http://dx.doi.org/10.18778/7525-936-0

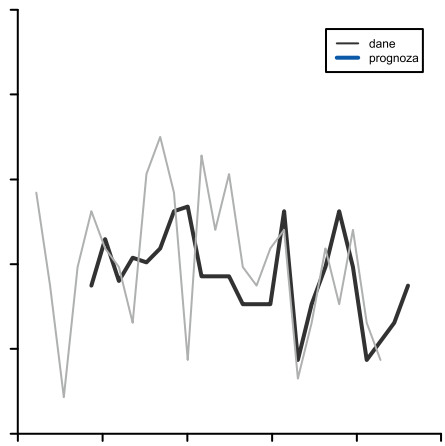
Adam Kucharski

Prognozowanie
szeregów czasowych
metodami ewolucyjnymi

40 LAT
WYDAWNICTWA
UNIWERSYTETU
ŁÓDZKIEGO

Adam Kucharski

Prognozowanie szeregów czasowych metodami ewolucyjnymi



WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

ŁÓDŹ 2013

Adam Kucharski – Katedra Badań Operacyjnych, Wydział Ekonomiczno-Socjologiczny
Uniwersytet Łódzki, 90-214 Łódź, ul. Rewolucji 1905 r. nr 41/43

RECENZENT
Tadeusz Trzaskalik

SKŁAD I ŁAMANIE
Adam Kucharski

PROJEKT OKŁADKI
Łukasz Orzechowski

© Copyright by Uniwersytet Łódzki, Łódź 2013

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego
Wydanie I. W.06283.13.0.H

ISBN (wersja drukowana) 978-83-7525-936-0
ISBN (ebook) 978-83-7969-163-0

Wydawnictwo Uniwersytetu Łódzkiego
90-131 Łódź, ul. Lindleya 8
www.wydawnictwo.uni.lodz.pl
e-mail: ksiegarnia@uni.lodz.pl
tel. (42) 665 58 63, faks (42) 665 58 62

Spis treści

Wstęp 7

Rozdział 1. Algorytmy genetyczne – charakterystyka metody 11
1.1. Klasyczny algorytm genetyczny 11
1.1.1. Matematyczne podstawy algorytmów genetycznych 16
1.1.2. Przykład działania klasycznego algorytmu genetycznego 21
1.2. Przegląd modyfikacji wpływających na konstrukcję chromosomu 30
1.2.1. Zmiany reprezentacji genotypu 30
1.2.2. Działania wpływające na porządek genów w chromosomie 33
1.2.3. Kodowanie rzeczywiste chromosomu 36
1.3. Przegląd modyfikacji operatorów algorytmu 40
1.3.1. Nieklasyczne metody selekcji chromosomów 40
1.3.2. Model uogólnionego krzyżowania 47
1.3.3. Wykorzystanie zjawiska nisz i gatunków 48
1.4. Przegląd modyfikacji zmieniających postać funkcji przystosowania 53
1.4.1. Skalowanie funkcji przystosowania 53
1.4.2. Algorytmy ze zmienną liczebnością populacji 55
1.4.3. Zadania z ograniczeniami 57
1.4.4. Zastosowanie algorytmów genetycznych do optymalizacji zadań wielokryteriowych 59
1.5. Strategie ewolucyjne 61
1.6. Algorytm genetyczny jako element algorytmów ewolucyjnych 63

Rozdział 2. Prognozowanie – własności i metody 65
2.1. Podstawowe pojęcia 65
2.2. Dekompozycja szeregu czasowego 71
2.3. Prognozowanie przy pomocy metod mechanicznych 75
2.3.1. Metody naiwne 75
2.3.2. Metoda średniej ruchomej 77
2.3.3. Metoda wygładzania wykładniczego Browna 81
2.3.4. Metoda Holta 82
2.3.5. Metoda Wintersa 84
2.4. Błędy prognoz *ex post* 88
2.5. Charakterystyka metod analitycznych 95

Rozdział 3. Algorytm genetyczny w prognozowaniu 99
3.1. Prognostyczny algorytm genetyczny 99

3.1.1. Uwzględnianie trendu w PAG 109
3.1.2. Uwzględnianie sezonowości 111
3.2. Algorytm bazujący na średniej ruchomej 115
3.3. Prognozowanie kaskadowe 119
3.3.1. Podejście bayesowskie w prognozowaniu kaskadowym 123
3.3.2. Prognozowanie kaskadowe – przykład 127

Rozdział 4. Prognozy wybranych kategorii przy pomocy PAG 133
4.1. Prognozy *ex post* i *ex ante* wybranych szeregów danych 133
4.1.1. Prognozy notowań i wolumenu obrotów danych giełdowych 133
4.1.2. Prognozy notowań kursów walut 156
4.1.3. Prognozy IRDN oraz inflacji 169
4.1.4. PAG a zmienne losowe 177
4.2. Efektywność PAG podczas przeszukiwania zbioru wariantów prognostycznych 181
4.2.1. Zmiany przystosowania najlepszego chromosomu populacji 183
4.2.2. Zmiany średniego przystosowania chromosomów populacji 197
4.3. Dobór parametrów PAG 208

Zakończenie 215

Bibliografia 217

Wstęp

Z prognozami stykamy się częściej niż może to się wydawać na pierwszy rzut oka. Pierwsze skojarzenie zwykle wiąże się z przepowiadaniem pogody, a tymczasem z prognoz (niekiedy mówimy o przewidywaniach) korzystamy na co dzień i to często nieświadomie. Wyobraźmy sobie, że zastanawiamy się nad wyborem trasy pokonywanej samochodem w mieście. Nim wyruszymy w drogę rozważamy na przykład czy i gdzie możemy trafić na korki o określonej porze dnia. Uwzględnienie spodziewanych utrudnień w ruchu spowodowanych godzinami szczytu stanowi właśnie przykład nieświadomego przewidywania. Nie planujemy jednak zajmować się bezrefleksyjnymi zachowaniami. Planujemy sięgnąć po metody naukowe, oparte na świadomym, racjonalnym działaniu. Spośród rożnych obszarów zainteresowania prognozowania jako nauki zdecydowaliśmy się na predykcję na podstawie szeregów czasowych. Skłoniło nas do tego kilka przyczyn.

Po pierwsze, w przypadku niektórych zjawisk decyzje należy podejmować na podstawie samych szeregów czasowych gdyż albo nie ma dość czasu na poszukiwanie zewnętrznych czynników wpływających na dane zdarzenie, albo takowych nie da się wskazać w jednoznaczny sposób. Na przykład decydent zainteresowany podjęciem szybkiej decyzji dotyczącej rynku finansowego może nie dysponować wystarczającą ilością czasu aby zbudować model ekonometryczny. Ma dostęp do szeregów notowań i niejednokrotnie tyle musi mu wystarczyć.

Po drugie uważamy, że wciąż nie powiedziano ostatniego słowa jeśli chodzi o metody, które można w przypadku szeregów czasowych wykorzystać. Mamy tu na myśli zarówno modyfikacje istniejących jak i opracowywanie nowych metod. Istnieje już wprawdzie całkiem całkiem pokaźny ich zbiór, ale wciąż można znaleźć słabo zbadane obszary zastosowań. Wiążą się one m.in. z uchylaniem założeń ograniczających zastosowania metod do szeregów czasowych o ściśle określonej dekompozycji. Dowiedziemy, że możliwym jest uogólnienie restrykcyjnych warunków stosowalności wybranych metod, aby osiągnąć większą elastyczność ich wykorzystania.

Po trzecie, opieranie się w prognozach na informacjach na temat mechanizmów odpowiedzialnych za zachowanie zjawiska, które obecne są w szeregu czasowym bynajmniej nie wyklucza tej drogi postępowania w porównaniu

z narzędziami o bardziej skomplikowanej konstrukcji. W konfrontacji z nimi (na przykład jedno- i wielorównaniowymi modelami ekonometrycznymi) takie postępowanie wydaje się zbyt uproszczone, żeby nie powiedzieć „trywialne". Praktyka jednak pokazuje, że metoda prostsza nie oznacza automatycznie metody gorszej. Dobrze to ilustruje porównanie metod naiwnych i wygładzania wykładniczego. W przypadku tych drugich optymalizacja parametrów potrafi doprowadzić do bardzo dokładnego dopasowania prognoz do danych z próby co jednak wcale nie gwarantuje równie wysokiej trafności predykcji poza samą próbą. Metody naiwne potrafią prognozować te same dane ze zbliżoną dokładnością a są o wiele łatwiejsze do zastosowania.

Kolejna przyczyna wiąże się z planowanym wykorzystaniem algorytmów genetycznych i ewolucyjnych. Wywodzą się one ze starej jak sama ludzkość chęci naśladowania lub odwzorowywania procesów, zjawisk i zdarzeń obserwowanych w naturze. Należy zauważyć, że tego typu narzędzia, choć rozwijane od kilku dziesięcioleci, wciąż słabo eksplorują niektóre obszary potencjalnych zastosowań. Zaliczymy do nich m.in. prognozowanie. Przeanalizowana na potrzeby tej pracy literatura dowodzi, że naprawdę niewiele publikacji traktuje o połączeniu algorytmów genetycznych i ewolucyjnych z prognozowaniem. Jeśli już do tego dochodzi to często celem staje się tak naprawdę opracowanie generatora reguł decyzyjnych operującego na już istniejących metodach. Przesadą byłoby oczywiście stwierdzenie, że wkraczamy w tej książce na zupełnie dziewicze tereny. Tym niemniej warto podkreślić, że zaproponowane w niej wykorzystanie algorytmu genetycznego do zastosowań prognostycznych stanowi oryginalny wkład Autora. Trzymana w ręku praca stanowi próbę przerzucenia pomostu pomiędzy prognozowaniem niestrukturalnym a wybranymi metodami optymalizacji heurystycznej. Postawiliśmy sobie za cel udowodnienie, że takie połączenie jest nie tylko możliwe, ale że niesie ze sobą również korzyści.

Nie mamy zamiaru deprecjonować stworzonych do tej pory istniejących metod i narzędzi. W istocie stanowią one cenną inspirację. Zdajemy sobie jednak sprawę z ich niedoskonałości w określonych okolicznościach. Z punktu widzenia prognozowania interesować nas będą szczególnie takie sytuacje, w których metody (umownie nazwane klasycznymi) mają kłopoty z trafnością predykcji i odwzorowaniem mechanizmów stojących za zachowaniem szeregu.

Podstawowym celem jaki sobie stawiamy jest opracowanie zmodyfikowanego algorytmu genetycznego, który miałby posłużyć do prognozowania szeregów czasowych charakteryzujących się skomplikowaną i/lub w jakiś sposób nietypową dekompozycją. Istnieje wiele metod przystosowanych do szeregów o konkretnym zachowaniu, ale istnieje też wiele zachowań nieprzystających do założeń tych metod. Uważamy, że elastyczność będąca

charakterystyczną cechą metod optymalizacji heurystycznej pozwala przezwyciężyć te ograniczenia.

Od strony algorytmu genetycznego lub ewolucyjnego chcieliśmy przedstawić, w jaki sposób można uwzględnić w nim wiedzę specyficzną dla zadania oraz jak poprawia to działanie systemu ewolucyjnego (nazwę cytujemy za Z. Michalewiczem (Michalewicz 2003)). W tym celu opracowaliśmy algorytm, który nazwaliśmy *Prognostycznym Algorytmem Genetycznym* (PAG) będący hybrydą algorytmów genetycznych i ewolucyjnych przystosowaną do tworzenia prognoz na podstawie szeregów czasowych.

Ponadto postanowiliśmy poszukać rozwiązania innej bolączki występującej w przypadku prognozowania niestrukturalnego. Metody zaliczane do tej grupy nie uwzględniają długookresowych procesów obecnych w szeregach czasowych, zwykle zadowalając się wykorzystywaniem obserwacji pochodzących z niedalekiej przeszłości. Problem w ich wypadku stanowią również prognozy *ex ante*, które sięgają bardzo blisko w przyszłość. Dlatego celem dodatkowym uczyniliśmy znalezienie rozwiązania tych wad właśnie przy pomocy PAG.

Pracę podzieliliśmy na cztery rozdziały. Pierwszy z nich poświęcony został zaprezentowaniu idei algorytmów genetycznych wraz ze stojącymi za nimi definicjami i twierdzeniami matematycznymi. Ponadto zawarliśmy tam opis wybranych modyfikacji, którym poddano te algorytmy w ciągu lat badań i analiz. Zmiany objęły takie elementy jak sposób kodowania, metody selekcji i inne. Konsekwencją wprowadzanych modyfikacji stało się uogólnienie algorytmów genetycznych do algorytmów ewolucyjnych. Na zakończenie tego rozdziału podjęliśmy próbę wskazania występujących między nimi różnic i podobieństw.

Rozdział drugi traktuje o prognozowaniu. Przybliżamy w nim założenia, funkcje i cechy charakterystyczne tej dziedziny nauki. W dalszej części skupiamy się na prezentacji metod prognostycznych przydatnych podczas predykcji opartej na szeregach czasowych. Zdecydowaliśmy się wyjść poza spotykane zwykle w literaturze samoistne ograniczenie do narzędzi o podręcznikowym charakterze. Oprócz nich zawarliśmy w tej części pracy przegląd wyników najnowszych badań dotyczących różnych punktów widzenia na metodologię prognoz oraz sposoby oceniania ich dokładności.

Trzeci rozdział poświęciliśmy połączeniu algorytmu genetycznego i prognozowania w jeden prognostyczny algorytm genetyczny. Omówiliśmy jego konstrukcję wraz z przykładami ilustrującymi zachowanie dla różnych typów szeregów czasowych. W polskiej literaturze brak prac o podobnej tematyce, a i w zagranicznej nie ma ich zbyt wiele. Zawarliśmy jednak omówienie prac anglojęzycznych, w których autorzy zmierzyli się ze wspomnianym zagadnieniem w większym lub mniejszym zakresie. Należy w tym miejscu podkreślić, że PAG stanowi oryginalny wkład badawczy Autora pracy, będący wynikiem

serii przemyśleń i badań. Pozostali autorzy skupiali się na jednym z elementów – algorytmie genetycznym lub ewolucyjnym, czy też jednym z aspektów prognozowania. W porównaniu z ich opracowaniami przedstawiamy metodę o bardziej kompleksowym charakterze.

W rozdziale czwartym zdecydowaliśmy się zamieścić wyniki obliczeń uzyskanych przy pomocy prognostycznego algorytmu genetycznego tworzącego tzw. prognozy kaskadowe dla szeregów czasowych pochodzących z różnych źródeł i prezentujących wybrane obszary działalności gospodarczej. Szeregi te ponadto charakteryzują się odmienną dekompozycją, utrudniającą lub wręcz uniemożliwiającą wykorzystanie klasycznych metod. PAG posłużył w ich wypadku do wyznaczenia prognoz *ex post* i *ex ante*. Ostatnie z wymienionych oceniliśmy również jako prognozy wygasłe, aby wykazać użyteczność przyjętego podejścia. Poza tym rozdział zawiera analizę efektywności funkcjonowania opisanego algorytmu pod kątem efektywności poszukiwania rozwiązania.

Różne sposoby uwzględniania specyfiki właściwej dla ewolucji występującej w przyrodzie upodobniły się do siebie w miarę upływu lat. Strategie ewolucyjne, algorytmy genetyczne czy programowanie ewolucyjne przejmowały nawzajem od siebie charakterystyczne dla nich rozwiązania. Czyniło je to coraz bardziej podobnymi do siebie zaś ich czyste, zbliżone do początkowych, formy są dziś rzadko spotykane. Obecnie każdą z wymienionych technik optymalizacji heurystycznej (wraz z kilkoma innymi) łączy się zazwyczaj w postaci jednego nurtu – algorytmów ewolucyjnych. Sam algorytm genetyczny uważany jest za część szerszego zbioru algorytmów ewolucyjnych (szerzej na ten temat piszemy w rozdziale pierwszym).

Z tego powodu proponowany algorytm należy rozpatrywać w szerszym kontekście. Jako wyróżnik klasycznego algorytmu genetycznego opracowanego przez J. Hollanda przyjmuje się m.in. binarne kodowanie i selekcję metodą koła ruletki. Problem polega na określeniu granicy, po przekroczeniu której algorytm genetyczny staje się algorytmem ewolucyjnym. W naszej procedurze, którą stworzyliśmy do celów prognostycznych można wskazać związki z propozycjami Hollanda. Na przykład użyty operator krzyżowania ma konstrukcję analogiczną do zastosowanej przez tego autora. Z drugiej strony chromosomy posiadają kodowanie rzeczywiste co łączy je z algorytmem ewolucyjnym. Trudno zatem jednoznacznie zaliczyć przyjętą przez nas drogę postępowania podczas predykcji szeregów czasowych do jednej, konkretnej rodziny metod. Zdecydowaliśmy więc, aby dla ujednolicenia terminologii określać algorytm właśnie mianem genetycznego.

Rozdział 1

Algorytmy genetyczne – charakterystyka metody

Algorytm opracowany przez nas na użytek prognozowania szeregów czasowych jest częściowo modyfikacją algorytmu genetycznego. Dlatego postanowiliśmy, aby na początek przybliżyć właśnie tę metodę postępowania poszukiwania rozwiązania. Zaczniemy od ujęcia zaproponowanego przez Hollanda po czym opiszemy przekształcenia jakim poddano ów algorytm obejmujące m.in. sposób kodowania, selekcję itp. Wiąże się to z późniejszym ich wykorzystaniem przedstawionym w książce. Z tego powodu rozdział ten, przynajmniej w początkowej części, będzie miał nieco dydaktyczny charakter. Dalsze podrozdziały i paragrafy poświęcone zostały przeglądowi wybranych modyfikacji metody zaproponowanej przez Hollanda. Po pierwsze zaprezentowane zostały transformacje różnych elementów algorytmu genetycznego, które zostały wykorzystane przy opracowaniu naszej własnej, prognostycznej wersji algorytmu. Poza tym krótko scharakteryzowaliśmy te modyfikacje, które mogłyby zaciekawić Czytelnika chcącego zainteresować się (oraz mamy nadzieję rozwijać) tematyką algorytmów genetycznych/ewolucyjnych.

1.1. Klasyczny algorytm genetyczny

W trakcie dalszej lektury Czytelnik często będzie natykał się na terminologię, która swoje początki bierze z nauk przyrodniczych, a konkretnie tej ich części, która zajmuje się badaniem mechanizmów ewolucji i genetyką. Tak więc problem, wobec którego stajemy konstruując algorytm wiąże się z pojęciem środowiska, w którym funkcjonuje pewna populacja osobników. Każdy z nich z punktu widzenia algorytmu jest jednym z możliwych rozwiązań. Zwykle jednostce takiej przypisuje się pewną wartość liczbową wyrażającą jakość reprezentowanego rezultatu zwaną przystosowaniem.

Zestaw cech osobnika określa genotyp zaś w efekcie oddziaływania między genotypem a środowiskiem otrzymujemy fenotyp czyli konkretne rozwiązanie na przykład punkt w przestrzeni rozwiązań (Arabas 2004, s. 15). Genotyp osobnika zawiera co najmniej jeden chromosom, który w naszych rozważaniach zwykle utożsamiany będzie z łańcuchem kodowym zawierającym takie informacje jak wspomniany już fenotyp lub inne ważne z punktu

widzenia ewolucji osobnika informacje. Pojedynczy element owego łańcucha nosi nazwę genu. Geny decydują o dziedziczeniu cech (np kolor oczu) i mogą występować w pewnej liczbie stanów zwanych allelami. Położenie genu w chromosomie określa się jako locus.

Algorytmy genetyczne zaliczają się do metod poszukiwania optymalnych wartości funkcji na zadanym zbiorze rozwiązań podobnie jak popularny algorytm simpleks. Niektóre terminy używane w klasycznej optymalizacji mają swoje swoje odpowiedniki w języku algorytmów genetycznych. Funkcja celu zadania optymalizacyjnego odpowiada funkcji przystosowania algorytmu genetycznego przy czym kierunek optymalizacji pozostaje bez zmian. W algorytmie genetycznym funkcja przystosowania jest zawsze maksymalizowana, ale da się to założenie obejść. Aby osiągnąć efekt w postaci minimalizacji funkcji kryterium można ją odjąć od pewnej stałej i poszukiwać maksimum tak przekształconej zależności. Jeżeli w zadaniu nie występują ograniczenia wówczas obie funkcje są tożsame. W wypadku pojawienia się dodatkowych warunków funkcję przystosowania modyfikuje się na przykład wprowadzając karę proporcjonalną do stopnia naruszenia ograniczeń (szerzej na ten temat piszemy w jednym z podrozdziałów). Same ograniczenia nie występują jawnie w dalszym toku algorytmu genetycznego.

Zmienne decyzyjne stają się zmiennymi algorytmu. Różnice pojawiają się w momencie przyjęcia określonego sposobu kodowania chromosomu. Oczywiście najbardziej intuicyjne staje się sięgnięcie po rzeczywiste wartości zmiennych decyzyjnych, ale na przykład klasyczny algorytm genetyczny operuje na ich reprezentacji w postaci kodu dwójkowego. Niesie to ze sobą konsekwencje, o których mowa będzie w dalszej części tego rozdziału. Istnieją też algorytmy, które służą tworzeniu zbiorów reguł decyzyjnych. W ich przypadku stosuje się kodowanie adekwatne do problemu. Może ono przykładowo obejmować sygnały typowe dla rynku kapitałowego czyli „kup”, „trzymaj”, „sprzedaj”.

Za twórcę nowoczesnej teorii dotyczącej algorytmów genetycznych uważa się Johna H. Hollanda (Goldberg 2003, s. 18). W latach 70 postanowił on stworzyć ścisły model, który opisywałby ewolucyjne i adaptacyjne zjawiska zachodzące w przyrodzie. Z czasem okazało się, że zasugerowane przez niego rozwiązania mogą znaleźć szereg zastosowań w najróżniejszych dziedzinach, a zaproponowany przez Hollanda model nosi dziś nazwę klasycznego algorytmu genetycznego (KAG). Pomimo biologicznego rodowodu obszar zastosowań dawno już opuścił „przyrodnicze” pielesze.

W metodzie tej przetwarzamy parametry zadania w postaci zakodowanej. Klasyczny algorytm genetyczny operuje na zbiorach łańcuchów (chromosomów), które mają stałą długość a składają się z ciągów zer i jedynek (genów). Dużym uproszczeniem jest przyjęcie w KAG iż jeden chromosom równa się jednemu osobnikowi. W środowisku naturalnym taką budowę ma-

ją jedynie najprostsze organizmy. Mimo wszystko tego typu ujęcie sprawdza się w większości zastosowań (jak się wkrótce okaże również podczas prognozowania) a ponadto ułatwia analizę sposobu przeszukiwania zbioru potencjalnych rozwiązań.

Zbiór wszystkich chromosomów utworzonych w celu znalezienia rozwiązania problemu nazywamy populacją (Goldberg 2003, s. 23). W rzeczywistości algorytm wykorzystuje kilka rodzajów populacji: bazową, potomną oraz tymczasową. Pierwsza populacja bazowa składa się losowo wygenerowanych osobników. W tymczasowej przechowuje się kopie osobników z populacji bazowej (Arabas 2004, s. 65) w związku z wykonywanymi operacjami. Populacja potomna powstaje poprzez zastąpienie populacji bazowej z poprzedniej po zakończeniu wszystkich przewidzianych procedurą operacji. Wszystkie wymienione populacje łączy jednakowa liczebność.

W KAG nie występują ograniczenia ani warunki brzegowe. W obrębie rozpatrywanej populacji wszystkie chromosomy są tej samej długości czyli mają jednakową liczbę genów określaną (podobnie jak liczbę chromosomów w populacji) *a priori* przez badacza w zależności od charakteru rozwiązywanego problemu.

Oparcie się na systemie zer i jedynek oznacza, że przed rozpoczęciem poszukiwania optymalnej wartości należy dokonać zakodowania zmiennych zadania do postaci binarnej. Następnie zaś rozkodować je, aby sprawdzić wynik. W czasie obliczeń algorytmu mamy do czynienia z operacjami wykonywanymi na ciągach i podciągach tego zbioru. Osobniki poddaje się działaniu pewnych procedur tzw. operatorów, które mają za zadanie naśladować zachowania spotykane w środowisku naturalnym. Sprowadzają się one do trzech podstawowych, a mianowicie (Goldberg 2003, s. 27):

— reprodukcji;
— krzyżowania;
— mutacji.

Jak łatwo zauważyć, nazwy powyższych procesów przypominają o ich biologicznym rodowodzie. Tak więc w czasie reprodukcji dochodzi do powielenia chromosomów. O szansach na to, aby dany łańcuch kodowy został powielony decyduje przede wszystkim wartość funkcji przystosowania zaś sam wybór dokonuje się poprzez losowanie (zazwyczaj jest to losowanie ze zwracaniem). Prawdopodobieństwa wylosowania chromosomów do reprodukcji różnią się między sobą. Polega to na tym, że osobniki o lepszym przystosowaniu mają większą szansę wprowadzenia potomka do następnego pokolenia. Stanowi to substytut zasady, że w naturalnym środowisku przeżywają osobniki najlepiej przystosowane (na przykład takie, które najlepiej unikają drapieżników).

Po wybraniu chromosomów do reprodukcji wykonujemy operację krzyżowania, które dla klasycznej wersji algorytmu nosiS nazwę krzyżowania

prostego (Goldberg 2003, s. 28). Wykonujemy je następująco: losujemy parę chromosomów-rodziców, podejmujemy decyzję o krzyżowaniu (prawdopodobieństwo krzyżowania jest jednym z parametrów algorytmu) po czym wybieramy tzw. punkt krzyżowania, a następnie dokonujemy zamiany miejscami podciągów znaków od tegoż punktu do końca łańcucha. Zamiana obejmuje oba chromosomy, a w jej efekcie powstają dwa nowe ciągi zer oraz jedynek. Oznacza to także, że dwoje rodziców pozwala na uzyskanie dwojga potomków. Istnieją również modyfikacje operatora krzyżowania, w wyniku których para rodziców może doczekać się innej liczby potomków, ale zostaną one omówione później.

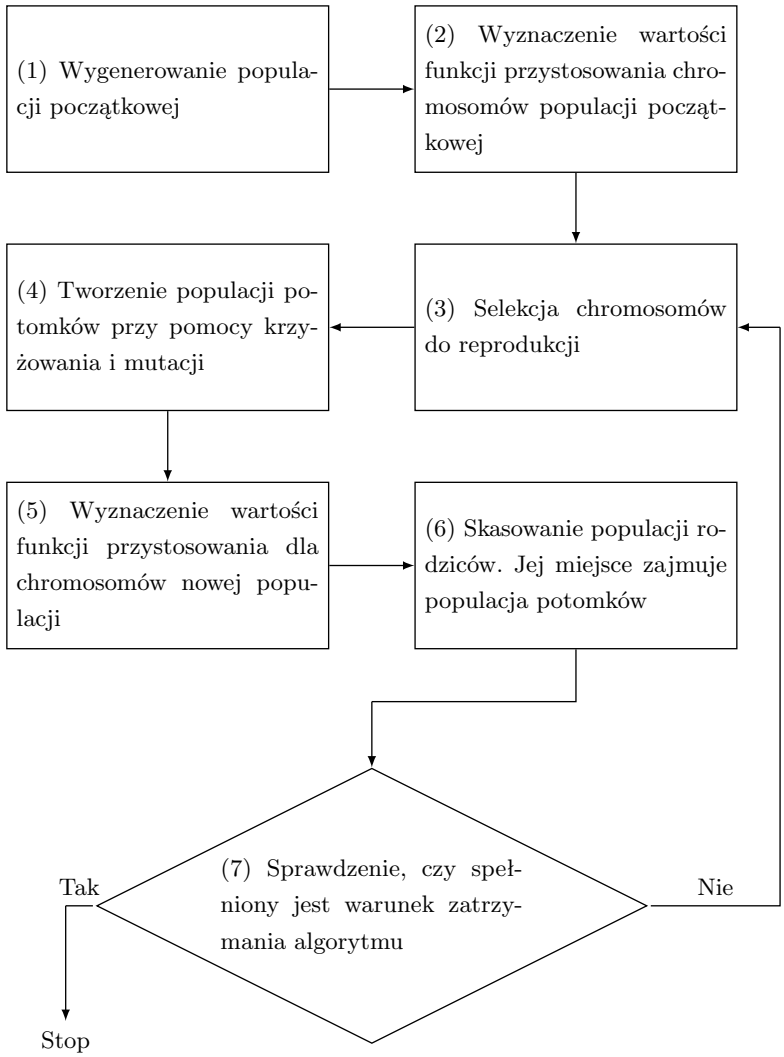
Trzecim ważnym procesem jakiemu poddaje się chromosomy jest mutacja. Wprowadzenie jej zostało podyktowane z jednej strony chęcią jak najbardziej wiernego oddania zjawisk występujących w przyrodzie. Z drugiej zaś, gdyby ograniczyć się tylko do reprodukcji i krzyżowania, zachodziłoby niebezpieczeństwo, że wyeliminowany zostanie potencjalnie wartościowy materiał genetyczny. Wprowadzenie mutacji ma zapobiec takim niekorzystnym sytuacjom. W klasycznym ujęciu mutacja w algorytmach genetycznych polega na zachodzącej z niewielkim prawdopodobieństwem przypadkowej zmianie niektórych genów. Mówiąc prościej – w losowy sposób dochodzi do zamiany zera na jedynkę bądź odwrotnie.

Populacja zostaje poddana działaniu trzech powyższych procedur oraz mechanizmu oceniającego jakość chromosomów. Sam proces poszukiwania rozwiązania rozpoczynamy od ustalenia punktu startowego. W przyrodzie problem taki nie występuje ponieważ ewolucja zawsze dokonuje się na już istniejącym zbiorze osobników. My takim zbiorem nie dysponujemy dlatego zmuszeni jesteśmy wygenerować losową populację początkową, od której rozpoczyna się proces poszukiwania optimum.

Poszczególne etapy, przez które przechodzi klasyczny algorytm genetyczny dadzą się przedstawić w postaci schematu blokowego, który znalazł się na rysunku 1.1.

Oddzielnym zagadnieniem jest kontrola parametrów sterujących przebiegiem procesu poszukiwania rozwiązania optymalnego. Zwykle z góry przyjmuje się takie wielkości jak: prawdopodobieństwa krzyżowania i mutacji, wielkość populacji czy liczbę pokoleń. Istnieją stosowne modyfikacje wpływające na zachowanie wymienionych parametrów. Część spośród nich omówimy w dalszej części tego rozdziału. Sposoby (właściwie strategie) ustalania parametrów sterujących obliczeniami da się podzielić na dwie podstawowe grupy (Montero i Riff 2011):

1. Strategie off-line – poszukujemy najlepszych wartości parametrów sterujących w kolejnych, próbnych uruchomieniach algorytmu potem pozostają one już stałe podczas zasadniczego przebiegu. Takie postępowanie ma



Rysunek 1.1: Schemat blokowy działania algorytmu genetycznego

Źródło: opracowanie własne

jednak tę wadę, że uzyskane parametry są optymalne jedynie z punktu widzenia prób wykonanych przez badacza.

2. Strategie on-line – parametry podlegają zmianie w trakcie obliczeń algorytmu zatem to od użytkownika zależy kiedy i o ile zmienić ich wartości. Należy liczyć się w tym przypadku ze wzrostem czasu obliczeń gdyż algorytm musi wykonywać dodatkowe operacje.

Montero i Riff zaproponowali, aby każdy chromosom dysponował własnym zestawem operatorów a zatem własnymi prawdopodobieństwami krzyżowania oraz mutacji. Operatory genetyczne ingerują nie tylko w strukturę

chromosomu tworząc potomków ponieważ zestaw parametrów potomka koryguje się o losową wartość pochodzącą z przedziału $\langle 0, 1\rangle$. Autorzy artykułu zastosowali więc strategię on-line podczas gdy w rozdziale trzecim skorzystamy ze strategii off-line.

1.1.1. Matematyczne podstawy algorytmów genetycznych

Klasyczny algorytm genetyczny da się opisać w sformalizowany, matematyczny sposób. Zaprezentowane pojęcia (Michalewicz 2003), po pewnych modyfikacjach, nadają się także do opisywania algorytmów w ich mniej klasycznej postaci w tym ewolucyjnych.

Oznaczmy przez $\mathbf{X}$ zbiór rozwiązań dopuszczalnych danego zagadnienia optymalizacyjnego.

Definicja 1.1. *Chromosomem nazywamy wektor postaci:*

$$\mathbf{x} = (x_1, \ldots, x_n) \in \mathbf{X} \tag{1.1}$$

stanowiący rozwiązanie dopuszczalne zadania optymalizacyjnego.

Definicja 1.2. *Populacją rozwiązań dopuszczalnych zadania optymalizacyjnego w* k*-tej iteracji nazwiemy podzbiór:*

$$P(k) = \left\{\mathbf{x}_1^k, \ldots, \mathbf{x}_m^k\right\} \quad P(k) \subset \mathbf{X} \tag{1.2}$$

Definicja 1.3. *Funkcją przystosowania F dla rozwiązania $\mathbf{x}_i^k \in P(k)$ nazywamy wartość normy z funkcji celu C dla tego rozwiązania:*

$$F(\mathbf{x}_i^k) = \left|\left|C(\mathbf{x}_i^k)\right|\right| \tag{1.3}$$

gdzie: $\underset{\mathbf{x}_i^k \in P(k)}{\forall} C(\mathbf{x}_i^k)$

Definicja 1.4. *Operatorem mutacji M_K nazywamy przekształcenie:*

$$M_K : \mathbf{X} \to \mathbf{X} \tag{1.4}$$

dokonujące losowej, z zadanym prawdopodobieństwem, zmiany k-tej składowej rozwiązania $\mathbf{x}_i^k$:

$$M_K\left(\mathbf{x}_i^k\right) = \mathbf{x}_i^{k+1} \tag{1.5}$$

gdzie:
$\mathbf{x}_i^k = (x_1, \ldots, x_k, \ldots, x_n)$ oraz $\mathbf{x}_i^{k+1} = (x_1, \ldots, x_k^*, \ldots, x_n)$

Definicja 1.5. *Operatorem krzyżowania K_L nazywamy przekształcenie:*

$$K_K : \mathbf{X} \times \mathbf{X} \to \mathbf{X} \times \mathbf{X} \tag{1.6}$$

dokonujące, z zadanym z góry prawdopodobieństwem, losowej wymiany składowych rozwiązań $\mathbf{x}_i^k$ i $\mathbf{x}_j^k$ względem składowej l:

$$K_K\left(\mathbf{x}_i^k,\mathbf{x}_j^k\right) = \left(\mathbf{x}_i^{k+l},\mathbf{x}_j^{k+l}\right) \tag{1.7}$$

gdzie:
$\mathbf{x}_i^k = (x_1, \ldots ,x_n)$ oraz $\mathbf{x}_j^k = (v_1, \ldots ,v_n)$
$\mathbf{x}_i^{k+l} = (x_1, \ldots ,x_l,v_{l+1}, \ldots ,v_n)$ oraz $\mathbf{x}_j^{k+l} = (v_1, \ldots ,v_l,x_{l+1}, \ldots ,x_n)$

W matematycznej teorii algorytmów genetycznych kluczowym zagadnieniem jest pojęcie schematu czyli wzorca opisującego podzbiór ciągów podobnych ze względu na ustalone pozycje (Goldberg 2003, s. 36). Schemat obejmuje takie układy genów, które mają największe szanse przetrwać kolejne mutacje i operacje krzyżowania. Założenie to implikuje wniosek, że algorytm preferuje schematy o małej długości, które są często wybierane oraz powielane w kolejnych iteracjach. Do rozwiązania optymalnego zbliżamy się zestawiając schematy niskiego rzędu i o małej rozpiętości.

Dla dalszych rozważań wprowadźmy symbol „$*$” na oznaczenie genu, który nie należy do schematu. Innymi słowy, symbol ten oznacza nieistotne z punktu widzenia schematu elementy łańcucha genów. Rozważmy chromosom o długości 6. Do schematu (1 1 0 0 1 $*$) pasują dwa następujące łańcuchy: (1 1 0 0 1 1) i (1 1 0 0 1 0). Można wykazać, że do każdego schematu pasuje 2^r łańcuchów gdzie r to liczba „gwiazdek”. Do każdego chromosomu o długości m pasuje 2^m schematów.

Definicja 1.6. *Chromosom jest dopasowany do schematu* S *kiedy:*
1. *Liczba elementów schematu równa się liczbie genów chromosomu.*
2. *Chromosom i schemat są identyczne na wszystkich pozycjach określonych czyli takich, którym w schemacie odpowiada „0” lub „1”.*

Definicja 1.7. *Rzędem (licznością) schematu* S *nazywamy liczbę pozycji określonych w tym schemacie.*

Definicja 1.8. *Długością schematu nazywamy liczbę określającą odległość między pierwszą a ostatnią pozycją w danym schemacie.*

Tabela 1.1 prezentuje kilka przykładów związku między rzędem a długością w schemacie.

Definicja 1.9. *Przystosowaniem schematu w iteracji k nazywamy liczbę:*

$$F(S,k) = \frac{\sum_{i=1}^{s} F(ch_i)}{s} \tag{1.8}$$

gdzie:
s – liczba wystąpień schematu w populacji
$F(ch_i)$ – wartość funkcji przystosowania dla i-tego chromosomu

Okazuje się (uwzględniając na razie tylko mechanizm selekcji), że liczba chromosomów pasujących do danego schematu w kolejnych iteracjach uzależniona jest od stosunku przystosowania danego schematu do średniej wartości funkcji przystosowania. Jeżeli przystosowanie schematu zostanie ocenione powyżej średniej wtedy uzyska on w kolejnym pokoleniu większą liczbę chromosomów do niego dopasowanych. Można także wykazać, że liczba wystąpień ciągu genów tworzących schemat o przystosowaniu powyżej średniej będzie w kolejnych iteracjach rosła wykładniczo.

Schematy, tak jak całe chromosomy, również podlegają selekcji, krzyżowaniu i mutacji. Operacje te mogą niekiedy dany schemat zniszczyć w związku z czym liczba schematów ulega zmianie. Prawdopodobieństwo zniszczenia schematu w wyniku krzyżowania wyraża wzór (1.9).

$$P_{zk} = \frac{D_s}{d_{ch} - 1} \tag{1.9}$$

gdzie:
D_S – długość schematu d_{ch} – długość chromosomu

Tabela 1.1: Rząd a długość schematu

Schemat	Rząd	Długość
0 1 1 0 1 0	6	6
* 1 * * * *	1	1
* 1 * * 0 *	2	4
* * 1 0 * *	2	2
1 * 0 0 * *	3	4

Źródło: opracowanie własne

Krzyżowanie zachodzi z pewnym prawdopodobieństwem p_k więc prawdopodobieństwo przetrwania schematu w procesie krzyżowania otrzymamy z zależności (1.10).

$$P_{pk} \geqslant 1 - p_k \frac{D_s}{d_{ch} - 1} \tag{1.10}$$

Wpływ krzyżowania na liczbę chromosomów dopasowanych do danego schematu w kolejnych iteracjach wyraża wzór (1.11).

$$s_{t+1} \geqslant s_t \frac{F(S,k)}{\overline{F}(S,k)} \left[1 - p_k \frac{D_s}{d_{ch} - 1}\right] \tag{1.11}$$

gdzie:
s_t,s_{t+1} – liczba wystąpień schematu S w iteracjach t i t+1
$\overline{F}(S,k)$ – średnia wartość funkcji przystosowania w k-tej iteracji

Na liczbę chromosomów, które będą pasować do danego schematu wpływa stosunek przystosowania schematu do średniej wartości funkcji przystosowania oraz sama długość schematu.

Na ewentualne przetrwanie schematów ma także wpływ wystąpienie operatora mutacji. Prawdopodobieństwo, że schemat nie zostanie zniszczony wskutek mutacji da się aproksymować wzorem (1.12).

$$P_{pm} \approx L_S p_m \tag{1.12}$$

gdzie:
L_S – rząd schematu S
p_m – prawdopodobieństwo mutacji

Zestawiając łącznie wpływ selekcji, krzyżowania i mutacji na przetrwanie schematów w ramach chromosomu otrzymujemy zależność (1.13).

$$s_{t+1} \geqslant s_t \frac{F(S,k)}{\overline{F}(S,k)} \left[1 - p_k \frac{D_s}{d_{ch} - 1} - L_S p_m\right] \tag{1.13}$$

Analizując występujące w algorytmie genetycznym schematy, J. Holland sformułował kilka założeń (Arabas 2004, s. 74) dotyczących KAG:

1. Populacja bazowa zawiera nieskończenie wiele osobników.
2. W populacji bazowej znajdują się chromosomy należące do wszystkich możliwych schematów.
3. Selekcja odbywa się na podstawie tzw. koła ruletki.
4. Stosowana jest reprezentacja binarna i krzyżowanie jednopunktowe.

Na podstawie powyższych założeń powstało tzw. twierdzenie o schematach (Michalewicz 2003, s. 80) obrazujące zależność dotyczącą liczby chromosomów dopasowanych do schematu S:

Twierdzenie 1.1. *Krótkie, niskiego rzędu oraz oceniane powyżej średniej schematy uzyskują wykładniczo rosnącą liczbę swoich reprezentantów w kolejnych pokoleniach algorytmu genetycznego.*

Z tego twierdzenia wynika następne, o równie fundamentalnym znaczeniu dla teorii algorytmów genetycznych. Jest to twierdzenie o tzw. hipotezie bloków budujących:

Twierdzenie 1.2. *Algorytm genetyczny poszukuje działania zbliżonego do optymalnego przez zestawianie krótkich, niskiego rzędu schematów o dużej wydajności działania zwanych blokami budującymi.*

Hipoteza bloków budujących wskazuje, że podczas działania algorytmu genetycznego w kolejnych pokoleniach chromosomy tworzone są z coraz lepszych, krótkich schematów. Choć mówimy tu o binarnej reprezentacji genotypu to wyjaśnia ona również funkcjonowanie chromosomów, których geny tworzą liczby rzeczywiste. Okazała się na przykład bardzo przydatna podczas konstrukcji operatorów przystosowanych do rozwiązywania problemu komiwojażera.

Sposób kodowania i operatory należy dobierać tak, aby bloki tworzące nie były tracone w toku przetwarzania kolejnych populacji. To właśnie ten wniosek wykorzystamy w rozdziale trzecim projektując kodowanie chromosomów.

Pojawienie się w toku przetwarzania kolejnych iteracji dobrze przystosowanego schematu może spowodować, że kolejne populacje zostaną przez niego zdominowane. Doprowadzi to do zjawiska przedwczesnej zbieżności. Ryzykujemy wtedy osiągnięcie lokalnego nie zaś globalnego ekstremum. Próbą poradzenia sobie z tym problemem są modyfikacje klasycznego algorytmu genetycznego.

Za twierdzeniem o schematach stoją bardzo silne i w praktyce nierealne założenia: nieskończona populacja, kodowanie binarne (kłopotliwe kiedy chcemy zwiększyć liczbę miejsc po przecinku dla zmiennych), bardzo dobre zróżnicowanie. Problem stanowi w tej sytuacji choćby określenie kryterium zatrzymania algorytmu. Dlatego użyteczność schematów jako narzędzia analizy algorytmów genetycznych jest niewielka choć stanowią ciekawy punkt wyjścia rozważań teoretycznych.

Relacjom matematycznym podlega także selekcja metodą koła ruletki (Lipowski i Lipowska 2012). Niech $w_{\max}$ oznacza maksymalną wartość funkcji przystosowania osiągniętą w populacji zaś w_i oznacza przystosowanie osobnika o numerze i. Załóżmy ponadto, że chromosom do reprodukcji wybierany jest w sposób losowy przy czym prawdopodobieństwo wylosowania jest jednakowe dla każdego z osobników i niezależne od wartości funkcji przystosowania.

Wylosowany chromosom trafia do dalszego postępowania z tzw. prawdopodobieństwem akceptacji równym $w_i/w_{\max}$. Prawdopodobieństwo wylosowania konkretnego osobnika wynosi $w_i/Nw_{\max}$ gdzie N oznacza całkowitą

wielkość populacji. Prawdopodobieństwo wylosowania tego osobnika w ciągu pewnej liczby prób równa się:

$$p'_i = \frac{w_i}{N w_{\max}}(1 + q + q^2 + \ldots) \tag{1.14}$$

gdzie:
$q = 1 - \dfrac{\bar{w}}{w_{\max}}$

Wartość q wyraża prawdopodobieństwo odrzucenia chromosomu natomiast $\bar{w}$ średnie przystosowanie populacji.

Szereg geometryczny dany wzorem (1.14) jest zbieżny i da się dowieść, że po zsumowaniu prawdopodobieństwo p'_i jest równe prawdopodobieństwu z klasycznego koła ruletki. Dzięki wykonanym przekształceniom jesteśmy jednak w stanie obliczyć liczbę prób τ potrzebną, aby dany osobnik został wybrany choć raz:

$$\tau = \frac{1}{N}\sum_{i=1}^{N}\frac{w_i}{w_{\max}}(1 + 2q + 3q^2 + \ldots) = \frac{w_{\max}}{\bar{w}} \tag{1.15}$$

Powyższe rozważania można rozszerzyć na przypadek kiedy prawdopodobieństwo akceptacji wynosi w_i/A gdzie A jest pewną stałą wyższą od $w_{\max}$. Rośnie wówczas prawdopodobieństwo odrzucenia oraz liczba prób potrzebna by wylosować danego osobnika. Obniża to efektywność poszukiwania rozwiązania przez algorytm.

