliniowych. Użycie członów liniowych jest sporadyczne, dotyczy nie więcej niż 10% modeli. Interesująca jest także pewna stabilność struktury modeli. Np. dla symulacji wykonywanej od ok. 30 do ok. 85 kroku, oraz od ok. 150 do ok. 195 kroku najlepsze były modele o powtarzalnej, identycznej strukturze (wykorzystujące te same zmienne objaśniające – tu wszystkie jako nieliniowe). To kolejne potwierdzenie tezy, że rzeczywistość jest niestety złożona. Żaden inwestor poszukujący relacji w łańcuchu przyczynowo-skutkowym nie jest w stanie, bez użycia inteligencji obliczeniowej, dostrzec tych ujawnionych wyżej związków. Względna stabilność, co prawda na poziomie

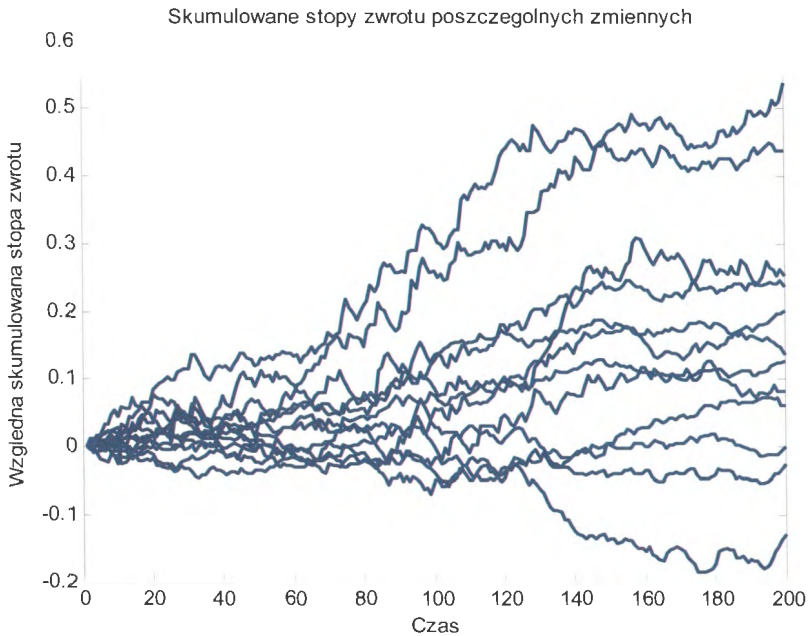
drugiego stopnia wielomianu, sugeruje możliwość ograniczenia przestrzeni danych tylko do tych uczestniczących częstotliwie. Rozważania te warto podjąć.



Rys. 84. Struktura najlepszych modeli nieliniowych rozpatrywanych w kolejnych 200 szeregach czasowych.

Wcześniej jednak będzie rozpatrywane podstawowe zadanie niniejszego rozdziału – czy wśród rozpatrywanych zmiennych są lepsze i gorsze w sensie zdolności do opisu regresyjnego w aspekcie prognostycznym. Wyniki badań z zastosowaniem nieliniowych modeli predykcyjnych opartych na GMDH, wykorzystywanych następnie w symulowanych systemach tradingowych przedstawiono na kolejnych rysunkach – rys. 85 i rys. 86.

Wykresy narastania kapitału przedstawione na rys. 85 nie odbiegają zbyt wiele od wyników badań modeli liniowych przedstawionych na rys. 77. Najlepsze zmienne są tu (dla modeli nieliniowych) jednak wyraźniej korzystniejsze, pozostałe są podobnie przeciętne. Są to zmienne o numerach 5 i 9 (*Gold* i *WIG20*). Jeżeli jednak celem zasadniczym badań jest znalezienie tych właśnie wyraźniej korzystniejszych zmiennych, to rozszerzenie badań o rozważanie modeli nieliniowych można uznać za zasadne i udane.

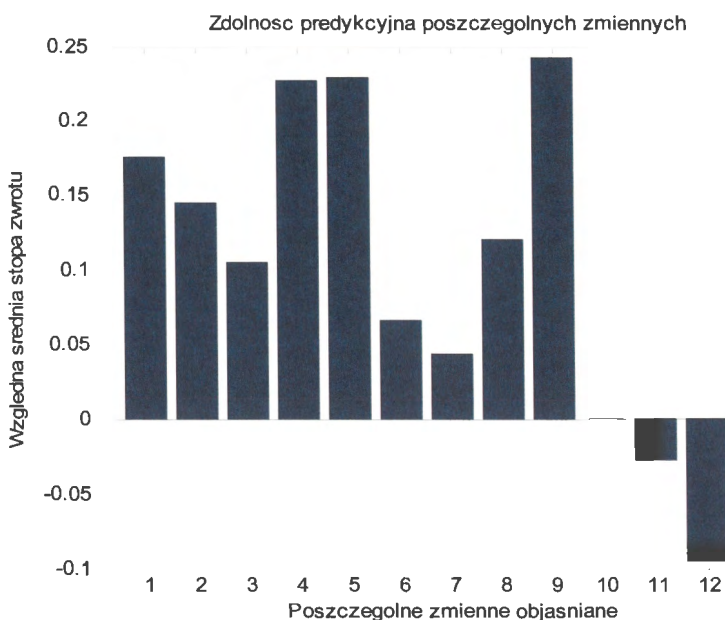


Rys. 85. Skumulowane stopy zwrotu różnych zmiennych objaśnianych opisywanymi modelami nieliniowymi.