1.1.2. Przykład działania klasycznego algorytmu genetycznego

Aby lepiej zrozumieć działanie algorytmu genetycznego prześledzimy jego poszczególne etapy (zilustrowane rysunkiem 1.1) na prostym przykładzie liczbowym stawiającym sobie za cel optymalizację prostej funkcji. Zainteresowanych Czytelników odsyłamy do książki T. Trzaskalika (Trzaskalik 2008), w której w podobny „ręczny” sposób zaprezentowano poszukiwanie rozwiązania problemu jednego komiwojażera. Należy zwrócić uwagę, że w praktyce dla zagadnienia tego istnieją dedykowane reprezentacje chromosomów oraz operatory. Wrócimy jeszcze do tego zagadnienia w dalszej części rozdziału.

Załóżmy, że poszukujemy maksimum następującej funkcji:

$$F(\mathbf{x}) = 5 + x_1^2 + 2x_2 \tag{1.16}$$

gdzie:
x_1,$x_2 \in \langle 0, 63\rangle$

$F(\mathbf{x})$ to w naszym przypadku funkcja przystosowania. Obie zmienne będą reprezentowane przez łańcuchy binarne. Dla uproszczenia rozważań ograniczymy się do poszukiwania rozwiązania wśród liczb całkowitych. Zawierać one będą po sześć bitów, wówczas ciąg (0 0 0 0 0 0) w kodzie dwójkowym równa się 0 zaś ciąg (1 1 1 1 1 1) jest równy 63. Cały chromosom składa się z dwunastu bitów, przy czym pierwsze sześć reprezentuje zmienną x_1 zaś następne – zmienną x_2. Poszczególne kroki algorytmu będą miały numerację zgodną ze schematem z rysunku 1.1.

(1) Wygenerowanie populacji początkowej

W pierwszym kroku tworzymy startową pulę chromosomów, od której rozpoczniemy poszukiwanie rozwiązania. Geny przybierają w losowy sposób wartość 0 lub 1. Do tego celu nadaje się na przykład rzut monetą, przyporządkowujący wyrzuceniu reszki zero, a orła jedynkę. Załóżmy, że tworzymy populację składającą się z dziesięciu osobników. Ponieważ każdy chromosom składa się z dwunastu genów, należy wykonać sto dwadzieścia rzutów. Przyjmijmy, że otrzymaliśmy w ten sposób następującą populację początkową:

$$
\begin{aligned}
Ch_1 &= (\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 1\ 0\ 0\ 0\ 0\ ) \\
Ch_2 &= (\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ ) \\
Ch_3 &= (\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ ) \\
Ch_4 &= (\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ ) \\
Ch_5 &= (\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ ) \\
Ch_6 &= (\ 0\ 0\ 0\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ ) \\
Ch_7 &= (\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 0\ 0\ 1\ 1\ ) \\
Ch_8 &= (\ 0\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 1\ 1\ 1\ ) \\
Ch_9 &= (\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ ) \\
Ch_{10} &= (\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ )
\end{aligned}
$$

(2) Wyznaczenie wartości funkcji przystosowania dla poszczególnych chromosomów populacji początkowej

W drugim kroku obliczmy wartości funkcji przystosowania (1.16) dla każdego chromosomu. Zmienne zostały zdekodowane z postaci dwójkowej do dziesiątkowej a następnie podstawione do funkcji przystosowania. Uzyskaliśmy dla niej 10 następujących wartości:

$$F(Ch_1) = F(26{,}48) = 777$$
$$F(Ch_2) = F(4{,}37) = 95$$

$$F(Ch_3) = F(37{,}42) = 1458$$
$$F(Ch_4) = F(19{,}12) = 390$$
$$F(Ch_5) = F(33{,}10) = 1114$$
$$F(Ch_6) = F(3{,}19) = 52$$
$$F(Ch_7) = F(17{,}19) = 332$$
$$F(Ch_8) = F(20{,}23) = 451$$
$$F(Ch_9) = F(26{,}15) = 711$$
$$F(Ch_{10}) = F(49{,}4) = 2414$$

Funkcja przystosowania jest najwyższa dla chromosomu Ch_{10} zaś najniższa dla Ch_6. Łączna suma wartości funkcji przystosowania wszystkich chromosomów wynosi 7108, średnia równa się 779,4 z odchyleniem 686,02. Dysponując punktem wyjścia w postaci populacji początkowej przechodzimy do następnego kroku postępowania.

(3) Selekcja chromosomów do reprodukcji

Na tym etapie dokonamy wyboru, które z chromosomów posłużą do wygenerowania nowej populacji potomków. Innymi słowy utworzymy zbiór osobników przeznaczonych do reprodukcji. Liczba wybranych chromosomów odpowiada liczebności populacji początkowej (w naszym przykładzie będzie ich więc sześć), a samego wyboru dokonuje się losowo choć prawdopodobieństwa tu występujące nie są dla poszczególnych osobników jednakowe. Należy bowiem w jakiś sposób zapewnić, aby najlepsze (z punktu widzenia funkcji przystosowania) chromosomy miały większe szanse na wylosowanie do reprodukcji. Stanowi to odzwierciedlenie zasady występującej w przyrodzie, w myśl której osobniki przykładowo najlepiej unikające drapieżników mają lepsze perspektywy na wydanie potomstwa.

Podobnie jak uczynił to Holland, użyjemy najprostszej z metod selekcji, mianowicie metody koła ruletki. Od klasycznej gry znanej z kasyn różni się ona jednak tym, że w naszym przypadku wypadnięcie niektórych wartości liczbowych jest bardziej prawdopodobne niż innych. W algorytmie genetycznym fragment koła odpowiadający danemu chromosomowi ma długość wprost proporcjonalną do jego przystosowania.

Najpierw sumujemy wartości funkcji przystosowania dla wszystkich osobników populacji. Sumę tę będziemy traktować jako całe koło ruletki. Każdemu z chromosomów przyporządkowujemy wycinek koła obliczany według wzoru (1.17).

$$\frac{F(Ch_i)}{\sum_{i=1}^{n} F(Ch_i)} \cdot 100\% \tag{1.17}$$

Występujące w powyższym wzorze $F(Ch_i)$ wyraża wartość funkcji przystosowania dla chromosomu o numerze i.

Otrzymane wycinki koła zawierają się w przedziale $\langle 0, 100\rangle$ i stają się prawdopodobieństwami wylosowania poszczególnych chromosomów. Na podstawie wartości funkcji przystosowania określiliśmy rozkład prawdopodobieństwa skokowej zmiennej losowej opisującej szansę znalezienia się chromosomu w następnym etapie postępowania. Dla wygody dalszych obliczeń skumulujemy otrzymane wycinki tworząc dystrybuantę wspomnianej zmiennej. W tabeli 1.2 znalazły się wycinki koła ruletki wraz z dystrybuantą dla naszego przykładu. Aby wyznaczyć osobnika do reprodukcji wykorzystamy

Tabela 1.2: Metoda koła ruletki dla chromosomów z przykładu

Chromosom	Wycinek koła ruletki [%]	Wycinki skumulowane [%]
Ch_1	9,97	9,97
Ch_2	1,22	11,19
Ch_3	18,71	29,89
Ch_4	5,00	34,90
Ch_5	14,29	49,19
Ch_6	0,67	49,86
Ch_7	4,26	54,12
Ch_8	5,79	59,91
Ch_9	9,12	69,03
Ch_{10}	30,97	100,0

Źródło: opracowanie własne

losowanie metodą odwracania dystrybuanty. Losujemy liczby z przedziału $\langle 0{,}100\rangle$ i porównujemy je z wartościami dystrybuanty. Wylosowaliśmy sześć następujących liczb z przedziału $\langle 0{,}100\rangle$ wskazujące, które chromosomy zostaną rodzicami:

17,12; 42,05; 35,98; 76,30; 90,12; 61,44; 33,81; 61,77; 74,17; 44,42

Porównujemy je ze skumulowanymi wycinkami z tabeli 1.2. Symbolem a oznaczymy wylosowaną liczbę zaś $P(Ch_i)$ wartość dystrybuanty dla chromosomu Ch_i. Jeżeli zachodzi: $a \leqslant P(Ch_1)$ wtedy do reprodukcji trafi chromosom Ch_1. W pozostałych przypadkach spełniony zostaje warunek:

$$P(Ch_{i-1}) < a \leqslant P(Ch_i)$$

Na przykład $42{,}05 < 49{,}19$ czyli do reprodukcji trafi chromosom Ch_5. Postępując w opisany sposób uzyskamy nową populację, którą podajemy poniżej (dodatkowo podajemy również oznaczenie chromosomu w populacji

poprzedniej). Zwróćmy uwagę, że do reprodukcji trafia tyle samo osobników ile liczyła sobie wyjściowa populacja.

$$
\begin{aligned}
Ch_1 &= (\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ )(Ch_3)\\
Ch_2 &= (\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ )(Ch_5)\\
Ch_3 &= (\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ )(Ch_5)\\
Ch_4 &= (\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ )(Ch_{10})\\
Ch_5 &= (\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ )(Ch_{10})\\
Ch_6 &= (\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ )(Ch_9)\\
Ch_7 &= (\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ )(Ch_4)\\
Ch_8 &= (\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ )(Ch_9)\\
Ch_9 &= (\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ )(Ch_{10})\\
Ch_{10} &= (\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ )(Ch_5)
\end{aligned}
$$

Selekcja metodą koła ruletki doprowadziła do tego, że chromosomy Ch_5 oraz Ch_{10} o wysokich wartościach funkcji przystosowania w nowej populacji pojawiły się po trzy razy. Powtarzał się też chromosom Ch_9 o przystosowaniu zbliżonym do średniej dla całej populacji.

Wśród wylosowanych znalazł się osobnik Ch_4, który nie miał zbyt wysokiego przystosowania. Wiąże się to z elementem losowości wprowadzonym przez ruletkę. Osobniki najlepiej przystosowane mają większą szansę znalezienia się w kolejnym etapie, ale nie przekreśla to ewentualnego wyboru pozostałych chromosomów. Z tego samego powodu chromosom Ch_3 o niezłym przystosowaniu wybrany został tylko raz. Uzyskujemy w ten sposób większe zróżnicowanie populacji co przydaje się podczas przeszukiwania zbioru rozwiązań.

(4) Tworzenie populacji potomków przy pomocy krzyżowania oraz mutacji

Chromosomy przeznaczone do reprodukcji losowo łączymy w pary. Niech w naszym przypadku prawdopodobieństwo krzyżowania $p_k = 0{,}25$ zaś mutacji $p_m = 0{,}01$. Wielkości te przyjmuje się zwykle *a priori* projektując algorytm i standardowo pozostają one stałe przez cały czas obliczeń choć, jak wspomnieliśmy wcześniej, można je modyfikować aż do momentu znalezienia wartości zapewniających najlepsze wartości funkcji przystosowania.

Połączenie chromosomów w pary osiągamy losując sześć liczb z przedziału $\langle 0, 1\rangle$. Jeżeli wylosowana liczba okaże się mniejsza lub równa p_k wtedy dany chromosom jest przeznaczany do krzyżowania. W przypadku, kiedy liczba wybranych chromosomów jest nieparzysta losujemy z wyjściowej

populacji jeszcze jednego osobnika. Załóżmy, że dla każdego chromosomu wylosowano następujące liczby:

$$0{,}3331;\ 0{,}2024;\ 0{,}4178;\ 0{,}0135;\ 0{,}7899;$$
$$0{,}6428;\ 0{,}9411;\ 0{,}2074;\ 0{,}7247;\ 0{,}1426$$

Do krzyżowania przeznaczone zostały następujące chromosomy bieżącej populacji: Ch_2, Ch_4, Ch_8 i Ch_{10} ponieważ w ich wypadku wylosowane wartości okazały się mniejsze od prawdopodobieństwa krzyżowania. Uzyskaliśmy więc cztery chromosomy rodziców. Jak widać lepsze od innych przystosowanie tych chromosomów dało rezultaty. Pary rodziców tworzy się także losowo. Załóżmy, że uzyskaliśmy następujące pary przeznaczone do krzyżowania: Ch_2 i Ch_4 oraz Ch_8 i Ch_{10}.

Kiedy już wiemy, które osobniki tworzą pary rodzicielskie przychodzi czas wylosowanie dla każdej nich punktu krzyżowania. Rozcięcie obu chromosomów rodziców następuje w tym samym miejscu. Załóżmy, że dla pierwszej pary wypadnie ono w punkcie $n = 5$ zaś dla drugiej w $n = 3$:

$$\begin{aligned}
Ch_2 &= (\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ |\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ ) \\
Ch_4 &= (\ 1\ 1\ 0\ 0\ 0\ |\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ ) \\
Ch_8 &= (\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ |\ \mathbf{0}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ ) \\
Ch_{10} &= (\ 0\ 1\ 1\ |\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ )
\end{aligned}$$

Chromosomy potomków powstają wskutek zamiany podciągów między chromosomami rodziców zaczynając od punktu $n + 1$. Pierwszy potomek otrzyma początkowe n genów od pierwszego z rodziców a resztę od drugiego. Drugi potomek pierwsze n genów otrzymuje od rodzica drugiego natomiast pozostałe od pierwszego. Powstałe po krzyżowaniu chromosomy potomków reprezentują następujące łańcuchy zer i jedynek[1]:

$$\begin{aligned}
Ch_2^* &= (\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ ) \\
Ch_4^* &= (\ 1\ 1\ 0\ 0\ 0\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ ) \\
Ch_8^* &= (\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ ) \\
Ch_{10}^* &= (\ 0\ 1\ 1\ \mathbf{0}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ \mathbf{1}\ \mathbf{0}\ )
\end{aligned}$$

[1] Oznaczono je symbolem „gwiazdki” dla odróżnienia od chromosomów rodziców.

Populacja, uwzględniająca nowopowstałe chromosomy potomków prezentuje się następująco:

$$
\begin{aligned}
Ch_1 &= (\ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ )\\
Ch_2^* &= (\ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ )\\
Ch_3 &= (\ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ )\\
Ch_4^* &= (\ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ )\\
Ch_5 &= (\ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ )\\
Ch_6 &= (\ 0\ \ 1\ \ 1\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 1\ \ 1\ )\\
Ch_7 &= (\ 0\ \ 1\ \ 0\ \ 0\ \ 1\ \ 1\ \ 0\ \ 0\ \ 1\ \ 1\ \ 0\ \ 0\ )\\
Ch_8^* &= (\ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 1\ \ 1\ \ 1\ )\\
Ch_9 &= (\ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ )\\
Ch_{10}^* &= (\ 0\ \ 1\ \ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 1\ \ 0\ \ 1\ \ 0\ )
\end{aligned}
$$

W kolejnym kroku poddamy chromosomy mutacji. Dla każdego z genów we wszystkich chromosomach losujemy liczbę z przedziału $\langle 0, 1\rangle$. Jeżeli okaże się ona mniejsza lub równa prawdopodobieństwu mutacji (przypomnijmy, że $p_m = 0{,}01$) wtedy gen podlega mutacji czyli dochodzi do zamiany bitu na przeciwny.

Załóżmy, że wskutek losowań do mutacji przeznaczone zostały chromosomy Ch_2^* i Ch_5^*. Dla chromosomu Ch_2^* wylosowaliśmy punkt mutacji $n = 9$. Przed mutacją chromosom miał postać (gen przeznaczony do mutacji został pogrubiony):

$$Ch_2^* = (\ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ )$$

Zaś po mutacji otrzymamy:

$$Ch_2^* = (\ 1\ \ 0\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ \mathbf{1}\ \ 1\ \ 0\ \ 0\ )$$

W przypadku chromosomu Ch_5 wylosowaliśmy punkt mutacji $n = 3$ co oznacza, że przed mutacją chromosom ma postać:

$$Ch_5 = (\ 1\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ )$$

Zaś po mutacji:

$$Ch_5 = (\ 1\ \ 1\ \ \mathbf{1}\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ \ 0\ \ 1\ \ 0\ \ 0\ )$$

Pozostałe chromosomy nie uległy mutacji. Po zakończeniu krzyżowania oraz mutacji otrzymujemy nową, zmienioną populację:

$$
\begin{aligned}
Ch_1 &= (\ 1\ 0\ 0\ 1\ 0\ 1\ 1\ 0\ 1\ 0\ 1\ 0\ ) \\
Ch_2 &= (\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ ) \\
Ch_3 &= (\ 1\ 0\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ ) \\
Ch_4 &= (\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ ) \\
Ch_5 &= (\ 1\ 1\ 1\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ ) \\
Ch_6 &= (\ 0\ 1\ 1\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ ) \\
Ch_7 &= (\ 0\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ 1\ 1\ 0\ 0\ ) \\
Ch_8 &= (\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 1\ 1\ 1\ ) \\
Ch_9 &= (\ 1\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ 0\ 1\ 0\ 0\ ) \\
Ch_{10} &= (\ 0\ 1\ 1\ 0\ 0\ 1\ 0\ 0\ 1\ 0\ 1\ 0\ )
\end{aligned}
$$

(5) Wyznaczenie wartości funkcji przystosowania dla chromosomów nowej populacji

Dla nowej populacji wyznaczamy wartości funkcji przystosowania[2]:

$$
\begin{aligned}
F(Ch_1) &= F(37{,}42) = 1458 \\
F(Ch_2) &= F(33{,}12) = 1118 \\
F(Ch_3) &= F(33{,}10) = 1114 \\
F(Ch_4) &= F(49{,}10) = 2426 \\
F(Ch_5) &= F(57{,}4) = 3262 \\
F(Ch_6) &= F(26{,}15) = 711 \\
F(Ch_7) &= F(19{,}12) = 390 \\
F(Ch_8) &= F(34{,}15) = 1191 \\
F(Ch_9) &= F(49{,}4) = 2414 \\
F(Ch_{10}) &= F(25{,}10) = 601
\end{aligned}
$$

Jeżeli porównamy najlepsze rozwiązania z poprzedniej i bieżącej populacji wtedy okaże się, że w drugiej z nich znaleźliśmy punkt o wyższej wartości funkcji przystosowania. Poprzednio równała się 2414, a teraz wynosi 3262.

Do 14685 wzrosła również suma funkcji przystosowania dla wszystkich chromosomów oraz średnia arytmetyczna wartości przystosowania (do 1468,5 przy odchyleniu standardowym 885,58).

[2] W tym celu ponownie zamieniamy chromosomy na liczby rzeczywiste i podstawiamy do funkcji (1.16) tak jak uczyniliśmy to dla populacji początkowej.

Zaobserwowaliśmy poprawę wartości funkcji przystosowania zarówno pod względem sumy jak i średniego poziomu. Poza tym najlepszy chromosom w drugiej iteracji (zwanej tu również pokoleniem) zapewnił wyższą wartość tej funkcji niż najlepszy chromosom z populacji poprzedniej. Nie zawsze tak musi być ponieważ zastosowane operatory mogą zniszczyć chromosom o najwyższej wartości funkcji przystosowania. Spadło przy tej okazji zróżnicowanie wśród chromosomów drugiej populacji.

Spadek zróżnicowania zwykle obserwujemy w dalszych niż drugie pokoleniach. Jest to zjawisko niekorzystne i zostanie omówione szerzej w dalszej części rozdziału. W praktyce krzyżowanie i mutacja potrafią zmniejszyć średnie przystosowanie zwiększając w zamian różnorodność osobników. Najsilniej na ten proces wpływa krzyżowanie, które obejmuje swoim działaniem całe bloki chromosomów. Mutacja dotyczy pojedynczych genów.

(6) Skasowanie populacji rodziców

Po zakończeniu procedur krzyżowania i mutacji powstaje nowa populacja składająca się z chromosomów potomków i pozostałych skopiowanych osobników. Kiedy wyznaczymy dla nich wartości funkcji przystosowania wówczas dotychczasowa populacja bazowa nie jest już potrzebna[3]. Jej miejsce zajmują chromosomy poddane działaniu krzyżowania oraz mutacji, stając się populacją biorącą udział w następnej iteracji algorytmu. Dla rozpatrywanego przykładu będzie nią populacja z punktu czwartego.

(7) Sprawdzenie warunku zatrzymania algorytmu

Jednym z parametrów wpływających na pracę algorytmu genetycznego jest warunek zatrzymania. W naszym przykładzie przedstawiamy mechanizm działania algorytmu jako takiego, w związku z czym poprzestajemy na wyliczeniach wykonanych dla pojedynczej iteracji (jednego pokolenia).

O zakończeniu postępowania może zadecydować liczba wykonanych przebiegów lub skala poprawy rozwiązania wynikająca z porównania dopasowania najlepszych chromosomów z dwóch kolejnych pokoleń. Wybór kryterium zakończenia algorytmu zależy od przeprowadzającego obliczenia. Ciekawe rozważania na temat doboru kryteriów zatrzymania algorytmu znajdziemy m.in. w (Ong Bun i Fukushima 2011). Autorzy tej pracy zwrócili uwagę, że natura w przeciwieństwie do nas nie poszukuje sposobu przerwania funkcjonowania populacji. Musimy jednak w jakiś sposób określić moment zakończenia obliczeń przy co wiąże się z tzw. problemem przedwczesnej zbieżności. Zbyt wczesne zatrzymanie obliczeń powoduje, że ryzykujemy znalezienie lokalnego maksimum zamiast globalnego. Praktyka stosowania algorytmów

[3] Pewien wyjątek od tej reguły stanowi metoda selekcji elitarnej, w której ze starszej populacji zachowuje się najlepszy chromosom na wypadek jego zniszczenia w dalszym postępowaniu.

genetycznych polega zazwyczaj na tym, że osoba prowadząca obliczenia określa moment i kryterium zatrzymania na przykład odpowiednio małe różnice przystosowania najlepszych osobników w kolejnych pokoleniach lub ustaloną liczbę pokoleń.

Ong Bun i Fukushima opracowali kryterium zatrzymujące pracę algorytmu nie wymagające od użytkownika wiedzy *a priori*. Wprowadzili tzw. operator mutagenezy, który różni się od klasycznej mutacji tym, że osobniki poddane jego działaniu mają eksplorować nowe obszary zbioru rozwiązań. Mutagenezę wywołujemy już po utworzeniu potomka a obejmuje ona pewną liczbę najgorzej przystosowanych osobników, które zmieniane są tak by trafiły do niezbadanych jeszcze obszarów zbioru.

Każdemu z genów przypisano m ustalonych wcześniej podobszarów zbioru rozwiązań. Autorzy zastosowali kodowanie rzeczywiste, ale geny nie przyjmują pojedynczych wartości tylko reprezentowane są przez macierze. Kiedy gen uzyskuje wartość z niezbadanej do tej pory części zbioru odpowiadającą mu wartość w wierszu macierzy genów zmieniamy z 0 na 1. Kiedy sprawdzone zostaną wszystkie wyróżnione obszary postępowanie algorytmu zostaje zakończone.

1.2. Przegląd modyfikacji wpływających na konstrukcję chromosomu

1.2.1. Zmiany reprezentacji genotypu

Pierwsza wersja prognostycznego algorytmu genetycznego wykorzystywała osobniki diploidalne (Kucharski 2007) a zatem posiadające dwa chromosomy. Pierwszy zawierał konkretną prognozę na dany okres, zaś drugi informację o okresie, z którego wywodziła się dana wartość. Dalsze badania zredukowały tę liczbę do jednego chromosomu i ostateczna wersja algorytmu wykorzystuje reprezentację haploidalną. Mimo tego uważamy, że opis modyfikacji algorytmu genetycznego warto rozpocząć od opisu innych niż haploidalne reprezentacji genotypu. Mogą one stanowić interesujący kierunek dalszych badań.

Genotyp to inaczej zbiór chromosomów, które określają strukturę danego osobnika. Do tej pory posługiwaliśmy się najprostszym przypadkiem czyli haploidem, w którym genotyp osobnika składa się z jednego chromosomu. Spotykamy się wprawdzie w przyrodzie z taką sytuacją, lecz dotyczy ona mało rozwiniętych, a co za tym idzie nieskomplikowanych organizmów.

Powszechniej występują formy życia mające znacznie bardziej rozbudowaną strukturę genetyczną. Posłużenie się we wprowadzeniu haploidem pozwala na łatwiejsze zrozumienie mechanizmów występujących w czasie

działania algorytmu genetycznego. Większą część problemów optymalizacyjnych daje się zresztą rozwiązać przy pomocy haploidów, jednakże istnieje trend do jak najwierniejszego odwzorowania procesów zachodzących w przyrodzie. W środowisku naturalnym zaś w „powszechnym użyciu" znajdują się genotypy diploidalne bądź poliploidalne.

W diploidzie genotyp składa się z pary chromosomów (tzw. chromosomów homologicznych), zaś każdy z nich koduje tę samą funkcję u osobnika. Genotypy poliploidalne zawierają więcej par chromosomów homologicznych (Goldberg 2003, s. 162).

Powstaje więc pytanie dlaczego dochodzi do redundancji. W końcu na wielokrotne zapamiętanie tej samej informacji potrzeba przecież więcej miejsca. Czemu w przyrodzie, w której nic nie dzieje się przypadkiem, dochodzi do przechowywania w pewnym sensie kopii posiadanej już informacji?

Odpowiedzią jest mechanizm dominacji, który decyduje jaki wariant genu zostanie wybrany. W chromosomie gen (bądź grupa genów czyli allel) może być dominujący lub recesywny. Z badań przeprowadzonych przez biologów wynika, że allel dominujący uzyskuje pierwszeństwo przed allelem recesywnym. Załóżmy, że mamy do czynienia z dwoma chromosomami, których budowę prezentujemy poniżej (dużymi literami oznaczono geny dominujące).

$$Ch_1 = (\ a \quad B \quad C \quad d \quad e \quad F\ )$$
$$Ch_2 = (\ A \quad b \quad C \quad d \quad e \quad F\ )$$

Geny dominujące i recesywne podlegają następującym regułom:

gen dominujący + gen dominujący → gen dominujący
gen dominujący + gen recesywny → gen dominujący
gen recesywny + gen recesywny → gen recesywny

Zjawisko polegające na tym, że gen dominujący pojawia się przed recesywnym na tej samej pozycji w chromosomie nosi nazwę ekspresji. Informacja przechowywana jest w dwóch wariantach, a do głosu dochodzą te geny, które spełniają reguły dominacji. Stąd chromosom powstały z Ch_1 i Ch_2, który powstanie po rozkodowaniu będzie miał postać:

$$Ch_3 = (\ A \quad B \quad C \quad d \quad e \quad F\ )$$

Istnienie genów recesywnych tłumaczy się jako sposób na zapamiętanie genów bądź ich kombinacji, które w przeszłości okazały się pożyteczne oraz niewykluczone, że ponownie wystąpi taki stan rzeczy. W wypadku, gdyby informacja została zniszczona w toku ewolucji a warunki środowiskowe wróciły do stanu przeszłego organizm musiałby po raz kolejny pokonywać mozolną drogę przystosowywania się niemal od zera. Innymi słowy genotypy

diploidalne i poliploidalne zapewniają szybszą adaptację do zmieniających się warunków otoczenia. Zresztą sama kwestia tego czy dany gen bądź allel jest dominujący czy nie również podlega ewolucji.

Zdając sobie sprawę z możliwości jakie daje zastosowanie wariantu diploidalnego próbowano go dość wcześnie zastosować w analizie algorytmów genetycznych. Za najprostszy i najbardziej przejrzysty uchodzi tutaj model Hollstiena-Hollanda (Goldberg 2003, s. 166).

R. Hollstien opisał model diploidalnego algorytmu genetycznego, który uwzględniał także możliwość mutacji współczynnika dominacji. W pierwszej metodzie używał on dwóch genów: funkcyjnego i modyfikatora. Gen funkcyjny był normalnym genem przyjmującym wartości „0” lub „1” zaś modyfikator decydował o tym, który gen dominował. Jednak analizując swój model Hollstien stwierdził, że da się go uprościć i tak powstał drugi wariant. Chromosomy są w nim konstruowane na podstawie następującego alfabetu genów: $\{0, 1, 2\}$. W tym ujęciu „2” staje się dominującym dla „1” (poprzednio gen dominujący oznaczaliśmy dużą literą w odróżnieniu od recesywnego, oznaczonego małą) a „1” recesywnym dla „0”. Wówczas „2” dominuje nad „1” i nad „0” zaś „0” dominuje nad „1”. Układ dominacji genów w takim modelu da się zapisać w postaci następującego schematu:

Tabela 1.3: Schemat dominacji w modelu Hollstiena

	0	1	2
0	0	0	1
1	0	1	1
2	1	1	1

Źródło: Hollstien [1971]

Tabela 1.3 ilustruje, które geny będą dominować w myśl modelu Hollstiena. Na przykład gen „2” dominuje nad „1” i dlatego na pozycjach (3,2) i (2,3) znalazły się jedynki.

Zaproponowane przez Hollanda podejście różniło się tym, że w jego alfabecie zamiast „1” występowało „10” i stąd znany jest pod nazwą model Hollstiena-Hollanda. Rozwijając dalej badania nad tym zagadnieniem obaj uczeni wykazali jednak, że tylko w niektórych przypadkach wprowadzenie diploidu poprawia działanie algorytmu genetycznego. Być może stało się tak dlatego, że przyjmowano stałą w czasie funkcję przystosowania. Jej uzmiennienie odpowiadałoby przemianom dokonującym się w środowisku, w którym działa algorytm.

Jak już nadmieniliśmy, istnieje duża grupa problemów, w których dobrze sprawdza się reprezentacja haploidalna. Dlatego w naszej pracy, o ile nie wskażemy, że jest inaczej, opieramy się na osobnikach posiadających je-

den chromosom. Z tego też powodu będziemy zamiennie używać terminów „osobnik” i „chromosom”.

1.2.2. Działania wpływające na porządek genów w chromosomie

Przybliżając działanie klasycznego algorytmu genetycznego wspomnieliśmy o przykładzie, który znalazł się w (Trzaskalik 2008). Autor wykorzystał tam w charakterze ilustracji problem jednego komiwojażera. Wybór ten był o tyle niezbyt szczęśliwy, że zagadnienie to oraz wiele z nim spokrewnionych (jak problem wielu komiwojażerów) kiepsko rozwiązują się przy użyciu kodowania binarnego i tradycyjnie rozumianych operatorów. Wymagają one wówczas dużych nakładów czasu i mocy obliczeniowej. Na użytek tej rodziny zadań opracowane zostały modyfikacje uwzględniające specyfikę problemu, które poprawiły efektywność poszukiwania rozwiązania optymalnego.

Jednym ze sposobów zmierzających do poprawy rozwiązań otrzymywanych przy pomocy algorytmu genetycznego (właściwie to w przypadku bardziej zaawansowanych zmian można już mówić o algorytmie ewolucyjnym) jest ingerowanie w strukturę genów tworzących chromosom.

Pierwszą z metod rekonfiguracji chromosomu zwana inwersją polega na „wycięciu” fragmentu chromosomu a następnie wstawieniu go w poprzednie miejsce, tyle że w odwrotnej kolejności (Goldberg 2003, s. 182) na przykład:

$Ch = (\ 1\ \ 0\ \ 1\ \ 1\ \ \mathbf{0}\ \ \mathbf{0}\ \ \mathbf{1}\ \ \mathbf{1}\ \ 0\ \ 1\ \ 1\ \ 0\ )$ – przed inwersją

$Ch^* = (\ 1\ \ 0\ \ 1\ \ 1\ \ \mathbf{1}\ \ \mathbf{1}\ \ \mathbf{0}\ \ \mathbf{0}\ \ 0\ \ 1\ \ 1\ \ 0\ )$ – po inwersji

Operatora inwersji używamy dla pojedynczych chromosomów. Po jej wykonaniu powstaje osobnik o nowym przystosowaniu oczywiście przy założeniu, że nowe położenie danego genu (lub całej grupy genów czyli allelu) zmienia jego funkcje.

W przyrodzie allele chromosomu po inwersji muszą wypełniać te same funkcje co przed nią więc oprócz reprezentacji bitowej należy zapamiętać dodatkowo początkową pozycję genu w chromosomie. Dlatego nasz przykład należałoby przedstawić nieco inaczej:

$Ch = (\ 1\ \ 2\ \ 3\ \ 4\ \ \mathbf{5}\ \ \mathbf{6}\ \ \mathbf{7}\ \ \mathbf{8}\ \ 9\ \ 10\ \ 11\ \ 12\ )$ – przed inwersją

$Ch^* = (\ 1\ \ 2\ \ 3\ \ 4\ \ \mathbf{8}\ \ \mathbf{7}\ \ \mathbf{6}\ \ \mathbf{5}\ \ 9\ \ 10\ \ 11\ \ 12\ )$ – po inwersji

Pod każdą z ponumerowanych pozycji ukrywa się oczywiście zero bądź jedynka. Tak dokonana zmiana porządku genów nie spowoduje zmiany wartości funkcji przystosowania.

Zastosowanie operatora inwersji powoduje wszakże problemy z krzyżowaniem. Załóżmy, że dla chromosomów z ostatniego przykładu punkt krzyżowania wypadnie po szóstym genie:

$$Ch_1^R = (\ \mathbf{1}\ \ \mathbf{2}\ \ \mathbf{3}\ \ \mathbf{4}\ \ \mathbf{5}\ \ \mathbf{6}\ \ |\ \ \mathbf{7}\ \ \mathbf{8}\ \ \mathbf{9}\ \ \mathbf{10}\ \ \mathbf{11}\ \ \mathbf{12}\ )$$
$$Ch_2^R = (\ 1\ \ 2\ \ 3\ \ 4\ \ 8\ \ 7\ \ |\ \ 6\ \ 5\ \ 9\ \ 10\ \ 11\ \ 12\ )$$

$$Ch_1^P = (\ \mathbf{1}\ \ \mathbf{2}\ \ \mathbf{3}\ \ \mathbf{4}\ \ \mathbf{5}\ \ \mathbf{6}\ \ 6\ \ 5\ \ 9\ \ 10\ \ 11\ \ 12\ )$$
$$Ch_2^P = (\ 1\ \ 2\ \ 3\ \ 4\ \ 8\ \ 7\ \ \mathbf{7}\ \ \mathbf{8}\ \ \mathbf{9}\ \ \mathbf{10}\ \ \mathbf{11}\ \ \mathbf{12}\ )$$

gdzie:
Ch^R – chromosom rodzica
Ch^P – chromosom potomka

Zajdzie wtedy niekorzystne zjawisko ponieważ chromosomy potomków nie zawierają niektórych spośród genów, z kolei inne pojawiają się więcej niż raz. Podejmowane próby poradzenia sobie z nieprawidłowościami powstającymi podczas krzyżowania doprowadziły w 1985 roku do zaprezentowania kilka modeli (Goldberg 2003, s. 184), które łączyły w sobie inwersję i krzyżowanie. Były to modele: PMX (*partially matched crossover*); OX (*order crossover*) oraz CX (*cycle crossover*).

Model PMX

W modelu tym losujemy dwa punkty krzyżowania. Obszar pomiędzy nimi nosi nazwę obszaru dopasowania (*matching section*). W toku postępowania geny z jednego i drugiego chromosomu zamieniają się pozycjami. Dopasowania ciągów dokonujemy tak długo aż otrzymamy prawidłowych potomków. Pozycje genów zdublowanych określają geny z obszaru dopasowania. Jako ilustracja niech posłuży poniższy przykład:

$$Ch_1^R = (\ 1\ \ 2\ \ 3\ \ |\ \ 8\ \ 5\ \ 6\ \ |\ \ 7\ \ 9\ \ 4\ )$$
$$Ch_2^R = (\ 7\ \ 8\ \ 5\ \ |\ \ 3\ \ 2\ \ 1\ \ |\ \ 4\ \ 9\ \ 6\ )$$

$$Ch_1^P = (\ 6\ \ 5\ \ 8\ \ 3\ \ 2\ \ 1\ \ 7\ \ 9\ \ 4\ )$$
$$Ch_2^P = (\ 7\ \ 3\ \ 2\ \ 8\ \ 5\ \ 6\ \ 4\ \ 9\ \ 1\ )$$

Model OX

Model OX jest podobny do PMX. Ponownie wyznaczamy obszar dopasowania pomiędzy wylosowanymi punktami krzyżowania i wymieniamy

geny między chromosomami rodziców, tyle że tym razem zapełniamy luki powstałe po przemieszczeniu elementów. Załóżmy, że rozpatrujemy parę rodziców z poprzedniego przykładu. Postępowanie w modelu OX przebiega następująco:

$$Ch_1^R = (\ 1\ \ 2\ \ 3\ \mid\ 8\ \ 5\ \ 6\ \mid\ 7\ \ 9\ \ 4\ )$$
$$Ch_2^R = (\ 7\ \ 8\ \ 5\ \mid\ 3\ \ 2\ \ 1\ \mid\ 4\ \ 9\ \ 6\ )$$

$$Ch_1^P = (\ 8\ \ 5\ \ 6\ \ 3\ \ 2\ \ 1\ \ 7\ \ 9\ \ 4\ )$$
$$Ch_2^P = (\ 3\ \ 2\ \ 1\ \ 8\ \ 5\ \ 6\ \ 4\ \ 9\ \ 7\ )$$

Model CX

W modelu tym pozycja genu w chromosomach potomków bierze się od jednego lub drugiego rodzica. Po raz kolejny rozpatrzymy znaną nam już parę rodziców:

$$Ch_1^R = (\ 1\ \ 2\ \ 3\ \ 8\ \ 5\ \ 6\ \ 7\ \ 9\ \ 4\ )$$
$$Ch_2^R = (\ 7\ \ 8\ \ 5\ \ 3\ \ 2\ \ 1\ \ 4\ \ 9\ \ 6\ )$$

W chromosomie pierwszego potomka na pierwszej pozycji wpisujemy pierwszy od lewej gen chromosomu pierwszego z rodziców:

$$Ch_1^P = (\ 1\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ )$$

Teraz kopiujemy gen pierwszy od lewej drugiego z rodziców, ponieważ jednak w chromosomie potomka pierwsza pozycja jest już zajęta więc gen wpiszemy na pozycję, którą zajmował on w chromosomie pierwszego rodzica:

$$Ch_1^P = (\ 1\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ 7\ \ \ldots\ \ \ldots\ )$$

Siódma pozycja u potomka Ch_1 jest już zajęta, więc „czwórkę” z pozycji 7 rodzica Ch_2 kopiujemy na pozycję 9 rodzica Ch_1:

$$Ch_1^P = (\ 1\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ 7\ \ \ldots\ \ 4\ )$$

Takie postępowanie powtarzamy dalej aż do zapełnienia wszystkich pozycji. W następnym kroku otrzymujemy:

$$Ch_1^P = (\ 1\ \ \ldots\ \ \ldots\ \ \ldots\ \ \ldots\ \ 6\ \ 7\ \ \ldots\ \ 4\ )$$

W tym momencie powinniśmy wybrać 1 chromosomu Ch_2^R, ale ten wybór został już wcześniej dokonany. Dlatego pozostałe pozycje zostaną uzupełnione o geny pochodzące od drugiego z rodziców.

Aby otrzymać drugiego potomka zaczynamy od genu na pierwszej pozycji rodzica Ch_2^R a następnie wykonujemy operacje analogiczne jak w przypadku potomka pierwszego. W efekcie tego otrzymamy następujące chromosomy potomków:

$$Ch_1^P = (\ 1\ \ 8\ \ 5\ \ 3\ \ 2\ \ 6\ \ 7\ \ 9\ \ 4\ )$$
$$Ch_2^P = (\ 7\ \ 2\ \ 3\ \ 8\ \ 5\ \ 1\ \ 4\ \ 9\ \ 6\ )$$

Trzy powyższe modele (czyli PMX, OX, CX) od czasu pierwszego zaprezentowania były modyfikowane przez kolejnych badaczy. Swoją przydatność (zwłaszcza PMX) udowodniły szczególnie w przypadku rozwiązywania problemu komiwojażera. Stanowią oczywiście ciekawą alternatywę nie tylko dla tego zagadnienia. Zmiana porządku genów w chromosomie zapewnia bowiem jednoczesne usprawnianie poszukiwania wartości funkcji oraz znajdowanie lepszych zestawów genów.

Wzrost znaczenia logistyki spowodowany postępującą globalizacją gospodarki przyczynił się do większego zainteresowania tematyką optymalizacji tras pojazdów przy różnych dodatkowych założeniach. Zainteresowany Czytelnik bez trudu znajdzie prace, w których zaprezentowano zarówno rozwinięcia powyższych jak i nowe klasy metod. My pozwolimy sobie jedynie na zarysowanie zagadnienia.

1.2.3. Kodowanie rzeczywiste chromosomu

W klasycznym algorytmie genetycznym chromosomy koduje się binarnie, wykorzystując łańcuchy zer i jedynek. Ma to swoje zalety: na bit informacji przypada najwięcej schematów, łatwiej stosuje się operatory takie jak krzyżowanie czy mutacja.

Kłopoty zaczynają się jednak w przypadku zadań z wieloma zmiennymi decyzyjnymi lub takich, w których wymagana jest duża dokładność obliczeń. Chromosomy składają się wtedy z długich łańcuchów co znacząco wydłuża czas potrzebny do znalezienia rozwiązania. Niekiedy wręcz stawia to pod znakiem zapytania samo wykorzystanie algorytmu genetycznego czy ewolucyjnego.

Wykorzystanie kodowania w postaci liczb rzeczywistych pozwala uniknąć powyższych wad oraz uzyskać bliższy związek algorytmu z przestrzenią zadania. W rozdziałach trzecim i czwartym korzystamy właśnie z takiej reprezentacji genotypu. Z tego powodu konieczne jest przybliżenie zagadnień jakie niesie ze sobą wprowadzenie do chromosomów liczb rzeczywistych.

Zmiana sposobu reprezentacji genotypu oznacza z jednej strony lepsze odwzorowanie warunków zadania, ale powoduje też zmiany w stosowanych w algorytmie operatorach krzyżowania i mutacji (Michalewicz 2003, s. 160).

Poniżej dokonujemy przeglądu uznanych przez nas za interesujące modyfikacji tych dwóch procedur. Zakładają one, że gen przyjmuje wartość ze zbioru liczb rzeczywistych.

Istnieje jednak również inne podejście do kodowania tzw. kodowanie drzewiaste, w którym chromosom ma strukturę drzewa: węzły zawierają informację zaś krawędzie wyrażają relacje między węzłami. Innymi słowy mamy do czynienia z rodzajem grafu. Ten typ kodowania wykorzystuje się dla szczególnych zastosowaniach, w których celem jest tworzenie i ewolucja reguł decyzyjnych. Przykład wykorzystania takiego kodowania znajdziemy w (Zheng, Dong i Zheng 2001) gdzie zadanie przydziału w systemach przydziału pamięci zostało przedstawione właśnie jako graf. Kodowanie drzewiaste znajdziemy również w pracy (Bérard 2005) gdzie analizowane jest zachowanie tego typu struktur w czysto losowym otoczeniu.

Poszukiwanie reguł decyzyjnych to temat chętnie poruszany w publikacjach poświęconych inwestycjom giełdowym. W (Allen i Karjalainen 1999) autorzy użyli kodowania drzewiastego do tworzenia reguł podpowiadających strategie inwestycyjne bazujące na analizie technicznej. W ich przypadku chromosomy miały zmienną długość co jest powszechną praktyką w tym rodzaju kodowania. Warto nadmienić, że tworzone reguły decyzyjne miały skomplikowaną budowę obejmującą zarówno funkcje matematyczne (jak średnia), logiczne oraz operatory arytmetyczne. Wprawdzie nurt ten nie znajduje się w obszarze naszych bezpośrednich zainteresowań, ale uznaliśmy iż warto o nim wspomnieć. Nieco podobną filozofię stosują Beenstock oraz Szpiro w (Beenstock i Szpiro 2001). Wprawdzie nie używają chromosomów o kodowaniu drzewiastym, ale ich algorytm składa równania ekonometryczne ze zmiennych, operatorów matematycznych i parametrów. Jako funkcję przystosowania wykorzystują współczynnik determinacji. Sęk w tym, że ich postępowanie bardzo przypomina data mining ponieważ zupełnie ignorują zależności merytoryczne między zmiennymi.

Krzyżowanie proste

Załóżmy, że dysponujemy następującą parą chromosomów rodziców:

$$Ch_1^R = (x_1, \ldots, x_n)$$
$$Ch_2^R = (y_1, \ldots, y_n)$$

Krzyżowanie proste do pewnego stopnia przypomina znane nam już krzyżowanie jednopunktowe. Wyznaczamy więc punkt cięcia i tworzymy potomków, u których geny znajdujące się za tym punktem uzyskują nowe wartości zgodnie z poniższym schematem:

$$Ch_1^P = (x_1, \ldots, x_p, \alpha y_{p+1} + (1-\alpha)x_{p+1}, \ldots, \alpha y_n + (1-\alpha)x_n)$$
$$Ch_2^P = (y_1, \ldots, y_p, \alpha x_{p+1} + (1-\alpha)y_{p+1}, \ldots, \alpha x_n + (1-\alpha)y_n)$$

Wyznaczenie odpowiedniej wartości α, która zapewni odpowiednią wymianę informacji może sprawiać problemy. Najprostsze rozwiązanie polega na tym, aby zacząć od $\alpha = 1$ i stopniowo je zmniejszać poszukując najlepszej wartości parametru gwarantującej, że potomkowie znajdą się w dziedzinie.