Obok bezwzględnych wartości skumulowanego narastania kapitału – tu przedstawianego jako wynik po 200 krokach, istotne jest jak zwykle zagrożenie z powodu niestabilności wzrostu kapitału wyrażone wariacją lub odchyleniem standardowym danej zmiennej. Dla zbadania tego wskaźnika określono go zgodnie z (61) dla wszystkich badanych zmiennych objaśnianych. Wyniki przedstawiono na rys. 86.

Z diagramu tego wynika, że interesujące zmienne, o praktycznym znaczeniu dla systemu transakcyjnego, to nie tylko zmienne nr 9 i 5 (o najwyższych wartościach skumulowanego kapitału z rys. 85), ale także zmienne o numerach 4 (GBP/USD), 1 (DAX), 2 (EUR/USD), 8 (*Oil*). Warto zwrócić także uwagę na to, że tzw. niezawodność predykcyjna (znormalizowana stopa zwrotu) uzyskana za pomocą modeli nieliniowych jest ogólnie wyższa niż poprzednio uzyskana za pomocą modeli liniowych (rys. 79).

W odróżnieniu od wyników badań modeli liniowych wróciły tu do łask ważne pary walutowe – EUR/USD i GBP/USD. Nie powinno to przesłaniać najważniejszej konkluzji – wg tych modeli bardzo perspektywiczną w sensie predykcyjnym zmienną jest nasz indeks WIG20 i cena złota.



Rys. 86. Znormalizowana stopa zwrotu poszczególnych zmiennych objaśnianych badana w okresie 200 kroków dobowych.

Badania te jednoznacznie wskazują na zasadność tezy o możliwości wyekstrahowania z przestrzeni danych pomiarowych takich zmiennych, które są lepiej niż inne, opisywane wzajemnie. W tym sensie tak rozumiana predykcja wieloraka ma praktyczny sens. Wnioski o przewadze niektórych zmiennych nad pozostałymi pozostają jednak raczej w obszarze myślenia indukcyjnego – modele regresyjne są przecież rezultatem statystycznego uzasadnienia zależności opisujących. Są więc wynikiem zauważenia statystycznie potwierdzonej reguły dotyczącej szeregu drobnych i pozornie całkowicie chaotycznych zdarzeń. W gąszczu tych zdarzeń często niemożliwe jest bez użycia maszyny dostrzeżenie logicznego związku pomiędzy przyczyną i skutkiem. Jeżeli jest jednak dostrzegany to może w znacznym stopniu poprawić optymizm inwestora, jeżeli tego związku inwestor nie dostrzeże, to jeszcze nic z tego nie wynika.

A może wynika raczej, że taka właśnie, w tej chwili, jest struktura statystyczna danych. To dla inwestora szansa na wykorzystanie wiedzy (np. o zaobserwowanych regułach). To, że za chwilę ta struktura będzie inna nie powinno przekreślać szansy na wykorzystanie wyekstrahowanej wiedzy.

Podsumowanie

Metoda GMDH sprzężona z mechanizmami półautomatycznego obrotu na platformach brokerskich tworzy bardzo interesujący i obiecujący system transakcyjny. Przykłady użycia w różnych obszarach rynków finansowych wskazują na jej uniwersalność w aspekcie miejsca użycia, ale też na ograniczone możliwości wykorzystania. Podstawowym ograniczeniem metody jest niska moc obliczeniowa współczesnych maszyn. W metodzie konieczne jest bowiem transformowanie przestrzeni predyktorów do nowych wielowymiarowych przestrzeni pochodnych. To z kolei wymusza czasochłonne przeszukiwanie przestrzeni zmiennych objaśniających w celu określenia najbardziej obiecujących modeli predykcyjnych. Dodatkowo, eksperymentalnie sprawdzono, że budowane modele tylko przez chwilę są zdolne do opisania przyszłości i wymagają ustawicznej adaptacji, zarówno poprzez zmianę parametrów modelu (tu wielomianowego) jak też poprzez zmianę struktury tego modelu. Stąd, trudne wyzwanie obliczeniowe – sedno tzw. inteligencji obliczeniowej, które sprowadzić można do poszukiwania kompromisu pomiędzy dokładnością i pracochłonnością algorytmu. Można tu także skonstatować, że przydatność metody daje się stwierdzić wyłącznie poprzez obliczenia, poprzez eksperymenty na fizycznej rzeczywistości reprezentowanej przez macierze predyktorów. Przyszłość tej metodologii – chyba w wykorzystaniu mocy rozproszonych systemów obliczeniowych zgodnie z ideą *cloud computing*, ale także w poszukiwaniu nowych specyficznych dla metody algorytmów heurystycznego przeszukiwania przestrzeni zmiennych objaśniających.

Zapotrzebowanie na moc obliczeniową i lepsze algorytmu przeszukiwania wynika z przekonania, że warto budować modele o dużej złożoności. W rozdziale 1 widać było związek pomiędzy dokładnością predykcji a złożonością modelu.

Zaimplementowana przez autora do sześciu warstw sieć obliczeniowa z trudem daje się wyjaśnić do poziomu czwartej warstwy tak, by można było odtworzyć pierwotne współczynniki równań regresji (te, związane z pierwotnymi, nieprzetworzonymi danymi) dla najlepszych modeli. Z uwagi na złożoność tego opisu przeniesiony on został do dodatku, gdzie zajął ok. 5 stron. Opis warstwy piątej byłby już 4-5 krotnie obszerniejszy, a warstwy szóstej zająłby już kilkadziesiąt stron. W praktycznej implementacji tych algorytmów autor nie naśladował Aleksieja Ivakhnenki, który zazwyczaj przedstawiał modele wielomianowe w formie

uwzględniającej pierwotne predyktory z wyższymi stopniami złożoności (np. jako człony kwadratowe, sześciennie itd.). Ponieważ eksperymentalnie w niniejszej pracy dowiedziono, że model jest poprawny tylko przez przysłowiową chwilę, to dociekanie, które z danych wejściowych mają wpływ na ciągle zmieniającą się strukturę modelu nie było w istocie ważne. Ważna była zdolność systemu do samoorganizacji, do ciągłej zmiany struktury modelu i parametrów. Adaptacyjność tę wielokrotnie uzyskiwano i potwierdzono empirycznie jej zasadność i skuteczność.

Intencją było umożliwienie implementacji tej metody w dowolnym środowisku programowym. Środowiskiem wybranym przez autora jest Matlab. Zainteresowany czytelnik może wybrać dowolne inne o podobnych lub lepszych funkcjonalnościach.

Warto zauważyć, że obliczenia te dają się zrównoleglić z uwagi na wielokrotnie zagnieżdżaną pętlę. To istotny aspekt rozpatrywanego ograniczenia. Współczesne metody zrównoleglenia obliczeń uwzględniają możliwość tworzenia farm komputerów, czy zastosowania szybkich procesorów graficznych w celach naukowych. To jeden z kierunków rozwoju rozpatrywanej metody GMDH. Drugi, to spokojne oczekiwanie na spełnienie się prawa Moora „gwarantującego” wykładniczy rozwój mocy obliczeniowej.

Zupełnie inne podejście, to użycie intelektu zamiast maszyn. To zapewne, na razie, dopóty górujemy nad maszynami, najlepsze przyspieszenie wzrostu skuteczności metody. To poszukiwanie usprawnień metody GMDH w obszarze poszukiwania lepszych baz dla reprezentacji modeli i odpowiedniejszych kryteriów, poszukiwanie lepszych metod przeszukiwania przestrzeni predyktorów.

W pracy stosowano kilka kryteriów jakości prognoz charakterystycznych dla praktycznego handlu. Był więc stosowany w analizie portfelowej wskaźnik Sharpe’a czy Sortino, był także wskaźnik uwzględniający awersję do maksymalnego obsunięcia kapitału. W fazach testowania stosowane były w przypadku poszukiwania modeli regresyjnych nie tylko tradycyjne kryteria metody najmniejszych kwadratów, ale także minimaksowe. Zwłaszcza kryteria związane z awersją do maksymalnego obsunięcia są, zdaniem autora, bardziej naturalne i związane z osobowością inwestora niż kryteria stosowane w modelach Markowitza czy Sharpe’a. Oczywiście zawsze lepiej jest znaleźć model, który przy danym poziomie ryzyka pozwoli na osiągnięciu wyższego niż niższego zysku. Zawsze lepiej jest osiągnąć założony zysk przy niższym niż wyższym poziomie ryzyka. Jednakże jak zmierzyć tę szczególną cechę inwestora, który wybiera poziom ryzyka „nierozsądnie”? Dlaczego istnieją na rynku tyśiące alternatyw wobec bezpiecznych obligacji?

Praca jest w jakiejś mierze wyrazem fascynacji autora dorobkiem zmarłego niedawno (w 2008 roku) ukraińskiego wybitnego uczonego prof. Aleksieja Ivakhnenki. Ivakhnenko nigdy nie koncentrował się na zadaniach predykcji na rynkach finansowych. Z przeglądu licznych źródeł tego Autora wynika, że do kilku publikacji dotyczących tej problematyki namówili Go raczej niemieccy partnerzy – Johann-Adolf Mueller i Frank Lemke. Kontakt ten nie przyniósł spektakularnych praktycznych rezultatów, ale nie zniechęcił też (np. Franka Lemkego, z którym autor niniejszej pracy ma bezpośredni kontakt) do kontynuacji usprawnień metody. Nigdy jednak nie był, w tej grupie uczonych, rozpatrywany aspekt handlu automatycznego, częstotliwego, o zupełnie nieoczekiwanych cechach. Wprowadzenie infobotów do zarządzania kontami klientów to, w przekonaniu autora, niedaleka przyszłość. To nieodłączny atrybut członka przyszłego społeczeństwa informacyjnego, który będzie poszukiwał specyficznego, outsourcingowego wsparcia osobistych finansów. To, z dużą dozą prawdopodobieństwa, będzie typowe zachowanie drobnego inwestora zniecierpliwionego natarczywością i nieodpowiedzialnością banków, zwłaszcza po fali bankructw tych ostatnich w okresie trwającego kryzysu finansowego.

Handel tradycyjny i w pełni automatyczny to dwa różne światy – w aspekcie teorii podejmowania decyzji, strategii matematycznych, psychologii inwestora, szybkości działania, wpływu technologii komunikacyjnych na teorie racjonalnego inwestowania (i zarządzania ryzykiem). Zastosowanie tradycyjnych koncepcji dla handlu automatycznego może być zupełnie bezzasadne. To tak, jakby rozpatrywać dwa środki komunikacji – wóz konny i samolot. Przyczepiwszy skrzydła z boków wozu konnego, konstatować ze zdziwieniem, że nie unosi się.