Krzyżowanie arytmetyczne

W krzyżowaniu arytmetycznym potomkowie są liniową kombinacją całych chromosomów rodziców, czyli:

$$Ch_1^P = \alpha Ch_1^R + (1-\alpha)Ch_2^R$$
$$Ch_2^P = \alpha Ch_2^R + (1-\alpha)Ch_1^R$$

Wartość α losuje się z przedziału $\langle 0,1 \rangle$. Dzięki temu nie ma problemu, aby uzyskać potomków należących do dziedziny funkcji przystosowania.

Krzyżowanie heurystyczne

W przypadku tego operatora krzyżowania wykorzystujemy wartość funkcji kryterium, aby określić kierunek poszukiwań. W rezultacie powstaje tylko jeden potomek i możliwa jest sytuacja, w której nie ma go wcale. Potomka tworzymy w oparciu o regułę:

$$Ch_3^P = r(Ch_2^R - Ch_1^R) + Ch_2^R$$

gdzie:
r – liczba wylosowana z przedziału $\langle 0,1 \rangle$

Chromosom rodzica Ch_2^R nie może być gorszy niż chromosom drugiego rodzica czyli Ch_1^R a więc zachodzą relacje:

$$FP(Ch_2^R) \geqslant FP(Ch_1^R) \quad \text{dla maksymalizacji}$$
$$FP(Ch_2^R) \leqslant FP(Ch_1^R) \quad \text{dla minimalizacji}$$

Jeżeli otrzymamy niedopuszczalnego potomka wtedy losujemy nową wartość r i tworzymy kolejnego osobnika. Proces ten powtarza się ustaloną liczbę razy. W razie niepowodzenia w kolejnych próbach rezygnujemy z tworzenia potomka z danej pary chromosomów.

Mutacja równomierna

Kodowanie rzeczywiste wymusza zmiany również w operatorze mutacji. Pierwszym z kilku sposobów jego modyfikacji jest mutacja równomierna. Zgodnie z ideą mutacji, losowo wybieramy gen x_p z danego chromosomu po czym losujemy (z równomiernym rozkładem prawdopodobieństwa) jego nową wartość z przedziału $\langle left(p), right(p) \rangle$, który spełnia dwa warunki:

1. Dla każdych dwóch punktów p_1 i p_2 należących do wypukłej przestrzeni S można wyznaczyć ich liniową kombinację także będącą punktem tej przestrzeni.
2. Jeżeli punkt p_0 należy do S oraz do prostej l wtedy prosta ta przecina brzegi S w dwóch punktach: $left_l^{p0}$ i $right_l^{p0}$.

Mutację równomierną warto stosować w początkowej fazie działania algorytmu oraz wtedy kiedy początkowa populacja składa się z wielu egzemplarzy tego samego chromosomu. Operator ten odsuwa niebezpieczeństwo przedwczesnej zbieżności do lokalnego optimum. Jeżeli w zadaniu występują ograniczenia oraz algorytm ma zadany punkt startowy wówczas w populacji początkowej pojawi się wiele kopii tego samego chromosomu. W takiej sytuacji ten operator sprawdza się szczególnie dobrze.

Mutacja brzegowa

Metoda ta jest podobna do mutacji równomiernej z tym, że mutowany gen równa się albo $left(p)$, albo $right(p)$ z jednakowym prawdopodobieństwem. Mutacja brzegowa sprawdza się kiedy rozwiązanie optymalne zadania leży na, bądź blisko brzegu zbioru rozwiązań dopuszczalnych.

Mutacja nierównomierna

Ten operator pozwala na dokładne dostrojenie się systemu. W wyniku użycia mutacji nierównomiernej tego typu otrzymujemy potomka, w którym zmutowany gen x_p^* wyznaczamy zgodnie z regułą:

$$x_p^* = \begin{cases} x_p + \triangle(k,right(k) - x_p), & \text{gdy wylosowano 0} \\ x_p - \triangle(k,x_p - left(p)), & \text{gdy wylosowano 1} \end{cases} \tag{1.18}$$

Funkcja $\triangle(k,y)$ przyjmuje wartości z przedziału $\langle 0, y\rangle$ przy czym ze wzrostem k (gdzie k to numer kolejnej iteracji algorytmu) rośnie prawdopodobieństwo, że wartość tej funkcji będzie bliska 0. Dzięki temu w początkowych pokoleniach przestrzeń rozwiązań jest przeszukiwana bardziej jednorodnie zaś w dalszych pokoleniach bardziej lokalnie.

Za (Michalewicz 2003, s. 143) podajemy przykład funkcji, którą można wykorzystać w przypadku tego operatora. We wzorze (1.19) r oznacza liczbę losowaną z przedziału $\langle 0,1\rangle$, K – najwyższy numer pokolenia (iteracji) zaś b to parametr określający stopień niejednorodności.

$$\triangle(k,y) = yr\left(1 - \frac{k}{K}\right)^b \tag{1.19}$$

1.3. Przegląd modyfikacji operatorów algorytmu

1.3.1. Nieklasyczne metody selekcji chromosomów

Klasyczny algorytm genetyczny wybiera chromosomy do reprodukcji na podstawie metody koła ruletki, w której prawdopodobieństwo wylosowania osobnika jest wprost proporcjonalne do jego wartości funkcji przystosowania. To proste i intuicyjne rozwiązanie, ale także nie pozbawione wad.

W przypadku funkcji jednomodalnych występuje niepożądane zjawisko spadku zróżnicowania w kolejnych pokoleniach populacji (nazywamy to zmniejszonym naciskiem selektywnym) co staje się szczególnie widoczne pod koniec postępowania. Coraz większa liczba potomków uzyskuje zbliżone przystosowanie.

Przypomnijmy sobie przykład działania KAG przedstawiony powyżej. Po jednokrotnym uruchomieniu algorytmu uzyskaliśmy poważny wzrost średniego przystosowania chromosomów, za co zapłaciliśmy dużym spadkiem zmienności wartości funkcji kryterium. Trzeba również zwrócić uwagę, że selekcja metodą koła ruletki nie gwarantuje, że najlepsze rozwiązanie z populacji k znajdzie się w populacji $k+1$.

Pewną pomocą w takiej sytuacji jest korekta nacisku selektywnego przez dodanie lub odjęcie od funkcji przystosowania stałej. Nie rozwiązuje to wszystkich naszych problemów. Rozważmy dwa algorytmy, których funkcje przystosowania F^1 i F^2 różnią się tylko o pewną stałą zaś reszta parametrów (populacja, kodowanie chromosomu itd.) jest identyczna. Załóżmy też, że w obu algorytmach pojawiły się dwa identyczne osobniki Ch_1, Ch_2 oraz przystosowanie Ch_1 jest wyższe od przystosowania Ch_2. Proporcje liczby potomków uzyskanych przez rozważane osobniki w pierwszym oraz drugim algorytmie będą się różnić właśnie z powodu dodania stałej. Ponadto okazuje się, że sama wartość stałej wpływa na przebieg poszukiwania rozwiązania.

Polecamy przyjrzenie się rezultatom eksperymentu opisanym w pracy (Arabas 2004, s. 115), z których wynika, że zwiększanie dodawanej stałej prowadzi do zmniejszania się szybkości zbieżności. W tej samej pracy znalazła się próba rozwiązania tego problemu dzięki ustalaniu stałej dla każdego pokolenia oddzielnie na poziomie nie przekraczającym kresu dolnego wartości funkcji przystosowania wszystkich chromosomów populacji.

Metodzie koła ruletki brakuje również odporności na przedwczesną zbieżność algorytmu. Na początku postępowania, kiedy występuje duże zróżnicowanie wartości funkcji przystosowania obserwujemy duży nacisk selektywny co prowadzi do przedwczesnej zbieżności. W konsekwencji ta metoda selekcji słabo sprawdza się w zagadnieniach, w których występują ekstrema lokalne gdyż często nie może opuścić ich obszaru przyciągania.

Mając świadomość ograniczeń metody koła ruletki już na etapie projektowania algorytmu do celów prognostycznych zakładaliśmy wykorzystanie nieklasycznych metod selekcji. Zaprezentujemy kilka innych podejść do selekcji chromosomów, opierając się przy tym na doświadczeniach K. de Jonga oraz A. Brindle (Goldberg 2003, s. 122-136).

Wyniki przedstawionych przez nich analiz wywarły największy wpływ na późniejsze rozważania na temat metod selekcji. Są w związku z tym często cytowane więc nie mogło ich zabraknąć i w naszej pracy tym bardziej, że niektóre z zaproponowanych przez wymienionych autorów modyfikacji okazały się przydatne w prognostycznym algorytmie genetycznym.

De Jong zbadał i dokonał porównania kilku metod selekcji chromosomów. Rozważał następujące metody:
1. Metodę elitarną (*elitist model*).
2. Metodę wartości oczekiwanej (*expected value model*).
3. Metodę elitarną i wartości oczekiwanej (*elitist expected value model*).
4. Metodę współczynnika przepełnienia (*crowding factor model*).

Jako pole testowe posłużyło mu kilka funkcji, które można zaliczyć (oczywiście nie jednocześnie) do następujących typów:
1. Ciągłe/nieciągłe
2. Wypukłe/niewypukłe.
3. Jednomodalne/wielomodalne.
4. Kwadratowe/niekwadratowe.
5. Niskowymiarowe/wielowymiarowe.
6. Deterministyczne/stochastyczne.

Poza powyższymi de Jong rozważał także model uogólnionego krzyżowania, ale nie dotyczył on *sensu stricte* problemu selekcji.

Wszystkie modele porównywał z metodą koła ruletki. Efektywność algorytmów opartych o poszczególne metody selekcji mierzył przy pomocy dwóch miar. Bieżącą efektywność algorytmu de Jong określał jako średnią wartość funkcji przystosowania ze wszystkich prób łącznie z bieżącą. Zbieżność algorytmu[4] oceniał za pomocą wartości przeciętnej (do danej próby włącznie) najlepszych wartości funkcji celu ze wszystkich prób.

Metoda elitarna

W metodzie elitarnej wybieramy i zapamiętujemy najlepszego osobnika z danej populacji. W następnym pokoleniu sprawdzamy czy znalazł się on wśród nowych chromosomów. Jeżeli nie, wówczas dołączamy zapamiętany łańcuch do bieżącej populacji jako jej kolejny element (usuwając przy okazji innego, na przykład najgorszego osobnika, aby zachować niezmienioną wielkość populacji).

[4] Jako kryterium zatrzymania postępowania.

Oczywiście nic nie stoi na przeszkodzie aby do następnego pokolenia przechodziło więcej najlepszych chromosomów z poprzedniej populacji. Należy jednak liczyć się z szybciej postępującym spadkiem nacisku selektywnego. Możemy doprowadzić do rychłego spadku zróżnicowania w dalszych pokoleniach a w konsekwencji do przedwczesnej zbieżności algorytmu.

Metoda ta daje lepsze rezultaty od koła ruletki w przypadku funkcji jednomodalnych. Dla funkcji wielomodalnych klasyczna metoda okazuje się być lepsza z punktu widzenia obu zaproponowanych przez de Jonga miar ponieważ metoda elitarna szybciej znajduje ekstrema lokalne niż globalne. Jest to jednak metoda wciąż wykorzystywana do poszukiwania rozwiązań konkretnych problemów (Gong, Yamazaki, Gen i Xu 1999) lub analiz zachowania algorytmów (Hanne 1999).

Metoda wartości oczekiwanej

Idea tej metody polega na próbie zredukowania stochastycznych błędów powstających podczas losowania za pomocą koła ruletki, dla której obserwujemy znaczącą różnicę między oczekiwaną a faktyczną liczbą potomków danego chromosomu spowodowaną występowaniem dużej wariancji. Przypomnijmy sobie przykład działania klasycznego algorytmu Hollanda. Pojawił się w nim chromosom Ch_{10} o wartości funkcji przystosowania znacznie przewyższającej przystosowanie pozostałych osobników. Trzykrotnie został on wylosowany do reprodukcji, ale jako rodzic pojawił sie tylko raz. Biorąc pod uwagę jego przewagę należało się spodziewać, że uzyska więcej potomków.

Z problemem uzyskiwania dużej liczby potomków przez niektóre, lepiej przystosowane chromosomy mierzy się metoda wartości oczekiwanej. Dla każdej k-tej iteracji oblicza się oczekiwaną liczbę potomków M jako zależność:

$$M = \frac{F(Ch_i)}{\overline{F}} \tag{1.20}$$

gdzie:
$F(Ch_i)$ – wartość funkcji przystosowania dla chromosomu Ch_i
$\overline{F}$ – średnia wartość funkcji przystosowania w populacji

Wartość M wyznacza się zanim przystąpimy do losowania przy pomocy metody koła ruletki. Jeżeli chromosom zostaje wylosowany do reprodukcji oraz krzyżowania M jest zmniejszane o 0,5 zaś gdy wylosowano go zarówno do reprodukcji jak i mutacji M zmniejsza się o 1. W momencie kiedy $M < 0$ nie bierze się już pod uwagę takiego chromosomu.

Metoda wartości oczekiwanej, po uśrednieniu wyników dla wszystkich rozpatrywanych funkcji testowych okazała się lepsza od koła ruletki oraz metody elitarnej.

Metoda elitarna i wartości oczekiwanej

Metoda ta powstaje z połączenia dwóch poprzednio opisanych metod. Po przeprowadzeniu stosownych eksperymentów okazała się ona lepsza od dwóch poprzednich w przypadku funkcji jednomodalnych, ale gorsza od metody wartości oczekiwanej dla funkcji wielomodalnych.

Metoda współczynnika przepełnienia

Metoda ta ma na celu osiągnięcie lepszych rezultatów (w porównaniu do poprzednio opisanych) jeśli chodzi o czas obliczeń i wartość funkcji przystosowania w przypadku funkcji wielomodalnych. Wprowadzamy tzw. współczynnik przepełnienia (CF) o zadanej z góry wartości. W momencie kiedy powstaje nowy osobnik należy z populacji rodziców wykasować jeden chromosom. W tym celu losujemy spośród rodziców podzbiór CF-chromosomów po czym kasujemy tego osobnika, który jest najbardziej podobny do potomka. Zgodnie z oczekiwaniami de Jonga metoda współczynnika przepełnienia rzeczywiście dawała lepsze rezultaty w przypadku funkcji wielomodalnych.

Zachęcona wynikami poprzednika, w 1981 roku A. Brindle zaproponowała własne metody selekcji:

1. Metodę deterministyczną.
2. Metodę stochastyczną na podstawie reszty bez zamiany.
3. Metodę stochastyczną bez zamiany.
4. Metodę stochastyczną na podstawie reszty z zamianą.
5. Metodę stochastyczną z zamianą.
6. Metodę turnieju stochastycznego.

Metoda deterministyczna

W metodzie tej dla każdego chromosomu rodzica obliczamy prawdopodobieństwo wylosowania do reprodukcji oraz oczekiwaną liczbę potomków e_i będącą iloczynem prawdopodobieństwa wylosowania do reprodukcji p_s oraz liczebności populacji:

$$e_i = p_s n \tag{1.21}$$

Każdy chromosom rodzica otrzymuje liczbę potomków równą części całkowitej e_i. Populację potomków uzupełnia się chromosomami rodziców przyjmując jako kryterium najlepsze wartości części ułamkowej e_i dla tychże chromosomów.

Metoda stochastyczna na podstawie reszty bez zamiany

W tej metodzie wyznaczamy oczekiwaną liczbę potomków w sposób analogiczny jak we wzorze (1.21). Jednak część ułamkowa jest traktowana jako prawdopodobieństwo wylosowania do reprodukcji. Populację potomków uzupełniamy chromosomami rodziców przy pomocy ważonego rzutu monetą jako wagi przyjmując ułamkową część e_i.

Metoda stochastyczna bez zamiany

Ta metoda jest analogiczna do metody wartości oczekiwanej de Jonga.

Metoda stochastyczna na podstawie reszty z zamianą

W metodzie tej wyznaczamy wartość e_i w znany nam już sposób. Następnie populację potomków uzupełniamy chromosomami rodziców wylosowanymi metodą koła ruletki, w której koło wyskalowano według części ułamkowej e_i wyznaczonej dla każdego chromosomu.

Metoda stochastyczna z zamianą

To po prostu inna nazwa metody koła ruletki.

Metoda turnieju stochastycznego

Chromosomy rodziców są losowane do reprodukcji jak w klasycznym algorytmie genetycznym. Z pary rodziców wybiera się ten o wyższej wartości funkcji przystosowania i powtarzamy to działanie tyle razy aż osiągniemy nową populację. Możemy jednak zwiększyć liczbę chromosomów biorących udział w turnieju. Wówczas z populacji bieżącej losujemy podpopulację złożoną z ustalonej, ale wyższej niż 2, liczby osobników. Na podstawie wartości przystosowania wyłaniamy spośród nich zwycięzcę, który zostaje skopiowany do populacji potomnej. Metoda turnieju stochastycznego, w którym biorą udział tylko 2 osobniki może być traktowana jako szczególny przypadek liniowej reprodukcji rangowej.

W przypadku zaproponowanych przez Brindle rozwiązań metoda stochastyczna na podstawie reszty z zamianą okazała się szczególnie popularna w późniejszych zastosowaniach.

Selekcja na podstawie rang

Wśród modyfikacji klasycznego procesu selekcji chromosomów do reprodukcji warto wymienić jeszcze takie, które porządkują chromosomy według założonych z góry wag. To oddzielna rodzina metod, w której prawdopodobieństwo reprodukcji osobnika podawane jest na podstawie rangi (Arabas 2004, s. 120) charakteryzującej jego jakość na tle populacji.

Selekcja na podstawie rang ma w założeniu zapobiec sytuacji, w której osobniki mające dużo wyższą w porównaniu z resztą populacji wartość funkcji przystosowania (tzw. „superosobniki") będą uzyskiwać liczne potomstwo. Prowadzi to do ograniczenia liczby potomków pozostałych chromosomów, eliminuje część materiału genetycznego i w konsekwencji prowadzi do przedwczesnej zbieżności w lokalnym optimum. Rangi ustalamy następująco:

1. Osobniki sortujemy nierosnąco według wartości funkcji przystosowania. Jeżeli dwa lub więcej chromosomów charakteryzuje się tym samym przystosowaniem, uzyskują one różne numery na liście.

2. Przydzielamy osobnikom kolejne liczby całkowite zaczynając od zera przy czym chromosomy o jednakowym przystosowaniu otrzymują te same rangi.

Po ustaleniu rang tworzymy zmienną losową, w której każdemu z chromosomów przypisujemy prawdopodobieństwo reprodukcji na podstawie jego rangi. Funkcję określającą owo prawdopodobieństwo dobieramy arbitralnie z zastrzeżeniem jednak, że będzie ona nierosnąca wraz z numerem rangi. Za J. Arabasem podamy dwa przykłady takich funkcji:

$$p_r(\mathbf{X}) = a + k\left(1 - \frac{r(\mathbf{X})}{r_{\max}}\right) \tag{1.22}$$

$$p_r(\mathbf{X}) = a + k(r_{\max} - r(\mathbf{X}))^b \tag{1.23}$$

gdzie:
$r(\mathbf{X})$ – zbiór rang dla osobników;
$r_{\max}$ – najwyższa spośród rang.

Parametry a i k dobieramy tak, aby zachodziło:

$$\sum p_r(\mathbf{X}) = 1 \tag{1.24}$$

$$0 \leqslant p_r(\mathbf{X}) \leqslant 1 \tag{1.25}$$

$$r(\mathbf{X}) \geqslant r(\mathbf{Y}) \Rightarrow p_r(\mathbf{X}) \leqslant p_r(\mathbf{Y}) \tag{1.26}$$

Prawdopodobieństwa przypisane do rang mogą zmieniać się z pokolenia na pokolenie lub pozostać stałe.

Podstawową trudnością jeśli chodzi o reprodukcję rangową jest ustalenie liczby kopii osobnika na podstawie rang. Próbę poradzenia sobie z tym problemem stanowi reprodukcja progowa będąca szczególnym przypadkiem reprodukcji rangowej. Wprowadzamy w niej parametr zwany wskaźnikiem nacisku selektywnego (Arabas 2004, s. 127) o następującej konstrukcji: im niższą przyjmuje wartość tym mniej osobników podlega reprodukcji a zarazem rośnie liczba potomków jednego osobnika. W oparciu o ten wskaźnik ustalamy próg określający prawdopodobieństwo reprodukcji.

Choć za reprodukcją opartą o rangi stoi chęć uniknięcia problemu przedwczesnej zbieżności algorytmu to jednak należy pamiętać, że przy okazji tracimy informacje niesione przez funkcję przystosowania (nie bierzemy pod uwagę względnych ocen różnych chromosomów) oraz zostaje naruszone twierdzenie o schematach.

Wymienione powyżej nieklasyczne metody selekcji różnią się pod względem otrzymywanych wyników w zależności od rodzaju problemu z jakim mają do czynienia co wykazały między innymi właśnie badania de Jonga.

Na przykład metoda elitarna sprawdzała się w przypadku funkcji unimodalnych lepiej niż multimodalnych. W ich przypadku lepsze rezultaty

dawała metoda współczynnika przepełnienia. Jednak nawet w ramach danej funkcji testowej poszczególne metody różnią się jeśli chodzi o wyniki. Generalnie, wprowadzenie modyfikacji do klasycznej metody selekcji okazało się opłacalne gdyż wzrastała efektywność działania algorytmu lub osiągano lepsze rezultaty.

Michalewicz cytuje w swojej książce (Michalewicz 2003, s. 88) zaproponowaną w literaturze przedmiotu ogólną klasyfikację metod selekcji. Według jednego z kryteriów metody owe podzielimy na:

1. Statyczne – prawdopodobieństwa wyboru są jednakowe dla wszystkich pokoleń.
2. Dynamiczne – prawdopodobieństwa wyboru nie są jednakowe dla wszystkich pokoleń.

Inny proponowany podział to:

1. Metody zachowujące – dla każdego osobnika istnieje niezerowe prawdopodobieństwo wyboru;
2. Metody wygaszające – niektóre osobniki charakteryzują się zerową szansą wyboru. Metody te dzielimy dalej na:
   a) lewe – nie dopuszczamy do selekcji najlepszych osobników, aby uniknąć przedwczesnej zbieżności;
   b) prawe – nie dopuszczamy do selekcji najgorzej przystosowanych osobników.

Niektóre procedury selekcji nazwiemy *wyłącznymi*, czyli takimi gdzie rodzice mogą reprodukować tylko w jednym pokoleniu. Można również wyróżnić selekcję *pokoleniową* (zbiór rodziców jest stały do momentu utworzenia wszystkich potomków dla następnego pokolenia) i *w locie* (potomek natychmiast zastępuje rodzica).

Autor ten opisuje również ciekawą zmianę w podejściu do tej procedury. Zaproponował on bowiem, aby zrezygnować z kroku selekcji. W to miejsce wybierać mamy niezależnie pewną liczbę (niekoniecznie różnych) chromosomów do reprodukcji oraz grupę różnych chromosomów do wyginięcia. Wyboru dokonujemy na podstawie względnego dopasowania chromosomów. Łańcuch o przystosowaniu powyżej średniej dla populacji ma większe szanse na reprodukcję. Z kolei gorzej oceniany osobnik ma większe szanse na wyginięcie. Na skutek przeprowadzonego postępowania w populacji znajdą się trzy grupy chromosomów: przeznaczone do reprodukcji, przeznaczone do wyginięcia i pozostałe (zwane neutralnymi).

Operatory genetyczne stosuje się do całych osobników, czyli część rodziców ulegnie krzyżowaniu, część mutacji a reszta na przykład inwersji. Zdaniem Michalewicza zaproponowana modyfikacja zapewnia lepsze wykorzystanie dostępnej pamięci gdyż unikamy pojawiania się wielokrotnych, dokładnych kopii tego samego chromosomu. Skutkiem tego staje się większe zróżnicowanie populacji oraz zwiększenie przestrzeni poszukiwań.

1.3.2. Model uogólnionego krzyżowania

W poprzednim paragrafie wspomnieliśmy, że de Jong w swoich badaniach rozpatrywał również model, który nie dotyczył procesu selekcji chromosomów do reprodukcji. W modelu tym, nazwanym przez autora modelem uogólnionego krzyżowania wprowadzony został parametr określający liczbę punktów krzyżowania (PK). Dla PK równego jeden otrzymujemy krzyżowanie znane z klasycznego algorytmu genetycznego. Może ono doprowadzić do zniszczenia korzystnych z punktu widzenia rozwiązania schematów. Wyobraźmy sobie, że w chromosomach rodziców wystąpiły schematy jak poniżej:

$$Ch_1 = (\ \mathbf{1}\ \ \mathbf{1}\ \ 0\ \ 1\ \ 1\ \ \mathbf{0}\ )$$
$$Ch_2 = (\ 0\ \ 0\ \ \mathbf{0}\ \ \mathbf{1}\ \ 0\ \ 1\ )$$

Potomek, który powstanie z takiej pary nie odziedziczy obu schematów a zatem pewne allele zostaną utracone. Poza tym efekt krzyżowania nie zależy od długości chromosomu jeżeli krzyżowanie dokonuje się tylko w jednym punkcie jak ma to miejsce w przypadku mutacji. W chromosomach składających się z dłuższych łańcuchów rośnie prawdopodobieństwo mutacji coraz większej ilości genów podczas kiedy niezależnie od długości chromosomu krzyżowanie jednopunktowe łączy zawsze dokładnie dwa łańcuchy.

Próbę rozwiązania tego problemu stanowi krzyżowanie dwupunktowe. Losujemy w nim dwa punkty cięcia a następnie wymieniamy między chromosomami jedynie środkowy fragment. Szybko jednak okaże się, że podobnie jak miało to miejsce w przypadku krzyżowania jednopunktowego mogą pojawić się schematy, które także ulegną zniszczeniu. To prowadzi nas do wprowadzenia większej liczby punktów krzyżowania. W krzyżowaniu wielopunktowym należy rozgraniczyć dwa przypadki: kiedy liczba punktów krzyżowania jest parzysta lub nieparzysta (Goldberg 2003, s. 134).

Przypadek parzystej liczby tych punktów zilustrowaliśmy rysunkiem 1.2, w którym przyjęliśmy cztery takie punkty. Analogicznie jak w wariancie jednopunktowym, dochodzi do wymiany genów miedzy chromosomami rodziców. Otrzymujemy dwóch potomków, u których podciągi genów pochodzących z chromosomów rodziców zawierają się między wybranymi punktami krzyżowania.

Na rysunku 1.3 przedstawiliśmy sytuację, w której dochodzi do krzyżowania wielopunktowego z nieparzystą ilością punktów krzyżowania. Zakładamy w tej sytuacji, że dodatkowy punkt przypada na początku łańcucha.

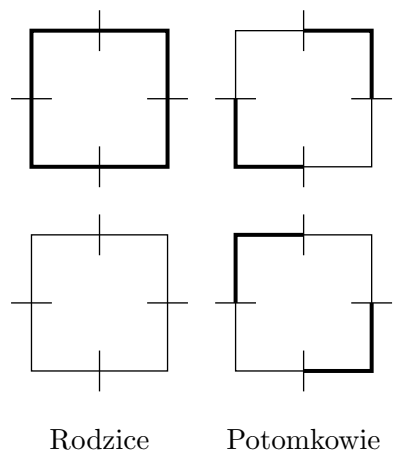
Liczbę punktów krzyżowania możemy ustalić z góry (tak uczynił de Jong) lub potraktować ją jako jeszcze jeden parametr zmieniający się podczas obliczeń przeprowadzanych przez algorytm.

Wprowadzenie więcej niż jednego punktu krzyżowania poprawia efektywność algorytmu tylko dla pewnej klasy zadań. W pozostałych wypadkach efektywność ta spada ze wzrostem liczby punktów krzyżowania. Dzieje się tak, ponieważ struktura chromosomu narażona zostaje na zniszczenie korzystnych dla dalszego przetwarzania schematów. Zetknęliśmy się z tym problemem tego konstruując prognostyczny algorytm genetyczny o czym będzie mowa w trzecim rozdziale.

1.3.3. Wykorzystanie zjawiska nisz i gatunków

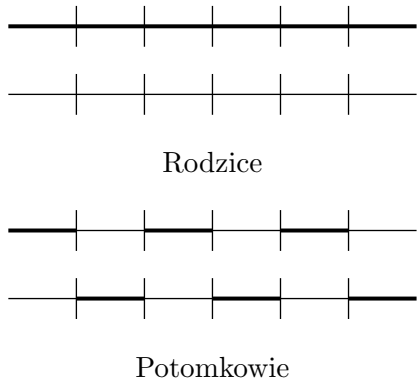
Jak wcześniej podkreśliliśmy, w teorii algorytmów genetycznych (i ewolucyjnych) istnieje tendencja do jak najwierniejszego naśladowania zjawisk i procesów zachodzących w przyrodzie. Sprawiło to, że podjęto próbę przeszczepienia na ten grunt kwestii występowania wyspecjalizowanych gatunków i zajmowanych przez nie nisz ekologicznych. Okazało się bowiem, że odwzorowanie biologicznego podziału organizmów na gatunki może przynieść korzyści algorytmom jako takim.

Włączenie wspomnianych zjawisk przydaje się w przypadku funkcji wielomodalnych gdzie mamy do czynienia z występowaniem lokalnych ekstremów, które traktujemy wówczas jako nisze. Gatunkami stają się się podzbiory chromosomów, których wartość funkcji przystosowania jest zbieżna do tychże ekstremów. Klasyczny algorytm w takiej sytuacji zlokalizuje co najwyżej ekstremum globalne podczas gdy możemy być zainteresowani poznaniem również lokalnych optimów. Zresztą wskazanie najlepszego rozwiązania w całym zbiorze może w tej sytuacji przekraczać możliwości KAG,



Rysunek 1.2: Schemat krzyżowania wielopunktowego przy parzystej liczbie punktów krzyżowania

Źródło: na podst. (Gwiazda 1995)



Rysunek 1.3: Schemat krzyżowania wielopunktowego przy nieparzystej liczbie punktów krzyżowania

Źródło: na podst. (Gwiazda 1995)

który jak wiemy bywa podatny na problem przedwczesnej zbieżności. Jeżeli na przykład znajdziemy się w pobliżu jednego z lokalnych maksimów to z dużą dozą prawdopodobieństwa właśnie ono zostanie przez algorytm potraktowane jako ostateczny wynik.

Z zagadnieniem tworzenia gatunków wiąże się jeszcze kwestia doboru partnerów do krzyżowania. Postępując w sposób klasyczny losujemy rodziców z całej populacji. Tymczasem w przyrodzie potomstwo powstaje tylko ze skojarzenia dwóch osobników tego samego gatunku. Oczywiście możemy przyjąć założenie, że tworzymy populację osobników jednego gatunku, ale stanowi to zbyt duże uproszczenie.

Metoda współczynnika przepełnienia zaproponowana przez de Jonga omówiona w poprzednim podrozdziale jest jednym ze sposobów wykorzystujących zjawisko nisz. Stanowi uogólnienie propozycji Cavicchio, w której potomek trafiał do nowej populacji pod warunkiem, że jego przystosowanie było lepsze niż rodzica. W przeciwnym wypadku usuwamy potomka zaś populacja pozostaje bez zmian. Mechanizm ten nazywamy preselekcją.

Wśród podejmowanych prac nad włączeniem zjawiska podziału na gatunki do algorytmu genetycznego zasługuje na uwagę podejście opracowane przez Goldberga i Richardsona (Goldberg 2003, s. 207).

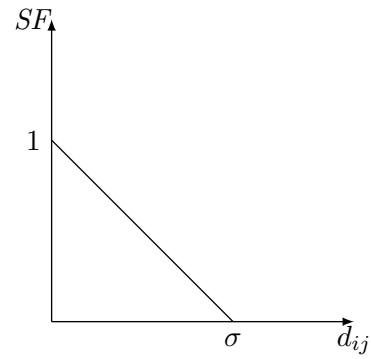
Wprowadzają oni pojęcie tzw. funkcji udziału (*SF – sharing function*). Ma ona określić sąsiedztwo danego chromosomu w populacji. Najprostszą funkcję udziału otrzymujemy stosując liniową zależność w oparciu o odległość między dwoma chromosomami d_{ij} zdefiniowaną wzorem (1.27).

Funkcja udziału przyjmuje wartości bliskie jedności jeśli chromosomy są do siebie bardzo podobne a bliskie zeru dla chromosomów znacząco różniących się od siebie. Selekcji dokonujemy na podstawie zmodyfikowanej

funkcji przystosowania danej wzorem (1.28). Rysunek 1.4 przedstawia graficzną ilustrację liniowej funkcji udziału.

$$d_{ij} = ||Ch_i + Chj|| \tag{1.27}$$

$$F_{SF} = \frac{F(Ch_i)}{\sum_{i=1}^{n} SF(d_{ij})} \tag{1.28}$$



Rysunek 1.4: Liniowa funkcja udziału

Źródło: na podst. (Goldberg 2003)

Występowanie w sąsiedztwie podobnych osobników powoduje wzajemne obniżanie przystosowania ograniczając nadmierny przyrost chromosomów jednego gatunku. Przeprowadzone eksperymenty wykazały, że wykorzystanie efektu tworzenia się nisz i gatunków daje dobre rezultaty dla szeregu zagadnień optymalizacyjnych. Należy jednak uważać, czy nie dochodzi do „migotania" populacji czyli zjawiska polegającego na oscylacji pomiędzy obszarami przyciągania różnych maksimów lokalnych (Arabas 2004, s. 213). Można tego uniknąć jeżeli mianownik z równania (1.28) będzie różny dla różnych osobników z niszy. Modyfikacja funkcji polega na tym, że dla dwóch osobników o identycznym genotypie większą wartość uzyska ten, który jako pierwszy został wzięty pod uwagę.

Istnieje pewien sposób wykorzystania algorytmu genetycznego w ogólnych założeniach przypominający wykorzystanie nisz i gatunków. Naśladując zjawiska znane z przyrody grupujemy zbliżone do siebie chromosomy natomiast tutaj wykorzystujemy algorytm genetyczny do grupowania podobnych obiektów w klastry. Tematyka wykorzystania klastrów zyskała na popularności w ciągu ostatnich kilkunastu lat. Jako przykłady niech posłużą prace Falkenauera ((Falkenauer 1992), (Falkenauer 1998)), które zainspirowały autorów (Agustín-Blas, Salcedo-Sanz, Jiménez-Fernández,

Carro-Calvo, Del Ser i Portilla-Figueras 2012). Poza tym można wymienić m.in. artykuły (Locatelli i Schoen 2003) czy (Wagner 1999).

Do danego klastra trafiają osobniki podobne do siebie z punktu widzenia przyjętego kryterium na przykład zajmujące to samo sąsiedztwo stąd podobieństwo do nisz. W tym przypadku kluczową rolę odgrywa pomiar odległości miedzy obiektami gdyż decyduje on o zaliczeniu do konkretnego klastra. Stosuje się różnego rodzaju metryki odległości (prostokątną, euklidesową czy Mahalonobisa). Rezultaty podziału na klastry (take jak odległości między nimi) różnią się w zależności od przyjętej metody oceny takiej jak suma kwadratów odległości danego wektora od centroidu klastra.

Jak wspomnieliśmy, do tworzenia klastrów można wykorzystać algorytm genetyczny. Oczywiście kodowanie chromosomu i operatory powinny zostać dostosowane do wymogów zadania. Pojedynczy chromosom składa się z dwóch części: pierwsza przechowuje numery klastrów, do których przynależą dane obserwacje; druga identyfikatory klastrów bieżącego rozwiązania. Część pierwsza składa się z ustalonej liczby genów, ale druga nie ma stałej długości. Załóżmy, że w populacji wystąpił chromosom Ch_j o następującej budowie:

$$Ch_j = (\ 1\ \ 3\ \ 2\ \ 1\ \ 4\ \ 1\ \ 1\ \ 2\ \ 3\ \ 2\ \ 1\ \ 3\ \ 4\ \ 2\ \ 1\ \ |\ \ 1\ \ 2\ \ 3\ \ 4\ )$$

Chromosom ten reprezentuje 4 klastry, na które składają się następujące zbiory danych wejściowych:

$$\{x_1, x_4, x_6, x_7, x_{11}, x_{15}\},\ \{x_3, x_8, x_{10}, x_{14}\},\ \{x_2, x_9, x_{12}\},\ \{x_5, x_{13}\}$$

Selekcja chromosomów opiera się na rangach wyznaczonych w trakcie dwuetapowej procedury. Etap pierwszy polega na posortowaniu osobników według ich przystosowania. W etapie drugim najlepszy osobnik uzyskuje rangę $R_j = N$, następny $N-1$ i tak dalej gdzie N jest wielkością populacji.

Dla każdego chromosomu obliczamy funkcję przystosowania wg wzoru:

$$F_j = \frac{2R_j}{N(N+1)} \tag{1.29}$$

Wartości funkcji danej wzorem (1.29) należą do przedziału od 0 do 1. Na ich podstawie skalujemy koło ruletki po czym losujemy chromosomy do krzyżowania.

W pierwszym etapie krzyżowania losujemy parę rodziców i dla każdego z nich po 2 punkty krzyżowania. Załóżmy, że rodzice i wylosowane punkty miały następującą postać (symbol $\bullet$ oznacza punkt krzyżowania):

$$Ch_1^R = (\ 1\ \ 3\ \ 2\ \ 1\ \ 4\ \ 1\ \ 1\ \ 2\ \ 3\ \ 2\ \ 1\ \ 3\ \ 4\ \ 2\ \ 1\ \ |\ \ 1\ \ \bullet\ \ 2\ \ 3\ \ \bullet\ \ 4\ )$$

$$Ch_2^R = (\ 3\ \ 1\ \ 2\ \ 1\ \ 3\ \ 2\ \ 2\ \ 1\ \ 3\ \ 1\ \ 2\ \ 3\ \ 2\ \ 2\ \ 2\ \ |\ \ \bullet\ \ 1\ \ 2\ \ \bullet\ \ 3\ )$$

Wybrane przez punkty krzyżowania geny rodzica 1 są kopiowane do chromosomu potomka:

$$Ch_j = (\ _\ \ 3\ \ 2\ \ _\ \ _\ \ _\ \ _\ \ 2\ \ 3\ \ 2\ \ _\ \ 3\ \ _\ \ 2\ \ _\ \ |\ \ 2\ \ 3\ )$$

Następnie kopiujemy geny 2 rodzica o ile wcześniej nie zostały przypisane przez 1 rodzica:

$$Ch_j = (\ _\ \ 3\ \ 2\ \ 1'\ \ _\ \ 2'\ \ 2'\ \ 2\ \ 3\ \ 2\ \ 2'\ \ 3\ \ _\ \ 2\ \ 2'\ \ |\ \ 2\ \ 3\ \ 1'\ \ 2'\ )$$

Pozostałe geny uzupełniamy losowymi wartościami:

$$Ch_j = (\ 3\ \ 3\ \ 2\ \ 1'\ \ 1'\ \ 2'\ \ 2'\ \ 2\ \ 3\ \ 2\ \ 2'\ \ 3\ \ 3\ \ 2\ \ 2'\ \ |\ \ 2\ \ 3\ \ 1'\ \ 2'\ )$$

Usuwamy ewentualne puste klastry, zmieniamy etykiety klastrów i stosownie do ich wartości zmieniamy geny przypisujące obserwacje do klastrów. Ostatecznie otrzymujemy potomka:

$$Ch_j = (\ 2\ \ 2\ \ 1\ \ 3\ \ 3\ \ 4\ \ 4\ \ 1\ \ 2\ \ 1\ \ 4\ \ 2\ \ 1\ \ 1\ \ 4\ \ |\ \ 1\ \ 2\ \ 3\ \ 4\ )$$

Zależy nam ponadto na tym, aby prawdopodobieństwo krzyżowania było wyższe w początkowych pokoleniach i niższe w końcowych. Osiągniemy to definiując je następująco:

$$p_k^i = p_k^l + \frac{i}{L_G}(p_k^l + p_k^u) \tag{1.30}$$

gdzie:
p_k^i – prawdopodobieństwo krzyżowania w populacji i;
p_k^l, p_k^u – zakładany dolny i górny kraniec przedziału prawdopodobieństwa krzyżowania;
L_G – całkowita liczba pokoleń.

W podobny sposób można obliczyć prawdopodobieństwo mutacji jednak ten operator działa inaczej niż w klasycznym algorytmie. Przede wszystkim występuje w jednej z dwóch odmian. Pierwsza z nich polega na tym, że wylosowany klaster jest dzielony na dwa. Jeden z wygenerowanych klastrów zachowuje swój dotychczasowy identyfikator a drugi dostaje nowy numer. Zmienia się więc długość drugiej części chromosomu. Wybór klastra zależy od jego wielkości, im jest on większy tym wyższe staje się prawdopodobieństwo jego wyboru. Załóżmy, że w uzyskanym poprzednio potomku do mutacji wylosowany został klaster 1. Po mutacji chromosom przyjmie następującą postać:

$$Ch_j = (\ 2\ \ 2\ \ 1\ \ 3\ \ 3\ \ 4\ \ 4\ \ 5\ \ 2\ \ 1\ \ 4\ \ 2\ \ 5\ \ 1\ \ 4\ \ |\ \ 1\ \ 2\ \ 3\ \ 4\ \ 5\ )$$

Drugi wariant mutacji polega na połączeniu dwóch istniejących klastrów. Załóżmy, że do połączenia wybrane zostały klastry 2 i 4:

$$Ch_j = (\ 2\ \ 2\ \ 1\ \ 3\ \ 3\ \ 2\ \ 2\ \ 1\ \ 2\ \ 1\ \ 2\ \ 2\ \ 1\ \ 1\ \ 2\ \ |\ \ 1\ \ 2\ \ 3\ )$$

Wyniki algorytmu da się poprawić stosując dodatkowo elitarny model selekcji. Jako, że grupujemy obserwacje w klastry przydatny okaże się również tzw. model wyspowy czyli podział chromosomów na podpopulacje, które ewoluują niezależnie od siebie. Jeżeli połączymy go z selekcją elitarną wówczas dopuszczamy migrację najlepszego osobnika między „wyspami".

1.4. Przegląd modyfikacji zmieniających postać funkcji przystosowania

1.4.1. Skalowanie funkcji przystosowania

Algorytmy genetyczne, zwłaszcza w swojej klasycznej wersji mają tendencję do przedwczesnej zbieżności do optimum lokalnego. Problem ten wynika z faktu, że po pewnej liczbie iteracji w populacji zaczynają przeważać chromosomy o ponadprzeciętnej wartości funkcji przystosowania, nawet jeśli nie jest to jeszcze wartość optymalna. Osobniki gorzej przystosowane nie wytrzymują konkurencji i szybko odpadają. Ich miejsce zajmuje coraz większa liczba potomków chromosomów, dla których funkcja przystosowania kształtuje się powyżej średniej dla populacji.

O ile w końcowej fazie działania algorytmu jest to zjawisko pożądane, o tyle spadek zróżnicowania wśród osobników tworzących populację stanowi przeszkodę na początku poszukiwań rozwiązania. Szczególnie odczuwa się to wówczas kiedy oprócz optimum globalnego występują jeszcze optima lokalne. Algorytm może wskazać jedno z nich jako rozwiązanie i zakończyć działanie.

Kłopoty ze zbieżnością są znane praktycznie od początku historii algorytmów genetycznych. Tyle samo czasu poszukuje się rozwiązania tego problemu. W pracy (Takahashi 1998) znajdziemy próbę podsumowania zagadnień związanych z osiąganiem zbieżności jednak autor bardzo mocno uprościł budowę chromosomu, który ograniczył do dwóch genów. Zrezygnował także z operatora mutacji. Dochodzi do podobnych wniosków jak inni badacze a uproszczony model pozwolił mu wyprowadzić matematyczne dowody stawianych tez.

Niektóre modyfikacje opisane w poprzednich rozdziałach powstały między innymi pod kątem rozwiązania problemu spadku zróżnicowania osobników. Istnieje wszakże jeszcze jedno podejście, które zasługuje na osobne

omówienie, a jest nim kwestia skalowania funkcji przystosowania. Na podstawie (Goldberg 2003, s. 139) wymienimy i krótko omówimy trzy typy skalowania funkcji:
1. Skalowanie liniowe.
2. Obcinanie na poziomie odchylenia standardowego[5].
3. Skalowanie potęgowe.

Skalowanie liniowe

Funkcję przystosowania F modyfikujemy następująco:

$$F^* = aF + b \tag{1.31}$$

Powyższa transformacja wymaga określenia parametrów a i b równania (1.31). Powinny one zostać dobrane tak, by średnia wartość nowej funkcji przystosowania równała się średniej wartości funkcji starej zaś maksymalna wartość F^* była krotnością F. Chromosomy o średniej wartości F^* otrzymają jednego potomka w kolejnej populacji, zaś te lepsze otrzymają potomków w liczbie odpowiadającej wielokrotności średniej. Poza tym maksymalna wartość F^* powinna być krotnością (najczęściej przyjmuje się dwukrotność) średniej wartości F. Parametry a i b nie zależą zwykle od zadania i obowiązują w całej populacji.

Jeżeli wspomniane warunki zostaną spełnione wtedy chromosom o wartości przystosowania równej średniej dla populacji otrzymuje przeciętnie jednego potomka zaś chromosom o przystosowaniu powyżej średniej otrzyma tylu potomków ile wynosi przyjęta krotność. Skalowanie liniowe nadaje się do zastosowania w zadaniach gdzie nie występują ujemne wartości funkcji przystosowania.