Rozwiązaniem dla wozu jest zmiana koni. Na bardzo szybkie.

Bibliografia

1. Banaszak Z., Józefczyk J. *Towards dedicated decision support tools*. Applied Computer Science vo. 1 no 1 2005, ATH, Bielsko-Biała 2005, pp. 161-181.
2. Bargiela A., Pedrycz W. *Recursive Information Granulation: aggregation and interpretation issues*. IEEE Transactions on Systems, Man and Cybernetics Part B – Cybernetics 33 (1), 96-112.
3. Borawski M. *Analysis of Properties of Mean Value - Pseudo -Variance Representation*. Polish Journal of Environmental Studies vol.17 No 3b, 2008, p. 39-43.
4. Breiman L., Friedman J.H. , Olshen R.A., Stone C.J., *Classification and Regression Trees*, Wadsworth International Group, Monterey, 1984.
5. Brock W., Lakonishok j., LeBaron B. *Simple technical trading rules and stochastic properties of stock returns*. Journal of Finance 47(1992), 1731-1764.
6. Brown G., Wyatt J., Harris R., Yao X. *Diversity creation methods: A Survey and Categorisation*. University of Birmingham, 2004.
7. Burda Z. *Fizyka i zarządzanie ryzykiem finansowym*. Postępy fizyki, tom 57, zes.3, 2006
8. Cai B.M., Cai C.X., Keasey K. *Market Efficiency and Returns to Simple Technical Trading Rules: Further Evidence form US, UK, Asian and Chinese Stock Markets*. Asia-Pacific Financial Markets, Springer 2005.
9. Chan Kim W., Mauborgne R. *Strategia błękitnego oceanu*. MT Biznes, Warszawa 2005
10. Chua S. *Sammy Chua's Day Trade*. Wiley Trading, 2007.
11. Cizek P., Härdle W, Weron R. *Statistical Tools for Finance and Insurance*, Springer-Verlag, Berlin 2005.
12. Czapiewski P., Samborska-Owczarek A. *Applicability Verification of Pattern – Based Decision Making Method in Real-life Conditions*. Polish Journal of Environmental Studies vol.16 No 4a, 2007, p. 57-62.
13. DeBontd W., Thaler R. *Does the stock market overreact?* Finance of Journal 40(3) 793-805
14. Dehuri S., Cho S.-B. *Multicriterion Pareto based particle swarm optimized polynomial neural network for classification: A review and state-of-the-art*. Computer Science Review 3(2009) 19-40.

15. Duch W. *Computational Intelligence and Where Is It Going*. In Challenges for Computational Intelligence (Ed. J. Kacprzyk). Springer. vol. 63/2007 s. 1-13.
16. Duda R.O., Hart P.E., Stork D.G., *Pattern Classification*, John Wiley, New York (2000).
17. Elder A. *Come Into My Trading Room*. Wiley Trading, 2002.
18. Elder J.F. *A review of Machine Learning, Neural and Statistical Classification* (eds. Michie, Spiegelhalter & Taylor; Ellis Horwood, 1994), *Journal of the American Statistical Association* 91, no. 433: 436-437.
19. Elton E.J, Gruber M.J. *Nowoczesna teoria portfelowa i analiza papierów wartościowych*. WIG-PRESS, Warszawa 1998.
20. Fama E. Market efficiency, long-term returns, and behavioral finance. *Journal of Financial Economics* 49 (1998) 283-306.
21. Fiszeder P. *Dynamiczna alokacja aktywów – model Markowitza, Rynki finansowe – prognozy a decyzje*, Acta Universitatis Lodziensis, Folia Oeconomica, 177, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2004.
22. Fiszeder P. *Forecasting the Volatility of Polish Stock Index – WIG20*, w: Forecasting Financial Markets. Theory and Applications, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2005.
23. Fiszeder P. *Modelling Financial Processes with Long Memory in Mean and Variance*, Dynamic Econometric Models 7, UMK, Toruń, 2006.
24. Friesen G.C.F, Weller P, Dunham L.M. *Price trends and patterns in technical analysis: A theoretical and empirical examination*. *Journal of Banking & Finance* 33(2009) 1089-1100.
25. Fujimoto K., Nakabayashi S. *Applying GMDH Algorithm to Extract Rules from Examples*. *Systems Analysis Modelling Simulation*, Volume 43, Number 10, 2003/10 pp. 1311-1319.
26. Gatnar E. *What is Data Mining*. *Statistics in Transition* vol.5 no. 5 p. 837-842
27. Gencar R. *Linear, non-linear and essential foreign exchange rate prediction with simple technical trading rules*. *Journal of International Economics* 47(1999) 91-107.
28. Grabczewski K. *Zastosowanie kryterium separowalności do generowania reguł klasyfikacji na podstawie baz danych*. pr. dokt. IBS, Warszawa, 2003.
29. Greshilov A.A., Stakun V.A., Stakun A.A. *Matematicheskiye metody postroeniya prognozov*. Radio i sviaz. Moskva 1997.