Obcinanie na poziomie odchylenia standardowego

Obcinanie na poziomie odchylenia standardowego stanowi rozszerzenie modelu skalowania liniowego. Zmodyfikowana funkcja przystosowania ma tu postać:

$$F^* = F - \left(\overline{F} - c\sigma\right) \tag{1.32}$$

gdzie:
σ – odchylenie standardowe funkcji przystosowania obliczone dla całej populacji

Jako stałą c przyjmuje się niewielką liczbę całkowitą, zwykle z przedziału[6] $\langle 1{,}3 \rangle$. Zamiast ujemnej wartości F^* przyjmujemy zero.

[5] Nazwa za Michalewiczem (Michalewicz 2003)
[6] Michalewicz proponuje przyjęcie przedziału $\langle 1{,}5 \rangle$

Skalowanie potęgowe

W tej metodzie po zmodyfikowaniu funkcja przystosowania ma postać:

$$F^* = F^k \tag{1.33}$$

Parametr k przyjmuje wartości bliskie 1. Autor tego rozwiązania, Gillies zaleca $k = 1{,}005$ choć w zależności od rozpatrywanego problemu może okazać się konieczne przyjęcie innej wartości.

Baker zaproponował, aby o liczbie potomków chromosomu decydowała nadana mu ranga. Rangowanie miałoby się odbywać na podstawie wartości funkcji celu. Im wyższe przystosowanie tym wyższą pozycję w rankingu zajmuje osobnik. Chromosomy o najlepszych wynikach otrzymują większą ilość potomków. Przeprowadzone eksperymenty nie potwierdziły jednak przewagi tej metody nad innymi.

1.4.2. Algorytmy ze zmienną liczebnością populacji

Jednym z najważniejszych parametrów jakie ustalamy projektując algorytm jest liczba chromosomów tworzących populację. W klasycznym podejściu wielkość populacji pozostaje stała przez cały proces obliczeniowy. Naciskiem selektywnym steruje zróżnicowanie, za które odpowiada m.in. prawdopodobieństwo krzyżowania. Włączenie do algorytmu selekcji elitarnej powoduje pojawienie się osobników potencjalnie nieśmiertelnych w tym sensie, że dotrwają one do ostatniego pokolenia w niezmienionej formie.

Z punktu widzenia przyrody stanowi to stanowczo zbyt sztywne założenie. W naturze liczba osobników składających się na daną populację zmienia się w zależności od warunków otoczenia. Z kolei z punktu widzenia mechanizmów algorytmu, przyjęcie zbyt małej liczebności sprawia, że problemem staje się nadmiernie szybkie zbieganie się algorytmu. Kiedy zaś zakładamy dużą ilość osobników w populacji – wydłuża się czas jaki musi upłynąć do znalezienia rozwiązania. Liczebność wpływa także na różnorodność chromosomów tworzących populację.

Wielkość populacji możemy regulować różnymi sposobami, na przykład przy pomocy oddzielnego algorytmu genetycznego. Jednak najbliższe naturalnym, ewolucyjnym zachowaniom występującym w przyrodzie jest podejście polegające na wprowadzeniu wieku poszczególnych chromosomów czyli ustalenie liczby pokoleń, po których osobnika usuwamy z populacji jak to zaproponowano w (Michalewicz 2003, s. 101). Przyjęcie dłuższego czasu życia równa się bardziej stabilnemu zachowaniu algorytmu. Niższa wartość tego parametru skutkuje wzrostem zdolności adaptacyjnych algorytmu.

W algorytmie o zmiennej liczebności populacji tworzymy z potomków populację pomocniczą. Jej liczebność jest proporcjonalna do całkowitej liczebności populacji. O selekcji decyduje wiek chromosomu zaś chromosomy

do reprodukcji wybieramy z jednakowym prawdopodobieństwem niezależnie od wartości funkcji przystosowania. Krzyżowanie i mutacja dokonywane są w klasyczny sposób.

Czas życia ustalony dla chromosomu obowiązuje aż do jego śmierci następującej wtedy, kiedy liczba pokoleń, przez które pozostawał żywy przekroczy założoną długość życia. Wówczas usuwamy osobnika z populacji. Liczebność populacji (LP) w $k+1$ iteracji wynosi:

$$LP_{k+1} = LP_k + LPP_k - Z_k \tag{1.34}$$

gdzie:
LPP_k – liczebność populacji pomocniczej
Z_k – liczba chromosomów wyeliminowanych w pokoleniu k

Pozostaje kwestia sposobu określenia parametru wyrażającego długość życia chromosomu. Przyjęcie go na stałym poziomie, obowiązującym przez cały czas działania algorytmu nie jest zbyt korzystne, ponieważ liczebność populacji wzrasta wtedy wykładniczo, a nadmierny wzrost liczebności populacji wydłuża czas działania algorytmu. Z tego powodu proponuje się, aby czas życia ustalać jedną z poniższych metod:

1. Metoda przyporządkowania proporcjonalnego.
2. Metoda przyporządkowania liniowego.
3. Metoda przyporządkowania biliniowego.

W pierwszej z nich, czas życia ustala się proporcjonalnie do wartości funkcji przystosowania. Nie można w niej jednak porównywać tych wartości z wartościami chromosomów z poprzednich iteracji.

Jako alternatywę dla przyporządkowania proporcjonalnego proponuje się metodę przyporządkowania liniowego gdzie czas życia ustala się w zależności od najlepszej wartości funkcji przystosowania. Jednak jeśli większa ilość chromosomów ma przystosowanie bliskie maksymalnemu wtedy dostają one długie czasy życia co w konsekwencji prowadzi do znacznego wzrostu liczebności populacji.

Metoda przyporządkowania biliniowego stoi w pewnym sensie w połowie drogi między poprzednio wspomnianymi. Osobniki o najlepszych wartościach funkcji przystosowania otrzymują bardziej zróżnicowane długości czasu życia w oparciu o średnie, najsłabsze oraz najlepsze dopasowanie.

Wybór między stałą a zmienną liczebnością populacji zależy od badacza oraz rozpatrywanego problemu. Zwykle jednak zmienna liczebność populacji oznacza wyższe koszty (jeśli chodzi o czas obliczeń) choć potrafi dawać lepsze wyniki. Bardziej szczegółowe rozważania na temat obliczania długości czasu życia oraz związanych z tym konsekwencji można znaleźć w (Arabas 2004).

1.4.3. Zadania z ograniczeniami

Modele optymalizacyjne zawierają zwykle oprócz funkcji celu także ograniczenia wynikające z treści rozpatrywanego problemu. Charakteryzują one wpływ otoczenia na podejmowaną decyzję, a przedstawiane są jako zestaw nierówności i równań ograniczających zbiór rozwiązań. Algorytm genetyczny w swojej klasycznej formie nie uwzględnia jednak poza funkcją przystosowania dodatkowych warunków. Ich włączenie wymaga wprowadzenia daleko idących zmian.

Pierwsza modyfikacja wiąże się z podejściem do ograniczeń jako takich. O ile w programowaniu liniowym pożądane jest występowanie ograniczeń jako równań, o tyle z punktu widzenia algorytmu genetycznego preferuje się nierówności. Załóżmy, że mamy ograniczenia w postaci równań, które da się zapisać następująco:

$$\mathbf{Ax} = \mathbf{b} \tag{1.35}$$

Założymy również, że istnieje wśród nich p równań niezależnych a więc p zmiennych możemy przedstawić za pomocą pozostałych $x_{i1}, \ldots, x_{ip}$ zmiennych (da się je tym samym wyeliminować z układu równań).

Ograniczenia zapiszmy jak poniżej:

$$\mathbf{l} \leqslant \mathbf{x} \leqslant \mathbf{u} \tag{1.36}$$

Nierówności zaś oznaczymy:

$$\mathbf{Cx} \leqslant \mathbf{d} \tag{1.37}$$

Macierz $\mathbf{A}$ dzielimy pionowo na dwie macierze $\mathbf{A}_1$ i $\mathbf{A}_2$ w ten sposób, że j-ta kolumna należy do macierzy $\mathbf{A}_1$ jeśli $j \in (i_1, \ldots, i_p)$. Zajdzie wtedy (Gwiazda 1995):

$$\mathbf{A}_1\mathbf{x}_1 + \mathbf{A}_2\mathbf{x}_2 = \mathbf{b} \tag{1.38}$$

$$\mathbf{x}_1 = \mathbf{A}_1^{-1}\mathbf{b} - \mathbf{A}_2^{-1}\mathbf{x}_2 \tag{1.39}$$

Analogicznie postępujemy z ograniczeniami i nierównościami. W efekcie tego powstają poniższe zależności:

$$\mathbf{l}_1 \leqslant \mathbf{A}_1^{-1}\mathbf{b} - \mathbf{A}_2\mathbf{x}_2 \leqslant \mathbf{u}_1 \tag{1.40}$$

$$\mathbf{C}_1\mathbf{x}_1 + \mathbf{C}_2\mathbf{x}_2 \leqslant \mathbf{d} \tag{1.41}$$

Po odpowiednich przekształceniach ograniczenia są zdefiniowane układem 1.42) – (1.44), który definiuje wypukłą przestrzeń rozwiązań dopuszczalnych.

$$\mathbf{l}_2 \leqslant \mathbf{x}_2 \leqslant \mathbf{u}_2 \tag{1.42}$$

$$\mathbf{A}_1\mathbf{l}_1 \leqslant \mathbf{b} - \mathbf{A}_2\mathbf{x}_2 \leqslant \mathbf{A}_1\mathbf{u}_1 \tag{1.43}$$

$$(\mathbf{C}_2 - \mathbf{C}_1\mathbf{A}_1^{-1}\mathbf{A}_2)\mathbf{x}_2 \leqslant \mathbf{d} - \mathbf{C}_1\mathbf{A}_1^{-1}\mathbf{b} \tag{1.44}$$

Najprostszym sposobem postępowania jest po prostu odrzucenie chromosomów, które nie spełniają nałożonych ograniczeń. Ich miejsce zajęłyby osobniki odpowiadające warunkom zadania. Zwykle jednak zakładamy stałą liczebność populacji co doprowadzić może do znacznego wydłużenia czasu obliczeń ponieważ musimy podjąć dodatkowy wysiłek znalezienia stosownej liczby pasujących chromosomów. W przypadku zmiennej liczebności populacji również trzeba liczyć się z negatywnymi skutkami działania polegającego na pozbywaniu się niektórych chromosomów. Znaczna redukcja populacji poprzez usunięcie części osobników może skończyć się przedwczesną zbieżnością algorytmu. Mniejsze populacje łatwiej zapełniają się dużą ilością potomków osobników o przystosowaniu powyżej średniej.

W literaturze znaleźć można inne metody postępowania z ograniczeniami. Część z nich polega na zmodyfikowaniu funkcji przystosowania przez wprowadzenie kary za niespełnienie nałożonych warunków. Funkcja przystosowania ma wtedy postać:

$$FP^* = FP + FK \tag{1.45}$$

FK oznacza funkcję kary, która przyjmuje wartość zero jeśli ograniczenia są spełnione. W przeciwnym wypadku wartość funkcji przystosowania będzie powiększona właśnie o FK. Poniżej przedstawimy kilka metod ustalenia kary za niespełnienie ograniczeń.

Jedna z nich zakłada, że kary zależą od l tzw. poziomów naruszenia, które określa się dla każdego z m ograniczeń. Im wyższe naruszenie tym wyższa kara. Osobniki z populacji ocenia się na podstawie następującej funkcji przystosowania:

$$FP = \sum_{j=1}^{m} K_{ij} FP_j^2 \tag{1.46}$$

Niestety, w tej metodzie należy określić $m(2l+1)$ parametrów.

Joines i Houck (Joines i Houck 1994) zaproponowali następującą postać funkcji:

$$FP^* = FP + (Ck)^{\alpha} \sum_{j=1}^{m} FP_j^{\beta} \tag{1.47}$$

W tej metodzie podaje się wartości parametrów C,α,β (k oznacza numer iteracji). Zwiększanie kary o pewien mnożnik powoduje, że pod koniec procesu przyjmuje on duże wartości.

Inne podejście przyjęli Schoenauer i Xanthakis (Schoenauer i Xanthakis 1993). Porządkujemy w nim ograniczenia i kolejno je przetwarzamy. Populację poddajemy tym operacjom do momentu osiągnięcia tzw. progu uporządkowania (φ) czyli ustalonego procentu populacji spełniającego dane

ograniczenie. Potem przechodzimy do kolejnego ograniczenia i powtarzamy procedurę. W ostatnim kroku optymalizujemy funkcję przystosowania bez uwzględniania kary. Trzeba jednak w tym miejscu podkreślić, że przeprowadzone eksperymenty wykazały, iż różne uporządkowanie ograniczeń prowadzi do różnych wyników.

W metodzie zaproponowanej z kolei w (Powell i Skolnick 1993) funkcję przystosowania modyfikuje się następująco:

$$FP^* = FP + r\sum_{j=1}^{m} FP_j + \lambda(k) \tag{1.48}$$

We wzorze (1.48) r to pewna stała, zaś $\lambda(k)$ gdzie k to numer iteracji służący do rozróżniania dopuszczalnych i niedopuszczalnych chromosomów przy pomocy specjalnego algorytmu. Dla osobników niedopuszczalnych powiększa się karę tak, aby ich wartość nie przekroczyła wartości funkcji przystosowania najgorszego chromosomu dopuszczalnego.

Niezależnie od postaci funkcji kary należy wszakże pamiętać, że stosowanie jej może doprowadzić do stworzenia populacji składającej się wyłącznie z osobników niespełniających ograniczeń.

Dodatkowym problem pojawia się kiedy ograniczenia mają nieliniowy charakter. Gran i Aral zaproponowali progresywny algorytm genetyczny (Gran i Aral 1999), który radzi sobie zarówno z równaniami jak i nierównościami. Wyprowadzili oni wzory pozwalające wyznaczyć otoczenie punktu, które spełnia nałożone ograniczenia. Sam progresywny algorytm genetyczny jest procedurą dwuetapową. Najpierw przybliżamy problem używając ograniczeń sprowadzonych do postaci liniowej. W drugim etapie algorytm poszukuje lokalnego optimum w wyznaczonej części zbioru rozwiązań spełniającej nałożone warunki.

Oprócz metod bazujących na wariantach funkcji kary istnieje także metoda algorytmów naprawczych, w której chromosom niespełniający ograniczeń przekształca się do takiej postaci, która te ograniczenia spełnia. Z tym podejściem wiąże się niebezpieczeństwo tego, że w bardziej rozbudowanych problemach algorytm zamiast poszukiwać optymalnego rozwiązania będzie wciąż poprawiał nieprawidłowe osobniki.

1.4.4. Zastosowanie algorytmów genetycznych do optymalizacji zadań wielokryteriowych

Algorytmy genetyczne pierwotnie opracowane zostały jako metody optymalizacji jednokyteriowej. W niektórych problemach optymalizacyjnych należy poszukiwać jednak rozwiązania, które musi spełniać więcej niż jedno kryterium. Po wprowadzeniu odpowiednich modyfikacji również i na tym obszarze da się zastosować algorytmy genetyczne. Zwykle realizowane cele

są na tyle odmienne, że należy je wyrazić przy pomocy różnych funkcji celu. W takiej sytuacji klasyczne metody optymalizacyjne jak również KAG, nie mogą zostać wykorzystane gdyż opierają się tylko na jednym kryterium wartości największej lub najmniejszej.

Konflikt występujący pomiędzy celami, które chcemy osiągnąć potrafi spowodować, że poprawiając wartości z punktu widzenia jednego kryterium pogarszamy inne. Postępowanie takie da się przedstawić w kategoriach rozwiązań dominujących i zdominowanych (Trzaskalik 2008, s. 189). Rzadko udaje się osiągnąć rozwiązanie lepsze od wszystkich pozostałych z punktu widzenia wszystkich kryteriów jednocześnie. Zwykle zadowalamy się rozwiązaniem niezdominowanym, dla którego nie da się poprawić wartości żadnego kryterium nie obniżając wartości innego.

O jakości rozwiązania zadań z kilkoma kryteriami decyduje pojęcie optymalności w sensie Pareto. W myśl tego podejścia rozwiązanie uznajemy za zdominowane jeżeli dla wszystkich celów $f_i(i = 1, \ldots, k)$ zachodzi:

$$f_i(\mathbf{x}) \leqslant f_i(\mathbf{y}), 1 \leqslant i \leqslant k \quad \text{dla maksymalizacji}$$
$$f_i(\mathbf{x}) \geqslant f_i(\mathbf{y}), 1 \leqslant i \leqslant k \quad \text{dla minimalizacji}$$

Mówimy wtedy, że wektor $\mathbf{x}$ jest częściowo mniejszy od wektora $\mathbf{y}$. Jeżeli rozwiązanie nie jest zdominowane przez żadne inne rozwiązanie dopuszczalne wtedy powiemy, że mamy do czynienia z rozwiązaniem niezdominowanym. Wektor niezdominowany stanowi rozwiązanie optymalne w sensie Pareto (paretooptymalne).

Najprostszą i najpopularniejszą w przypadku algorytmów genetycznych metodę optymalizacji wielokryteriowej znajdziemy w (Schaffer 1984). Metoda zaproponowana przez Schaffera wymaga, aby chromosomy populacji podzielić na podpopulacje, których tworzy się tyle ile jest kryteriów optymalizacji zgodnie z zasadą mówiącą, że jedna podpopulacja odpowiada za jeden cel. Osobniki do reprodukcji wybiera się oddzielnie dla każdej podpopulacji. Krzyżowanie odbywa się jednak na chromosomach pochodzących ze wszystkich podpopulacji.

Wadą tej metody okazał fakt eliminowania rozwiązań średnich czyli takich, które spełniały wszystkie kryteria. Do następnych pokoleń przechodziły chromosomy najlepsze z punktu widzenia poszczególnych kryteriów, niekoniecznie zgodne z pozostałymi celami.

Próbę poradzenia sobie z tym problemem zaproponował Goldberg. Selekcji do reprodukcji dokonuje się na podstawie rankingu. Wszystkie niezdominowane chromosomy otrzymują w nim pozycję pierwszą. Spośród pozostałych (czyli po pominięciu tych, które znalazły się na szczycie rankingu) znowu wybiera się chromosomy niezdominowane i nadaje im pozycję drugą. Postępujemy w ten sposób aż do wyczerpania populacji. Większą

różnorodność populacji może zapewnić użycie opisanych wcześniej metod wykorzystujących zjawisko nisz i gatunków.

Inny przykład optymalizacji w sensie Pareto wykorzystujący algorytm ewolucyjny znajdziemy w pracy (Ducange, Lazzerini i Marcelloni 2010) gdzie autorzy poszukiwali optimum dla przypadku trzech kryteriów wykorzystując optymalizację rozmytą.

Ciekawy sposób wykorzystania algorytmu genetycznego do optymalizacji wielokryteriowej znajdziemy w (Aiello, La Scalia i Enea 2012). Autorzy poszukiwali rozwiązania problemu rozlokowania działów, które zajmują znaną powierzchnię na terenie firmy w taki sposób, aby z jednej strony zminimalizować koszty ich funkcjonowania a z drugiej osiągnąć maksymalną efektywność pracy. Całkowitą dostępną powierzchnię dzielą na prostokąty różnej wielkości. Stosują przy tym osobniki diploidalne i rozpatrują cztery warianty funkcji przystosowania w tym takie, które biorą pod uwagę zajmowanie przez jeden dział kilku obszarów obok siebie jak i takie, w których jeden dział rozlokowany zostaje na kilku oddzielnych obszarach.

Rozważania na temat zbieżności użycia algorytmów do problemów charakteryzujących się więcej niż jednym celem znajdziemy też w (Hanne 1999).

1.5. Strategie ewolucyjne

W rozważaniach, które znalazły się w rozdziałach trzecim i czwartym ważną rolę odgrywa strategia ewolucyjna (1+1), która również zalicza się rodziny metod naśladujących zachowania spotykane w naturze podobnie jak algorytmy genetyczne. Dlatego w niniejszym rozdziale nie może zabraknąć charakterystyki strategii ewolucyjnych.

Strategie ewolucyjne wzięły swój początek od prac Rechenberga i Schwefela (Schwefel w (Michalewicz 2003)), którzy rozważali problem optymalizacji urządzeń mechanicznych drogą permutacji rozwiązania początkowego (Arabas 2004, s. 18). Uzyskane wyniki zachęciły ich do uogólnienia metody na przypadek jednoczesnego przetwarzania wielu rozwiązań.

Najprostszy wariant to tzw. strategia ewolucyjna (1+1) czyli SE-(1+1). Populacja składa się w niej tylko z jednego osobnika, który podlega działaniu tylko jednego operatora – mutacji. Po przeprowadzeniu mutacji porównujemy starego i nowego osobnika[7] na podstawie wartości funkcji przystosowania. Do kolejnej iteracji postępowania trafia lepszy z dwóch chromosomów. Sama mutacja różni się jednak od tej znanej z algorytmu genetycznego. Do

[7] Stąd wzięła się nazwa metody, ponieważ jednocześnie występują w niej dwa chromosomy.

każdego genu dodajemy losową modyfikację o określonym rozkładzie. Niech Ch_t oznacza osobnika przed zaś Ch_{t+1} po mutacji:

$$Ch_{t+1} = Ch_t + \varepsilon_t \tag{1.49}$$

gdzie ε_t jest zmienną losową o rozkładzie $N(0, \sigma)$.

Wartość σ wyznacza zasięg mutacji (Arabas 2004, s. 84). Jej podwyższenie wywołuje większe zmiany mutowanych genów. W przypadku funkcji wielomodalnych warto początkowo przyjmować wyższe σ aby szybciej zlokalizować optimum globalne. Pod koniec postępowania opłaca się zmniejszyć zasięg mutacji, żeby poprawić zbieżność.

Jak widać, należy pogodzić dwa przeciwstawne dążenia ustalając wartość σ, która nie ulega zmianie w czasie trwania całego procesu obliczeniowego. Rachenberg zaproponował tak zwaną *regułę 1/5 sukcesów* (Rachenberg w (Michalewicz 2003, s. 194)): *Stosunek φ udanych mutacji do wszystkich mutacji powinien wynosić 1/5. Jeżeli φ jest większe od 1/5 to zwiększ wariancję operatora mutacji, w odwrotnym przypadku zmniejsz ją.* Regułę tę stosować należy co k pokoleń i formalnie można ją zapisać:

$$\sigma_{t+1} = \begin{cases} c_d\sigma_t, & \text{dla } \varphi(k) < 1/5 \\ c_i\sigma_t, & \text{dla } \varphi(k) > 1/5 \\ \sigma_t, & \text{dla } p_s(k) = 1/5 \end{cases} \tag{1.50}$$

gdzie:
$\varphi(k)$ – współczynnik sukcesów operatora mutacji w poprzednich k pokoleniach;
$c_i > 1$, $c_d < 1$ – parametry sterujące szybkością zmian wariancji[8]

Dalszy rozwój strategii ewolucyjnych doprowadził do wykształcenia strategii $(\mu+\lambda)$ oraz (μ,λ). Pierwsza z nich stanowi rozszerzenie strategii (1+1). Występuje w niej μ osobników tworzących λ potomków. Tymczasowa populacja złożona z $\mu + \lambda$ chromosomów jest zmniejszana do μ najlepszych osobników przez operator selekcji.

W strategii (μ,λ) nową populację μ osobników tworzymy tylko spośród λ potomków. Życie każdego osobnika trwa jedno pokolenie a strategia ta lepiej sprawdza się w zadaniach, w których optimum zmienia się w czasie lub funkcja kryterium jest zaszumiona.

W strategiach $(\mu + \lambda)$ i (μ,λ) nie stosuje się reguły 1/5 sukcesów zaś parametr σ podlega procesowi ewolucji. Co więcej, oprócz mutacji w strategiach tych stosuje się operator krzyżowania zwanego tu rekombinacją. Może ona przyjąć jedną z dwóch form:

[8] Schwefel użył następujących wartości: $c_d = 0{,}82$ oraz $c_i = 1/0{,}82 = 1{,}22$.

— dyskretną – każdy gen (składnik chromosomu) pochodzi od pierwszego lub drugiego rodzica;
— pośrednie – geny potomków oblicza się jako średnie wartości genów pochodzących od rodziców.

Zmianie uległa też mutacja, która swoim zasięgiem objęła również parametr σ, który zmienia swoją wartość w odróżnieniu od SE-(1+1). Włączenie rekombinacji oraz zmiany w mutacji mają na celu umożliwienie samoczynnej adaptacji zasięgu mutacji. Nie potrzebujemy więc już reguły 1/5 sukcesów. Osiągany przez wprowadzone modyfikacje efekt adaptacji jest konsekwencją premiowania lepszych osobników.

Strategie ewolucyjne sprawdzają się szczególnie dobrze w zadaniach numerycznych. Jak niebawem wszakże udowodnimy, w połączeniu z innymi metodami optymalizacji przydają się również do rozwiązywania innych typów problemów. Okazały się użyteczne w przypadku korekty zestawu metod wykorzystywanych przez prognostyczny algorytm genetyczny.

1.6. Algorytm genetyczny jako element algorytmów ewolucyjnych

Jak wspomnieliśmy na początku tego rozdziału, algorytmy genetyczne zaliczają się do szerszej grupy zwanej algorytmami ewolucyjnymi podobnie jak strategie ewolucyjne i programowanie ewolucyjne. Wspólnym mianownikiem wszystkich wymienionych metod jest chęć odwzorowania (dokładniej rzecz ujmując symulowania) procesów obserwowanych podczas ewolucji.

Pewne elementy są dla nich wspólne. Po pierwsze w każdej z metod rozpoczynamy od losowej populacji osobników, którą poddajemy działaniu selekcji oraz określonych operatorów po to aby stopniowo poprawiać otrzymywane rozwiązania. W (Rutkowska, Piliński i Rutkowski 1999, s. 210) autorzy podają sposób kodowania (binarne a zmiennoprzecinkowe) jako różnicę między algorytmami genetycznymi a pozostałymi metodami. Jednak takie podejście wydaje się zbyt radykalne. Jak wykazaliśmy wcześniej, możliwe jest wykorzystanie kodowania rzeczywistego w algorytmie genetycznym przy odpowiedniej modyfikacji operatorów.

To właśnie sposób postępowania z populacją i operatorami najbardziej różni wymienione metody. W strategiach ewolucyjnych tworzymy populację pośrednią, na którą składają się wszyscy rodzice oraz pewna liczba potomków po czym redukujemy ją do liczebności populacji bazowej przez eliminację najgorzej przystosowanych osobników. Tak powstaje kolejne pokolenie.

Podczas selekcji w algorytmie genetycznym najlepsze osobniki mogą zostać wybrane kilkakrotnie podczas gdy w strategiach ewolucyjnych nie dochodzi do takiej sytuacji. Wybieramy w nich określoną liczbę najlepiej

przystosowanych osobników bez powtórzeń. Wynika to z faktu, że w algorytmie genetycznym selekcja ma charakter losowy. Odwrotna jest również kolejność operatorów (Michalewicz 2003, s. 198). W strategii ewolucyjnej para rodziców pozwala utworzyć potomka, którego poddajemy mutacji po czym włączamy w skład populacji pośredniej, której liczebność w końcu redukujemy. W algorytmie genetycznym najpierw wybieramy populację pośrednią a krzyżowanie i mutację stosujemy tylko do niektórych osobników.

Prawdopodobieństwa krzyżowania i mutacji są w algorytmie genetycznym z góry ustalone i niezmienne. Z kolei w strategiach ewolucyjnych podlegają one adaptacji stosownie do potrzeb postępowania tak jak na przykład adaptujące się prawdopodobieństwo mutacji opisane w (Mascheck 2010).

Ta praca jest ciekawa również o tyle, że operator mutacji jest tam jedynym użytym, ale autor cały czas odwołuje się do algorytmów genetycznych. Algorytm genetyczny i strategia ewolucyjna odmiennie poza tym podchodzą do kwestii występowania ograniczeń w problemach.

W programowaniu ewolucyjnym używa się jedynie operatora mutacji zaś selekcji dokonuje się metodą turniejową (Michalewicz 2003, s. 201). Wartości dopasowania uzyskujemy poprzez skalowanie funkcji celu.

Ciekawym pomysłem jest też ekonomiczny algorytmu genetyczny (zaproponowany m.in. w (Friedman 1998) lub (Riechmann 2001)), w którym funkcję przystosowania utożsamiamy z wypłatą wygranej w grze ewolucyjnej spełniającej następujące warunki:

1. Strategie o wyższej wypłacie wypierają te o niższej.
2. Występuje inercja.
3. Gracze nie wpływają na swoje przyszłe działania w świadomy sposób.

Gra ewolucyjna jest dynamicznym modelem opisującym interakcje zachodzące między poszczególnymi strategiami. Z tego powodu przystosowanie osobnika (zwanego też agentem) zależy nie tylko od przyjętej przez niego strategii, ale także od strategii innych agentów.

Lata badań i rozwijania różnych podejść do symulacji ewolucji biologicznej zaowocowały zapożyczeniami, które stopniowo zacierają dotychczasowe różnice. Jako przykład niech posłużą kodowanie rzeczywiste, sterowanie parametrami algorytmu genetycznego przy pomocy innego sub-algorytmu czy wprowadzenie do strategii ewolucyjnej krzyżowania (wczesne jej wersje korzystały tylko z mutacji). Należy się chyba spodziewać dalszego przenikania algorytmów ewolucyjnych co powinno dać jeszcze większą elastyczność ich zastosowań.

Rozdział 2

Prognozowanie – własności i metody

Znane przysłowie mówi, że „starego psa nie da się nauczyć nowych sztuczek". Takie też wrażenie można odnieść przeglądając książki i inne opracowania na temat prognozowania szeregów czasowych. Praktycznie każda z nich zawiera ten sam zestaw metod, z których większość nie zmieniła się od pół wieku bo tyle mniej więcej liczą sobie na przykład prace Holta. Modele i metody opisane poniżej mają obecnie status „podręcznikowych". Umieszczenie metody w podręczniku nie znaczy jednak, że zupełnie wyszła ona z użycia. Świadczy za to zwykle o dobrym poznaniu jej własności co przydaje się podczas porównań z innymi metodami.

W tym rozdziale zawarliśmy m.in. charakterystykę prognozowania jako dziedziny nauki. Ta część z konieczności ma nieco wykładowy charakter, ale nie widzimy sensu wyważania otwartych drzwi. Definicje zaprezentowanych pojęć od dawna funkcjonujące w literaturze nie wymagają naszym zdaniem ponownego formułowania. Tym niemniej mają znaczenie w kontekście późniejszych rozważań i dlatego się tu znalazły.

Metody i modele, które opisaliśmy dotyczą predykcji szeregów czasowych i nie zostały wybrane przypadkowo. Część z nich wciąż ma potencjał rozwoju co udowodnimy w rozdziale następnym. Pozostałe zaś nadają się jako benchmark pozwalający ocenić przydatność prognostycznego algorytmu genetycznego. Kolejna metoda prognozowania szeregów czasowych powinna bowiem wnosić coś nowego – pozwolić uchylić założenie lub poprawić jakość wyników.

Podobnie jak w rozdziale pierwszym postanowiliśmy zawrzeć w tej części książki nieco wiedzy podstawowej na temat predykcji szeregów czasowych. Wierzymy, że dzięki temu analizy zamieszczone w kolejnych rozdziałach staną się przystępne niezależnie od stopnia zaawansowania wiedzy czytelnika.

2.1. Podstawowe pojęcia

Prognozy wykorzystujemy częściej niż to się wydaje na pierwszy rzut oka. Wybieramy kierunek studiów – zakładamy czym w przyszłości chcielibyśmy się zajmować; fundusz emerytalny – określamy jak wysoką emeryturę

spodziewamy się otrzymywać. Ściślej rzecz biorąc, w takich momentach mamy do czynienia raczej z przewidywaniem. Przykładem przewidywania jest wnioskowanie statystyczne, w którym na podstawie wyników badania próby statystycznej wypowiadamy się na temat całej populacji. Jeżeli wnioskujemy o zdarzeniach, które dopiero mają zajść w przyszłości wówczas mamy do czynienia z przewidywaniem przyszłości.

Nie każde przewidywanie nazwiemy prognozowaniem. Nie jest nim na przykład wróżba (nawet jeśli się spełni) ponieważ brak jej racjonalnych przesłanek. Prognozowanie (zwane również predykcją) opiera się na racjonalnych, zwykle naukowych podstawach.

Prognoza odnosi się do określonego obiektu (kraju, przedsiębiorstwa itp.), w którym zachodzą pewne zjawiska (demograficzne, gospodarcze, inne) dające się opisać za pomocą zmiennych ilościowych lub jakościowych (również o charakterze losowym). Konkretne wartości prognoz obarczone są niepewnością z uwagi na fakt, że interesujące nas zmienne dopiero zrealizują się w przyszłości.

Wyobrażenie o przyszłości jakie posiadamy jest z definicji ułomne i obciążone błędem skoro wnioskujemy na temat nieznanych realizacji na podstawie dotychczasowych obserwacji. „Prognozując, staramy się dociec co znajdziemy za rogiem, nie wyglądając za ten róg" (Gajda 2001, s. 134). Dopiero kiedy prognoza wygasa (czyli po poznaniu faktycznej realizacji zmiennej) dowiadujemy się jaką dokładnie wartość miał popełniony błąd.

Brakuje jednej, konkretnej i powszechnie uznawanej definicji prognozowania. Jednak niezależnie od podejścia do zdefiniowania tego procesu, dochodzimy do dwóch ważnych wniosków:

1. Prognozując dokonujemy sądu na temat przyszłego zajścia określonego zdarzenia.
2. Na podstawie posiadanej wiedzy o przeszłości próbujemy dowiedzieć się czegoś o przyszłości.

Warto zwrócić uwagę, że w logice klasycznej wszystkie sądy są prawdziwe lub fałszywe natomiast w przypadku prognozowania mówimy, że są one trafne lub nietrafne (chybione). Kryterium pozwalającym je rozróżnić czynimy rozbieżność wyników predykcji w odniesieniu do zrealizowanej wartości zmiennej. Prognozę uznajemy za chybioną, kiedy wspomniana rozbieżność nie pozwala nam zaakceptować uzyskanych wyników.

Jeśli chodzi o drugi z warunków, prognozując spodziewamy się, że przyszłość będzie podobna do przeszłości. Skoro w przeszłości zaobserwowaliśmy, że między zmiennymi zachodził związek to zakładamy iż nadal się on utrzyma (Gajda 2001, s. 136). W przypadku niektórych metod czynimy mocne założenie, że związek ten dotyczy okresów odległych w czasie o pewną stałą.

Na podstawie powyższych uwag powiemy, że prognoza jest sądem posiadającym następujące własności (Cieślak 1997, s. 17):

— formułujemy go z wykorzystaniem dorobku nauki;
— odnosi się do określonej przyszłości;
— jest weryfikowalny empirycznie;
— jest niepewny, ale akceptowalny.

Ze względu na sposób wyrażenia zmiennej, prognozy dzielimy na ilościowe i jakościowe (Cieślak 1997, s. 23). Pierwsze z wymienionych mogą mieć charakter prognoz punktowych (podajemy konkretną, przyszłą wartość zmiennej) lub przedziałowych (określamy przedział liczbowy z określonym prawdopodobieństwem obejmujący przyszłą wartość zmiennej).

Prognozy wspomagają proces decyzyjny i w związku z tym spełniają różnorodne funkcje (Gajda 2001, s. 134):
— preparacyjną – prognoza przygotowuje nas do podjęcia działania. Na jej podstawie podejmujemy decyzje przy czym nie zawsze prognozy takie dotyczyć będą zmiennych, na które mamy wpływ (np. pogoda);
— aktywizującą – stają się nimi prognozy preparacyjne, które po opublikowaniu wpływają na prognozowane zjawisko, ponieważ podejmowane są działania sprzyjające lub przeciwdziałające ich realizacji. Wyróżnimy:
  — prognozy samorealizujące – kiedy prognoza wywołuje działania wychodzące jej naprzeciw;
  — prognozy samounicestwiające – prognoza, która ostrzega przed niekorzystnym zjawiskiem powoduje, że podejmuje się działania mające na celu zapobieżenie spodziewanym skutkom.

Prognozy są również klasyfikowane ze względu na różne kryteria. Najczęściej spotykamy podział ze względu na horyzont czasowy, którego dotyczą. Wyróżnimy wówczas:
— prognozy krótkookresowe;
— prognozy średniookresowe;
— prognozy długookresowe.

Co jednak rozumiemy przez krótki, a co przez długi okres to zależy od charakteru prognozowanego zjawiska. Czym innym jest bowiem prognoza na miesiąc w przyszłość dla danych dziennych, a czym innym dla miesięcznych.

Generalnie rzecz biorąc, prognoza krótkookresowa dotyczy przedziału czasu, w którym zachodzą jedynie zmiany ilościowe. W prognozie średniookresowej zaczynamy obserwować zmiany jakościowe zaś w długookresowej mamy do czynienia nie tylko zmianami ilościowymi, ale i poważnymi przeobrażeniami jakościowymi[1]. Zeliaś w swojej książce (Zeliaś 1997, s. 23) podaje inne kryteria stosowane podczas klasyfikacji na przykład podział prognoz na regionalne, krajowe, międzynarodowe i światowe.

[1] Niektórzy autorzy np. Zeliaś (por. (Zeliaś 1997)) horyzont prognozy traktują bardzo dosłownie. Wówczas prognozy krótkookresowe zamykają się w obrębie roku, a średniookresowe – 5 lat.

Wachlarz metod wykorzystywanych do otrzymywania prognoz jest szeroki i wciąż powstają kolejne jak choćby metody opisane w kolejnym rozdziale. W sensie ogólnym dadzą się one pogrupować w następujące kategorie:

1. Metody niestrukturalne – prognozujemy zjawisko na podstawie danych na temat dotychczasowego kształtowania się zmiennej. Dane te zazwyczaj zebrane są w postaci szeregu czasowego.
2. Metody strukturalne – podstawą jest zazwyczaj model przyczynowo-skutkowy. Przykładem mogą być jedno- i wielorównaniowe modele ekonometryczne.
3. Metody analogowe – przewidujemy przyszłość określonej zmiennej na podstawie danych o podobnych zmiennych, co do których istnieją zbyt słabe podstawy by przypuszczać, że są one przyczynowo powiązane ze zmienną prognozowaną. Istota tych zmiennych może być taka sama (np. liczba ludności w dwóch krajach) bądź różna (np. populacja muszek owocówek zamkniętych w słoju i liczba ludności zamieszkującej górzysty, odcięty od świata region).
4. Metody heurystyczne – polegają na wykorzystaniu opinii ekspertów opartej na ich intuicji i doświadczeniu.

Prognozowanie niestrukturalne zakłada, że w przeszłości zmiennej zawiera się informacja o mechanizmach generujących jej przyszłe wartości. Niektóre z metod (tzw. metody mechaniczne) nie wymagają zatem dysponowania dużymi próbami. Z kolei inne, w których na przykład estymuje się parametry, potrzebują odpowiednio długich szeregów czasowych. Metodom niestrukturalnym zarzuca się jednak zwykle, że ignorują teorie (ekonomiczne) związane z prognozowanymi zjawiskami. Uchylając niektóre założenia można przynajmniej częściowo uwzględnić te zarzuty.

W prognozowaniu strukturalnym wykorzystujemy model strukturalny odzwierciedlający pewną teorię (na przykład ekonomiczną). W grę wchodzą tu związki przyczynowo-skutkowe lub symptomatyczne pomiędzy zmiennymi. Do pierwszego przypadku zaliczają się modele ekonometryczne, a do drugiego modele trendu, autoregresyjne itp. Dysponujemy więc możliwością włączenia założeń teoretycznych opisujących wpływ wybranych czynników na prognozowaną zmienną, o ile takowe przesłanki istnieją.

Prognozując na podstawie modeli strukturalnych prognozy konstruujemy wg jednej z poniższych zasad (Gajda 2001, s. 141–143):

1. Zasada prognozowania wg wartości oczekiwanej – jest łatwa w realizacji oraz daje prognozy nieobciążone. Predyktorem jest warunkowa wartość oczekiwana (regresja zmiennej objaśnianej względem zmiennych objaśniających). W tym przypadku pojedyncze prognozy mogą być obarczone błędem, ale przy wielokrotnym powtarzaniu prognoz w niezmienionych warunkach (o ile to, rzecz jasna, możliwe) średni błąd ciągu prognoz będzie równy zero.

2. Zasada prognozowania według największego prawdopodobieństwa – kiedy nie możemy sięgnąć po eksperyment (a tak dzieje się w naukach ekonomicznych) trudno odtworzyć warunki w jakich miałby powstać ciąg prognoz. Prognoza dla danych warunków wyznaczana jest tylko raz. Wtedy przypadku najbardziej użyteczna będzie prognoza o najwyższym prawdopodobieństwie realizacji czyli na poziomie dominanty rozkładu prognoz. Jednak dla wyraźnie asymetrycznych rozkładów zasada potrafi generować skrajne wartości zmiennej.
3. Zasada prognozowania wg mediany – jest szczególnie użyteczna w przypadku rozkładów prognoz o tzw. grubych ogonach gdyż wykorzystuje medianę rozkładu prognozowanej zmiennej.
4. Zasada prognozowania minimalizująca oczekiwaną stratę – jest stosowana gdy błąd popełniany podczas prognozowania wiąże się z określoną stratą co opisujemy przy pomocy stosownej funkcji o nieujemnych wartościach. Wybieramy wówczas prognozę związaną z jak najmniejszą oczekiwaną stratą.

W praktyce najczęściej korzystamy z mieszanki prognozowania niestrukturalnego i strukturalnego. Dzieje się tak ponieważ niektóre teorie ekonomii okazują się niejednoznaczne lub bazują na zbyt mocnych założeniach.

Bywa też tak, że budowa modelu strukturalnego okazuje się zbyt kosztowna i czasochłonna z punktu widzenia naszych potrzeb. Przykładem jest sytuacja, gdy potrzebujemy przyszłych wartości zmiennych egzogenicznych modelu ekonometrycznego. Przy pomocy metod niestrukturalnych otrzymujemy realizacje tych zmiennych, które następnie wykorzystujemy, aby wyznaczyć prognozy zmiennych endogenicznych modelu.

W (Zeliaś 1997, s. 25) znajdziemy rozbudowany schemat prezentujący sklasyfikowane różne metody prognostyczne z nieco innego (matematyczno-statystycznego) punktu widzenia. Wciąż jednak powstają nowe metody. Na przykład, jeżeli przechodzimy z wysokiej częstotliwości danych na niższą, kiedy dla każdej jednostki czasu po agregacji dysponujemy określonym rozkładem wartości, przydać się może zaprezentowana w (Arroyo i Maté 2009) metoda prognozowania na podstawie histogramów. Załóżmy, że dysponujemy zbiorem realizacji obserwowanej zmiennej dla każdego momentu czasu lub zmienna ta odznacza się pewnym stopniem zmienności. Autorzy artykułu wymieniają dwie takie sytuacje:

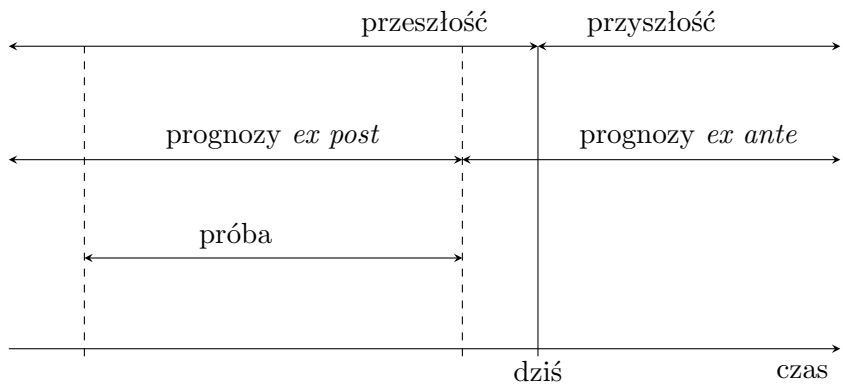
— zmienna jest mierzona w czasie dla każdej jednostki lub grupy jednostek zaś nas interesuje zachowanie całej grupy;
— zmienna wyrażona jest przy pomocy danych o określonej częstotliwości, ale chcemy ją analizować w niższej częstotliwości.

Rozwiązaniem problemu przedstawiania tego typu danych są histogramy, z kolei prognozy dotyczą szeregów czasowych rozkładów obserwowanej

zmiennej w czasie. W ich efekcie również uzyskujemy rozkład reprezentowany przez histogram.

Okresy przeszłości i przyszłości w naszym rozumieniu nie pokrywają się z ich analogiami z punktu widzenia mechanizmów predykcyjnych. Różnice z tym związane prezentuje schemat na rysunku 2.1. Moment wyznaczenia prognozy determinuje podstawowy podział na: *ex post* i *ex ante*. Granicę między nimi określa ostatnia chronologicznie obserwacja, którą wykorzystujemy do wykonania prognozy. W rzeczywistości dane nie pojawiają się od razu, lecz z pewnym opóźnieniem. Trudno w połowie stycznia dysponować wartością za cały ten miesiąc, a w praktyce opóźnienia są jeszcze większe. Dlatego część naszej przeszłości staje się przyszłością dla modelu i dlatego właśnie obszar prognoz *ex ante* na rysunku jest taki a nie inny.

Powstaje oczywiście pytanie: po co robić prognozy *ex post*, skoro znamy realizacje zmiennych z tych okresów? Otóż prognozy *ex post* pozwalają ocenić jakość zastosowanej metody, dzięki porównaniu wartości rzeczywistych z prognozowanymi. W momencie, kiedy dane wreszcie „dogonią” okres *ex ante*, prognoza taka wygasa. Dla metod strukturalnych prognoza *ex post* tylko częściowo pokrywa się z prognozą wygasłą. Na przykład modele ekonometryczne wymagają uprzedniej znajomości wartości zmiennych egzogenicznych. W przypadku predykcji szeregów czasowych obszary prognoz *ex post* i wygasłych pokrywają się ze sobą.