Bibliografia

30. Hashem S. , Schmeiser B. *Improving Model Accuracy using Optimal Linear Combinations of Trained Neural Networks*. IEEE Transactions on Neural Networks, 1995.
31. Hashem S. , Schmeiser B. *Improving Model Accuracy using Optimal Linear Combinations of Trained Neural Networks*. IEEE Transactions on Neural Networks, 1995.
32. Hia Jong Teoh, Tai Liang Chen, Ching Hsue Cheng, Hsing Hui Chu. *A hybrid multi-order fuzzy time series for forecasting stock markets*. Expert Systems with Applications 36 (2009) 7888-7897.
33. Hia Jong Teoh, Tai-Liang Chen, Ching-Hsue Cheng. *Frequency-Weighted Fuzzy Time-Series Based on Fibonacci Sequence for TAIEX Forecasting*. Emerging Technologies in Knowledge Discovery and Data Mining. Springer 4819/2007.
34. Ivakhnenko A., Ivakhnenko G. *Problems of Further Development of the Group Method of Data Handling Algorithms*. Part I. Pattern Recognition and Image Analysis vol.10 No.2, 2000, pp. 187-194.
35. Ivakhnenko A., Ivakhnenko G. *Problems of Further Development of the Group Method of Data Handling Algorithms, Part I*. Pattern Recognition and Image Analysis vol.10 No.2, 2000, pp. 187-194.
36. Ivakhnenko A., Ivakhnenko G., Mueller J. *Self- organization of Neural Network with Active Neurons*. Pattern Recognition and Image Analysis 1999 v.4 no. 2 pp. 185-196.
37. Ivakhnenko A., Stepashko V. *Pomekhoustoicivost modelirovanija*, Icjier 1986 , Naukovaja Dumka.
38. Ivakhnenko A.G., *An Inductive Sorting Method for the Forecasting of Multi-dimensional Random Processes and Events with the Help of Analogs Forecast Complexing*, Pattern Recognition and Image Analysis, 1991, vol. 1, no.1, pp.99-108.
39. Ivakhnenko A.G., *Dolgosrochnoe Prognozirovanie i Upravlenie Slozhnymi Sistemami* (Long-Term Forecasting and Control of Complex Systems), Kiev: Tekhnika, 1975.
40. Ivakhnenko A.G., Ivakhnenko G.A., Müller J.A., *Self-Organization of Neural Nets with Active Neurons*, Pattern Recognition and Image Analysis, 1994, vol. 4, no. 2. pp. 185-196.
41. Ivakhnenko A.G., Ivakhnenko G.A., Müller J.A., *Self-Organization of Optimum Physical Clustering of the Data Sample for Weakened Description and Forecasting of Fuzzy Objects*, Pattern Recognition and Image Analysis, 1993, vol. 3, no. 4, pp. 415-422.

42. Ivakhnenko A.G., Ivakhnenko G.A., *Obzor zadac po algoritmam Metoda Gruppovogo Uceta Argumentov* . Pattern Recognition and Image Analysis, 1995 vol. 5 no. 4 pp. 527-535.
43. Ivakhnenko A.G., Ivakhnenko G.A., *Perceptron Synthesis according to Clustering-Balance Criterion*. Pattern Recognition and Image Analysis, 1995, vol. 5, no. 3, pp. 337-341.
44. Ivakhnenko A.G., Ivakhnenko G.A., *Simplified Linear Programming Algorithm as Basic Tool for Open-Loop Control*, Systems Analysis, Modelling, and Simulation, (SAMS), 1995, vol.18-19, pp.315-319.
45. Ivakhnenko A.G., Müller J.A., *Parametric and Non-parametric Selection Procedures in Experimental Systems Analysis*, Systems Analysis, Modeling, and Simulation, 1992, vol. 9, pp. 157-175.
46. Ivakhnenko A.G., Müller J.A., *Problems of Computer Clustering of the Data Sampling of Objects under Study*, Sov. J. Automation Inf. Sci., 1991, vol. 24, no. 1, pp. 58-67.
47. Ivakhnenko A.G., *Samoobujascijesja systemy raspoznavanija i avtomatycznego upravlenija*. Tekhnika, Kijev 1969.
48. Ivakhnenko A.G., Stepashko V.S., *Pomekhoustoichivost' Modelirovaniya* (Noise Immunity of Modeling), Kiev: Naukova Dumka, 1985.
49. Ivakhnenko A.G., Yurachkovskij Yu.P., *Modelirovanie Slozhnykh Sistem po Eksperimental'nym Dannym* (Modeling of Complex Systems from Experimental Data), Moscow: Radio i Svyaz', 1987.
50. Ivakhnenko A.G., Zaichenko J.P., Dimitrov V.D. *Priniatije resenij na osnove samoorganizacii*. Sovetskoe Radio, Moskwa 1976.
51. Ivakhnenko G.A., *Self-Organization of Neuronet with Active Neurons for Effects of Nuclear Test Explosions Forecastings*. System Analysis Modeling Simulation (SAMS), 1995, vol.20, pp.107-116.
52. Jajuga K. *Metody statystyczne w finansach*. Statsoft Polska, 2003.
53. Jajuga K., Jajuga T. *Jak inwestować w papiery wartościowe*. PWN, Warszawa 1994.
54. Jajuga K., Kuziak K., Markowski P. *Inwestycje finansowe*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego we Wrocławiu, Wrocław 1998.
55. Jing Dang, Brabazon A., Edelman D., O'Neill M. *An Introduction to Natural Computing in Finance*. LNCS, PublisherSpringer Berlin / Heidelberg Volume 5484/2009.