Rysunek 2.1: Prognozy *ex post* a prognozy *ex ante*

Źródło: opracowanie własne

Proces prognozowania wymaga przejścia przez szereg etapów, które zostały sformułowane w (Cieślak 1997, s. 57–61) w postaci następujących kroków postępowania:

1. Sformułowanie zadania prognostycznego – określamy obiekt, zjawisko, prognozowane zmienne a także cel, horyzont i wymagania odnośnie dopuszczalności prognozy.

2. Sformułowanie przesłanek prognostycznych – określamy czynniki wpływające na prognozowaną zmienną (na przykład zmienne egzogeniczne) wraz z kierunkiem wpływu oraz gromadzimy dane.
3. Wybór metody prognozowania (predyktora) – określamy przepis, według którego powstanie prognoza czyli decydujemy, którą z metod (niestrukturalnych, strukturalnych itd.) wybierzemy,
4. Wyznaczenie prognozy – postępujemy zgodnie ze schematem przewidzianym dla wybranej metody.
5. Ocena dokładności prognozy – określamy błąd prognoz *ex post* oraz (jeśli ich wyznaczenie było możliwe) oczekiwany tj. przed wygaśnięciem błąd prognoz *ex ante*.

2.2. Dekompozycja szeregu czasowego

Podstawą prognozowania, niezależnie od przyjętej metody jest zawsze szereg czasowy. Szeregi takie zawierają obserwacje uporządkowane chronologicznie – od najstarszej do najnowszej dostępnej przy czym ostatnia dana zwykle nie pokrywa się z momentem przeprowadzania badania. Przyjęło się przy tym, aby przez y_t oznaczać konkretną realizację zmiennej y w okresie (momencie) t.

Niezależnie od tego jakiego zjawiska dotyczy, każdy szereg czasowy składa się z pewnych składowych, które dadzą się pogrupować w kilka jednorodnych pod względem zachowania kategorii. Najbardziej ogólny podział polega na wyróżnieniu składowej systematycznej jako efektu oddziaływania na zmienną stałego zestawu czynników oraz składowej losowej. Kiedy wskazujemy szczegółowo poszczególne elementy składające się na szereg, mówimy wówczas o dekompozycji szeregu czasowego, w którym możemy wyodrębnić (Cieślak 1997, s. 62):

— wahania przypadkowe – losowe zmiany zmiennej o zróżnicowanej sile;
— tendencję rozwojową (trend) – długookresową skłonność do jednokierunkowych zmian (wzrostu lub spadku) wartości zmiennej jako efekt działania stałego zestawu czynników;
— stały (przeciętny) poziom zmiennej – występuje jeśli nie ma w szeregu tendencji rozwojowej, a wartości oscylują wokół pewnego stałego poziomu;
— wahania cykliczne – długookresowe, rytmiczne wahania wartości zmiennej wokół trendu lub stałego poziomu;
— wahania sezonowe – wahania wartości wokół trendu lub stałego poziomu mające skłonność do powtarzania się w określonym czasie nie przekraczającym roku.

Aby zaobserwować wahania cykliczne, musimy dysponować długimi (nierzadko liczącymi wiele lat) szeregami danych. Tylko wtedy zyskujemy pewność, że obserwujemy powtarzające się zjawisko a nie na przykład obecność jednego z rodzajów tendencji rozwojowej.

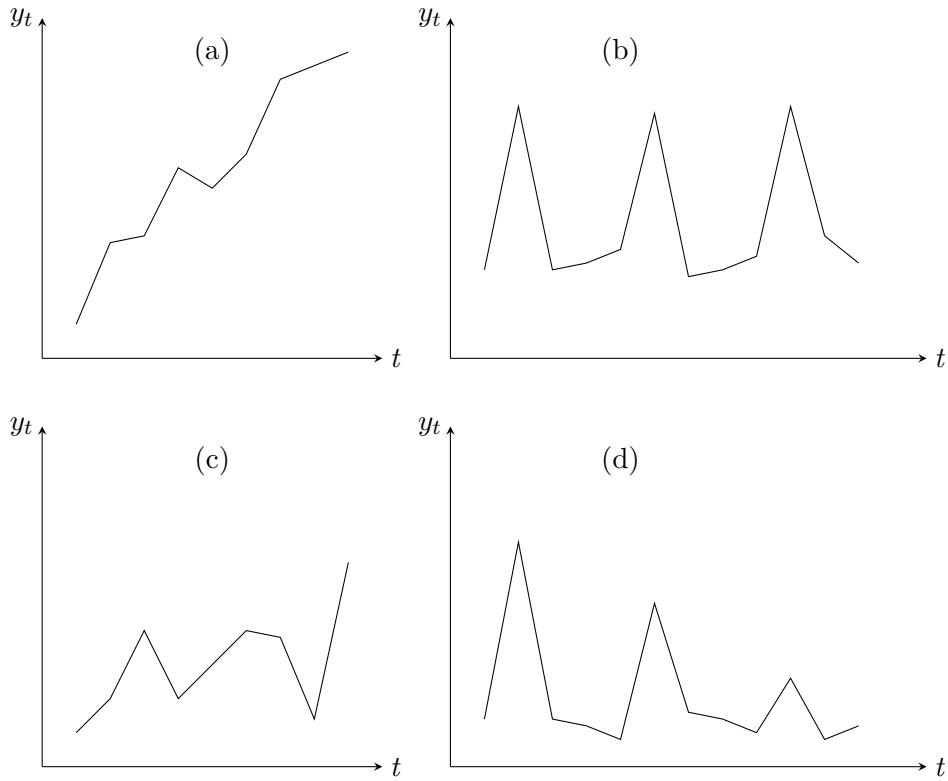
Spokrewnione z nimi wahania sezonowe występują w wersji addytywnej lub multiplikatywnej. Różnica polega na sposobie zachowania tych wahań względem trendu występującego w całym szeregu. Kiedy jest to mniej więcej stała różnica mówimy o wahaniach addytywnych, kiedy zaś wahania sezonowe wykazują się stałym ilorazem w stosunku do trendu wtedy są to wahania multiplikatywne.

Wymienione przed chwilą elementy dekompozycji mogą dla danego szeregu występować praktycznie w dowolnych konfiguracjach (np. małe wahania losowe, stały poziom zmiennej i wahania sezonowe). Sama dekompozycja ma bardzo duże znaczenie przy wyborze metody prognozowania gdyż pozwala z góry określić warunki w jakich dana metoda zostanie użyta. Wykresy na rysunku 2.2 ilustrują kilka możliwych wariantów dekompozycji.

Dekompozycji szeregu dokonujemy na podstawie wykresu dwuwymiarowego, który pozwala nie tylko wyodrębnić poszczególne składowe, ale również wskazać obserwacje nietypowe oraz punkty zwrotne, w których dochodzi do zmiany kierunku trendu. Obecność jednych i drugich wpływa negatywnie na dokładność prognoz szczególnie w przypadku krótkich szeregów czasowych. Generalnie im krótszy szereg tym mniej jednoznaczne stają się wnioski płynące z dekompozycji. Obserwacja uznana duże wahanie przypadkowe może w rzeczywistości zapowiadać w dłuższej perspektywie odwrócenie tendencji.

Swego rodzaju uzupełnieniem, a nawet rozszerzeniem tradycyjnej dekompozycji jest procedura, którą znajdziemy w pracy (Adya, Collopy, Armstrong i Kennedy 2012). Autorzy użyli prognozowania opartego na regułach, aby opracować zautomatyzowane heurystyki wykrywające w szeregach czasowych takie zjawiska jak: obserwacje odstające, zmiany poziomu zmiennej, zmiany głównego trendu, niestabilność bieżącego trendu, nietypową ostatnią obserwację czy formę funkcji. Wymieniają oni w sumie 28 zjawisk spotykanych w szeregach czasowych. Czynią założenie, że część z tych zachowań da się zidentyfikować automatycznie natomiast w przypadku niektórych konieczna jest pomoc eksperta posiadającego odpowiednie doświadczenie. Łączą różne metody i modele, a następnie dobierają dla nich wagi. Zanim jednak dojdzie do prognozowania szereg poddaje się następującej procedurze dekompozycyjnej:

1. Identyfikacja formy funkcyjnej szeregu tj. określenia rodzaju trendu.
2. Wyszukiwanie obserwacji odstających, które następnie zastępuje się wartościami uśrednionymi.



Rysunek 2.2: Przykłady dekompozycji szeregu czasowego: (a) – Wahania przypadkowe oraz trend liniowy, (b) – Wahania sezonowe addytywne i stały poziom zmiennej, (c) – Duże wahania przypadkowe i stały poziom zmiennej, (d) – Wahania sezonowe multiplikatywne i stały poziom zmiennej.

Źródło: opracowanie własne

3. Określenie punktów braku kontynuacji trendu wynikającej na przykład z przyjęcia przez zjawisko innej (wyższej bądź niższej) trajektorii.
4. Sprawdzenie czy ostatnia obserwacja nie zachowuje się nietypowo.
5. Zbadanie zmian współczynnika kierunkowego trendu – pozwala określić stabilność głównego trendu obserwowanego w szeregu.
6. Zbadanie stabilności bieżącego, krótkookresowego trendu.

Jak widać duże znaczenie w zaproponowanym podejściu do dekompozycji ma zachowanie trendu oraz pojawianie się obserwacji odstających. Nie powinno to dziwić ponieważ nagłe zmiany zachowania szeregu bardzo obniżają jakość prognoz.

W szeregach danych natykamy się ponadto na obserwacje odstające. Koehler i inni zaproponowali (Koehler, Snyder, Ord i Beaumont 2012), aby takie nietypowe wartości zaliczyć do jednaj z trzech kategorii: addytywne obserwacje nietypowe; przesunięcie poziomu; zmiana przejściowa. Pierwsze

z nich to w rzeczywistości pojedyncze dane o zachowaniu odmiennym od obserwowanego w szeregu trwające zwykle jeden okres. Kolejny typ oznacza sytuację, w której szereg czasowy całkowicie zmienia swoją trajektorię (na wyższą lub niższą od dotychczasowej). W ostatniej kategorii mamy do czynienia ze skokiem poziomu zmiennej, po której obserwacje stopniowo wracają do poprzedniego stanu. Celem autorów było opracowanie modelu przestrzeni stanów będącego zmodyfikowaną wersją modelu Wintersa, który uwzględnia obecność nietypowych obserwacji. Wprowadzili zatem zmienną przełącznikową, która wyraża typ obserwacji nietypowej a do estymacji parametrów modelu posłużyli się symulacją.

Progności nie mogą opierać się tylko na pojawianiu się powtarzalnych wzorców zachowań danych. Wykorzystują (a przynajmniej powinni tak robić) dodatkowe informacje, których wartości jednak nie zawsze są w stanie zweryfikować. Remus i pozostali (Remus, O'Connor i Griggs 1998) wzięli pod uwagę wpływ informacji o nieznanej poprawności na dokładność prognoz. Mówiąc wprost wykonali symulacje wpływu niesprawdzonych pogłosek na jakość predykcji używając do tego celu wygenerowanych szeregów danych. Stwierdzili, że w otoczeniu punktu zwrotnego nawet taka niedokładna informacja może mieć znaczenie a prognozy nie są wcale gorsze od tych wykonanych bez dodatkowych informacji. Zauważyli również, że wpływ nawet nieprawidłowych informacji zanika bardzo szybko w czasie a zatem o ewentualnym punkcie zwrotnym szeregu dowiemy się niemal w ostatniej chwili.

Innym sposobem dekompozycji szeregu jest metoda STL (Seasonal--Trend using Loess) opisana w (Cleveland, Cleveland, McRae i Terpenning 1990). To procedura, która tak jak poprzednio ma za zadanie wyodrębnić wahania losowe, trend i sezonowość. Jej podstawą jest skomplikowana obliczeniowo metoda wygładzania zwana Loess (*locally weighted scatterplot smoothing*). Wygładzanie szeregu odbywa się w niej poprzez dopasowanie ciągu wielomianów, których parametrów poszukujemy używając ważonej metody najmniejszych kwadratów.

Wahania przypadkowe można próbować eliminować, zaś trend oraz sezonowość wyodrębniać z szeregu, używając do tego celu metod, które dzielą się na mechaniczne i analityczne. Następny podrozdział zawiera przegląd tych metod. Część z nich stała się punktem wyjścia dla opisanego w rozdziale trzecim prognostycznego algorytmu genetycznego.

Bunn i Vsssipoulos zajęli się estymacją wskaźników sezonowości dla celów krótkookresowych prognoz sprzedaży (Bunn i Vassilopoulos 1999). Opisują dwie metody wyznaczania tzw. grupowych indeksów sezonowości (GSI – *grouped seasonal indices*) stosowanych dla grup produktów: indeks sezonowości obliczany jest dla każdego produktu należącego do grupy po czym uśrednia się je dla całej grupy; indeks sezonowości, którego punktem

wyjścia jest zagregowany popyt na produkty grupy dla każdego okresu czasu przeliczany następnie na wskaźniki dla każdego produktu należącego do grupy. Jednak autorzy ci pominęli resztę elementów dekompozycji, skupiając uwagę wyłącznie na sposobach przybliżenia efektów sezonowych.

2.3. Prognozowanie przy pomocy metod mechanicznych

2.3.1. Metody naiwne

Metody naiwne są najprostszymi spośród metod mechanicznych. Bazują na założeniu, że w przyszłości nie nastąpią zmiany w dotychczasowym oddziaływaniu czynników wpływających na prognozowaną zmienną (Cieślak 1997, s. 67). W związku z tym stosuje się je pod warunkiem braku dużych wahań prognozowanej zmiennej. Przy pomocy metod naiwnych dokonuje się jedynie krótkookresowych prognoz (na przykład na jeden, dwa okresy do przodu).

Najprostszy przypadek stanowi metoda naiwna bez zmian zwana również metodą naiwną prostą. Zakładamy w niej, że przyszłe realizacje zmiennej nie będą zbytnio różnić się od ostatniej wartości. W związku z tym prognozą na okres t staje się wartość z okresu $t-1$. Wymagamy przy tym, aby w szeregu występowały wyłącznie małe wahania losowe, stały poziom zmiennej i nie powinna pojawiać się sezonowość.

Niech symbol gwiazdki oznacza prognozę *ex post* zaś litera „P" prognozę *ex ante*. Wówczas prognoza zmiennej y na okres t $(t = 1, \ldots, n)$ metodą naiwną prostą jest równa:

$$y_t^* = y_{t-1} \tag{2.1}$$

$$y_t^P = y_n \tag{2.2}$$

Przykład zawarty w tabeli 2.1 pomoże w zrozumieniu istoty metody naiwnej bez zmian. W okresach 7 i 8 powstają już prognozy *ex ante*, a tworzy je ostatnia dostępna obserwacja powtórzona tyle razy ile wynosi zakładany horyzont prognoz. W rzeczywistości raczej trudno oczekiwać, że z poddanej predykcji zmiennej znikną wahania losowe. Dlatego metodą naiwną prostą wykonujemy prognozy wyłącznie krótkookresowe.

Jeśli w modelu występuje tendencja rozwojowa (a pozostałe założenia są zgodne z metodą naiwną bez zmian), wówczas sięgamy do metody naiwnej z trendem. Jako prognozę przyjmujemy zmienną na poziomie z okresu

Tabela 2.1: Przykład prognozy metodą naiwną bez zmian

Okres (t)	Dane (y_t)	Prognoza
1	102	
2	101	102
3	103	101
4	102	103
5	101	102
6	103	101
7		103
8		103

Źródło: opracowanie własne

poprzedniego powiększoną o ostatni dostępny przyrost (Cieślak 1997, s. 68). Dla przypadku trendu liniowego wyrazimy ją wzorem:

$$y_t^* = y_{t-1} + (y_{t-1} - y_{t-2}) \tag{2.3}$$

$$y_t^P = y_n + h(y_n - y_{n-1}) \tag{2.4}$$

gdzie: $h \geqslant 1$.

Kiedy zmienna przyrasta (wykładniczo) o stały procent w kolejnych okresach wówczas użyjemy modelu:

$$y_t^* = y_{t-1}\frac{y_{t-1}}{y_{t-2}} \tag{2.5}$$

$$y_t^P = y_n\left(\frac{y_n}{y_{n-1}}\right)^h \tag{2.6}$$

W tabeli 2.2 znalazł się przykład prognozy metodą naiwną z poprawką liniową. Należy zwrócić uwagę na to, że trend, który pojawia się w prognozach *ex ante*, nie ulegnie już odwróceniu. To argument za używaniem tej metody w prognozach krótkookresowych.

Kolejnym przypadkiem z grupy metod naiwnych jest metoda naiwna z sezonowością. Sięgamy po nią jeżeli w szeregu występują: małe wahania przypadkowe, stały poziom zmiennej i sezonowość addytywna.

Od strony technicznej jest ona podobna do metody naiwnej prostej, tyle że jako prognozy przyjmuje wartości zaobserwowane sezon wcześniej np. w odpowiednim kwartale roku poprzedniego. W odróżnieniu jednak od prognozy naiwnej prostej, potrzebujemy dłuższego szeregu czasowego, aby upewnić się, że mamy do czynienia z sezonowością a nie na przykład z ob-

Tabela 2.2: Przykład prognozy metodą naiwną z poprawką liniową

Okres (t)	Dane (y)	Przyrost	Prognoza
1	101		
2	102	1	
3	104	2	103
4	104	0	106
5	103	-1	104
6	105	2	102
7	108	3	107
8		3	111
9		3	114

Źródło: opracowanie własne

serwacjami nietypowymi. Jeżeli wymienione warunki są spełnione, prognozę wyznaczamy następująco:

$$y_t^* = y_{t-m} \tag{2.7}$$

$$y_t^P = y_{n-m+h} \tag{2.8}$$

gdzie: $h \leqslant m$.

W tabeli 2.3 znalazł się przykład użycia metody naiwnej z sezonowością, w której występują wahania sezonowe kwartalne. Z tego powodu pierwsza prognoza *ex post* wyznaczona została dla piątego okresu.

Metody naiwne są proste, szybkie i tanie. Zadowalają się niewielką liczbą obserwacji. Służą również jako punkt odniesienia dla innych metod. Jeżeli dają one gorsze rezultaty od prognoz naiwnych uważa się je za mało użyteczne. W przypadku metod naiwnych nie można jednak określić błędu prognoz *ex ante*. Tym niemniej jakość prognoz wykonanych z ich pomocą może być zaskakująco dobra w porównaniu z narzędziami bardziej skomplikowanymi.

2.3.2. Metoda średniej ruchomej

Występujące w szeregu czasowym wahania przypadkowe mogą mieć zarówno małą jak i dużą amplitudę. W tym drugim wypadku metody naiwne, włączające do prognozy skutki ostatnio zrealizowanego wahania losowego, przestają się sprawdzać. Jeżeli było ono duże, prognoza zostaje obarczona również dużym błędem. Metoda średniej ruchomej próbuje zaradzić temu problemowi.

Średnia ruchoma działa jak filtr, eliminując z szeregu wahania krótkookresowe. Wyróżnione zostają (w obrębie próby) wahania długookresowe

Tabela 2.3: Przykład prognozy metodą naiwną z sezonowością

Okres (t)	Dane (y)	Prognoza
1	101	
2	108	
3	102	
4	103	
5	102	101
6	109	108
7	101	102
8	102	103
9		102
10		109
11		101
12		102

Źródło: opracowanie własne

(Gajda 2001, s. 181). W efekcie dane ulegają „wygładzeniu". Metodę tę stosujemy w szeregach bez sezonowości i trendu, za to z dowolnymi wahaniami przypadkowymi.

Model średniej ruchomej prostej zakłada wykorzystanie określonej liczby sąsiednich obserwacji do wyznaczenia prognozy na dany okres. Na użytek prognozowania[2] przyjmuje się, że wartość zmiennej w okresie prognozy będzie równa średniej arytmetycznej z jej k poprzednich wartości gdzie k to stała wygładzania określana z góry na początku obliczeń. Prognoza przy pomocy średniej ruchomej prostej da się wyrazić wzorem:

$$y_t^* = \frac{1}{k} \sum_{i=t-k}^{t-1} y_i \tag{2.9}$$

$$y_t^P = \frac{1}{k} \sum_{i=n-k+1}^{n} y_i \tag{2.10}$$

Prognozując dalej niż na jeden okres *ex ante* wykonujemy tzw. prognozę dynamiczną czyli do obliczeń włączamy prognozy wykonane dla okresów wcześniejszych tak aby cały czas operować na k elementach.

Ze wzrostem stałej wygładzania rośnie efekt wygładzenia. Oparcie się na większej liczbie wyrazów wygładza silniej szereg, ale powoduje wolniejszą reakcję na zmiany poziomu prognozowanej zmiennej (Cieślak 1997, s. 71).

[2] W statystyce i innych naukach średnią ruchomą oblicza się dla środkowego spośród k okresów. Nie służy ona jednak wówczas do celów prognostycznych.

Mniej wyrazów w średniej oznacza z kolei szybsze odzwierciedlanie zmian wartości z ostatnich okresów, ale na prognozę silniej wpływają wahania przypadkowe. Trudno jednoznacznie wskazać obiektywne kryterium, które pomogłoby dobrać stałą wygładzania. W literaturze znajdziemy jedynie sugestie odnośnie wartości tego parametru.

Mimo iż dobór stałej jest arbitralny, należy zwrócić uwagę na fakt utraty części informacji z początkowych k okresów. Jest to spowodowane sposobem, w jaki wyznaczamy pierwszą prognozę *ex post*. Dlatego należy unikać wygładzania zbyt silnego, ograniczającego możliwość zweryfikowania jakości prognoz.

Tabela 2.4: Przykład prognoz metodą średniej ruchomej prostej

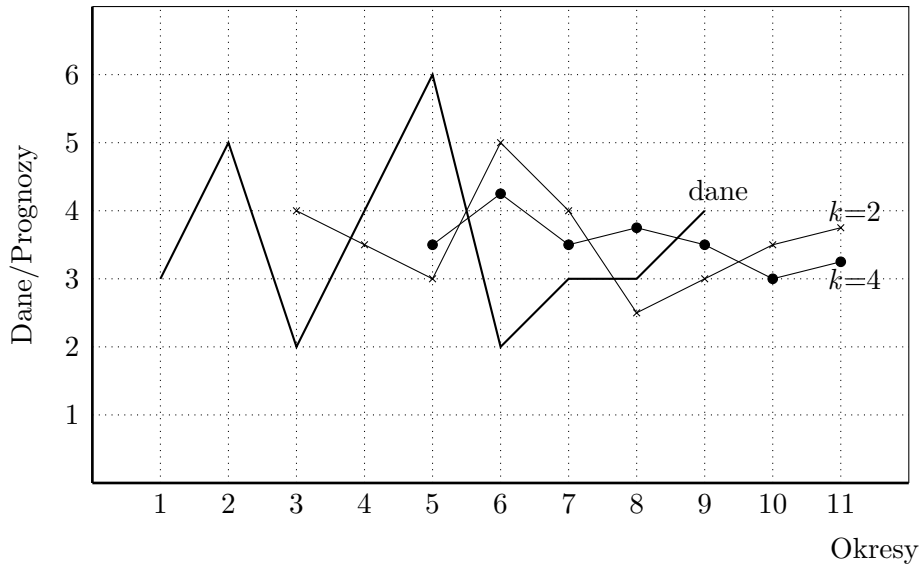
		Stała wygładzania	
Okres (t)	Dane (y)	k=2	k=4
1	3		
2	5		
3	2	4	
4	4	3,5	
5	6	3	3,5
6	2	5	4,25
7	3	4	3,5
8	3	2,5	3,75
9	4	3	3,5
10		3,5	3
11		3,75	3,25

Źródło: opracowanie własne

W tabeli 2.4 znajdują się przykładowe prognozy otrzymane przy pomocy dwóch średnich ruchomych o stałych wygładzania równych odpowiednio 2 oraz 4. Zwróćmy uwagę na dwie rzeczy: po pierwsze wzrost stałej wygładzania oznacza utratę większej liczby obserwacji z początkowych okresów, po drugie prognozy na okres 11 mają charakter dynamiczny. Na przykład dla stałej k=2 prognoza w tym okresie powstaje poprzez wykorzystanie wartości rzeczywistej z okresu 9 i prognozy na okres 10.

Wykres na rysunku 2.3 prezentuje w jaki sposób wysokość stałej k wpływa na efekt wygładzenia prognoz na przykładzie danych zawartych w tabeli 2.4. Przy stałej wygładzania równej 4 wygładzeniu ulegają nawet wysokie wahania przypadkowe, których wpływ wciąż jest widoczny w prognozach dla niższego k.

W metodzie średniej ruchomej prostej wszystkie obserwacje wykorzystywane do wyznaczenia prognozy mają takie samo znaczenie niezależnie



Rysunek 2.3: Efekt wygładzania w średniej ruchomej prostej

Źródło: opracowanie własne

od tego, z jak odległej przeszłości pochodzą. Logicznym wydaje się, że na bieżący poziom zjawiska większy wpływ powinny mieć niedawno zaistniałe czynniki. Z drugiej strony te z nich, które zaznaczyły swoją obecność w bardziej odległej przeszłości, w mniejszym stopniu kształtują wartości prognozowanej zmiennej (Cieślak 1997, s. 71).

Załóżmy, że w związku z powyższymi uwagami obserwacje nowsze będą mieć większą wagę niż starsze. Nazywamy to efektem postarzania informacji i na tym założeniu bazuje metoda średniej ruchomej ważonej:

$$y_t^* = \sum_{i=t-k}^{t-1} y_i w_i \tag{2.11}$$

$$y_t^P = \sum_{i=n-k+1}^{n} y_i w_i \tag{2.12}$$

Występujące w metodzie wagi w_i spełniać muszą następujące założenia:

$$0 \leqslant w_1 \leqslant w_2 \leqslant \ldots \leqslant w_k \leqslant 1 \quad \sum_{i=1}^{k} w_i = 1$$

Pierwsze z założeń to właśnie efekt postarzania informacji. Im dalej sięgamy w przeszłość (czyli im wyższa stała wygładzania), tym wagi stają się niższe.

W metodzie średniej ruchomej ważonej należy ustalić nie tylko stałą wygładzania, ale i poziomy poszczególnych wag. Może to sprawić sporą

trudność, ponieważ często brak wartościowych przesłanek pozwalających określić, w jakim stopniu dane pochodzące ze wskazanych okresów mają wpływać na wartość prognozy.

Warto również wspomnieć, że w metodzie średniej ruchomej ważonej efekt wygładzenia zależy nie tylko od stałej wygładzania, ale i od wartości samych wag. Wyższe wagi dla starszych obserwacji powodują silniejsze wygładzanie prognoz.

W tabeli 2.5 znalazł się przykład prognoz metodą średniej ruchomej ważonej, w której zastosowano trzy wagi wynoszące odpowiednio: $w_1 = 0{,}05$, $w_2 = 0{,}15$ i $w_3 = 0{,}8$. Wyjściowe dane są te same jak w tabeli 2.4. Dla okresu 11 wykonać należy prognozę dynamiczną, w której najwyższa waga przypada na prognozę dla okresu 10.

Tabela 2.5: Przykład prognoz metodą średniej ruchomej ważonej

Okres (t)	Dane (y)	Prognoza
1	3	
2	5	
3	2	
4	4	2,5
5	6	3,75
6	2	5,5
7	3	2,7
8	3	3
9	4	2,95
10		3,8
11		3,79

Źródło: opracowanie własne

Niekiedy (na przykład gdy stwierdzimy występowanie w szeregu trendu liniowego) używa się podwójnej średniej ruchomej. Wygładzony (średnią prostą lub ważoną) szereg wygładza się powtórnie. W obu metodach średnich ruchomych uwzględnia się tylko k ostatnich wartości zmiennej prognozowanej.

2.3.3. Metoda wygładzania wykładniczego Browna

Modele wygładzania wykładniczego można potraktować jako uogólnienie metod naiwnych (Gajda 2001, s. 185). Filtruje się w nich zakłócenia, aby zredukować nadmierny wpływ ostatnich obserwacji, a jednocześnie zachować korzyści jakie daje wykorzystanie ostatniej i przedostatniej realizacji zmiennej.

Da się udowodnić (Gajda 2001, s. 186), że występująca w metodzie Browna waga wynosi $\frac{1}{k}$ gdzie k to stała wygładzania znana ze średniej ruchomej. Dlatego metody tej używamy do szeregów o takich samych elementach dekompozycji jak w przypadku średniej. Ma ona jednak tę zaletę, że nie tracimy tak wielu informacji wyznaczając pierwszą prognozę *ex post*, która zawsze powstaje dla okresu drugiego. W metodzie Browna obowiązuje tzw. parametr wygładzania $\alpha \in \langle 0, 1 \rangle$ a prognozy za jej pomocą otrzymujemy następująco:

$$y_t^* = \alpha y_{t-1} + (1-\alpha) y_{t-1}^* \tag{2.13}$$

$$y_t^P = y_n + (1-\alpha) y_{t-1}^P \tag{2.14}$$

Analizując powyższe wzory zauważymy, że prognoza na dany okres jest średnią ważoną ostatniej dostępnej realizacji zmiennej oraz prognozy wyznaczonej na okres poprzedni (wagą jest parametr wygładzania). Prognozy z modelu średniej ruchomej są podobne do tych z modelu Browna, w którym $\alpha = \frac{2}{k+1}$ (Gajda 2001, s. 187).

Wartość (α) dobieramy eksperymentalnie, zazwyczaj minimalizując średnią z popełnianych błędów prognoz *ex post*. Im α bliższe 0 tym prognoza y_{t+1}^* staje się bliższa y_t^*, a co za tym idzie, szereg ulega silniejszemu wygładzeniu.

Problem stanowi drugi okres *ex post*, od którego rozpoczynamy proces prognozowania. Nie dysponujemy bowiem prognozami wykonanymi dla okresu pierwszego a co za tym idzie brak nam drugiego składnika wzoru. Jako startową wartość y_2^* przyjmujemy najczęściej początkową wartość zmiennej prognozowanej (y_1) lub średnią arytmetyczną rzeczywistych wartości zmiennej prognozowanej z próbki wstępnej np. z 3 pierwszych wartości.

Poszukując parametru α tak, aby błędy *ex post* były jak najniższe nie zawsze uzyskujemy w pełni zadowalające wyniki. W praktyce dobiera się $\alpha \leqslant 0{,}5$. Jeżeli okaże się, że parametr ten powinien być jednak większy od 0,5 wtedy najprawdopodobniej w szeregu występuje trend lub sezonowość.

W tabeli 2.6 oraz na wykresie na rysunku 2.4 znalazły się przykładowe prognozy dla dwóch wariantów modelu Browna różniących się stopniem wygładzenia.

2.3.4. Metoda Holta

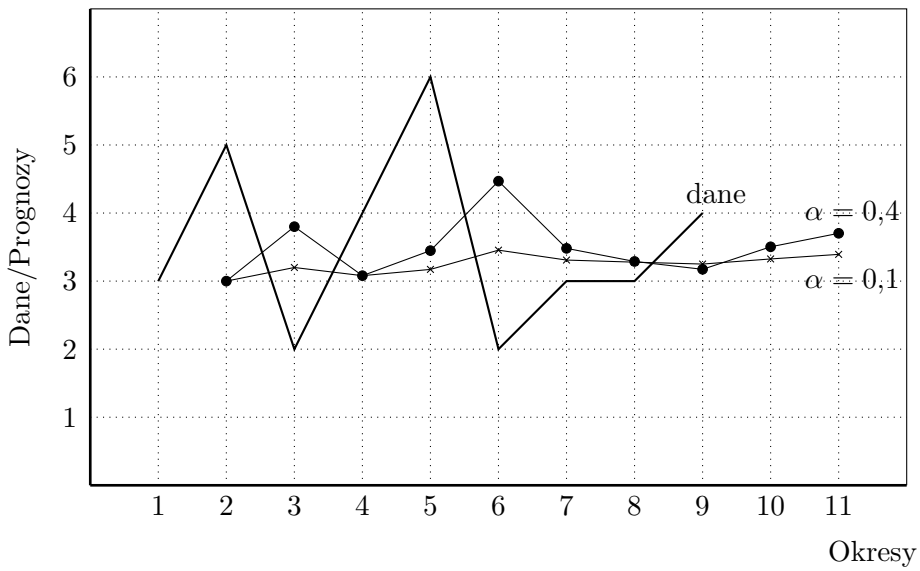
Metodę tę można traktować jako uogólnienie metody naiwnej z trendem liniowym[3]. Stosuje się ją dla szeregów, w których występują małe wahania przypadkowe i trend liniowy, lecz nie ma w nich sezonowości. Wygładzenie obejmuje zarówno poziom zmiennej jak i jej przyrost. W metodzie Holta

[3] W (Dittman 2004, s. 83) znajdziemy wersję modelu Holta dla szeregów zawierających trend hiperboliczny.

Tabela 2.6: Przykład prognoz metodą Browna

		Parametr wygładzania	
Okres (t)	Dane (y)	$\alpha = 0{,}1$	$\alpha = 0{,}4$
1	3		
2	5	3	3
3	2	3,2	3,8
4	4	3,08	3,08
5	6	3,172	3,448
6	2	3,4548	4,4688
7	3	3,3093	3,4813
8	3	3,2784	3,2888
9	4	3,2505	3,1733
10		3,3255	3,5040
11		3,3929	3,7024

Źródło: opracowanie własne



Rysunek 2.4: Prognozy i wygładzanie w metodzie Browna

Źródło: opracowanie własne

ustalamy dwa parametry: $\alpha \in \langle 0, 1\rangle$ – wygładzający wahania losowe oraz $\beta \in \langle 0, 1\rangle$ – parametr wygładzający trend. W odróżnieniu od poprzednio poznanych metod, w modelu Holta prognozy wyznaczamy w drugim kroku postępowania. Najpierw wygładzamy szereg za pomocą równań (2.15) oraz (2.16) (Gajda 2001, s. 189). Pierwsze usuwa z szeregu wahania przypadkowe

natomiast drugie służy do usuwania trendu. Prognozy *ex post* i *ex ante* obliczamy odpowiednio przy pomocy formuł (2.17) i (2.18).

$$F_t = \alpha y_t + (1-\alpha)(F_{t-1} + S_{t-1}) \tag{2.15}$$

$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1} \tag{2.16}$$

Prognozy *ex post* i *ex ante* dla modelu Holta:

$$y_t^* = F_{t-1} + S_{t-1} \tag{2.17}$$

$$y_t^P = F_n + hS_n \tag{2.18}$$

Po raz kolejny potrzebujemy wartości startowych (F_1 i S_1). Za F_1 można przyjąć y_1 lub średnią z kilku pierwszych wartości szeregu, natomiast za S_1 podstawimy pierwszą dostępną różnicę $y_2 - y_1$. Podobnie jak w przypadku metody Browna, parametry znajdujemy numerycznie minimalizując wybrany błąd prognoz *ex post*. W tabeli 2.7 zawarte zostały przykładowe prognozy metodą Holta z parametrami: $\alpha = 0{,}3$ i $\beta = 0{,}4$.

Tabela 2.7: Przykład prognozy metodą Holta

Okres (t)	Dane (y)	F_t	S_t	Prognoza
1	101	101	1	
2	102	102	1	102
3	104	103,3	1,12	103
4	104	104,294	1,0696	104,42
5	103	104,6545	0,7860	105,3636
6	105	105,3083	0,7331	105,4405
7	108	106,6290	0,9681	106,0415
8	109	108,0180	1,1365	107,5972
9	111	109,7081	1,3579	109,1545
10				111,0661
11				112,4240

Źródło: opracowanie własne

2.3.5. Metoda Wintersa

W metodzie Wintersa w szeregu tolerujemy wahania losowe, trend (liniowy) oraz sezonowość (zarówno addytywną jak i multiplikatywną). Zakładamy, że efekty sezonowe w skali roku sumują się do zera (lub ich iloczyn równy jest jedności). Wygładzeniu ulegają zatem: poziom zmiennej, przyrost oraz sezonowość (Gajda 2001, s. 191).

W metodzie Wintersa występują trzy parametry przyjmujące (podobnie jak we wcześniej opisanych metodach wygładzania wykładniczego) wartości z przedziału $\langle 0, 1\rangle$:
— α – parametr wygładzający wahania przypadkowe;
— β – parametr wygładzający trend;
— γ – parametr wygładzający sezonowość.

Jeśli efekt danego sezonu jest stały w czasie stosujemy model addytywny, a jeśli jest to stały udział efektów sezonowych w wartości zmiennej wersję multiplikatywną.

Model Wintersa – wariant addytywny:

$$F_t = \alpha(y_t - C_{t-r}) + (1-\alpha)(F_{t-1} + S_{t-1}) \tag{2.19}$$
$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1} \tag{2.20}$$
$$C_t = \gamma(y_t - F_t) + (1-\gamma)C_{t-r} \tag{2.21}$$

Model Wintersa – wariant multiplikatywny:

$$F_t = \alpha\frac{y_t}{C_{t-r}} + (1-\alpha)(F_{t-1} - S_{t-1}) \tag{2.22}$$
$$S_t = \beta(F_t - F_{t-1}) + (1-\beta)S_{t-1} \tag{2.23}$$
$$C_t = \gamma\frac{y_t}{F_t} + (1-\gamma)C_{t-r} \tag{2.24}$$

gdzie:
F_t – ocena wartości średniej (czyści z wahań losowych);
S_t – ocena przyrostu trendu;
C_t – ocena sezonowości.
r – długość cyklu sezonowego (dla 1/2 roku r=2, dla 1/4 roku r=4 itp.)

Prognozy dla modelu addytywnego:

$$y_t^* = F_{t-1} + S_{t-1} + C_{t-r} \tag{2.25}$$
$$y_t^P = F_n + hS_n + C_{t+h-r} \tag{2.26}$$

Prognozy dla modelu multiplikatywnego:

$$y_t^* = (F_{t-1} + S_{t-1})C_{t-r} \tag{2.27}$$
$$y_t^P = (F_n + hS_n)C_{t+h-r} \tag{2.28}$$

Podobnie jak w metodzie Holta, za F_1 przyjmujemy na przykład y_1, zaś za S_1 pierwszą dostępną różnicę. Pierwsze wartości równania wygładzającego z wahań sezonowych $(C_1, \ldots, C_r)$ określamy w zależności od rodzaju tych wahań:

— dla modelu addytywnego:

$$C_m = y_m - \bar{y} \tag{2.29}$$

— dla modelu multiplikatywnego:

$$C_m = \frac{y_m}{\bar{y}} \tag{2.30}$$

$$\bar{y} = \frac{1}{r}\sum_{i=1}^{r} y_i \tag{2.31}$$

W tabeli 2.8 zaprezentowaliśmy przykład prognoz otrzymanych przy pomocy addytywnego modelu Wintersa (sezonowość kwartalna) z parametrami wynoszącymi odpowiednio: $\alpha = 0{,}8$, $\beta = 0{,}9$ i $\gamma = 0{,}4$.

Tabela 2.8: Przykład prognoz metodą Wintersa

Okres (t)	Dane (y)	F_t	S_t	C_t	Prognoza
1	99			-4	
2	107			4	
3	106			3	
4	100	99	8	-3	
5	103	107	8	-4	103
6	108	106,2	0,08	3,12	119
7	109	106,056	-0,1216	2,9776	109,28
8	101	104,3869	-1,5144	-3,1548	102,9344
9	102	105,3745	0,7374	-3,7498	98,8725
10	110	106,7264	1,2904	3,1814	109,2319
11	111	108,0213	1,2945	2,9780	110,9944
12	103	106,7869	-0,9815	-3,4076	106,1610
13					102,0557
14					108,0055

Źródło: opracowanie własne

Lawton zwrócił uwagę na to, że w metodzie Wintersa pomimo dobrych prognoz całego szeregu prognozy jego składowych obarczone mogą być wyższymi błędami (Lawton 1998). Dla wariantu addytywnego zaproponował korektę wyników prognoz w sytuacji gdy zmienia się trend w trakcie obserwacji szeregu na przykład przechodząc w stały poziom zmiennej. Wtedy przez kilka okresów po zmianie model generuje wyższe błędy prognoz zaś efekty sezonowe nie sumują się, choć powinny, do zera. Błędy pojawią się także podczas wygładzania wahań losowych.

Można oczywiście nie robić nic. Jeśli prognozy szeregu jako całości są dobrej jakości i dysponujemy dostateczną liczbą obserwacji wtedy wyniki uzyskane metodą Wintersa w końcu ustabilizują się. Prognozy *ex ante* mają charakter krótkookresowy więc wyższe odchylenia w przeszłości w minimalnym stopniu wpłyną na wartości bieżące.

Drugim sposobem jest korekta warunków sezonowych tak, aby efekty sezonowe sumowały się do zera. Robi się to po prostu odejmując stosowną wartość. Ignorujemy jednak wtedy korektę poziomu zmiennej co skończy się wyższymi błędami przynajmniej do momentu aż model dostosuje się do nowego zachowania szeregu. Pojawia się jednak pytanie kiedy i jak często dokonywać korekty. Lawton zaleca wykonywanie jej co okres co osiąga modyfikując trzecie z równań modelu Wintersa mnożąc parametr γ przez $(r-1)/r$ gdzie r wyraża długość cyklu sezonowego.

Taylor zaproponował m.in. w (Taylor 2010) oraz (Taylor i Snyder 2001) rozszerzenie modelu Wintersa na przypadek prognozowania danych o wysokiej częstotliwości.

Wśród propozycji znalazła się modyfikacja modelu Wintersa zwana modelem IC (*Intraday Cycle*) dla przypadku szeregów, w których na sezonowość dzienną w ramach tygodnia nakłada się sezonowość godzinowa. Aby zwiększyć dokładność prognoz włącza on do modelu reszty z prognozy jako proces autoregresyjny przy czym celowo pomija równanie wygładzające trend, za to wykorzystuje dwa równania wygładzające wahania sezonowe, oddzielnie dla każdego z dwóch rozważanych poziomów sezonowości. Proponowany model (po pewnych przekształceniach) prezentuje się następująco:

$$y_t = l_{t-1} + d_{t-1} + w_{t-1} + \phi e_{t-1} + \varepsilon_t \tag{2.32}$$

$$e_t = y_t - (l_{t-1} + d_{t-m_2} + w_{t-m_2}) \tag{2.33}$$

$$l_t = \alpha(y_t - d_{t-m_1} - w_{t-m_2}) + (1-\alpha)l_{t-1} \tag{2.34}$$

$$d_t = \delta(y_t - l_t - w_{t-m_2}) + (1-\delta)d_{t-m_1} \tag{2.35}$$

$$w_t = \omega(y_t - l_t - d_{t-m_1}) + (1-\omega)w_{t-m_2} \tag{2.36}$$

W modelu danym równaniami (2.32)-(2.36) przyjęto następujące oznaczenia: l_t – równanie wygładzające z wahań losowych; d_t – równanie wygładzające z wewnętrznych (godzinowych) wahań sezonowych; w_t – równanie wygładzające z zewnętrznych (dziennych) wahań sezonowych; e_t – reszty pochodzące z prognoz *ex post*; m_1, m_2 – liczba okresów w cyklach sezonowych zewnętrznym i wewnętrznym.

Problem w podejściu Taylora i innych autorów zajmujących się modyfikacjami modelu Wintersa na przypadek nakładających się sezonowości stanowi liczba cykli, które trzeba brać pod uwagę. W każdej metodzie wygładzania wykładniczego potrzebne jest określenie punktu startowego, ale tutaj

wymagana jest znajomość co najmniej kilkuset wartości. Aby zmniejszyć tę niedogodność proponuje się wyodrębnianie podobnych wzorców zachowań szeregu dla tych samych pór różnych dni.

W (Corberán-Vallet, Bermúdez i Vercher 27) znajdziemy wielowymiarowy model wygładzania wykorzystujący analizę bayesowską. Ta praca jest o tyle ciekawa, że prognostyczny algorytm genetyczny przy wyznaczaniu prognoz *ex ante* również opiera się na prawdopodobieństwie warunkowym zaproponowanym przez Bayesa.

2.4. Błędy prognoz *ex post*

Wiemy już, że wykonując prognozę liczymy się z możliwością popełnienia błędu. Potrafimy go zmierzyć dopiero kiedy upłynie okres czasu, na który ustalono prognozę. Z tego powodu ocena jakości dokonywana jest przede wszystkim dla prognoz *ex post* choć w przypadku niektórych metod (np. modeli ekonometrycznych) istnieje możliwość oszacowania tzw. oczekiwanego błędu prognozy dla okresów *ex ante*.