56. Jinhong You, Xian Zhou, Li-Xing Zhu. *Inference on a regression model with noised variables and serially correlated errors*. Journal of Multivariate Analysis 100(2009) 1182-1197.
57. Kacprzyk J., Wilbik A., Zadrozny S. *A linguistic quantifier based aggregation for a human consistent summarization of time series*. In: J. Lawry, E. Miranda, A. Bugarin, S. Li, M. A. Gil, P. Grzegorzewski, O. Hryniewicz (Eds.): *Soft Methods for Integrated Uncertainty Modeling*. Springer-Verlag, Berlin Heidelberg 2006, ss. 183-190, Seria: *Advances in Soft Computing*.
58. Kacprzyk J., Wilbik A., Zadrozny S. *Using a genetic algorithm to derive a linguistic summary of trends in numerical time series*. In: *Proceedings of 2006 International Symposium on Evolving Fuzzy Systems*, Ambleside, UK, September 7-9, 2006, IEEE Press, ss. 137-142.
59. Kacprzyk J., Zadrozny S. *Fuzzy linguistic data summaries as a human consistent, user adaptable solution to data mining*. In: Gabrys B., Leiviskä K., Strackeljhan J. (Eds.): *Do smart adaptive systems exist?* Springer Verlag, Heidelberg New York 2005, ss. 321-339, Seria: *Studies in fuzziness and soft computing*.
60. Kohavi R. *A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection*. Computer Science Department, Stanford University, 1995.
61. Kompa K., Matuszewska Janica A. *Examination of Warsaw Stock Exchange Indexes Behaviour: Applications of Rolling Windows Variance Ratio Test*. Polish Journal of Environmental Studies vol.17 No 3b, 2008, p. 150-154.
62. Korbicz J., Marcin Mrugalski M. *Confidence estimation of GMDH neural networks and its application in fault detection systems*. International Journal of Systems Science, 2008, Vol. 39, no 8, pp. 783-800.
63. Krawczak M. *Theoretical foundations of neural networks prediction*. W: Studziński J., Drelichowski L., Hryniewicz O., Kacprzyk J. (red.) *Technologie informatyczne w zarządzaniu, systemy wspomaganie decyzji*. IBS PAN, Warszawa 2000, pp. 245-254.
64. Krawczak M., Jakubowski A., Konieczny P., Kulikowski R., Miklewski A., Szkatuła G. *Aktywne zarządzanie inwestycjami finansowymi – wybrane zagadnienia*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2003.
65. Krogh A., Vedelsby J., *Neural Network Ensembles, Cross Validation and Active Learning*. NIPS, 7 of 1995.
66. Kruttsinger J. *Systemy transakcyjne. Sekrety mistrzów*. Wig-Press, Warszawa 1999.
67. Kulikowski R. *Wspomaganie inwestycji obarczonych ryzykiem*. W: Krawczak M., Jakubowski A., Konieczny P., Kulikowski R., Miklewski A., Szkatuła

- G.: *Aktywne zarządzanie inwestycjami finansowymi – wybrane zagadnienia*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2003, ss. 253-268.
68. Kuś J., Korbicz J. *Statyczne i dynamiczne sieci GMDH w Sieci neuronowe* (eds. Duch, Korbicz, Rutkowski, Tadeusiewicz) w serii pod red. M. Nałęcza *Biocybernetyka i inżynieria medyczna*, PAN, Warszawa 2000.
 69. Kutner R. *Niegaussowskie procesy stochastyczne i niedebye'owska relaksacja oraz zdarzenia rzadkie w świecie realnym*. Uniwersytet Warszawski, 2006.
 70. Lemke F., Mueller J.A. *Self-Organizing Data Mining A Portfolio Trading System*. *Journal of Computational Intelligence in Finance*, 1997/05 pp 12-26.
 71. LeSage J., Sudipto Banerjee, Fischer M.M., Congdon P. *Spatial Statistics: Methods, models & computation*. *Computational Statistics and Data Analysis* 53(2009) 2781-2785.
 72. Łuniewska M. *The Construction and Evaluation of Sectorial Portfolios*. *Polish Journal of Environmental Studies* vol.16 No 4a, 2007, p. 161-165.
 73. Madala H.R., Ivakhnenko A. G. *Inductive Learning Algorithms for Complex Systems Modelling*. CRC Press 1994.
 74. Makridakis S., Wheelwright S., Hyndman R. *Forecasting: Methods and Applications*. John Wiley and Sons Inc. 1998.
 75. Marjak H. *The Architecture Selection for Neural Network – a Financial Criterion or a Prediction Criterion*. *Polish Journal of Environmental Studies* vol.17 No 3b, 2008, p. 255-261.
 76. McLachlan G.J. *Discriminant Analysis and Statistical Pattern Recognition*, Wiley Interscience, New York, 1992.
 77. Merz Ch. J. *Classification and Regression by Combining Models*. Ph.D. thesis, University of California, Irvine, 1998.
 78. Mohsen Mehrara, Ali Moeini, Mehdi Ahrari, Ali Erfanifard. *Investing the efficiency in oil futures market based on GMDH approach*. *Expert Systems with Applications* 36(2009) 7479-7483.
 79. Mrugalski, M., Arinton, E., Korbicz, J. *Sieci neuronowe typu GMDH w diagnostyce systemów przemysłowych*. *Pomiary, Automatyka, Kontrola*. 2002, R. 48, nr 2, s. 10-14.
 80. Mueller J., Lemke F. *Self-organizing Data Mining. An Intelligent Approach to Extract Knowledge from Data*. Berlin, 1999.
 81. Muriel A., *Short-term predictions in forex trading*, *Physica A* 344 (2004) 190-193