Podstawową miarą oceny błędów *ex post* jest reszta z prognozy wyznaczana jako przyrost bezwzględny (2.37) lub względny (2.38) (por. m.in. (Cieślak 1997, s. 47)).

$$e_t = y_t - y_t^* \tag{2.37}$$

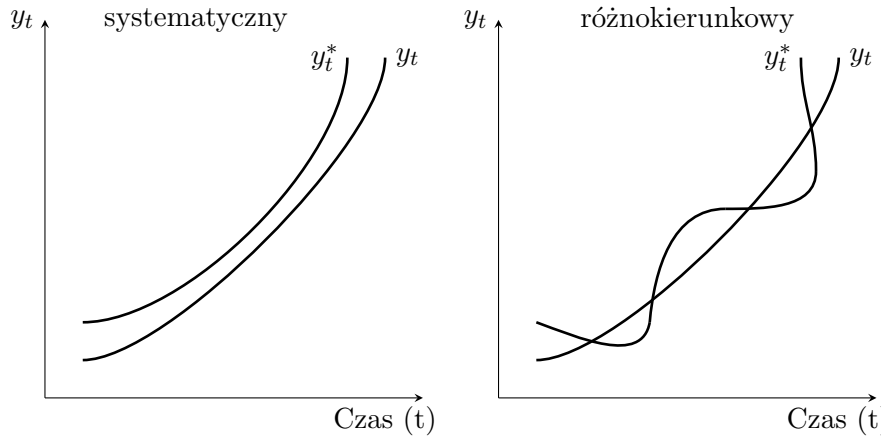
$$e_t = \frac{y_t - y_t^*}{y_t} \tag{2.38}$$

Znak reszty nie jest z góry określony. Dla prognoz przeszacowanych reszty przyjmują wartości ujemne, a dla niedoszacowanych dodatnie. Lepsza z dwóch to ta prognoza, dla której występują mniejsze błędy jednak porównywanie poziomów reszt z kolejnych okresów to proces mozolny. Dlatego oceniając jakość dłuższych szeregów czasowych, za lepszą uznajemy tą metodę, dla której mniejsze błędy występują pod koniec próby.

Ogólnie rzecz biorąc, z punktu widzenia zachowania szeregu prognoz względem szeregu danych błędy *ex post* dadzą się podzielić na dwie grupy:
1. Systematyczne.
2. Różnokierunkowe.

W przypadku błędu systematycznego prognozy są permanentnie przeszacowane lub niedoszacowane. Z błędem różnokierunkowym mamy zaś do czynienia jeśli część prognoz znajdzie się powyżej a część poniżej rzeczywistych realizacji zmiennych. Różnice między nimi ilustruje rysunek 2.5.

Pojedyncze reszty nadają się do oceny jakości prognoz w sytuacji gdy analizowane szeregi nie są długie. Tylko wtedy jesteśmy w stanie przeanalizować oddzielnie każdy okres. W praktyce takie postępowanie okazuje się



Rysunek 2.5: Błąd systematyczny i różnokierunkowy w prognozie *ex post*

Źródło: opracowanie własne

zbyt uciążliwe, dlatego opracowano miary uśredniające otrzymane wyniki. Występują one (podobnie jak same reszty) w dwóch wariantach: absolutnym i względnym. Poniżej prezentujemy charakterystykę wybranych miar na podstawie książki (Welfe 2003, s. 231–235). Występująca we wzorach wartość S oznacza całkowitą liczbę okresów, dla których obliczono prognozy *ex post*[4].

Błąd średni

$$ME = \frac{1}{S}\sum_{t=1}^{S}(y_t - y_t^*) \tag{2.39}$$

Jest to wielkość mianowana i z tego powodu nie nadaje się do porównywania szeregów wyrażonych w różnych jednostkach. Interpretujemy ją jako średnie przeszacowanie lub niedoszacowanie prognozy.

Średni błąd absolutny

$$MAE = \frac{1}{S}\sum_{t=1}^{S}|y_t - y_t^*| \tag{2.40}$$

Ta miara również ma miano takie samo jak prognozowana zmienna, ale jej znak jest z góry określony (dodatni). Interpretujemy ją jako średni co do modułu błąd prognozy.

[4] Nie dla każdej metody jest ona tożsama z ilością obserwacji występujących w szeregu. Przypomnijmy, że na przykład użycie metody średniej ruchomej oznacza poświęcenie części danych z początkowych okresów.

Błąd średniokwadratowy

$$MSE = \frac{1}{S}\sum_{t=1}^{S}(y_t - y_t^*)^2 \tag{2.41}$$

$$RMSE = \sqrt{MSE} \tag{2.42}$$

Miano *MSE* jest podniesione do kwadratu. Z tego powodu interpretujemy zazwyczaj jego wartość po wyciągnięciu pierwiastka czyli *RMSE*.

Średni błąd procentowy

$$MPE = \frac{1}{S}\sum_{t=1}^{S}\frac{y_t - y_t^*}{y_t} \tag{2.43}$$

Pierwsza z grupy miar względnych. Jak wszystkie pozostałe nie ma określonego znaku a sam błąd interpretujemy jako średnie względne (procentowe) przeszacowanie lub niedoszacowanie prognozy.

Średni absolutny błąd procentowy

$$MAPE = \frac{1}{S}\sum_{t=1}^{S}\left|\frac{y_t - y_t^*}{y_t}\right| \tag{2.44}$$

Interpretuje się go jako średni co do wartości bezwzględnej błąd popełniany podczas prognozy. Przyjęło się traktować jako dobre takie prognozy, dla których *MAPE* nie przekracza 5%.

Procentowy błąd średniokwadratowy

$$MSPE = \frac{1}{S}\sum_{t=1}^{S}\left(\frac{y_t - y_t^*}{y_t}\right)^2 \tag{2.45}$$

$$RMSPE = \sqrt{MSPE} \tag{2.46}$$

Wszystkie błędy względne możemy wyrazić w procentach. Nie mają one miana i nadają się do porównywania prognoz szeregów występujących w różnych jednostkach.

Między miarami błędów prognoz *ex post* zachodzą pewne związki, które da się wykorzystać w ocenie jakości użytej metody. Ważne jednak, aby pamiętać, że porównania dokonujemy tylko w obrębie tej samej grupy (bezwzględnych lub względnych) błędów.

Błąd średni może przyjmować wartości zarówno dodatnie jak i ujemne co oznacza, że reszty o przeciwnych znakach będą znosić się wzajemnie.

Nazywamy to zjawisko kompensacją reszt. W praktyce spotykamy się z nim dość powszechnie. Problem pojawia się, kiedy zarówno dodatnie jak i ujemne reszty przyjmują duże wartości. Zaniża to błąd *ME* (podobnie jak błąd *MPE*) sugerując dużo lepsze prognozy niż w rzeczywistości (silna kompensacja reszt). Dlatego błąd ten porównuje się z odpowiadającym mu błędem absolutnym. Zbliżone (w skrajnym przypadku identyczne) wartości *ME* oraz *MAE* (bądź *MPE* i *MAPE*) świadczą o systematycznym przeszacowaniu lub niedoszacowaniu prognoz.

Różne wartości oznaczają występowanie reszt różnokierunkowych. Szczególnie duża dysproporcja stanowi sygnał o kompensowaniu się dużych reszt dodatnich i ujemnych. W takiej sytuacji należy zrezygnować z oceny jakości za pomocą błędu średniego i oprzeć się tylko na błędzie absolutnym.

Inną niekorzystną sytuacją jest pojawianie się reszt z prognozy o nietypowo dużych wartościach przy czym mogą to być nawet pojedyncze (w skali całej prognozy) przypadki. Stanie się tak na przykład wtedy, kiedy zastosujemy metodę Holta do szeregów zachowujących się wykładniczo. Zestawienie błędu średniego i absolutnego nie pozwoli wychwycić tego faktu. Dlatego sięgamy po błędy średniokwadratowe. Porównujemy je z odpowiednimi błędami absolutnymi. Istotna różnica między nimi oznacza, że w okresie *ex post* wystąpiły reszty o nietypowej wysokości.

Tabela 2.9: Przykład błędów prognoz *ex post* dla średniej ruchomej

Okres (t)	Dane (y)	Prognoza	Reszty (e_t)	$\|e_t\|$	e_t^2
1	3				
2	5				
3	2				
4	4				
5	6	3,5	2,50	2,50	6,2500
6	2	4,25	-2,25	2,25	5,0625
7	3	3,5	-0,50	0,50	0,2500
8	3	3,75	-0,75	0,75	0,5625
9	4	3,5	0,50	0,50	0,2500
10		3	ME	MAE	MSE
11		3,25	-0,1	1,3	2,4750
					RMSE
					1,5732

Źródło: opracowanie własne

W tabeli 2.9 znalazły się błędy prognoz *ex post* obliczone dla danych pochodzących z przykładu ilustrującego wykorzystanie średniej ruchomej prostej. Obliczyliśmy wprawdzie miary bezwzględne, ale podobną analizę

da się wykonać również dla mierników wykorzystujących reszty względne. Ich analiza pokazuje, że prognozy są średnio przeszacowane o 0,1 jednostki. Jednak porównanie wartości *ME* z *MAE*, które wyraźnie różnią się między sobą, pozwala wysnuć wniosek, że doszło do znaczącej kompensacji reszt dodatnich i ujemnych. Rzeczywiście, w okresach piątym i szóstym mieliśmy do czynienia najpierw ze znacznym niedoszacowaniem a później przeszacowaniem prognoz. Z kolei porównanie średniego błędu absolutnego z *RMSE* sugeruje, że w szeregu reszt z prognoz *ex post* pojawiły się nietypowe wartości za co odpowiadają wyniki dla okresów 5 i 6.

Wymienione do tej pory miary z pewnością przodują w zastosowaniach praktycznych. Zamiast średniej arytmetycznej możemy jednak użyć mediany reszt z prognozy modyfikując w ten sposób każdy z powyższych mierników. Postępowanie takie niesie ze sobą dobrze znane ze statystyki konsekwencje, takie jak niewrażliwość na obserwacje skrajne czy ograniczony wpływ zmienności szeregu na interpretację.

Tashman zwraca uwagę (Tashman 2000) na fakt, że w sytuacji, kiedy przynajmniej część z nich przyjmuje wartości bliskie zeru reszty względne charakteryzują się silną asymetrią. Właśnie w takiej sytuacji przydają się miary oparte na medianie. Przyjrzyjmy się jak prezentują się medianowe wersje błędów z przykładu w tabeli 2.9.

Tabela 2.10: Błędy prognoz *ex post* w wariancie medianowym dla przykładu

Miara błędu	MdE	MdAE	MdSE	MdSE
Wartość	-0,5	0,75	0,5625	0,75

Źródło: opracowanie własne

Analizując wyniki z tabeli 2.10 stwierdzamy, że w połowie przypadków przeszacowanie prognozy nie przekroczyło 0,5 jednostki. Niezależnie od znaku 50% reszt było nie większych od 0,75 jednostki. Z uwagi na własności mediany, wszystkie cztery miary ignorują wartości z okresów 5 i 6.

Opisane miary błędów nie wyczerpują całkowicie tematu. W literaturze (m.in. (Theodosiou 2011)) możemy znaleźć inne sposoby oceny trafności prognoz *ex post*. Wymienimy tu tzw. symetryczny *MAPE* (*sMAPE*) zaproponowany w połowie lat osiemdziesiątych przez S. Armstronga. Jest to średnia arytmetyczna modułu reszt bezwzględnych podzielona przez średnie wartości rzeczywistych i prognoz dla danego okresu (wzór (2.47)).

$$sMAPE = \frac{1}{S}\sum_{t=1}^{S}\frac{|y_t - y_t^*|}{(y_t + y_t^*)/2} \tag{2.47}$$

Wartość *sMAPE* zawiera się w przedziale od 0 do 200% co stwarza pewne problemy podczas interpretacji. Dlatego możemy spotkać się z podejściem, w którym w mianowniku 2.47 występuje suma zamiast średniej.

Nazwa „symetryczny" *MAPE* jest myląca ponieważ miara ta nie traktuje jednakowo przeszacowanych i niedoszacowanych prognoz. W przypadku tych drugich *sMAPE* okazuje się wyższy co stanowi podstawowy stawiany mu zarzut. Można jednak spotkać się z wykorzystaniem *sMAPE* w prognozowaniu na podstawie sieci neuronowych.

Współczynnik Theila

Współczynnik Theila w porównaniu z wymienionymi wcześniej miarami ma bardziej złożony charakter. Obliczamy go następująco:

$$I^2 = \frac{\sum_{t=1}^{S}(y_t - y_t^*)^2}{\sum_{t=1}^{S} y_t^2} \tag{2.48}$$

Dla prognoz idealnych $I^2 = 0$. Wartość dana wzorem 2.48 jest tym wyższa im bardziej prognozy odbiegają od danych rzeczywistych. Sam współczynnik można rozbić na następujące składowe:

$$I^2 = I_1^2 + I_2^2 + I_3^2 \tag{2.49}$$

gdzie:

$$I_1^2 = \frac{(\bar{y} - \bar{y^*})^2}{\frac{1}{S}\sum_{t=1}^{S} y_t^2} \qquad I_2^2 = \frac{(s_r - s_p)^2}{\frac{1}{S}\sum_{t=1}^{S} y_t^2} \qquad I_3^2 = \frac{2 s_r s_p (1-r)}{\frac{1}{S}\sum_{t=1}^{S} y_t^2} \tag{2.50}$$

$$s_r = \sqrt{\frac{1}{S}\sum_{t=1}^{S}(y_t - \bar{y})^2} \quad s_p = \sqrt{\frac{1}{S}\sum_{t=1}^{S}(y_t^* - \bar{y^*})^2} \quad r = \frac{cov(y_t, y_t^*)}{s_r s_p} \tag{2.51}$$

Poszczególne składowe wzoru (2.49) wyrażają odpowiednio:

I_1^2 – obciążoność prognozy (stopień nieodgadnięcia średniej wartości zmiennej prognozowanej);

I_2^2 – niedostateczną elastyczność (stopień nieodgadnięcia wahań zmiennej);

I_3^2 – niedostateczną zgodność prognoz z rzeczywistym kierunkiem zmian zmiennej prognozowanej.

Aby uzyskać informacje o udziałach poszczególnych przyczyn w błędzie całkowitym obliczamy:

$$\hat{I}_1^2 = \frac{I_1^2}{I^2} \qquad \hat{I}_2^2 = \frac{I_2^2}{I^2} \qquad \hat{I}_3^2 = \frac{I_3^2}{I^2} \tag{2.52}$$

Dla modeli strukturalnych, w których dysponujemy wynikami z regresji istnieje możliwość obliczenia oczekiwanego błędu prognoz *ex ante*. Jest to rodzaj modyfikacji wariancji reszt z modelu. Dla równania o dowolnej liczbie zmiennych egzogenicznych zapiszemy (Gajda 2001, s. 154):

$$s_\tau^2 = \sigma^2 \left[\mathbf{x}_\tau (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{x}_\tau^T + 1\right] \tag{2.53}$$

gdzie:
τ – numer okresu *ex ante*, na który dokonujemy prognozy.
σ^2 – wariancja reszt z modelu.

Błędy prognoz *ex post* są ważnym jeśli nie najważniejszym sposobem oceny jakości predykcji. Kto wie jednak, czy równie istotna nie jest jakość prognoz poza próbą. Jak możemy przeczytać w (Tashman 2000) progności generalnie zgadzają się, że jakość poza próbą ma bardzo duże znaczenie. Praktyka pokazuje jednak, że często przeceniają dokładność prognoz w obrębie próby przywiązując mniejszą wagę do zachowania przyjętego modelu poza nią. Sam model i jego parametry dobiera się na podstawie danych historycznych a przecież metody sprawdzające się w okresie *ex post* nie muszą równie dobrze funkcjonować podczas okresów *ex ante*. Reszty z prognoz odzwierciedlają dostosowanie modelu do posiadanych danych. Umykają im jednak pewne niuanse z przeszłości czy przyszłości jak choćby oczekiwana zmiana trendu.

Ponadto tradycyjne podejście zakłada wykonywanie prognoz na jeden okres naprzód naraz. Prowadzi to do powiększania się błędów w dalszych okresach i może w konsekwencji ograniczyć użyteczny horyzont predykcji jak ma to miejsce na przykład w przypadku metod naiwnych.

Tashman proponuje podział danych historycznych na dwie części: dopasowaną i testową. Pierwsza pełni dotychczasową rolę czyli służy dopasowaniu metody do szeregu zaś druga przeznaczona została do testowania faktycznej jakości prognoz. Okres testowy pełni *de facto* rolę prognoz wygasłych dla metody.

Jakość prognoz wykonanych dla zbioru testowego da się ocenić na dwa sposoby. W pierwszym z nich wyznaczamy jeden szereg prognoz poczynając od okresu pierwszego po okresie wyznaczonym jako granica między dwiema podpróbami. Jeżeli taki okres oznaczymy jako T zaś do końca całego szeregu zostały licząc od niego 4 obserwacje wtedy jesteśmy w stanie wyznaczyć

prognozy na okresy $T+1$, $T+2$, $T+3$ i $T+4$. Następnie dla tychże okresów obliczamy błędy prognoz, które można potem uśrednić.

Takie zachowanie ma jednak wady. Aby ocenić prognozy w okresie testowym oraz aby ocena ta niosła ze sobą konkretną wartość potrzebujemy odpowiednio długich szeregów danych. Tylko większa liczba prognoz wygasłych daje pewność co przydatności użytej metody. Poza tym pojawienie się choćby jednej obserwacji odstającej negatywnie wpłynie na ocenę dopasowania prognoz do danych. Po trzecie wreszcie, uśredniona miara błędów bierze pod uwagę jednocześnie wszystkie odchylenia czyli pochodzące z okresów zarówno bliskich jak i dalekich od punktu T. Tym niemniej postępowanie w ten sposób wciąż może być użyteczne jeśli dysponujemy tylko jednym szeregiem obserwacji dla prognozowanego zjawiska i nie jesteśmy w stanie uzyskać kolejnych replikacji wykorzystanej metody.

Drugim sposobem pomiaru błędów prognoz poza okresem T jest przyjęcie zmiennego początku prognoz w próbie testowej. Uzyskamy kilka szeregów prognoz wygasłych. Przyjmując dotychczasowe oznaczenia najpierw otrzymamy cztery prognozy w okresach od $T+1$ do $T+4$. Następnie cały szereg dzielimy ponownie przy czym okresy danych dopasowanych rozszerzamy włączając do niego dotychczasowy punkt $T+1$. Uzyskamy w ten sposób trzy prognozy $T+2$, $T+3$ i $T+4$. W kolejnym kroku dołączamy okres $T+2$ (powstaną dwie prognozy poza próbą służącą dopasowaniu metody) a w końcu okres $T+3$. W tym konkretnym wypadku daje to 10 prognoz, których jakość możemy ocenić przy pomocy znanych nam już miar błędów. Jeżeli N oznacza liczbę okresów następujących po punkcie T wtedy opisane przed chwilą postępowanie dostarczy $N(N+1)/2$ prognoz.

Podczas testowania prognoz poza próbą szczególnego znaczenia nabiera punkt podziału szeregu. Tashman zaleca, aby długość szeregu wydzielonego do testu równała się co najmniej horyzontowi prognozy *ex ante*, którą mamy zamiar docelowo wykonać. Do tematyki wykorzystania prognoz wygasłych do oceny wybranej metody prognozowania wrócimy jeszcze w rozdziale poświęconym prognozowaniu kaskadowemu.

2.5. Charakterystyka metod analitycznych

U podstaw metod analitycznych[5], do których zaliczamy m.in. modele trendu i autoregresyjne leży założenie, że istnieje mechanizm generujący wartości zmiennej, który zawiera się w zrealizowanych wartościach. Opierając się na tych założeniach buduje się modele o różnym stopniu skomplikowania. Przedstawimy ogólną charakterystykę niektórych z nich.

[5] W literaturze (Gajda 2001, s. 196) możemy spotkać także termin modele szeregów czasowych.

Modele trendu

W przypadku gdy zmienna przejawia skłonność do systematycznych zmian w czasie na przykład stałych wzrostów lub spadków mówimy, że zawiera trend czasowy. Do jego opisu wykorzystujemy sztuczną zmienną reprezentującą upływ czasu. Zazwyczaj oznaczana ona jest przez t a jako swoje wartości przyjmuje numery kolejnych okresów ($t = 1,2,\ldots,n$). Oprócz zmiennej sztucznej w funkcji występuje składnik losowy (ε_t), który wprowadzony addytywnie pozwala zapisać ogólną postać funkcji następująco:

$$y_t = f(t) + \varepsilon_t \tag{2.54}$$

Jeżeli założymy, że ε_t ma wartość oczekiwaną zero, stałą wariancję oraz nie podlega autokorelacji wtedy parametry równania (2.54) da się oszacować metodą najmniejszych kwadratów (MNK) (Gajda 2001, s. 197). Można więc powiedzieć, że modele trendu stoją na pograniczu metod prognozowania szeregów czasowych oraz modeli ekonometrycznych. Z jednej strony bowiem potrzebujemy jedynie informacji o wartościach zmiennej objaśnianej (szeregu czasowym), a z drugiej szacujemy parametry przy pomocy MNK charakterystycznej dla modeli strukturalnych.

Postać funkcyjna samego modelu bywa różna, a jej wybór zależy od przesłanek dotyczących mechanizmu rozwojowego zmiennej, zazwyczaj określanego na podstawie analizy wykresu. Z uwagi na wykorzystanie sztucznej zmiennej, istnieje duża mnogość funkcji, które można dopasować do szeregu czasowego. Przegląd najczęściej używanych postaci funkcji zawierają się m.in. (Gajda 2001, s. 198–199) czy (Cieślak 1997, s. 78–80). Funkcjonalność modeli tego rodzaju da się rozszerzyć wprowadzając zmienne zero-jedynkowe co pozwala m.in. na uwzględnienie wpływu efektów sezonowych.

Dopasowanie konkretnej funkcji trendu wymaga niekiedy sporej ilości obserwacji. Kiedy ich brak, do szeregu potrafimy zwykle dopasować więcej niż jedną funkcję. W takiej sytuacji wybieramy tą o najprostszej postaci analitycznej. Jako kryterium rozstrzygające o wyborze postaci funkcji używamy zazwyczaj współczynnika determinacji[6] (R^2). Innym sposobem mogą być proponowane przez Dittmana (Dittman 2004, s. 66–68) testy statystyczne jako sposób identyfikacji trendu.

Jedną z możliwych opcji jest sięgnięcie po funkcję wielomianową. Jednak jak zauważa Zeliaś (Zeliaś 1997, s. 75) przydatność tych modeli podlega ograniczeniom. Zwykle brak racjonalnego uzasadnienia, że poza zakresem próby przebieg wielomianu będzie rzeczywiście odpowiadał tendencji rozwojowej badanej zmiennej. Z tego powodu używa się wielomianów co

[6] R^2 przyjmuje wartości z przedziału $\langle 0, 1\rangle$ i opisuje dopasowanie modelu do danych rzeczywistych

najwyżej piątego, szóstego stopnia. Jeśli uzyskane prognozy nadal nie mają satysfakcjonującej dokładności, należy rozważyć zmianę postaci funkcji.

Używając metody najmniejszych kwadratów do szacowania parametrów trzeba pamiętać o sprowadzaniu równania do postaci liniowej względem parametrów (w przeciwnym wypadku nie możemy użyć tej metody). Prognozę otrzymujemy przez ekstrapolację funkcji trendu czyli podstawianie numeru okresu, na który wyznaczamy prognozę. Horyzont prognoz należy dobierać ostrożnie ponieważ zwykle trudno uwierzyć w niezmienną, utrzymującą się w długim okresie tendencję. Same prognozy tworzone są przy następujących założeniach:

1. Postać analityczna funkcji nie ulegnie zmianie.
2. Parametry w okresie prognozy nie zmienią się.

Pamiętajmy, że wybór postaci analitycznej funkcji trendu staje się niezwykle istotny przy prognozach o dalszym horyzoncie. Pomyłka podczas estymacji parametrów lub w wyborze samej funkcji może skończyć się dramatycznie wysokimi błędami podczas ekstrapolacji poza zbiór obserwacji. Zjawisko to jest szczególnie widoczne w przypadku funkcji wykładniczych czy wielomianów. Zdarza się wówczas, że prognozy *ex ante* przyjmują nieprawdopodobnie wysokie bądź niskie wartości.

Modele autoregresyjne

Zdarza się, że dla niektórych zjawisk obserwujemy zależność od ich wartości z poprzednich okresów. Sięgamy wówczas po tzw. modele autoregresyjne, w których prognozy opisujemy przy pomocy funkcji, w której jako zmienne objaśniające występują wartości zmiennej objaśnianej z okresów poprzedzających okres badany (Cieślak 1997, s. 93). Ten typ modeli wykorzystywanych również do prognozowania szeregów czasowych zyskał ostatnimi czasy na popularności czego dowodem są liczne publikacje dotyczące wzmiankowanej tematyki.

Załóżmy, że ε_t jest czysto losowym procesem o zerowej średniej i wariancji σ^2. Zdefiniujmy następujący proces dotyczący szeregu czasowego (Maddala 2008, s. 585):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t \tag{2.55}$$

Model (2.55) nazywamy procesem autoregresyjnym rzędu p i oznaczamy jako AR(p). Równanie to wyraża regresję y_t względem p jego przeszłych wartości. Choć powyższe równanie ma postać liniową, można oczywiście sięgnąć również po zależności o nieliniowym charakterze. Jego parametry szacuje się zazwyczaj metodą najmniejszych kwadratów pomimo tego, ze nie daje ona najbardziej efektywnych estymatorów. Z tego powodu sięga się także po metodę największej wiarygodności.

Prognoza powstaje poprzez podstawienie do modelu wartości zmiennej objaśnianej z poprzednich okresów, przy czym predykcja na dalsze okresy ma charakter dynamiczny czyli wykorzystuje prognozy obliczone dla poprzednich okresów.

Modele autoregresyjne łączy się z modelami średniej ruchomej. Załóżmy, że ε_t spełnia warunki jak poprzednio. Procesem średniej ruchomej rzędu q (MA(q)) nazywamy proces dany wzorem (Maddala 2008, s. 583):

$$y_t = \beta_0\varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q} \tag{2.56}$$

gdzie: $\beta_0 = 1$

Zapiszmy kombinację obu modeli czyli model ARMA(p,q):

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_p y_{t-p} + \varepsilon_t + \beta_1\varepsilon_{t-1} + \ldots + \beta_q\varepsilon_{t-q} \tag{2.57}$$

W przypadku modeli ARMA szczególną uwagę zwraca się na stacjonarność szeregu. Proces stochastyczny nazywamy słabo stacjonarnym jeżeli jego wartość oczekiwana oraz wariancja są stałe tj. nie zależą od momentu czasu, w którym je mierzymy (Gajda 2004, s. 139). Z kolei w szeregu niestacjonarnym wariancja lub kowariancja mogą być zmienne w czasie. Modele rozszerzone o stacjonarność nazywamy zintegrowanymi rzędu d (ARIMA(p,d,q)). Jak łatwo zauważyć, ostatni sposób zapisu ma najbardziej ogólny i uniwersalny charakter. Na przykład gdy d i q równają się zero, otrzymujemy model AR rzędu p.

Kłopot sprawia określenie stopnia opóźnienia części autoregresyjnej oraz średniej ruchomej a także postaci funkcji. Jeżeli jednak proces jest stacjonarny wówczas wyboru opóźnienia można dokonać na podstawie funkcji autokorelacji cząstkowej.

Modele autoregresyjne możemy budować dla różnych rozkładów opóźnień. Jeżeli ograniczymy się jedynie do wpływu wartości zmiennej z okresu bezpośrednio poprzedzającego uzyskujemy tzw. model błądzenia losowego (*random walk*). W swej najprostszej postaci (Gajda 2001, s. 209) ma on postać:

$$y_t = y_{t-1} + \varepsilon_t \tag{2.58}$$

Będącą szczególnym przypadkiem najprostszego modelu autoregresyjnego:

$$y_t = \alpha y_{t-1} + \varepsilon_t \tag{2.59}$$

Równanie (2.58) to właśnie model błądzenia losowego. Co do parametru α z równania (2.59) to czynimy założenie, że jest on mniejszy od jedności. Wówczas wpływ zakłócenia z upływem czasu zmniejsza się. W naszej pracy pozwalamy sobie jedynie zasygnalizować powyższe zagadnienia jako związane z procesem autoregresyjnym. Zainteresowanych odsyłamy do literatury.

Rozdział 3

Algorytm genetyczny w prognozowaniu

W pewnych sytuacjach nie możemy uzyskać satysfakcjonujących prognoz. Wiąże się to najczęściej z niedogodnymi z punktu widzenia dekompozycji elementami szeregu czasowego (np. układanie się obserwacji w serie czy pojawianie się punktów zwrotnych). Bywa, że problem stanowi długość szeregu – zbyt krótki daje szansę jedynie na krótkookresowe prognozy, możliwości jakie daje długi szereg nie wykorzystuje z kolei znakomita większość narzędzi predykcji. Jeżeli zawodzą metody rodem z którejś ze standardowych grup, sytuacja staje się nie do pozazdroszczenia. Dlatego warto poszukiwać innych rozwiązań, na przykład sięgając po algorytm genetyczny.

Punktem wyjścia dalszych rozważań stały się modele niestrukturalne. Choć od strony technicznej proste, nie oznacza to, że automatycznie dają prognozy gorszej jakości niż w przypadku bardziej skomplikowanych podejść. Przypomnijmy jeszcze, iż u podstaw prognozowania niestrukturalnego leży założenie, że cała potrzebna do wykonania prognozy informacja znajduje się w samym szeregu czasowym. Założenie to ułatwia konstrukcję algorytmu genetycznego dedykowanego predykcji.

Algorytmy genetyczne/ewolucyjne można również połączyć z metodami prognozowania strukturalnego. Przykład takiego wykorzystania znajdziemy w (Beenstock i Szpiro 2001), gdzie z bloków budujących odpowiadających różnym elementom modelu ekonometrycznego (zmiennym objaśniającym, opóźnieniom zmiennej objaśnianej, operatorom matematycznym itp.) tworzone są równania prognozujące nieliniowe szeregi czasowe. W artykule (Venkatesan i Kumar 2002) przeczytamy o użyciu algorytmu genetycznego do poszukiwania estymatorów parametrów opisujących zachowania rynku podczas fazy wzrostu liczby abonentów sieci komórkowych.

3.1. Prognostyczny algorytm genetyczny

W oparciu o reguły rządzące metodami naiwnymi zbudujemy, a następnie przetestujemy algorytm genetyczny, który ma tworzyć prognozy szeregów czasowych zawierających wahania przypadkowe, stały poziom zmiennej lub trend liniowy oraz sezonowość. Szczególną uwagę zwrócimy na metodę

naiwną bez zmian oraz naiwną z sezonowością, w których prognoza na dany okres kształtuje się na poziomie wartości zaobserwowanej w okresie poprzednim bądź wcześniejszym. Dzięki temu dobrze nadają się one do zaimplementowania w algorytmie genetycznym, który miałby tworzyć prognozy.

Warto zwrócić uwagę, że metody będące inspiracją da tworzonych chromosomów, podlegają silnym ograniczeniom, jeśli chodzi o dekompozycję szeregu czasowego. Opisany w dalszej części rozdziału algorytm jest pod tym względem o wiele bardziej elastyczny, pomimo tego że nie zmieniamy ogólnej idei mechanizmu generowania prognoz. Nadal będziemy poszukiwać pośród przeszłych realizacji takiej wartości, która posłuży do otrzymania prognoz na bieżący okres.

Użyty algorytm stanowi modyfikację klasycznego algorytmu genetycznego (KAG) i został nazwany prognostycznym algorytmem genetycznym (PAG). Różnica najbardziej widoczna jest w przypadku budowy pojedynczego osobnika. Zastosowaliśmy kodowanie rzeczywiste w miejsce binarnego. Pojedynczy gen wyraża rząd opóźnienia obserwacji będącej prognozą na dany okres, a cały chromosom jest jednym z możliwych wariantów prognostycznych obejmujących cały szereg. W konsekwencji wprowadziliśmy zmiany w operatorach krzyżowania i mutacji.

W pierwszym wypadku losujemy punkt krzyżowania dla wybranych rodziców a następnie dokonujemy wymiany informacji między chromosomami w sposób znany z reprezentacji binarnej KAG. Oznacza to, że geny rodziców wymieniane są za punktem krzyżowania z tym, że teraz mamy do czynienia z wartościami rzeczywistymi zamiast binarnych.

Nieco więcej zmian zastosowano w przypadku mutacji. Wprawdzie nadal polega ona na zmianie wartości wybranego genu, lecz jej zasięg został ograniczony przez parametr opisujący maksymalne opóźnienie przyjęte dla prognozy (szerzej o nim będzie mowa w dalszej części pracy). Poza tym prawdopodobieństwa krzyżowania i mutacji podlegały adaptacji, którą opisano w (Agustín-Blas et al. 2012).

Przyjmijmy na chwilę punkt widzenia Riechmanna (Riechmann 2001), który porównał algorytm genetyczny do dynamicznej gry ewolucyjnej. Analizował on przydatność tego narzędzia w zastosowaniach ekonomicznych, dla których chromosom (zwany przez autora agentem) wyraża pewną strategię. W naszym przypadku pojedynczy osobnik odpowiadający wariantowi prognostycznemu również może być uznany za agenta stosującego pewną strategię, której cel stanowi jak najbardziej dokładna prognoza. Podobnie postąpił Maschek (Mascheck 2010), u którego chromosom reprezentuje zbiór reguł decyzyjnych w okresie t.

Funkcję przystosowania stanowić może dowolna miara oceniająca trafność prognoz. Najbardziej oczywistym (choć nie jedynym) kierunkiem poszukiwać kryterium są błędy prognoz *ex post* – dokładniej ich moduły lub

kwadraty. Błędy średnie *ME* i *MPE* podlegają kompensacji reszt i z tego powodu nie nadają się na funkcję przystosowania.

Analiza próbnych uruchomień algorytmu wykazała, że *RMSPE* dobrze sprawdza się w tej roli ponieważ zapewnia bardziej jednoznaczne (w porównaniu średnich błędów absolutnych) wyniki. Okazało się, że wykorzystanie *MAPE* stwarza pewne problemy. Kilkukrotne uruchomienie procedury przy tych samych parametrach algorytmu dawało niekiedy osobniki o tej samej wartości funkcji przystosowania, lecz różnym genotypie. Nie stanowiłoby to może dużego problemu, wszak naszym celem jest otrzymywanie prognoz *ex ante*, lecz mediany (podobnie jak dominanty) opóźnień w kolejnych podejściach przyjmowały różne wartości. Wyniki otrzymywane przy użyciu jako kryterium *RMSPE* okazały się bardziej stabilne. W tym wypadku otrzymywaliśmy chromosomy o podobnych genotypach.

Funkcja przystosowania będąca miarą dopasowania prognoz *ex post* do danych rzeczywistych winna przyjmować jak najniższe wartości. Klasyczny algorytm genetyczny zakłada maksymalizację funkcji przystosowania. Aby spełnić nasze oczekiwania co do zachowania algorytmu musimy przekształcić standardową postać kryterium. Ponieważ używane przez nas miary przyjmują tylko wartości nieujemne (a w praktycznych zastosowaniach jedynie dodatnie), wartość funkcji przystosowania transformujemy wg formuły:

$$\min F(x) = \max(-F(x) + C) \tag{3.1}$$

gdzie:
$F(x)$ – wartość wybranego, uśrednionego błędu prognoz *ex post*
C – stała, określana *a priori* na wejściu

Poziom stałej C ma spore znaczenie, jeśli chodzi o zastosowaną funkcję przystosowania. Musi ona zapobiec pojawieniu się ujemnych wartości tej funkcji i stosownie do tego należy ustalić wartość C.

Choć większość zastosowań algorytmów genetycznych i ewolucyjnych ma charakter właśnie maksymalizacji funkcji przystosowania, istnieją przykłady zastosowań, których celem jest minimalizacja przyjętego na przykład (Zou, Kong i Li 2002) lub (Locatelli i Schoen 2003).

Podczas testowania algorytmu wprowadziliśmy jeszcze jedną modyfikację w stosunku do KAG. Selekcja chromosomów do reprodukcji odbywała się metodą elitarną i wartości oczekiwanej. Jako kryterium zatrzymania algorytmu przyjęliśmy ustalaną na początku obliczeń liczbę pokoleń.

Podstawą prognoz *ex post* (a przy okazji zasadniczą różnicą między klasyczną metodą naiwną a jej modyfikacją wprowadzoną w algorytmie) jest parametr, określany dalej jako t_m. Wyraża on maksymalny, dopuszczalny numer opóźnienia danej rzeczywistej będącej prognozą dla aktualnego okresu. Taka definicja oznacza, że prognozę *ex post* dla okresu o numerze i może

tworzyć dowolna obserwacja, nie starsza jednak niż $i - t_m$. Sam parametr t_m może przyjmować wartości z przedziału $\langle 1, n-2 \rangle$ gdzie n oznacza liczbę posiadanych obserwacji. Wartość t_m ustalamy *a priori* na początku postępowania i pozostaje on niezmienny do końca obliczeń. Przyjęliśmy, że punktem startowym będzie prognoza wykonana na okres trzeci. Chodziło o zapewnienie algorytmowi przynajmniej minimalnej swobody w ustalaniu wartości pierwszego genu. Rozpoczynanie od okresu drugiego oznaczałoby *de facto* każdorazowo kopiowanie do niego pierwszej obserwacji szeregu. Gdyby odbiegała ona co do swojej wartości od pozostałych danych już na początku prognozy zostałyby obarczone dużym błędem.

Wyniki predykcji dla pojedynczego okresu niekoniecznie muszą sięgać w przeszłość na największą możliwą głębokość. Tabela 3.1 ilustruje zbiory wartości mogących utworzyć prognozy dla hipotetycznego szeregu czasowego złożonego z 6 obserwacji, w którym przyjęto $t_m = 3$. Dla okresu trzeciego t_m może być równe 1 lub 2, dla okresu czwartego będzie to jedna z liczb od 1 do 3 itd.

Tabela 3.1: Zbiory potencjalnych prognoz *ex post* uzyskanych PAG w kolejnych okresach

Okres (t)	Wartość rzeczywista	Zbiór potencjalnych prognoz dla okresu
1	y_1	
2	y_2	
3	y_3	y_1, y_2
4	y_4	y_1, y_2, y_3
5	y_5	y_2, y_3, y_4
6	y_6	y_3, y_4, y_5

Źródło: opracowanie własne

Wartość prognozy w okresie t równa się jednej z obserwacji pochodzących sprzed $t - t_m$ okresów, np. w okresie trzecim będzie to y_{t-1} lub y_{t-2}, stosownie do wartości genu. W tabeli 3.2 prezentujemy jeden z możliwych zestawów prognoz *ex post* dla tego samego co poprzednio hipotetycznego szeregu. Trzecia kolumna zawiera wartość maksymalnego dopuszczalnego opóźnienia obserwacji będącej prognozą na dany okres. To oczywiście tylko jeden z potencjalnych wariantów, podany w charakterze przykładu.

W literaturze znajdujemy inne próby operowania obserwacjami z szeregu do tworzenia prognoz w jakiś sposób wykorzystujących optymalizację heurystyczną, które razem tworzą swego rodzaju rodzinę metod. Na przykład w artykule (Pasley i Austin 2004) występuje parametr wyrażający liczbę najświeższych obserwacji zmiennej użytą do skonstruowania progno-

zy. Autorzy wykorzystują ciąg wartości pochodzących z następujących po sobie okresów, a następnie sięgają po sieć neuronową.

W pracy (Lukoseviciute i Ragulskis 2010) znajdziemy próbę użycia algorytmu ewolucyjnego do określenia optymalnego zestawu opóźnień szeregu czasowego. Chromosom ma postać macierzy, a funkcja przystosowania opiera się na atraktorze. Trzeba jednak zwrócić uwagę, że autorzy ograniczyli się jedynie do szeregów nie dłuższych niż 30 obserwacji, a sam algorytm miał bardzo uproszczoną konstrukcję. Podobne rozwiązania wykorzystano w (Ragulskis, Lukoseviciute, Navickas i Palivonaite 2011).

Tabela 3.2: Przykładowe realizacje prognoz *ex post* dla PAG

Okres (t)	Prognoza *ex post*	Wartość t_m
1		
2		
3	y_1	2
4	y_1	3
5	y_4	1
6	y_3	3

Źródło: opracowanie własne

Interesujące podejście zaprezentowano w artykule (Granger i Jeon 2003). Autorzy wykonywali symulacje na modelach autoregresyjnych i średniej ruchomej, używając zarówno wartości opóźnionych, jak również pochodzących z okresów przyszłych względem tego, dla którego liczona jest prognoza.

W opisywanym przez nas postępowaniu wzrost liczby osobników w populacji pozwala na poprawę rezultatów. Podobnie rzecz ma się w razie powiększania liczby pokoleń. W obu przypadkach wytłumaczenie jest podobne i ma związek ze sposobem tworzenia chromosomu. Poszczególne osobniki różnią się bowiem między sobą ułożeniem genów. Zwiększanie liczby osobników lub pokoleń daje w rezultacie potencjalnie większą liczbę przeglądanych wariantów prognoz.

Zastosowanie kodowania rzeczywistego w PAG nie wyklucza krzyżowania chromosomów w sposób znany z reprezentacji binarnej, ponieważ geny przechowują informację na temat wartości opóźnienia obserwacji będącej prognozą na dany okres.

Rozważmy następujący przykład. Założymy w nim, że dysponujemy parą rodziców R_1 i R_2, z których każdy składa się z dziesięciu genów. Przyjmujemy poza tym, że maksymalny, dopuszczalny rząd opóźnienia dla prognozy *ex post* czyli t_m wynosi 4, punkt krzyżowania zaś wypadł pomiędzy trzecim a czwartym genem.

Podobnie jak miało to miejsce w przypadku krzyżowania binarnego dokonujemy wymiany genów rodziców za punktem krzyżowania. Niech zatem chromosomy rodziców mają postać:

$$R_1 = (\ \mathbf{2}\ \ \mathbf{3}\ \ \mathbf{3}\ \mid\ \mathbf{2}\ \ \mathbf{4}\ \ \mathbf{3}\ \ \mathbf{2}\ \ \mathbf{4}\ \ \mathbf{4}\ \ \mathbf{4}\ )$$

$$R_2 = (\ 1\ \ 2\ \ 3\ \mid\ 1\ \ 3\ \ 3\ \ 1\ \ 1\ \ 3\ \ 4\ )$$

Po wymianie genów otrzymamy następujących potomków P_1 i P_2:

$$P_1 = (\ \mathbf{2}\ \ \mathbf{3}\ \ \mathbf{3}\ \mid\ 1\ \ 3\ \ 3\ \ 1\ \ 1\ \ 3\ \ 4\ )$$

$$P_2 = (\ 1\ \ 2\ \ 3\ \mid\ \mathbf{2}\ \ \mathbf{4}\ \ \mathbf{3}\ \ \mathbf{2}\ \ \mathbf{4}\ \ \mathbf{4}\ \ \mathbf{4}\ )$$

Jak widać, nie uzyskaliśmy nieprawidłowych (czyli np. większych od 4) wartości genów, a co za tym idzie otrzymane prognozy również będą zgodne z założonym poziomem t_m.

Operator mutacji wymaga nieco więcej uwagi z racji tego, że pierwsze $t_m - 1$ genów nie pochodzi z całego dostępnego przedziału[1]. Pierwszy poddany mutacji gen może zmienić się z 1 na 2 lub odwrotnie. Następny będzie równy 1, 2 lub 3 itd. Trzymając się podanego przed chwilą przykładu, dopiero trzeci gen podlega mutacji w całym dostępnym przedziale. Zatem programując algorytm należy pamiętać o sterowaniu zasięgiem mutacji.

Przyjęty sposób tworzenia prognoz *ex post* powoduje, że możemy określić maksymalną możliwą liczbę powstających szeregów (jeden chromosom równa się jednemu szeregowi prognoz). Znamy dzięki temu rozmiary zbioru (liczbę różniących się między sobą chromosomów), w którym poszukujemy rozwiązania, a który w algorytmach genetycznych nosi nazwę populacji. Maksymalną liczbę możliwych do uzyskania różnych chromosomów da się wyznaczyć analitycznie:

$$L_O = t_m^{n-tm}(t_m - 1)! \tag{3.2}$$

gdzie:
L_O – liczba różnych osobników,
t_m – maksymalne dopuszczalne opóźnienie,
n – liczba obserwacji szeregu czasowego.

W tabeli 3.3 znalazły się przykładowe liczebności populacji w zależności od założonych parametrów. Na tej podstawie wyciągamy wniosek, że wraz z wydłużaniem się szeregu czasowego i wzrostem parametru t_m gwałtownie powiększa się zbiór rozwiązań. Z punktu widzenia algorytmu genetycznego oznacza to konieczność operowania dużymi populacjami, co znacząco wpływa na czas obliczeń.

[1] Tej wady pozbawiona jest, omówiona w następnej części rozdziału reprezentacja chromosomu bazująca na średniej ruchomej.

Tabela 3.3: Maksymalna liczba osobników w populacji przy zastosowanym algorytmie

Liczba obserwacji	$t_m = 3$	$t_m = 4$	$t_m = 5$
10	4 374	24 576	75 000
11	13 122	98 304	375 000
12	39 366	393 216	1 875 000
13	118 098	1 572 864	9 375 000

Źródło: opracowanie własne

Analiza tabeli 3.1 wykazuje, że ostatnia dostępna wartość rzeczywista nie jest wykorzystywana w procesie wyznaczania prognoz *ex post*. Służy ona za to do obliczania prognoz *ex ante*. Tego typu postępowanie nie jest niczym nowym w predykcji szeregów czasowych i występuje chociażby w metodach naiwnych czy średniej ruchomej. Prognostyczny algorytm genetyczny różni się tylko tym, że najpierw musimy zdekodować chromosom do postaci, w której prognoza staje się dla nas czytelna.