Bibliografia

82. Niederliński A. *Systemy komputerowe automatyki przemysłowej – zastosowania*. WNT, Warszawa 1985.
83. Nowak R. *Statystyka dla fizyków*. WN PWN, Warszawa 2002.
84. Onwubolu G. *Hybrid Self-Organizing Modeling Systems*. Springer 2009.
85. Owens S., Lizotte O. *When to trade*. FX Engines, 2004.
86. Palit A., Popovic D. *Computational Intelligence in Time Series Forecasting*. Springer 2005.
87. Park B.J., Oh S.K., Ahn T.C., Pedrycz W. *A study on multilayer fuzzy polynomial interference system based on an extended GMDH algorithm*. Proc. 8th IEEE Inf. Conference on Fuzzy Systems, Seoul 1999.
88. Pawlak Z. *Rough sets and intelligent data analysis*. Information Sciences, 147, 2002, 1-12.
89. Pawlak Z., Polkowski L., Skowron A. *Rough Set Theory*. In: B. Wan, (ed.), *Encyclopedia of Computer Science and Engineering*, Wiley, New York 2004.
90. Pedrycz W. *Computational Intelligence: An Introduction*. CRC Press 1997.
91. Penrose R. *Droga do rzeczywistości*. Prószyński i Ska. Warszawa 2006.
92. Piotrowski E. *Dwoistości wartości kapitału*. Politechnika Białostocka 2002.
93. Piotrowski E., Śładkowski J. *Quantum Diffusion of Prices and Profits*. Physica A, 345(2005) s. 185-195.
94. Piper J. *The Way to Trade*. Harriman Modern Classic, 2006.
95. Raghuraj R.K., Lakshminarayanan S. *Variable Predictive Models – A new multivariate classification approach for pattern recognition application*. Elsevier, Pattern Recognition vol. 42/1, 2009 p.7-17 [sciencedirect.com].
96. Raghuraj R.K., Lakshminarayanan S. *VPMCD: Variable Interaction Modeling Approach for Class Discrimination in Biological Systems*. FEBS Letters, 581, 826-830, 2007.
97. Rejer I. *Fuzzy Expert Model of Index WIG20*. Polish Journal of Environmental Studies vol.17 No 3b, 2008, p. 365-370.
98. Rua A., Nunes C. *International comovement of stock market: A wavelet analysis*. Journal of Empirical Finance, 2009.
99. Rutkowski L. *Metody i techniki sztucznej inteligencji*, WNT, Warszawa 2006.
100. Satchwell Ch. *Pattern Recognition and Trading Decisions*. McGraw Hill, New York 2005.

101. Sawaragi, Y., Soeda, T., Tamura, H. et al., *Statistical Prediction of Air Pollution Levels Using Non-Physical Models*, Automatica (IFAC), 1979, vol. 15, no. 4.
102. Sawaragi, Y., Soeda, T., Tamura, H. et al., *Statistical Prediction of Air Pollution Levels Using Non-Physical Models*, Automatica (IFAC), 1979, vol. 15, no. 4, pp. 441-452.
103. Schwager J.D. *Analiza techniczna rynków terminowych*. WIG-Press, Warszawa 2002.
104. *Self-Organizing Methods in Modeling*, Statistics: Textbooks and Monographs, Farlow, S.J., Ed., New York: Marcel Dekker Inc., 1984, vol. 54.
105. Shahwan T. , Lemke F. *Forecasting Commodity Prices for Predictive Decision Support Systems*. EFITA/WCCA 2005, Via Real, Portugal.
106. Sharpe W.F. *The Sharpe Ratio*. The Journal of Portfolio Management Fall, 1994 s.49-58
107. Sheng-Hsun Hsu, Po-An Hsieh, Ting-Chih Chih, Kuei-Chu Hsu. *A two-stage architecture for stock price forecasting by integrating self-organizing map and support vector regression*. Expert Systems with Applications 36(2009) 7947-7951.
108. Shleifer A. *An inefficient market: An introduction to behavioral finance*. Clarendon Lectures. Oxford University Press 2000.
109. Sortino F., Van der Meer R. *Downside Risk*. The Journal of Portfolio Management Fall., 1991 s.27-31.
110. Sung-Kwun Oh, Pedrycz W. *The design of self-organizing Polynomial Neural Networks*. Information Sciences, 141 (2002) pp.237-258.
111. Szymański S. *Integracja metod klasyfikacji odkrywania wiedzy w systemie ekspertowym*. pr. dokt. Wydział Informatyki Politechniki Szczecińskiej, 2006.
112. Tadeusiewicz R. *Sieci neuronowe*. Akademicka Oficyna Wydawnicza RM, Warszawa 1993.
113. Tak-chung Fu, Fu-lai Chung, Luk R., Chak-man Ng. *Representing Financial Time Series based on data point importance*. In: Engineering Applications of Artificial Inteligence 21(2008) p.277-300.
114. Tarczyński W, Łuniewska M. *Dywersyfikacja ryzyka na polskim rynku kapitałowym*. Placet, Warszawa 2004.
115. Tarczyński W, Łuniewska M. *Metody wielowymiarowej analizy porównawczej na rynku kapitałowym*. PWN, Warszawa 2006.