Mechanizm tworzenia prognoz *ex ante* przechodzi następujące etapy:

1. Wyznaczamy prognozę *ex post*, którą staje się chromosom zapewniający najbardziej korzystną wartość funkcji przystosowania.
2. Określamy wartość opóźnienia wskazującego, które dane rzeczywiste wezmą udział w tworzeniu prognoz *ex ante*. Do tego celu mogą posłużyć np. dominanta lub mediana wartości opóźnień najlepszego chromosomu.
3. Wyznaczamy prognozę *ex ante* stosownie do wyników dekompozycji.

Etap trzeci różnie wygląda w zależności od tego, jakie elementy dekompozycji zostaną wyodrębnione w szeregu. W razie wystąpienia stałego poziomu zmiennej nasze postępowanie staje się analogiczne do metody naiwnej prostej (ewentualnie naiwnej z sezonowością). Prognozą *ex ante* może zostać wartość rzeczywista z ostatniego okresu lub któregoś z okresów wcześniejszych, wskazanych przez dominantę bądź medianę opóźnień. Ostatnia dana rzeczywista wyznacza kres prognoz *ex ante*.

W kolejnych okresach wartości zaczną się powtarzać, jak to ma miejsce w metodach naiwnych. Dla szeregów, w których stwierdzono trend liniowy naśladujemy postępowanie znane z metody naiwnej z poprawką liniową omówione w następnej części rozdziału.

Powyższe postępowanie dobrze sprawdza się przy stosunkowo krótkich szeregach. Jeżeli zakres danych ma jednak odpowiednio dużo obserwacji wówczas możemy podzielić go na segmenty a prognozy *ex ante* oprzeć na analizie bayesowskiej. Wrócimy do tej kwestii w następnym rozdziale.

Załóżmy, że chcemy wykonać prognozę szeregu czasowego, w którym występują wahania przypadkowe, stały poziom zmiennej, nie stwierdzamy natomiast sezonowości. W tym celu sięgniemy po dwie metody: naiwną bez

zmian oraz prognostyczny algorytm genetyczny (PAG) po czym zestawimy ze sobą uzyskane za ich pomocą wyniki.

W tabeli 3.4 znalazły się dane rzeczywiste, prognoza metodą naiwną oraz wykonana przy pomocy algorytmu genetycznego (już po rozkodowaniu). Parametry algorytmu w tym przykładzie wyglądały następująco:
— $t_m = 4$;
— wielkość populacji: 1000 osobników;
— liczba pokoleń: 50;
— prawdopodobieństwo krzyżowania: adaptacyjne z przedziału $\langle 0, 0{,}3 \rangle$;
— prawdopodobieństwo mutacji: adaptacyjne z przedziału $\langle 0, 0{,}1 \rangle$.

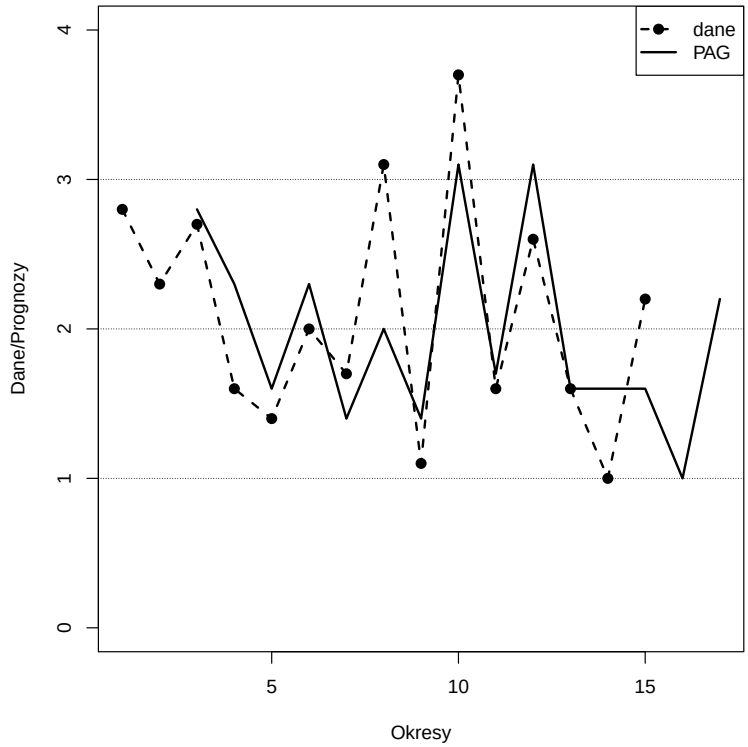
Porównania obu metod dokonamy wykorzystując jako kryterium oceny *RMSPE*, który przy okazji był funkcją przystosowania algorytmu. Najczęściej powtarzającą się wartością t_m okazało się 2 co oznacza, że prognozy *ex ante* dla okresów 16 i 17 równają się wartościom rzeczywistym sprzed dwóch okresów. Na rysunku 3.1 zamieściliśmy wykres przedstawiający dane rzeczywiste i prognozy wygenerowane przez PAG[2].

Tabela 3.4: Porównanie prognoz otrzymanych metodą naiwną i algorytmem genetycznym

Nr okresu	Dane rzecz.	Prognozy: met. naiwna	Prognozy: PAG
1	2,8		
2	2,3	2,8	
3	2,7	2,3	2,8
4	1,6	2,7	2,3
5	1,4	1,6	1,6
6	2	1,4	2,3
7	1,7	2	1,4
8	3,1	1,7	2
9	1,1	3,1	1,4
10	3,7	1,1	3,1
11	1,6	3,7	1,7
12	2,6	1,6	3,1
13	1,6	2,6	1,6
14	1	1,6	1,6
15	2,2	2,2	1,6
16		2,2	1
17		2,2	2,2
	RMSPE	0,7371	0,2730

Źródło: opracowanie własne

[2] Zarówno obliczenia, jak i wykres wykonaliśmy w pakiecie R.



Rysunek 3.1: Prognozy otrzymane PAG bazującym na metodzie naiwnej

Źródło: opracowanie własne

Zwróćmy uwagę na dużą różnicę w dopasowaniu prognoz *ex post* do danych rzeczywistych. Metoda naiwna wygenerowała dwukrotnie wyższy błąd średni. Po części odpowiada za to sam szereg czasowy, w którym pojawiają się wahania przypadkowe zbyt duże jak na wymagania takiego narzędzia. Na jaw wychodzi wada, jaką w tej sytuacji jest opieranie się na bezpośredniej (poprzedni okres) przeszłości. Prognozując przy pomocy PAG mocno ograniczamy wpływ owej przypadłości. Aby bowiem zmniejszyć błędy możemy sięgnąć po dowolną wartość znajdującą się nie dalej niż wynosi przyjęty poziom t_m. Można zaryzykować stwierdzenie, że zaprezentowany algorytm stanowi uogólnienie metody naiwnej, która staje się jego szczególnym przypadkiem o t_m równym 1.

Wciąż jednak rozwiązania wymagają pewne kwestie, wśród których wymienimy pojawianie się pośród danych obserwacji o nietypowych wartościach. Rzutują one na dekompozycję szeregu, a co za tym idzie również na jakość otrzymywanych prognoz. W pierwszej kolejności musimy odpowiedzieć sobie na pytanie: które z obserwacji uznamy za odstające od pozostałych? Możemy w tym celu posłużyć się procedurą opisaną na przykład

w pracy (Konarzewska i Karwacki 2005) wykorzystywaną w procesie tzw. estymacji odpornej[3].

W pierwszym kroku wyznaczamy macierz rzutowania:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T \tag{3.3}$$

Występującą we wzorze 3.3 macierz obserwacji na zmiennych objaśniających $\mathbf{X}$ zastąpimy wektorem obserwacji szeregu czasowego y_t, który obejmował okresy[4] od 3 do n. Przekształcona macierz prezentuje się następująco:

$$\mathbf{H}_y = \mathbf{y}(\mathbf{y}^T\mathbf{y})^{-1}\mathbf{y}^T \tag{3.4}$$

Dla interesującego nas szeregu wyznaczamy prognozy metodą średniej ruchomej prostej o stałej wygładzania równej 2. Na ich podstawie obliczamy reszty z prognozy *ex post*, które następnie standaryzujemy wg formuły:

$$e_t^* = \frac{e_t}{s\sqrt{1-h_t}} \tag{3.5}$$

gdzie:
e_t – reszta z prognozy *ex post*;
s – odchylenie standardowe reszt z prognozy;
h_t – element przekątnej głównej macierzy $\mathbf{H}_y$, który odpowiada prognozie o numerze t.

Reszty wyznaczone wg formuły 3.5 posłużą do określenia, które obserwacje należy uznać za nietypowe. Jako kryterium potwierdzające tę tezę przyjmiemy wartość $|e_t^*| > 2$ (podobnie jak uczyniono w pracy (Konarzewska i Karwacki 2005)).

Obserwację uznaną za nietypową należy z wektora danych wykluczyć. Ponieważ jednak mamy do czynienia z szeregami czasowymi, konieczne jest zastąpienie jej inną wielkością. W związku z tym nietypowa dana może zostać zamieniona na średnią arytmetyczną z obserwacji poprzedniej i następnej. Tak postąpili na przykład autorzy publikacji (Adya et al. 2012). Dodatkowo, szczególną uwagę zwracali oni na to, czy ostatnia obserwacja nie odstaje zbyt mocno od pozostałych.

Jednakże usunięcie obserwacji nietypowych, choć obniża wartość błędu prognozy, nie zawsze zmniejsza zmienność szeregu tak, jak należałoby tego oczekiwać[5]. Pozbywamy się pojedynczych obserwacji odstających od reszty

[3] Opisane postępowanie nie jest jedynym dostępnym. Na przykład Dittman zaproponował wykorzystanie rozstępu kwartylowego do identyfikacji obserwacji odstających.

[4] Podyktowane to zostało przyjęciem założenia, że pierwszą prognozę *ex post* wyznaczono na 3 okres.

[5] Ma to związek z faktem, iż uznanie danej wartości za nietypową dokonuje się względem reszty obserwacji w szeregu.

danych, ale na wyniki nadal może negatywnie wpływać duża zmienność pozostałych obserwacji. Powodować ją mogą te dane, których nie uznano za odbiegające od normy, lecz mimo to leżące w oddaleniu od głównej trajektorii szeregu.

Jeżeli zdecydujemy się na pozostawienie obserwacji nietypowych nietkniętych, istnieje możliwość potraktowania ich jak ekstremów lokalnych. Wówczas sięgamy po rozwiązania stosowane w takich przypadkach w algorytmie genetycznym. Są to sposoby służące odwzorowaniu występującego w przyrodzie podziału na nisze ekologiczne i dostosowane do nich gatunki. Szczególnie przydatna w naszej sytuacji okazuje się preselekcja, w której potomek o lepszym przystosowaniu zastępuje w nowo tworzonej populacji rodzica o przystosowaniu gorszym. W przeciwnym wypadku to potomek jest kasowany, a jego miejsce zajmuje na powrót rodzic.

Preselekcja działa podobnie jak zaproponowana w (Montero i Riff 2011) definicja miary sukcesu jako różnicy w przystosowaniu rodzica i potomka. Osobniki są nagradzane lub karane w zależności od wartości kryterium.

3.1.1. Uwzględnianie trendu w PAG

Chromosom w wersji prognozującej szereg zawierający trend nie różni się budową od przedstawionego poprzednio. Geny nadal zawierają numery wartości t_m dla kolejnych okresów. Zmienia się jednak interpretacja genów. Służą one nie tylko ustaleniu okresu, z którego pochodzi prognoza, ale również przyrostu korygującego stosowną wartość. Ponownie punktem odniesienia stanie się metoda naiwna – tym razem naiwna z poprawką. Rozpatrzymy dwa alternatywne podejścia. Prognozy *ex post* w wariancie przyrostu o liniowym charakterze zapiszemy następująco:

$$y_t^* = y_{t-1} + (y_{t-1} - y_{t-t_m}) \tag{3.6}$$

$$y_t^* = y_{t-t_m} + (y_{t-t_m} - y_{t-t_m-1}) \tag{3.7}$$

Pierwsze z równań 3.6–3.7 jest niewielką modyfikacją tradycyjnie rozumianej metody naiwnej z poprawką. Algorytm poszukuje takiego opóźnienia $t - t_m$, które ustali wartość przyrostu zapewniającego jak najniższy błąd liczony względem obserwacji bezpośrednio poprzedzającej okres prognozy. Możemy tu dostrzec analogię do PAG modyfikującego metodę naiwną bez zmian. Nie ograniczamy się do przyrostu mierzonego między dwoma sąsiednimi wartościami, lecz zwiększamy (w ramach założonych wcześniej granic) obszar poszukiwań zmniejszając wpływ niedawnych wahań losowych.

Ciekawą cechą zastosowanego rozwiązania jest to, że przyrost korygujący wartość z okresu poprzedniego może być równy zero nie tylko gdy dojdzie do przypadkowej zbieżności elementów składowych różnicy. Sam algorytm

dopuszcza, gdy to korzystne, włączenie obserwacji bezpośrednio poprzedzającej w miejsce y_{t-t_m}. Uzyskujemy wówczas metodę naiwną bez zmian, co może się przydać w razie gdyby tuż przed okresem wyznaczenia prognozy wystąpiło nietypowe wahanie losowe.

Drugi z modeli zawartych w 3.7 idzie znacznie dalej. Parametr t_m determinuje nie tylko przyrost, lecz staje się punktem wyjścia całej prognozy. Poszukujemy więc dwóch sąsiadujących ze sobą obserwacji rzeczywistych, które po zaproponowanych przekształceniach dadzą jak najbardziej korzystne prognozy *ex post*.

Jeżeli w szeregu stwierdzimy nieliniowe przyrosty wówczas prognozy *ex post* przybiorą postać:

$$y_t^* = y_{t-1}\frac{y_{t-1}}{y_{t-t_m}} \tag{3.8}$$

$$y_t^* = y_{t-t_m}\frac{y_{t-t_m}}{y_{t-t_m-1}} \tag{3.9}$$

Prognozy *ex ante* obliczamy analogicznie jak w metodzie naiwnej tj. k-krotnie korygujemy wybraną wartość ostatnim dostępnym przyrostem. Wprawdzie nie ma tutaj ograniczenia horyzontu, lecz należy pamiętać, żeby prognozy miały charakter krótkookresowy.

Przyjrzyjmy się teraz, jak obie wersje algorytmu zachowują się w przypadku szeregu z trendem liniowym. W tabeli 3.5 znalazły się szeregi danych rzeczywistych, prognoz metodą naiwną z trendem oraz prognoz otrzymanych PAG. Pierwszy model oznaczymy symbolem v_1 zaś drugi v_2. Parametry algorytmu były takie same jak we wcześniejszym przykładzie. Dla prognoz *ex ante* t_m wyznaczone na podstawie najlepszego chromosomu równało się odpowiednio: 1 oraz 2. Wykres na rysunku 3.2 prezentuje zestawienie danych i prognoz pochodzących z algorytmu genetycznego.

Ponownie, jak miało to miejsce przy okazji porównania z metodą naiwną prostą, uwagę zwraca duża różnica w średnim błędzie prognoz *ex post* na korzyść PAG. Druga z wersji daje prognozy nieco lepszej jakości niż pierwsza co wynika z faktu, że ma ona nieco więcej swobody w ich tworzeniu. Kiedy jednak przyjrzymy się wykresom na rysunkach 3.2 i 3.3 zauważamy inne zachowanie obu metod.

Pierwszy ze wspomnianych wykresów wskazuje, że doszło do usunięcia mniejszych wahań przypadkowych i częściowego wygładzeniem większych. Tego typu zjawisko obserwowaliśmy do tej pory na przykład w metodzie średniej ruchomej. Nam udało się włączyć je do algorytmu prognostycznego a nawet w pewnym sensie do metody naiwnej z trendem.

Szereg prognoz z rysunku 3.3 nie jest równie gładki jak jego poprzednik. Można o nim powiedzieć, że jednakowo potraktował wszystkie wahania losowe odfiltrowując je i pozostawiając trend obserwowany w szeregu.

Tabela 3.5: Porównanie prognoz otrzymanych metodą naiwną z trendem liniowym i algorytmem genetycznym

Nr okresu	Dane rzecz.	Prognozy: met. naiwna	PAG v_1	PAG v_2
1	11			
2	14			
3	13	17	14	11
4	18	12	15	17
5	17	23	18	17
6	25	16	21	23
7	26	33	25	23
8	28	27	27	27
9	26	30	28	27
10	28	24	27	30
11	29	30	29	30
12	31	30	32	30
13	30	33	31	32
14	34	29	31	32
15	37	38	38	38
16		40	40	38
17		43	43	40
	RMSPE	0,2209	0,0817	0,0687

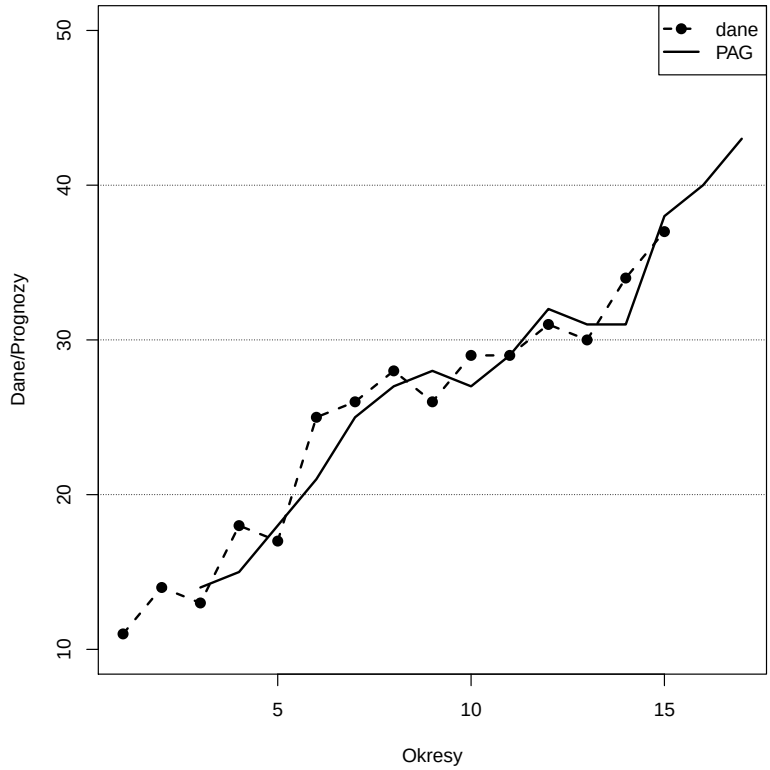
Źródło: opracowanie własne

W razie zaobserwowania, oprócz trendu, dużych wahań losowych (kiedy nie da się stosować nie tylko metody naiwnej, ale i niektórych z pozostałych klasycznych narzędzi prognozowania) możemy przyjąć t_m na wyższym poziomie. Wówczas PAG ujawnia swą elastyczność gdyż dobiera prognozy z okresów bardziej odległych w czasie. Podobnie sprawa będzie miała się jeżeli w danych wystąpi sezonowość. Szerzej tym problemem zajmiemy się w kolejnej części tego rozdziału.

3.1.2. Uwzględnianie sezonowości

Oddzielnej uwagi wymaga kwestia występowania sezonowości w szeregu czasowym. Generalnie rzecz biorąc, nasze postępowanie jest podobne m.in. do metody naiwnej z sezonowością, w której (w odróżnieniu od metody prostej) jako prognozę *ex post* wybieramy obserwację odległą o liczbę okresów składających się na jeden cykl sezonowy.

W algorytmie prognostycznym parametr t_m powinien mieć wartość równą co najmniej liczbie okresów tworzących cykl wahań sezonowych. Innymi



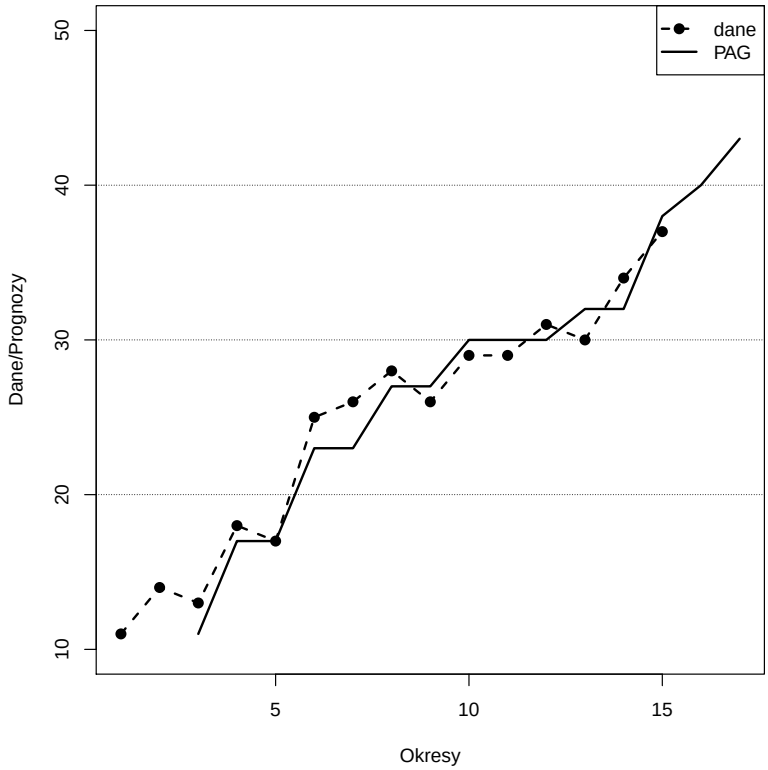
Rysunek 3.2: Prognozy szeregu zawierającego trend liniowy uzyskane modelem v_1

Źródło: opracowanie własne

słowy, jeśli efekt sezonowy pojawia się w szeregu co 4 okresy (zwykle kwartały), wtedy maksymalne dopuszczalne opóźnienie musimy przyjąć na co najmniej takim samym poziomie. Chodzi o to, aby PAG miał możliwość wykorzystania obserwacji znajdujących się w tej samej fazie cyklu co prognozowany okres.

Wahania sezonowe występują w dwóch podstawowych odmianach: addytywnej i multiplikatywnej. Zaprezentowane poniżej przykłady dotyczą pierwszej z wymienionych wersji. Metoda naiwna z sezonowością nie powinna być używana w przypadku stwierdzenia wariantu multiplikatywnego, ponieważ prognozy *ex post* w okresach odchylających się sezonowo będą systematycznie niedoszacowane lub przeszacowane. Algorytm da się jednak zastosować pod warunkiem jednak modyfikacji polegającej na przykład na pomnożeniu wartości prognoz (po rozkodowaniu) przez pewną stałą wyrażającą zmianę amplitudy wahań sezonowych w czasie.

Na potrzeby obliczeń wygenerowaliśmy szeregi charakteryzujące się pożądanymi własnościami. We wszystkich występują małe wahania przypad-



Rysunek 3.3: Prognozy szeregu zawierającego trend liniowy uzyskane modelem v_2

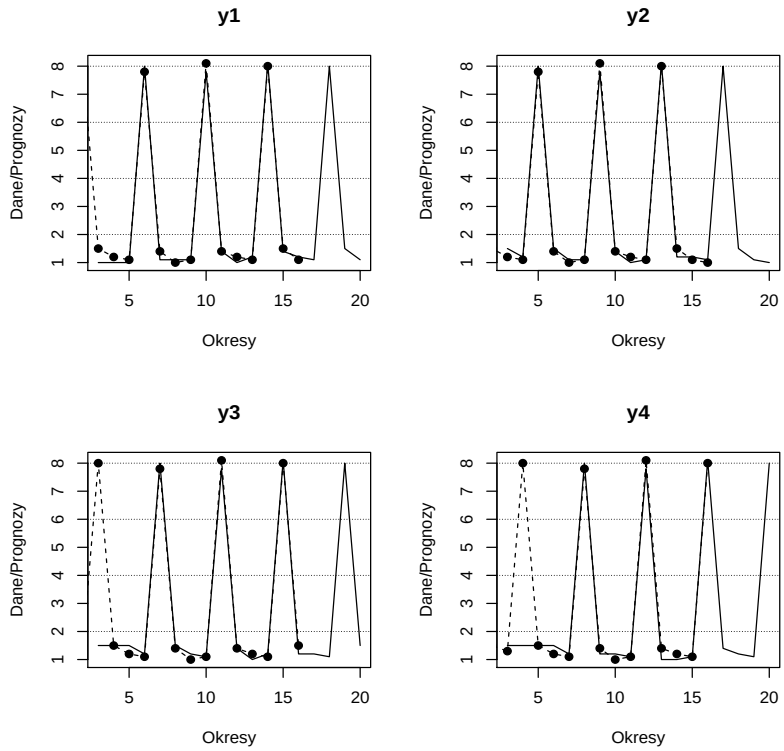
Źródło: opracowanie własne

kowe, stały poziom zmiennej oraz wyraźna sezonowość addytywna. Okres wahań sezonowych wynosi 4. Przy tej okazji jednak postanowiliśmy sprawdzić jak zachowuje się algorytm w sytuacji, kiedy obserwujemy zarówno pełne, jak i niepełne okresy wahań sezonowych na początku szeregu.

Parametry sterujące algorytmem pozostawiliśmy na poziomie z dotychczasowych przykładów. Najbardziej istotny z nich, czyli t_m równał się 4, a więc dokładnie tyle, ile wynosił okres wahań sezonowych[6]. Na rysunku 3.4 znalazły się wszystkie cztery możliwe przypadki z punktu widzenia momentu pojawienia się wahań sezonowych wraz z wyznaczonymi dla nich prognozami. Błąd *RMSPE* dla kolejnych szeregów wyniósł odpowiednio: 0,1347, 0,1115, 0,247 i 0,2573.

Analiza wykresów na rysunku 3.4 wskazuje, że dla szeregów oznaczonych y_1 i y_2 obie metody dały dość podobne, niezłe wyniki. Jednak w pozostałych dwóch przypadkach (szeregi y_3 i y_4) uległy one znacznemu pogorszeniu

[6] Czyli była to najmniejsza wartość t_m jaką należało przyjąć.



Rysunek 3.4: Prognozy szeregów o różnych typach sezonowości

Źródło: opracowanie własne

pomimo tego, że PAG dostosował się do wahań sezonowych. Zostało to spowodowane sposobem wyznaczania początkowych prognoz oraz zachowaniem samych szeregów.

Należy pamiętać, że dla zmodyfikowanego algorytmu genetycznego pierwszą prognozę otrzymujemy dla okresu trzeciego i jest nią obserwacja numer 1 lub 2. PAG nie dysponuje jeszcze maksymalnym zbiorem wartości, z których tworzy prognozy. Z tego powodu na wielkość błędu wpływa silne niedoszacowanie w początkowych okresach *ex post*. Warto zwrócić uwagę, że w kolejnych okresach otrzymano prognozy dobrej jakości. Jeżeli dla szeregu y_3 obliczymy błąd zaczynając od czwartego zaś dla y_4 od piątego okresu to otrzymamy odpowiednio: 0,1222 oraz 0,1422 czyli wartości zbliżone wyników dla pierwszych dwóch szeregów.

Jeśli chodzi o analogie do wspomnianej metody naiwnej to występować one będą również w obszarze prognoz *ex ante*. Zgodnie z zaproponowanym rozwiązaniem, dominanta opóźnień wyrażanych przez parametr t_m wynosi

4. Powoduje to, że zarówno algorytm genetyczny, jak i metoda naiwna dadzą w tym wypadku takie same prognozy *ex ante*.

W tym miejscu powstaje pytanie o zasadność sięgania po zdecydowanie bardziej skomplikowaną i pracochłonną metodę. Przyczyny są dwie. Pierwsza z nich to fakt, że wprowadzenie wysokich wartości t_m pozwala wychwycić wahania sezonowe w sytuacji, kiedy nie są one tak wyróżnione, jak w przypadku stworzonych powyższych szeregów. Poza tym, algorytm genetyczny radzi sobie w sytuacji, kiedy nie można użyć metod naiwnych. Chodzi tu mianowicie o pojawianie się dużych wahań przypadkowych pomiędzy okresami sezonowymi oraz obecność w szeregu trendu.

3.2. Algorytm bazujący na średniej ruchomej

Innym proponowanym przez nas podejściem do prognozowania przy pomocy algorytmu genetycznego jest oparcie go na średniej ruchomej. Zaprezentowane rozwiązanie stanowi przy okazji modyfikację opisanego wcześniej algorytmu bazującego na metodach naiwnych.

Zachowanie szeregów wygładzanych za pomocą średniej ruchomej determinuje stała wygładzania. Wraz z jej wzrostem rośnie efekt wygładzenia, eliminując wpływ części wahań przypadkowych z przeszłości. Przyjmowanie wysokiej stałej wygładzania prowadzi wszakże do nasilenia się kompensacji reszt z prognozy *ex post* bądź tworzenia serii błędów systematycznych.

Rozważmy przypadek średniej ruchomej prostej, w której stała wygładzania przyjmuje dowolną wartość z pewnego, zadanego z góry przedziału. Nie wymagamy przy tym, aby w kolejnych okresach utrzymywać jednakową wartość tego parametru. Łatwo dojść w tym miejscu do wniosku, że proponowana swoboda doprowadzić musi do wyboru między dużą ilością potencjalnych prognoz co tak jak poprzednio stwarza pole do popisu dla algorytmu genetycznego.

Parametr t_m wciąż interpretujemy jako maksymalny, dopuszczalny numer opóźnienia danej rzeczywistej wykorzystywanej do otrzymywania prognoz na bieżący okres. Ustalamy go na z góry zadanym poziomie przed rozpoczęciem obliczeń i do ich końca pozostaje on niezmieniony. Tym razem jednak zakładamy, że stała wygładzania nie może przekroczyć wartości t_m. Wynika stąd, że zachodzi zależność: $1 \leqslant k \leqslant t_m$. Przyjęcie akurat takich granic przedziału podyktowała chęć zapewnienia algorytmowi jak największej swobody w doborze parametrów. Naszym celem było uelastycznienie wartości k tak, aby szereg prognoz adaptował się do zmian szeregu danych. W dalszej części rozdziału prześledzimy przypadek zmniejszenia wspomnianego przedziału poprzez dodatkowe ograniczenie go z dołu.

Załóżmy, że dysponujemy szeregiem czasowym złożonym z sześciu obserwacji zaś $t_m = 3$. W tej sytuacji pierwszą prognozę *ex post* wyznaczymy dla czwartego okresu, ale niekoniecznie użyjemy do tego celu wszystkich trzech obserwacji poprzedzających. Przykładowo, jeżeli wybierzemy stałą wygładzania równą 2 obliczymy średnią arytmetyczną z obserwacji y_2 i y_3. Jest to wszakże tylko jedna z możliwości, ponieważ k w danym okresie może być równe 3 lub nawet 1. Sytuacja ta powtarzać się będzie w każdym kolejnym okresie.

Tabela 3.6: Przykłady prognoz algorytmu ze średnią ruchomą, w której $t_m = 3$

Okres (t)	Wartość rzeczywista	Stała wygładzania	Prognoza
1	y_1		
2	y_2		
3	y_3		
4	y_4	2	$(y_2 + y_3)/2$
5	y_5	1	y_4
6	y_6	3	$(y_3 + y_4 + y_5)/3$

Źródło: opracowanie własne

W tabeli 3.6 zawarte zostały przykładowe prognozy *ex post* dla opisanego powyżej przypadku. W trzeciej kolumnie znalazła się jedna z możliwych prognoz *ex post*. W okresach od 4 do 6 dopuszczamy stałą wygładzania, która jednakże mieści się w przedziale $\langle 1,3 \rangle$. W porównaniu z podejściem opisanym w poprzednim podrozdziale, pierwsza prognoza nie będzie wyznaczana za każdym razem w dla tego samego (trzeciego) okresu. Tym razem decyduje t_m, które działa na podobieństwo tradycyjnej stałej wygładzania. Oznacza to, że w tabeli 3.6 prognozy *ex post* zaczną pojawiać się od czwartego okresu.

Podobnie jak to miało miejsce wcześniej, jeden chromosom odpowiada jednemu z możliwych wariantów prognostycznych. Ich ogólną liczbę możemy obliczyć w oparciu o wzór 3.2. Pojedynczy gen zawiera wartość stałej wygładzania ze wspomnianego wyżej przedziału.

Funkcja celu ma taką samą postać jak w algorytmie bazującym na metodach naiwnych. Podobnie rzecz ma się z selekcją oraz krzyżowaniem. Jeśli chodzi o mutację, to wartość wylosowanego genu przed i po użyciu operatora nie mogło dzielić w czasie więcej niż t_m okresów. Po raz kolejny sięgnęliśmy też po preselekcję. Inaczej jednak przebiega rozkodowanie chromosomu ze względu na konieczność obliczenia średnich dla kolejnych okresów.

Prognozy *ex ante* otrzymamy w sposób typowy dla średniej ruchomej z jedną różnicą: dla wybranego chromosomu dysponujemy w rzeczywistości

zestawem różnych wartości stałej wygładzania i dlatego użyjemy dominanty lub mediany szeregu stałych, aby w oparciu o nie utworzyć prognozy.

Tabela 3.7: Porównanie prognoz otrzymanych metodą średniej ruchomej i prognostycznym algorytmem genetycznym

Nr okresu	Dane rzecz.	Prognozy: średnia ($k = 4$)	Prognozy: PAG
1	2,8		
2	2,3		
3	2,7		
4	1,6		
5	1,4	2,35	2,15
6	2	2	2
7	1,7	1,925	1,7
8	3,1	1,675	1,85
9	1,1	2,05	2,05
10	3,7	1,975	2,1
11	1,6	2,4	2,4
12	2,6	2,375	2,65
13	1,6	2,25	2,1
14	1	2,375	1,93
15	2,2	1,7	1,73
16		1,85	1,6
17		1,663	1,9
	RMSPE	0,6043	0,4905

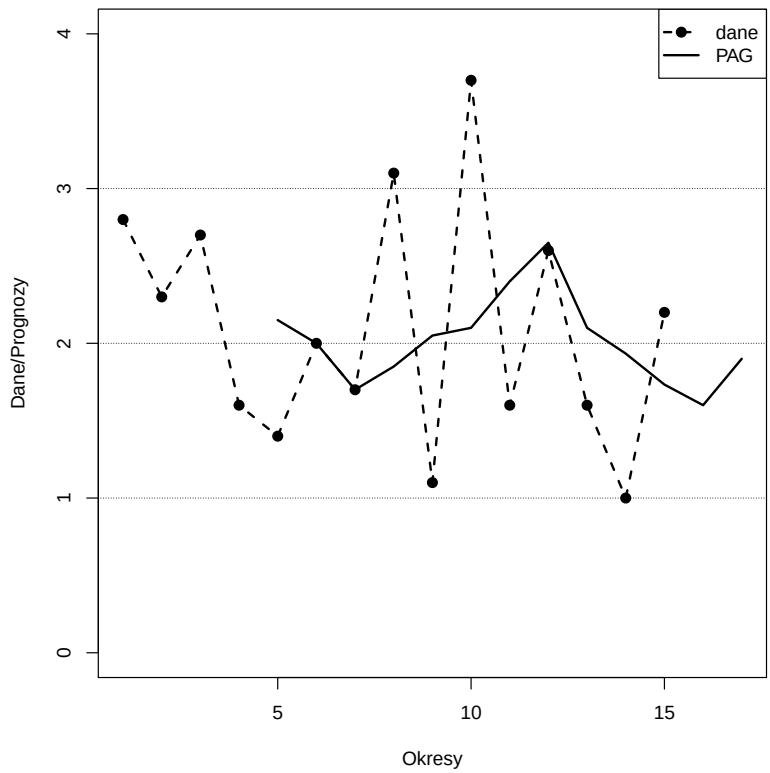
Źródło: opracowanie własne

Przedział, z którego pochodzą wartości stałych wygładzania niekoniecznie musi ograniczać się do przedziału $\langle 1, t_m \rangle$. Jeżeli podwyższymy lewy kraniec przedziału, wtedy osiągniemy rezultat odpowiadający zwiększeniu efektu wygładzania. Takie postępowanie ogranicza trochę swobodę działania algorytmu. Pewna rzecz zasługuje jednak na uwagę. Otóż cechę charakterystyczną średniej ruchomej stanowi utrata części obserwacji z początku szeregu potrzebnych do obliczenia pierwszej prognozy *ex post*. Aby uzyskać podobny efekt wygładzenia w klasyczny sposób, konieczne byłoby zwiększenie stałej wygładzania, co skończyłoby się utratą informacji z okresów początkowych. Dla krótszych szeregów pojawia się konflikt pomiędzy dążeniem do pomniejszenia wpływu wahań przypadkowych a rezygnacją z części zawartych w nich danych.

Prezentowane podejście pozwala rozwiązać ten problem. Zwróćmy bowiem uwagę, że za efekt wygładzenia odpowiada lewy kraniec przedziału.

Nie spowoduje to skrócenia szeregu wygładzonego gdyż moment wyznaczenia pierwszej prognozy zależy od t_m czyli prawego krańca przedziału, który nie ulega zmianie.

W charakterze ilustracji przytoczonych rozważań ponownie sięgniemy po dane z tabeli 3.4. Tym razem do porównania wybraliśmy metodę średniej ruchomej prostej o stałej wygładzania $k = 4$. Naszym kryterium oceny nadal pozostaje pierwiastek średniokwadratowego błędu procentowego. Opóźnienie $t_m \in \langle 2{,}4 \rangle$, natomiast pozostałe parametry sterujące algorytmem pozostają bez zmian.



Rysunek 3.5: Prognozy otrzymane PAG bazującym na średniej ruchomej

Źródło: opracowanie własne

W tabeli 3.7 znalazły się prognozy klasyczną metodą średniej ruchomej prostej oraz przy pomocy algorytmu genetycznego. Na rysunku 3.5, jak poprzednio, zaprezentowaliśmy szeregi: danych rzeczywistych oraz prognoz ostatnią z wymienionych metod. W najlepszym, chromosomie t_m najczęściej równało się 2 i taką stałą przyjęliśmy podczas wyznaczania prognoz *ex ante*. Dla okresu 17 wykonana została prognoza dynamiczna.

Należy zwrócić w tym miejscu uwagę na fakt, że wobec długości szeregu (15 okresów) zastosowaliśmy całkiem duży stopień wygładzenia i to w obu metodach. Stąd wysokie wartości błędów *ex post*, które jednak dla PAG przyjęły znacznie niższe wartości. Z tego powodu *RMSPE* w tym przypadku również okazało się niższe.

3.3. Prognozowanie kaskadowe

Niedoskonałością wielu metod prognozowania szeregów czasowych jest pomijanie ostatniej obserwacji podczas wyznaczania prognoz *ex post*. Zwykle staje się ona punktem wyjścia pierwszej prognozy poza okres czasu objęty danymi. Ponadto użyta metoda może wymagać poświęcenia części początkowych obserwacji, aby uzyskać punkt startowy, jak to ma miejsce na przykład w przypadku wszystkich metod wygładzania wykładniczego. Jeżeli więc nie dysponujemy bardziej obszernym zestawem danych ryzykujemy, że wyprognozowane wartości nie uwzględnią części mechanizmów stojących faktycznie za badanym zjawiskiem. Prognostyczny algorytm genetyczny, choć bardziej elastyczny w porównaniu z klasycznymi metodami, także nie jest całkowicie wolny od tej wady.

PAG poszukuje prognoz najlepszej jakości sprawdzając różne zestawy opóźnień zmiennej prognozowanej. Jak wykazaliśmy wcześniej, przestrzeń rozwiązań, którą należy pod tym kątem przeszukać gwałtownie powiększa się ze wzrostem t_m oraz wydłużaniem szeregu. Wymagane są więc większe populacje i podwyższona liczba pokoleń, aby algorytm zyskał dodatkową szansę na znalezienie chromosomu gwarantującego prognozy o akceptowalnej dokładności. Pomocne stają się modyfikacje wprowadzane do operatorów takie jak zmiana rodzaju selekcji czy preselekcja. Jednakże pomimo tego czasochłonność i pracochłonność obliczeń wciąż rosną.

Z drugiej strony dłuższy szereg czasowy niesie ze sobą więcej informacji o procesach go kształtujących. Dopiero odpowiednio duża ilość obserwacji pozwala uzyskać pewność, że mamy do czynienia z sezonowością, a nie przypadkowymi obserwacjami odstającymi. Niekiedy jeszcze więcej danych potrzeba do określenia, czy trend zachowuje liniowy charakter. Trudno pogodzić chęć jak najszybszego a jednocześnie jak najmniej pracochłonnego przeprowadzenia obliczeń z koniecznością zawarcia w danych maksymalnej dozy informacji. Problem narasta, jeżeli nie możemy jednoznacznie określić dekompozycji szeregu, gdyż obserwujemy na przykład zmiany tendencji centralnej w obrębie całego zestawu danych.

Jednym ze sposobów poradzenia sobie z sytuacją, w której mechanizm odpowiadający za kształtowanie się prognozowanej zmiennej podlega zmianom w czasie są modele adaptacyjne. Zaliczamy do nich m.in. model trendu

pełzającego (Cieślak 1997, ss. 85-87). Dobór wariantów PAG do segmentów oraz późniejsze obliczenia wykonywane przez algorytm sprawiają, że nasza metoda także zyskuje adaptacyjny charakter. Więcej informacji na ten temat znajdziemy poniżej i w następnym rozdziale.

Inne podejście to podział szeregu na podokresy poddane oddzielnemu działaniu narzędzi prognostycznych. Następnie prognozy łączymy w całość i wnioskujemy o przyszłym zachowaniu zmiennej. Wszelako rodzi się w tym momencie kilka pytań. Po pierwsze, w jaki sposób zestaw metod dla poszczególnych przedziałów czasu ma dać prognozy *ex ante*? Problem ów nie wystąpi jedynie wtedy, kiedy do wszystkich przedziałów czasu użyjemy tej samej metody. Po drugie, czy wybrane metody powinny pozostawać ze sobą w jakimś związku, na przykład poprzez zestaw wag?

Po trzecie wreszcie, zdecydowana większość metod niestrukturalnych wymaga rezygnacji z części początkowych obserwacji. Między prognozami dla kolejnych przedziałów tworzą się więc luki powodując brak ciągłości prognoz wyznaczanych dla kolejnych podokresów.

W literaturze znajdziemy przykłady prób łączenia kilku metod w celu predykcji jednego szeregu. Sikora i Piramuthu (Sikora i Piramuthu 2005) wykorzystali algorytm genetyczny oparty na technikach *data mining*, wyposażony w samoadaptującą się selekcję chromosomów na bazie odległości Hausdorffa. Charakterystyki algorytmu ewoluują wraz z nim zaś sam algorytm opracowuje optymalny zestaw reguł połączonych operatorami logicznymi. Wprawdzie celem autorów nie było wyznaczenie prognoz, ale mamy tu do czynienia z tworzeniem struktury pomniejszych procesów składających się na jeden większy.

Stricte prognostyczny cel przyświecał Zou i Yang (Zou i Yang 2004), którzy stworzyli algorytm o nazwie AFTER. Zestawia on wybrane metody prognozowania na podstawie testów statystycznych lub kryterium Akaike, po czym tworzy prognozę łączną wykorzystując ważone prognozy z wytypowanych modeli. Same metody w poszczególnych podokresach oceniane są przy pomocy błędu MSPE obliczanego rekurencyjnie, po uzyskaniu prognozy na jeden okres naprzód.

Z kolei w artykule (Wang, Wang, Zhang i Guo 2012) podjęto próbę prognozowania indeksu giełdowego *Shenzhen Integrated Index* przy pomocy hybrydowego modelu łączącego wygładzanie wykładnicze (metoda Browna), ARIMA i sieć neuronową. Autorzy sięgnęli po algorytm genetyczny, który służył do ustalania wag dla poszczególnych metod w opisywanym modelu. Wagi nakładane były na prognozy otrzymane za pomocą każdej z trzech składowych modelu.

Dong i Pedrycz (Dong i Pedrycz 2008) przy pomocy modeli rozmytych podzielili szereg czasowy na klastry. Prognozowanie odbywa się na podstawie przygotowanego przez autorów alfabetu symboli wyrażających różne sygna-

ły (wzrost, spadek zmiennej itp.). Przyznają jednak, że sposób podziału szeregu to wciąż kwestia otwarta. Sami robili to metodą prób i błędów.

Średnią ważoną prognoz znajdziemy także w pracy (Wichard 2011), której autor połączył kilka metod dla szeregu czasowego z sezonowością. Jako narzędzie predykcji posłużyła sieć neuronowa wykorzystująca metodę najbliższych sąsiadów.

Wspomniany w rozdziale 1 artykuł, którego autorami są m.in. Agustín i Salcendo-Sanz nie traktował wprawdzie o prognozowaniu, lecz warto go przypomnieć ze względu na wykorzystanie algorytmu genetycznego do podziału zbioru na klastry składające się z podobnych obiektów. Autorzy przy pomocy algorytmu, w którym stosują selekcję na podstawie rang tworzonych dla chromosomów oraz selekcję elitarną tworzą grupy jednorodnych osobników. Proponują też operatory krzyżowania i mutacji przystosowane do tego zagadnienia. Omawiają różne kryteria przynależności do klastra również sumę kwadratów reszt. Całość testują głównie na losowo generowanych zbiorach danych.

Metodę zaprezentowaną w tym rozdziale nazwaliśmy *prognozowaniem kaskadowym*. Zakładamy w niej, że w pierwszym kroku badacz dzieli szereg na segmenty. Wyróżnikiem podziału może być pojawienie się nietypowych obserwacji, zmiana głównego trendu, zmiana poziomu przyjmowanych wartości lub jakiekolwiek inne kryterium uzasadniające taki podział. W ostateczności jako powód podziału uznamy także chęć zwiększenia dokładności prognoz, wynikającą ze skrócenia szeregu danych. Krótszy szereg oznacza konieczność przeszukania mniejszej przestrzeni rozwiązań.