Bibliografia

116. Triseev, Y. P., *Approaches to the Solution of Mathematical Programming Problems on the Basis of Heuristic Self-Organization*, Sov. J. Automation Inf. Sci., 1987, vol. 20, no. 2, pp. 30-37.
117. Tsoumakas G., Katakis J., Vlahavas I. *Effective Voting of Heterogeneous Classifiers*. Department of Informatics, Aristotele University of Thessaloniki, 2004.
118. Tsymbal A., Pechenizkiy M., Cunningham P. *Diversity in search strategies for Ensemble Feature Selection*. Department of Computer Science, Trinity College Dublin, 2003.
119. Vapnik V. *Statistical Learning Theory*, Wiley-Interscience, New York, 1998.
120. Wagner N., Michalewicz Z., Khouja M., Mc Gregor R.R. *Forecasting with a Dynamic Window of Time: The DyFor Genetic Program Model*. Springer Berlin/Heidelberg vol.3490/2005, 205-215.
121. Wagner N., Michalewicz Z., Khouja M., Mc Gregor R.R. *Time Series Forecasting for Dynamic Environment: The DyFor Genetic Program Model*. IEEE Transaction on Evolutionary Computation, vol. 11 n. 4 /2007, s.433-453.
122. Weiss S., Kulikowski C., *Computer Systems that Learn: Classification and Prediction Methods from Statistics, Neural Networks, Machine Learning and Expert Systems*, San Francisco, 1991.
123. Weron A., Weron R. *Inżynieria finansowa: Wycena instrumentów pochodnych, Symulacje komputerowe, Statystyka rynku*. WNT, Warszawa 2005.
124. Wierzbicki A. *Badania operacyjne i systemowe w epoce cywilizacji wiedzy*. Instytut Badań Systemowych PAN seria Badania systemowe tom 64, Warszawa 2008, s. 13-35.
125. Wiliński A. *Analiza szeregów czasowych w ujęciu bayesowskim dla funkcji dyskryminacyjnej opartej na rozkładzie gaussowskim*. XI Sesja Informatyki Wydział Informatyki Politechniki Szczecińskiej, Wydawnictwo Informa, Szczecin 2006.
126. Wilinski A. *GMDH Based Computational Intelligence Method for Prediction of Financial Markets*. "Intellectual Systems of Decision Making and The Problems of Computational Intelligence (ISDMCI'2008)", Eupatoria, Tom 1 cz.1, 2008.
127. Wiliński A. *Risk Estimation of Consecutive Losses in Automated Currency Trading*. Polish Journal of Environmental Studies vol.16 No 4a, 2007, p. 380-385.
128. Wiliński A., Czapiewski P. *GMDH-based Multiple Prediction of Financial Markets*. Polish Journal of Environmental Studies vol.17 No 3b, 2008 p.73-78.

129. Wilinski A., Czapiewski P., Sienkiewicz S. *Model interfejsu graficznego w systemie wspomagania decyzji w obrocie instrumentami finansowymi*. X Sesja Informatyki Wydział Informatyki Politechniki Szczecińskiej, Wydawnictwo Informa, Szczecin 2005.
130. Witkowska D. *Artificial Neural Networks Application to Forecast the Participation Units of the Open-End Investment Fund*. Polish Journal of Environmental Studies vol.17 No 3b, 2008, p. 523-529.
131. Witkowska D., Matuszewska A., Kompa K. *Wprowadzenie do ekonometrii dynamicznej i finansowej*. Wydawnictwo SGGW, Warszawa 2008.
132. Yungho Leu, Chien-Pang Lee, Yie-Zu Jou. *A distance-based fuzzy time series for exchange rates forecasting*. Expert Systems with Applications 36(2009) 8107-8114.
133. Zhang Yudong, Wu Lenan. *Stock market prediction of S&P 500 via combination of improved BCO approach and BP neural network*. Expert Systems with Applications 36(2009) 8849-8854.
134. Zhohnarskij, A. A., *Agglomerative Cluster Analysis Procedures for Multidimensional Objects: A Test for Convergence*. Pattern Recognition and Image Analysis, 1992, vol. 2, no. 4, pp. 388-390.
135. Weisstein, Eric W. *Moore-Penrose Matrix Inverse*. Wolfram MathWorld [online] <http://mathworld.wolfram.com/Moore-PenroseMatrixInverse.html> [dostęp: 02/2008]
136. Stokalski B. Cloud computing – informatyka bez kłopotów [dostęp 05/2008] <http://www.infovidematrix.pl/inspiracje/?p=61#more-61>
137. www.e-forex.ro [dostęp 08/2008]
138. www.oanda.com [dostęp 05/2007 – 09/2008]
139. www.investopedia.com [dostęp 02/2005-10/2008]
140. www.wanadoo.com [dostęp 02/2006-11/2008]
141. www.gmdh.com [dostęp 06/2007-11/2008]
142. www.fxengines.com [dostęp 12/2007 – 02/2009]

Zasadniczym celem niniejszej monografii jest potwierdzenie tezy autora, że fuzja matematyki i technologii egzemplifikowana powiązaniem metod predykcji ze strategią inwestycyjną opartą na mechanizmach technologicznych platform brokerskich daje nową synergetyczną wartość.

Praca koncentruje się na pragmatycznym powiązaniu predykcji z możliwościami jej internetowej (automatycznej) realizacji. Wymagało to uwzględnienia szeregu niespodziewanych ograniczeń natury technologicznej. W pracy przedstawiono wyniki wielu eksperymentów w przestrzeniach historycznych różnych instrumentów finansowych.

ISSN 0208-8029

ISBN 9788389475237

Instytut Badań Systemowych PAN

Tel. Centrala 022-38 10 100 / fax 022-38 10 105 e-mail: ibs@ibspan.waw.pl