Każdemu segmentowi przypisana zostaje konkretna metoda prognozowania. Do wyboru będą to: prognostyczny algorytm genetyczny wykorzystujący metodę naiwną bez poprawki, metodę naiwną z trendem w wersjach 1 i 2 oraz średnią ruchomą.

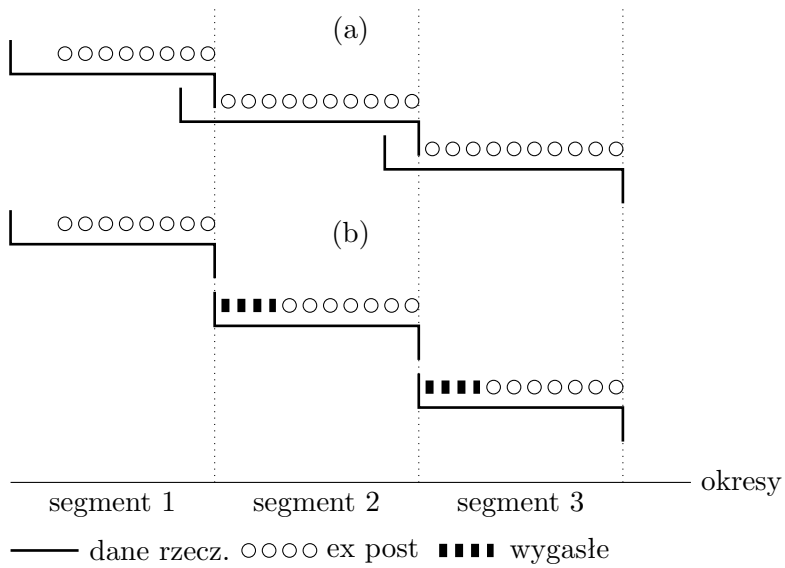
Wszystkie metody składające się na prognozowanie kaskadowe wymagają ustalenia punktu startowego otrzymywanego przez poświęcenie kilku pierwszych obserwacji prognozowanego szeregu. Dlatego poczynając od drugiego segmentu pomiędzy poszczególnymi zestawami prognoz pojawią się luki powodujące utratę ciągłości predykcji. Zaproponujemy i porównamy dwa rozwiązania tego problemu:

1. Ostatnie kilka obserwacji segmentu bezpośrednio poprzedzającego dołączamy do obserwacji na początku segmentu następnego (będziemy także używali określeń: prognozowanie kaskadowe 1 lub kaskada 1).
2. Na początku segmentu drugiego i dalszych dołączamy do obserwacji prognozy *ex ante* uzyskane metodą obowiązującą w segmencie poprzednim (prognozowanie kaskadowe 2 lub kaskada 2).

W pierwszym z wymienionych rozwiązań, dodane na początku segmentu dane pozwolą wystartować metodzie, gdyż posłużą do wygenerowania

prognoz w okresach przypadających na lukę pomiędzy segmentami. Tym samym zakresy danych sąsiednich segmentów zazębiają się co pozwala utrzymać ciągłość prognoz. Zachowujemy również ich spójność, ponieważ dane, które posłużyły do oceny dokładności jednej metody stają się zaczątkiem prognoz kolejnej z nich. Weryfikujemy więc przydatność wybranych metod pod kątem wpływu końcowych obserwacji co należy uznać za bardzo korzystne, gdyż wiele metod nie uwzględnia wpływu ostatniej obserwacji na jakość prognoz *ex post*.

Drugi sposób wypełniania luk wychodzi z innego punktu widzenia. Na podstawie wyników z segmentu poprzedniego wyznaczamy prognozy *ex ante* w ilości odpowiadającej liczbie opuszczonych okresów segmentu bieżącego. Prognozy te dołączamy do zestawu danych obecnego podokresu. W ten sposób pierwsze prognozy *ex post* powstaną z poprzednio wyznaczonych prognoz, a przez to staną się prognozami dynamicznymi. Podczas oceniania trafności będą już prognozami wygasłymi a ich jakość przełoży się na dokładność bieżącej predykcji.



Rysunek 3.6: Prognozowanie kaskadowe: (a) – wariant wykorzystujący ostatnie obserwacje segmentu poprzedniego, (b) – wariant wykorzystujący prognozy wygasłe

Źródło: opracowanie własne

Rysunek 3.6 ilustruje różnice w uzupełnianiu luk między pierwszym oraz drugim wariantem prognozowania kaskadowego. PAG stara się wybrać okresy zapewniające predykcję o jak najniższych błędach. Z jednej strony mamy więc dynamiczne łączenie prognoz dla kolejnych segmentów w je-

den ciąg. Z drugiej strony, w skali całego szeregu czasowego uwzględniamy zdolność wybranych metod do tworzenia prognoz *ex ante*.

Włączenie prognoz wygasłych do mechanizmu oceny dokładności predykcji jest niezwykle kuszące. Nie zrezygnujemy jednak z wariantu dołączającego dane rzeczywiste. Prognozy wykonywane poza zakres danych z zasady obarczone są wyższymi błędami. Dlatego pierwsze z podejść do uzupełniania luk pomiędzy podokresami potraktujemy jako punkt odniesienia. Jeżeli wariant uwzględniający prognozy wygasłe da zbliżone lub nawet lepsze wyniki, będzie to świadczyło o dobrym odwzorowaniu procesów kształtujących prognozowane zjawisko.

3.3.1. Podejście bayesowskie w prognozowaniu kaskadowym

W dotychczasowych rozważaniach na temat prognoz kaskadowych nie skupialiśmy się zbyt mocno na sposobie wyznaczania prognoz *ex ante*. Stało sie tak z dwóch powodów. Po pierwsze, prognozy *ex post* pozwalają na bardziej rozbudowane analizy co wynika z faktu porównywania ich z obserwacjami pochodzącymi z szeregu czasowego. Predykcja poza zakres dostępnych wartości zmiennej oznacza, że podobna analiza stanie się możliwa dopiero w momencie pojawienia się prognoz wygasłych.

Drugi powód wiąże się bezpośrednio z samym mechanizmem prognostycznym zastosowanym w PAG. Uważamy otóż, że w przypadku użycia algorytmu ograniczonego tylko do jednej z czterech proponowanych wersji mediana lub dominanta opóźnień zawartych w chromosomie jest wystarczającym rozwiązaniem kwestii prognoz *ex ante*. Wykorzystujemy zresztą to założenie w prognozowaniu kaskadowym 2.

Prognozowanie kaskadowe daje wszakże szersze pole do popisu. Przede wszystkim stosujemy je w przypadku o wiele dłuższych szeregów czasowych. W rezultacie dysponujemy większą ilością informacji na temat mechanizmów odpowiadających za przeszłe i przyszłe zachowanie interesującej nas zmiennej. Ponadto posiadamy dodatkową informację związaną z podziałem samego szeregu na segmenty co kieruje naszą uwagę w stronę statystyki bayesowskiej.

Tak zwane „podejście bayesowskie" stanowi osobny nurt w statystyce będący w opozycji do klasycznego (częstościowego) rozumienia tego przedmiotu. Statystyka bayesowska opiera się na wykorzystaniu dodatkowej wiedzy *a priori* na temat analizowanego zagadnienia obok zaobserwowanych dla próby wyników (Aczel 2000, s. 783).

Celem tej pracy nie jest prezentacja dyskusji na temat różnych filozofii podejścia do prawdopodobieństwa oraz zdarzeń losowych. Ograniczymy się tylko do przybliżenia tego zagadnienia w stopniu wystarczającym

na potrzeby dalszego wykorzystania w ramach algorytmu prognostycznego. Zainteresowanych odsyłamy do fachowej literatury.

Związek pomiędzy prawdopodobieństwami wynikającymi z informacji *a priori* oraz tymi pochodzącymi z obserwacji składających się na próbę statystyczną wyraża twierdzenie Bayesa pochodzące z 1763 roku. Twierdzenie to pozwala obliczyć prawdopodobieństwo zajścia jednego zdarzenia pod warunkiem, że zaszło inne zdarzenie. Dla zdarzeń oznaczonych A i B zapiszemy (Aczel 2000, s. 103):

$$P(B|A) = \frac{P(A|B)P(B)}{P(A|B)P(B) + P(A|\overline{B})P(\overline{B})} \tag{3.10}$$

Prawdopodobieństwa $P(B)$ i $P(\overline{B})$ są to prawdopodobieństwa *a priori* (bezwarunkowe). $P(B|A)$ nazywamy prawdopodobieństwem *a posteriori* zdarzenia B.

Uogólniając rozważania na przypadek zajścia jednego z wielu wykluczających się zdarzeń i ich apriorycznych prawdopodobieństw dla przypadku zmiennej losowej skokowej uzyskamy następującą formę twierdzenia Bayesa (Aczel 2000, s. 786):

$$P(\theta|x) = \frac{P(x|\theta)p(\theta)}{\sum_i P(x|\theta_i)p(\theta_i)} \tag{3.11}$$

gdzie:
θ – nieznany parametr populacji, który ma być szacowany na podstawie danych pochodzących z próby;
x – wyniki obserwacji (dane).

Warunkowe prawdopodobieństwa przyporządkowane poszczególnym elementom zbioru wszystkich możliwych wartości zmiennej losowej tworzą funkcję wiarygodności. Spójrzmy na wzór (3.11). Na jego podstawie powiemy, że funkcja wiarygodności przyporządkowuje warunkowe prawdopodobieństwa $P(x|\theta)$ pojawienia się wyniku obserwacji x możliwym wartościom nieznanego parametru θ.

Zakładamy także, że parametry populacji są zmiennymi losowymi. Takiej zmiennej przypisujemy pewien aprioryczny rozkład prawdopodobieństwa. Dysponując wynikami obserwacji możemy ustalić funkcję wiarygodności a zatem poznać prawdopodobieństwa uzyskania tych danych przy różnych wartościach parametru θ generowanych przez rozkład aprioryczny. Za pomocą twierdzenia Bayesa obliczymy więc aposterioryczne prawdopodobieństwo dla parametru θ. Uwzględnia ono tak posiadane wyniki obserwacji jak i informację *a priori*.

Twierdzenie Bayesa da się odnieść również do zmiennych losowych ciągłych. Niech $f(\theta)$ wyraża gęstość apriorycznego prawdopodobieństwa pa-

rametru θ zaś $f(x|\theta)$ warunkową gęstość wyniku obserwacji x przy danej wartości θ. W tym momencie $f(x|\theta)$ oznacza funkcję wiarygodności. Twierdzenie Bayesa dla rozkładów ciągłych ma postać:

$$f(\theta|x) = \frac{f(x|\theta)f(\theta)}{\int_{-\infty}^{+\infty} f(x|\theta)f(\theta)d\theta} \tag{3.12}$$

Podstawowym zarzutem stawianym podejściu beyesowskiemu przez jego krytyków jest subiektywne przyjmowanie informacji *a priori*. Jeżeli okaże się ona niedokładna lub tendencyjna dołączenie jej do danych prowadzi do mniej dokładnych wniosków niż w podejściu klasycznym. Co innego jednak jeśli informacja *a priori* pochodzi z poprzednich badań lub ma jakiekolwiek inne przesłanki wskazujące na jej nieprzypadkowy charakter. Wówczas dokładność wnioskowania rośnie.

Zwolennicy statystyki bayesowskiej twierdzą, że prawdopodobieństwo *a priori* wyraża miarę racjonalnego przekonania, że dane zdarzenie będzie miało miejsce. Aby wzbogacić owo przekonanie wykonujemy badanie, które przekształca prawdopodobieństwo *a priori* w prawdopodobieństwo *a posteriori*. Poza tym podejście bayesowskie da się zastosować zarówno w przypadku zjawisk powtarzalnych jak i niepowtarzalnych. Klasyczna statystyka gorzej radzi sobie w tym drugim wypadku.

Postanowiliśmy wykorzystać możliwości stwarzane przez analizę bayesowską do wyznaczenia prognoz *ex ante* na podstawie naszego algorytmu. Dominanta czy mediana opóźnień dla wybranego chromosomu sprawdza się dla krótszych szeregów czasowych. Wykorzystujemy to zresztą w prognozach kaskadowych. Nie da się jednak ukryć, że horyzont tych prognoz jest mocno ograniczony. Aby go wydłużyć, potrzebujemy czegoś więcej.

Podstawowym problemem w analizie bayesowskiej jest ustalenie informacji *a priori*. Naszym zdaniem w przypadku prognoz kaskadowych taką informacją jest rozkład t_m dla ostatniego segmentu. Zachowanie szeregu danych w tym przedziale czasu ma duże znaczenie dla następujących bezpośrednio po nim prognoz *ex ante*.

Zgodnie z zasadą postarzania informacji zachowanie zmiennej w okresach bezpośrednio poprzedzających prognozę najsilniej wpływać będzie na wartości w kolejnych okresach. Czynimy założenie, że są to przesłanki o nieprzypadkowym charakterze. Informacją *a priori* dla prognoz *ex ante* staje się również metoda, przy pomocy której obliczymy same prognozy. Będzie nią ta z metod, którą wskaże algorytm dla ostatniego segmentu.

Dłuższe szeregi i podział na segmenty stosowany w prognozowaniu kaskadowym pozwalają zatem uwzględnić informację *a priori*. Stanie się nią rozkład wartości parametru t_m w ostatnim segmencie. Z uwagi na konstrukcję chromosomu będzie to rozkład empiryczny dyskretnej zmiennej losowej,

w którym prawdopodobieństwa realizacji poszczególnych wariantów zmiennej wyrażać się będą poprzez częstości ich występowania w ostatnim segmencie szeregu. Jako dane z próby posłużą nam geny najlepszego chromosomu jaki wystąpi w ostatniej wyznaczonej populacji[7].

Z drugiej strony chcemy, aby prognozy *ex ante* uwzględniały również związki długookresowe. Rzecz jasna rozkład t_m możemy wyznaczyć dla całego szeregu czasowego. Zależy nam jednak na wzmocnieniu wpływu obserwacji z niedalekiej przeszłości. Dlatego właśnie korygujemy rząd opóźnienia t_m wyznaczony dla całego szeregu o prawdopodobieństwo jego wystąpienia w ostatnim segmencie. Po odpowiednich obliczeniach otrzymamy prawdopodobieństwa *a posteriori* wystąpienia danej wartości parametru t_m. Przyjmijmy następujące oznaczenia:

$P(t_m)$ – prawdopodobieństwo (jako częstość) występowania danej wartości t_m w całym chromosomie;
$P(t_m|S)$ – prawdopodobieństwo (jako częstość) występowania danej wartości t_m w ostatnim segmencie (wartości te tworzą funkcję wiarygodności);
$P(S|t_m)$ – prawdopodobieństwo *a posteriori*.

Wersje PAG składające się na prognozy kaskadowe zaliczają się ogólnie rzecz biorąc do jednaj z dwóch grup: przystosowanych do szeregów ze stałym poziomem zmiennej lub zawierających trend. Rozróżnimy zatem dwa warianty włączenia prawdopodobieństw *a posteriori* do prognoz *ex ante* jako ważonej średniej ostatnich t_m obserwacji lub ważonego przyrostu t_m ostatnich obserwacji. Niech y_n oznacza ostatnią obserwację szeregu czasowego. Prognozę *ex ante* na k kolejnych okresów dla pierwszego z wymienionych wariantów opiszemy wzorem:

$$y_{n+k} = \sum_{tm} P(S|t_m) y_{n-tm+k} \tag{3.13}$$

Obliczając same prognozy *ex ante* nie stosujemy już reguły postarzania informacji. Wagami stają się prawdopodobieństwa *a posteriori*, które nie podlegają takiemu procesowi. Wysokość wagi zależy bowiem od częstości pojawiania się danej wartości parametru t_m w chromosomie. Rzecz jasna, poczynając od drugiego okresu *ex ante* prognozy zyskują charakter dynamicznych.

Przy zachowanych powyższych oznaczeniach prognoza *ex ante* w przypadku gdy w ostatnim segmencie stwierdzimy trend liniowy da się opisać wzorem:

$$y_{n+k} = y_n + k \sum_{tm} P(S|t_m)(y_{n-tm+k} - y_{n-tm+k-1}) \tag{3.14}$$

[7] Przyjmując selekcję elitarną będzie to najlepszy chromosom we wszystkich rozpatrywanych pokoleniach.

Ostatnia obserwacja w szeregu korygowana jest ważoną sumą ostatnich t_m przyrostów, w której wagami są prawdopodobieństwa *a posteriori*.

3.3.2. Prognozowanie kaskadowe – przykład

Przedstawimy teraz przykład działania prognozowania kaskadowego. Ze względu na jego ilustracyjny charakter, szereg danych liczy tylko 26 obserwacji. Należy mieć na uwadze, że metoda ta daje się stosować w przypadku znacznie bardziej obszernych zestawów danych co zostanie zaprezentowane w rozdziale następnym.

Sam szereg podzielony został na dwa segmenty. Pierwszy obejmuje początkowe 12 obserwacji, drugi resztę. W obu przypadkach występują nieduże wahania losowe (współczynnik zmienności dla całego szeregu wyniósł około 1,9%). W pierwszym podokresie występuje stały poziom zmiennej, w drugim słaby trend malejący. Początkowo jednak wahania mają większą zmienność co wykorzystaliśmy jako kryterium podziału. Wybór narzędzi predykcji dokonywać się będzie spośród czterech wariantów PAG odpowiadających odpowiednio: metodzie naiwnej bez zmian o $t_m = 4$, średniej ruchomej dla $t_m \in \langle 2, 4 \rangle$ oraz dwóm, opisanym wcześniej odmianom uwzględniającym trend ($t_m = 4$). Poszukiwanie prognoz rozpoczniemy od drugiej ze wspomnianych metod czyli średniej ruchomej, która posłuży za punkt wyjścia tak w pierwszym, jak i drugim segmencie. Ilustracja obejmie oba typy prognoz kaskadowych wraz z wyznaczaniem prognoz *ex ante* wykorzystujących analizę bayesowską.

Wprowadzimy wszakże pewną innowację. Jak napisaliśmy wcześniej każdemu z wyodrębnionych segmentów należy przypisać metodę prognozowania. W szeregach zachowujących się w przewidywalny, podręcznikowy sposób nie jest to zadanie trudne. Po pierwsze dekompozycja nie zawsze daje oczywiste wyniki szczególnie jeśli w szeregu obserwujemy nakładanie się na siebie kilku jej elementów jak na przykład duże wahania losowe i trend. Po drugie budowa naszego algorytmu genetycznego ma swoją specyfikę i dobrze byłoby dać mu swobodę w doborze jednej z czterech metod.

Zestaw metod wybranych dla poszczególnych segmentów został zatem poddany działaniu strategii ewolucyjnej (1+1) (Arabas 2004, s. 83-84). Polega ona na tym, że z zadanym z góry prawdopodobieństwem (u nas równym 0,1) dokonujemy mutacji wektora metod. Jeżeli nowy zestaw metod daje lepsze prognozy, zastępują one metody użyte poprzednio. W ten sposób algorytm stopniowo sam poprawia jakość predykcji i koryguje nasze ewentualne potknięcia w przypadku niezbyt trafnego doboru narzędzi prognostycznych.

Ogólną jakość wektora metod oceniliśmy przy pomocy ważonej średniej z wartości RMSPE dla poszczególnych segmentów, w której wagami zostały udziały liczby okresów prognoz przypadających na dany segment w stosunku

do całkowitej liczby prognoz *ex post*. Za lepszy uznawaliśmy ten zbiór metod, dla którego ważone kryterium było niższe.

Optymalizacja w prognozowaniu kaskadowym dokonuje się na dwóch poziomach. Wewnętrznym, obejmującym poszukiwanie chromosomu zapewniającego najlepsze przystosowanie w ramach wyróżnionego segmentu i zewnętrznym, odnoszącym się do poszukiwania optymalnego połączenia metod prognostycznych.

Liczbę mutacji w strategii przyjęliśmy jako równą 10. Pozostałe parametry algorytmu (jednakowe dla obydwu segmentów i obu typów kaskady) są następujące:
— wielkość populacji – 100 osobników;
— liczba pokoleń – 50;
— prawdopodobieństwo krzyżowania zawierało się w przedziale $\langle 0,0{,}3\rangle$ zaś prawdopodobieństwo mutacji w przedziale $\langle 0{,}05, 0{,}1\rangle$.

Obliczenia wykonane zarówno dla prognozowania kaskadowego z dołączonymi obserwacjami rzeczywistymi z segmentu poprzedniego jak i wykorzystującego prognozy wygasłe znalazły się w tabeli 3.8. Są to najlepsze wartości RMSPE dla wyróżnionych segmentów i każdego z wariantów algorytmu. Ostatni wiersz przedstawia ważoną wartość kryterium. Ponadto w ostatniej kolumnie, w celach porównawczych zawarte zostały wartości wspomnianego błędu prognoz dla metody Browna[8]. Zdecydowaliśmy się na wygładzanie wykładnicze ze względu na możliwość optymalizacji parametru wygładzania wg tego samego kryterium co w PAG. Obliczenia wykonane zostały dla początkowego zestawu metod przypisanych do segmentów.

Tabela 3.8: Jakość prognoz kaskadowych dla szeregu o 2 segmentach

	Prognoza kaskadowa		Metoda
RMSPE	wersja 1	wersja 2	Browna
segment 1	0,0194	0,0194	0,0219
segment 2	0,0168	0,0135	0,0191
ważone	0,0177	0,0156	0,0204

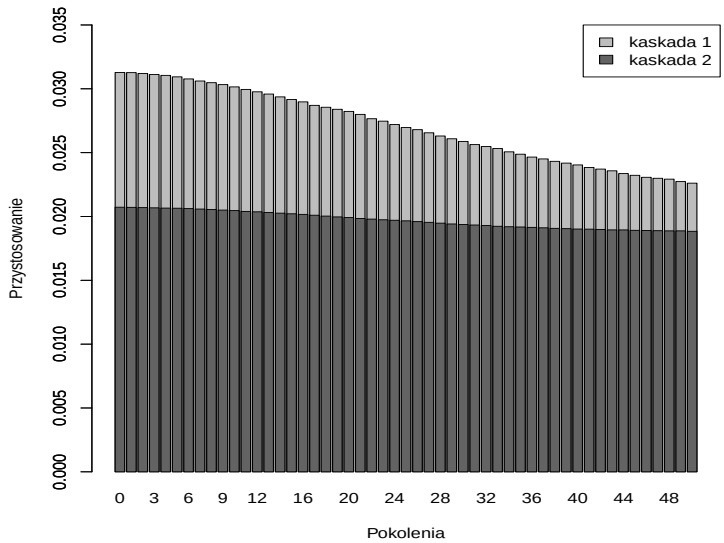
Źródło: opracowanie własne

W przypadku kaskady 2 strategia ewolucyjna zweryfikowała dokonany przez nas początkowy dobór metod i dla drugiego podokresu, pojawił się PAG oparty na trendzie liniowym drugiego rodzaju. RMSPE tak dla segmentów jak i ważone w przypadku obu wersji prognozowania kaskadowego znalazł się poniżej 2% przy czym kaskada 2 dała średnio rzecz biorąc nie-

[8] Optymalizacja wykonana została w dodatku Solver arkusza MS Excel. Ponadto użycie w drugim segmencie metody Holta pogorszyło ważone kryterium.

co niższy błąd prognoz *ex post* dla najlepszego osobnika. Metoda Browna wypada w tym porównaniu gorzej od algorytmu prognostycznego.

Z pokolenia na pokolenie w jednym i drugim segmencie następowała poprawa średniego przystosowania populacji. Podobne zachowanie obserwowaliśmy dla ważonej wartości RMSPE. Zauważymy jednak na wykresie na rysunku 3.7, że poprawa wartości kryterium okazała się być znacznie silniejsza dla drugiej wersji prognozowania kaskadowego w porównaniu do wyników wersji pierwszej. W jej przypadku obserwujemy nie tylko niższe przeciętne przystosowanie chromosomów w każdym z pokoleń, ale też postępy w poszukiwaniu lepszych wyników charakteryzuje się mniejszą dynamiką. Jak przekonamy się w ostatnim rozdziale, będzie to miało wpływ na zachowanie PAG w szeregach różniących się zmiennością.



Rysunek 3.7: Zmiany ważonego średniego przystosowania populacji uzyskane w kolejnych pokoleniach

Źródło: opracowanie własne

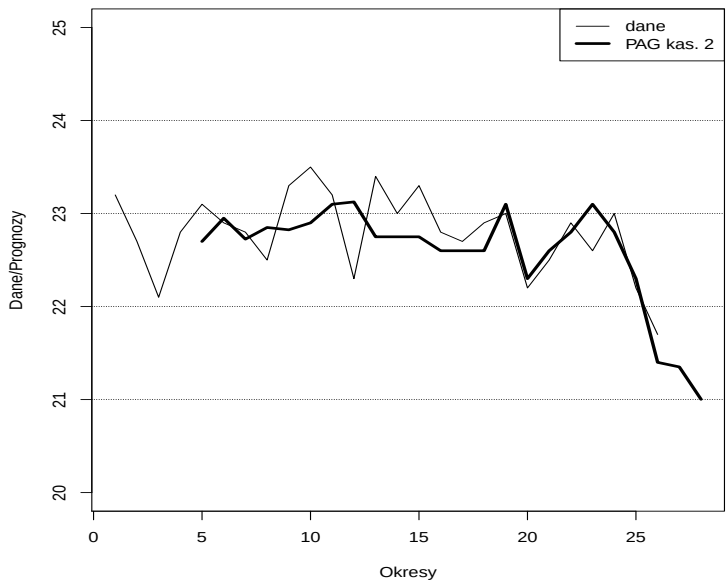
Analizując wyniki, które zamieszczone zostały w tabeli 3.8 można zauważyć, że dla drugiego segmentu lepsze wyniki uzyskaliśmy prognozując na podstawie drugiej odmiany prognozowania kaskadowego choć należy przyznać, że różnica pomiędzy nimi nie jest duża. Ważone kryterium dla kaskady 1 było zaledwie o 0,21 p% wyższe. Znacznie większe różnice występują na niekorzyść metody Browna. Poza tym strategia ewolucyjna dla kaskady 2 w drugim podokresie wskazała metodę uwzględniającą trend w odróżnieniu od kaskady 1. Rzutuje to na sposób wyznaczania prognoz *ex ante*.

Tabela 3.9: Obliczanie prawdopodobieństw *a posteriori* prognoz z przykładu – kaskada 2

t_m	$P(t_m)$	$P(t_m\|S)$	$P(t_m)P(t_m\|S)$	$P(S\|t_m)$
1	0,1818	0,2857	0,0519	0,2051
2	0,2727	0,2143	0,0584	0,2308
3	0,1818	0,2143	0,0390	0,1538
4	0,3636	0,2857	0,1039	0,4103
		$\sum$	0,2532	

Źródło: opracowanie własne

Tabela 3.9 zawiera wyniki cząstkowe oraz prawdopodobieństwa *a posteriori* obliczone na podstawie składu najlepszego chromosomu znalezionego przez PAG. W drugiej kolumnie znalazły się częstości odpowiednich wartości t_m z całego chromosomu, w trzeciej częstości z ostatniego segmentu zaś ostatnią wypełniają prawdopodobieństwa *a posteriori* będące skorygowanymi wartościami z drugiej kolumny.



Rysunek 3.8: Zachowanie prognoz kaskadowych dla drugiego wariantu ich wyznaczania

Źródło: opracowanie własne

Wskaźnik podobieństwa struktur wartości t_m między kaskadą 1 i kaskadą 2 obliczony dla prawdopodobieństw *a posteriori* wyniósł 72,1%. Liczba ta daje pogląd na temat podobieństwa struktur poziomów t_m obserwowanych w przypadku obu metod. Jak widać można je uznać za podobne co nie

powinno dziwić gdyż po pierwsze prognozami objęty został ten sam szereg, a po drugie każdy z wariantów prognozowania kaskadowego korzysta z tego samego zbioru metod PAG. Z drugiej strony wartość wskaźnika podobieństwa sugeruje, że w przypadku prognoz *ex ante* należy spodziewać się różnic istotnie wpływających na trafność predykcji. Jasnym więc chyba się staje znaczenie wyboru dokonywanego na podstawie prognoz *ex post*.

Ponieważ kaskada 2 w ostatnim segmencie wskazała na jedną z metod zawierających trend więc zgodnie ze wzorem 3.14 ostatnia dostępna obserwacja korygowana jest ważoną sumą ostatnich t_m przyrostów szeregu.

Na rysunku 3.8 zamieściliśmy porównanie danych rzeczywistych i prognoz z drugiego wariantu prognozowania kaskadowego w tym *ex ante* dla dwóch kolejnych okresów. Szereg prognoz w porównaniu do danych empirycznych okazał się dość mocno wygładzony. Dadzą się przy tym zauważyć spore różnice w końcowych okresach pierwszego segmentu, które nawiasem mówiąc wystąpiły także w wersji wykorzystującej dane z segmentu poprzedniego o czym świadczy identyczny poziom RMSPE w tabeli 3.8 dla segmentu 1 i obu kaskad. Łatwo możemy się domyślić, że zachowanie prognoz na wykresie dla drugiego segmentu wiąże się z wykorzystaniem prognoz wygasłych.

Rzut oka na wykres na rysunku 3.8 uświadamia nam również, jak istotny jest wpływ punktów podziału szeregu czasowego na poddawane predykcji segmenty. Niekiedy przesunięcie o kilka obserwacji potrafi zamienić wynik dekompozycji. Zalecamy więc uwagę w tym aspekcie, gdyż oddziaływanie wspomnianego podziału wykracza poza jakość prognoz *ex post* i przekłada się na zachowanie prognoz *ex ante*.

Rozdział 4

Prognozy wybranych kategorii przy pomocy PAG

W poprzednim rozdziale skupiliśmy się na przybliżeniu tego w jaki sposób zaproponowana hybryda algorytmu genetycznego i ewolucyjnego łączy się z prognozowaniem w jedną, zintegrowaną metodę. Przedstawiliśmy jak skonstruować chromosom będący podstawą PAG oraz zaprezentowaliśmy zachowanie algorytmu dla szeregów o różnej dekompozycji. Szeregi użyte w przykładach nie zaliczały się jednak do specjalnie długich, a zatem nie dało się w nich zaobserwować długookresowych zależności. Nie podlegały skomplikowanej dekompozycji. Nadawały się za to do zilustrowania własności metody. Nasz cel w bieżącym rozdziale to wykorzystanie i włączenie do prognoz znacznie większej ilości informacji. Opisane poprzednio prognozowanie kaskadowe pozwala rozwiązać ten i jeszcze jeden problem, którym jest gwałtowny wzrost liczby wariantów prognostycznych wymagających sprawdzenia.

4.1. Prognozy *ex post* i *ex ante* wybranych szeregów danych

4.1.1. Prognozy notowań i wolumenu obrotów danych giełdowych

Ta część rozdziału poświęcona została wykorzystaniu prognostycznego algorytmu genetycznego i obejmujących go prognoz kaskadowych w przypadku wybranych zjawisk ekonomicznych dających się sprowadzić do wspólnego mianownika procesów finansowych. Interesowały nas szeregi czasowe o własnościach utrudniających predykcję za pomocą klasycznych narzędzi. Trafność otrzymanych prognoz ocenialiśmy w okresach *ex post* porównując je z klasycznymi metodami. Ponadto dla PAG sprawdziliśmy również dokładność prognoz wygasłych.

Jako pierwsze weźmiemy pod uwagę notowania indeksu WIG. Wybraliśmy go nie tylko ze względu na fakt, że informuje o przeszłej i bieżącej kondycji warszawskiej giełdy. Również dlatego, że poziom WIG traktowany jest jako wyznacznik nastrojów i kondycji całej gospodarki. Uznaliśmy więc za interesujące podjęcie próby wyprognozowania jego notowań.

Zdajemy sobie sprawę, że współczesne teorie rynków finansowych podkreślają, oprócz roli stopy zwrotu i ryzyka, znaczenie płynności w analizach giełdowych. Z tego powodu w dalszej części rozdziału znalazły się także prognozy wolumenu obrotów.

Pierwszy z rozważanych szeregów danych obejmował dzienne kursy zamknięcia z okresu od 1.07.2011 do 17.01.2012 r., a więc nieco ponad pół roku (138 sesji). Notowania WIG podzieliliśmy na 3 podokresy (zwane dalej również segmentami). O podziale zadecydowało zachowanie szeregu prześledzone na wykresie.

Pierwszy segment trwał od 1.07.2011 do 10.08.2011 r. obejmując 29 obserwacji kiedy to dał się zaobserwować systematyczny spadek notowań. Początkowo słaby, gwałtownie przyspiesza pod koniec tego okresu. Towarzyszą mu małe wahania losowe.

Drugi, nieco dłuższy segment, sięga od sesji następującej po zakończeniu segmentu poprzedniego tj. 11.08.2011 r. do 4.10.2011 r. włącznie. W ciągu owych 36 sesji wyróżniliśmy stały poziom zmiennej z dużymi wahaniami losowymi i kilkoma lokalnymi szczytami notowań jak na przykład 31 sierpnia.

Wreszcie trzeci, najdłuższy segment przypadający na drugą połowę badanego okresu charakteryzował się stałym poziomem zmiennej (brakiem trendu) z wahaniami o bardzo różnorodnej amplitudzie. Zachowanie szeregu w tym podokresie przypominało trend wielomianowy, na który nakładały się wahania losowe. Nie znaleźliśmy w jego przypadku powodów, które uzasadniałyby dalszy podział.

Przejdźmy do omówienia parametrów algorytmu. Przyjęliśmy, że parametry sterujące przebiegiem obliczeń będą jednakowe dla każdego segmentu. Wielkość populacji ustaliliśmy na poziomie 100 osobników (chromosomów). Oznacza to, że w każdym pokoleniu dysponujemy stoma wariantami prognoz[1]. Liczba pokoleń równała się 150 zaś kryterium oceny jakości prognoz *ex post* była wartość RMSPE.

Jeśli chodzi o prawdopodobieństwo krzyżowania (p_k), to rozważyliśmy dwa warianty wyraźnie różniące się szansą na uzyskanie potomka. Obliczenia przeprowadziliśmy przy $p_k \in \langle 0,0{,}3 \rangle$ oraz $p_k \in \langle 0,0{,}6 \rangle$. Zmiany prawdopodobieństwa krzyżowania zachodziły w sposób podobny do tego jaki opisano w (Agustín-Blas et al. 2012) czyli rosło ono pod koniec postępowania, aby zapobiec zdominowaniu populacji przez zbyt podobne do siebie osobniki.

Przyjęcie i porównanie dwóch wariantów p_k miało na celu sprawdzenie jak na trafność prognoz wpływa dwukrotnie większa szansa utworzenia potomka. W stosunku do prawdopodobieństwa mutacji (p_m) zastosowaliśmy podobny zabieg tj. należało ono do przedziału $\langle 0,0{,}1 \rangle$, ale w jego wypadku był to jedyny wykorzystany wariant.

[1] Oczywiście niektóre z nich mogą się powtarzać, zwłaszcza pod koniec postępowania.

Rozważyliśmy także dwa zestawy parametrów sterujących dopuszczalnym, maksymalnym opóźnieniem obserwacji stającej się prognozą (t_m)). Przypomnijmy, że algorytm stara się dopasować jedną z czterech metod (rozwinięcie metody naiwnej prostej, średniej ruchomej i dwie wersje metody naiwnej z poprawką na obecność trendu). W związku z tym przyjęliśmy $t_m = 10$ w pierwszej serii obliczeń. Biorąc pod uwagę, że tydzień giełdowy liczy sobie (w pewnym uproszczeniu) 5 dni możemy założyć, że sięgamy w przeszłość maksymalnie 2 tygodnie wstecz. Na użytek PAG opartego na średniej ruchomej ustaliliśmy $t_m \in \langle 4, 10 \rangle$. Odpowiada to niezbyt silnemu wygładzeniu prognoz *ex post*.

Druga seria obliczeń różniła się parametrami opóźnień. Założyliśmy, że $t_m = 5$ oraz $t_m \in \langle 3, 5 \rangle$. Tak więc w tym przypadku prognozy powstają w oparciu o wartości mniej odległe w czasie, a ich ewentualny stopnień wygładzenia jest silniejszy niż poprzednio.

Jako punkt wyjścia prognoz kaskadowych ustaliliśmy następujący wektor metod: $[\,3\;2\;1\,]$. Zatem początkowo segment pierwszy prognozowany będzie przez PAG wykorzystujący pierwszy wariant poprawki ze względu na trend. Drugi podokres poddany został działaniu algorytmu zbudowanego na bazie średniej ruchomej, zaś trzeci – metody naiwnej prostej. Prawdopodobieństwo mutacji w strategii ewolucyjnej wynosiło 0,1, natomiast liczba przebiegów samej strategii równała się 10.

Pierwsze prognozy wykonywane są dla początkowego, przyjętego przez badacza zestawu metod. Algorytm przechodzi przez zadaną liczbę pokoleń używając metody przypisanej do danego segmentu. Następnie wektor metod poddawany jest mutacji po czym procedura powtarza się.

Możemy mieć do czynienia z dwiema sytuacjami: mutacja doprowadziła do zmiany przynajmniej jednego jednego elementu wektora metod lub nie doszło do żadnej zmiany. Wbrew pozorom obie sytuacje niosą ze sobą korzyści, choć są one różne. Zmiana składowej wektora daje szansę sprawdzenia czy inny zbiór narzędzi predykcji pozwala osiągnąć lepsze wyniki. Pozostawienie zestawu metod bez zmian pozwala z kolei sprawdzić go na innej populacji osobników co przydaje się z uwagi na rozmiary zbioru rozwiązań. Odpowiada to kilkukrotnemu uruchomieniu algorytmu z tymi samymi parametrami.

Ze względu na to, iż w przyjętej strategii ewolucyjnej występuje jedynie operator mutacji, nie ma zaś selekcji jako takiej[2] ograniczyliśmy się do zapamiętywania najlepszego dotychczasowego rozwiązania. W następnej iteracji sprawdzamy, czy kolejne okazało się lepsze. Jeżeli tak – staje się ono rozwiązaniem bieżącym, jeżeli nie – przechodzimy do następnej iteracji i następnego zestawu metod.

[2] W strategii ewolucyjnej dotychczasowy osobnik zostaje zastąpiony nowym tylko wtedy gdy w efekcie mutacji udało się poprawić wyniki.

Jako kryterium oceny strategii ewolucyjnej posłużyliśmy się ważoną wartością RMSPE z wagami uzależnionymi od długości danego podokresu.

Tabela 4.1: RMSPE dla prognoz notowań WIG przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0107	0,0093	0,0110	0,0083
Segment 2	0,0176	0,0149	0,0201	0,0181
Segment 3	0,0179	0,0153	0,0182	0,0147
Ważone kryt.	0,0164	0,0140	0,0173	0,0144
Metody	[3 1 3]	[3 1 1]	[3 1 1]	[3 1 1]

Źródło: opracowanie własne

Tabela 4.2: RMSPE dla prognoz notowań WIG przy $t_m = 5$

	Kaskada 1		Kaskada 2 D	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0096	0,0089	0,0101	0,0090
Segment 2	0,0205	0,0194	0,0220	0,0206
Segment 3	0,0133	0,0122	0,0143	0,0128
Ważone kryt.	0,0146	0,0136	0,0156	0,0142
Metody	[3 1 1]	[4 1 1]	[3 1 1]	[3 1 1]

Źródło: opracowanie własne

W tabeli 4.1 znalazły się wartości RMSPE dla poszczególnych segmentów oraz ważona wartość dla całego szeregu czasowego. Zestawiliśmy do porównania dwie wersje prognozowania kaskadowego, każdą dla dwóch wariantów prawdopodobieństwa krzyżowania. Kaskadę w wersji drugiej przeanalizowaliśmy przy założeniu, że prognozy *ex ante* dla segmentów drugiego i trzeciego powstały na podstawie dominanty zbioru wartości parametru t_m gdyż te na podstawie mediany wykazywały się wyższymi błędami.

Analizując ważoną wartość kryterium ze strategii ewolucyjnej możemy zauważyć, że prognozy *ex post* dla całego szeregu są bardzo przyzwoite. Znaleźliśmy się znacznie poniżej 5% progu, który w literaturze uznaje się za umowną granicę bardzo dobrych prognoz. Jakość predykcji zmierzona dla kaskady 2 jest tylko minimalnie gorsza od kaskady w pierwszej wersji. Nadal jednak są to bardzo dobre wyniki.

Przyglądając się w tabeli 4.1 rezultatom dla poszczególnych segmentów wyciągamy wniosek, że najmniej kłopotliwy z punktu widzenia trafności

prognoz okazał się pierwszy z nich. Właśnie dla tego podokresu uzyskaliśmy najmniejsze średnie błędy względne. Za najtrudniejszy do wyprognozowania należy uznać drugi segment. Przypomnijmy, że dochodziło tam do znacznych zmian wielkości wahań losowych.

Oceniając różnice dla zastosowanych wartości prawdopodobieństwa krzyżowania trudno zająć jednoznaczne stanowisko gdyż różnice między ważonymi kryteriami nie są duże. Wyższe p_k wpływa w pozytywny sposób na jakość prognoz *ex post* uzyskanych prognostycznym algorytmem genetycznym z $t_m = 10$, ale powoduje wzrost czasu obliczeń.

Niewielkie różnice obserwujemy zestawiając ostateczne wektory metod otrzymane dla wszystkich, zawartych w tabeli 4.1 wariantów prognoz. Prognozowanie kaskadowe 1 zakończyło obliczenia z różnymi zestawami w zależności od prawdopodobieństwa krzyżowania. Okazały się one również odmienne od wyjściowego wektora. Tworząc go sugerowaliśmy się oczywiście dekompozycją szeregu przypadającego na dany segment, ale strategia ewolucyjna zdołała jeszcze poprawić nasz wybór.

Pamiętajmy, że w tym wariancie dołączamy do segmentów drugiego oraz dalszych kilka ostatnich wartości z segmentu poprzedniego. Pomiędzy wartościami granicznymi sąsiednich segmentów zazwyczaj nie występują duże różnice. Nawet w sytuacji gdy szereg przechodził od trendu do stałego poziomu zmiennej lub odwrotnie, pierwsze kilka obserwacji nowego segmentu charakteryzowało się wartościami zbliżonymi do tych, które kończą segment poprzedni.

Z kolei prognozowanie kaskadowe 2 opiera się na odmiennym założeniu. W jego myśl brakujące początkowe wartości danego segmentu zostają uzupełnione prognozami uzyskanymi metodą przypisaną do segmentu poprzedniego. Wpływa to na wybór metody, która z jednej strony musi dopasować się do zachowania bieżącego segmentu, a z drugiej dostosować się do przewidywanej dekompozycji segmentu następnego. Jak się okazało algorytm pozostał przy tym samym zestawie metod w obu przypadkach dotyczących kaskady 2.

W tabeli 4.2 zawarte zostały wyniki opisujące jakość prognoz *ex post* przy $t_m = 5$. Oprócz niższej wartości tego parametru zastosowaliśmy średnią ważoną mocniej wygładzającą szereg niż miało to miejsce poprzednio.

Porównując wyniki z tabel 4.1 i 4.2 możemy zauważyć, że ważone kryterium przyjęło mniejsze wartości dla nowego poziomu t_m i dla $p_k \in \langle 0, 0{,}6 \rangle$.

Podobnie jak miało to miejsce poprzednio, najlepiej algorytm radził sobie w segmencie pierwszym, najgorzej w drugim. Przy tej okazji ujawnia się jeszcze jedna zaleta PAG. Pomaga on wskazać podokresy szeregu czasowego, których zachowanie utrudnia uzyskanie dobrych prognoz.

Dla celów porównawczych obliczyliśmy prognozy *ex post* metodami klasycznymi. Tam gdzie to było możliwe staraliśmy się używać wygładzania

wykładniczego, które daje szansę optymalizacji parametrów ze względu na przyjęte kryterium czyli RMSPE. Obliczyliśmy również ważone kryterium w analogiczny sposób jak w prognozowaniu kaskadowym. Tabela 4.3 prezentuje osiągnięte tą drogą rezultaty.

Tabela 4.3: RMSPE dla prognoz notowań WIG metodami klasycznymi

	Metoda	RMSPE
Segment 1	Holta	0,0120
Segment 2	naiwna prosta	0,0236
Segment 3	Browna	0,0137
Ważone kryt.		0,0161

Źródło: opracowanie własne

Metody klasyczne wypadają nieco gorzej od PAG zwłaszcza w porównaniu z wynikami dla wyższego prawdopodobieństwa krzyżowania. Zbliżone lub nawet lepsze rezultaty obserwujemy tylko w niektórych przypadkach użycia $t_m = 10$.

Kolejnym szeregiem poddanym działaniu prognostycznego algorytmu genetycznego jest wolumen obrotów indeksu WIG. Tego typu zmienną wykorzystuje się do oceny płynności aktywów dostępnych na rynku giełdowym. Płynność stanowi trzecią, obok stopy zwrotu i ryzyka, podstawową charakterystykę inwestycji w instrumenty finansowe. Dany walor uznajemy za płynny jeżeli można nim szybko obracać w dużych ilościach bez istotnego wpływu na wysokość notowań.

Jednym z bardziej popularnych sposobów określania płynności jest relacja wartości obrotów do kapitalizacji. Jeżeli wartość obrotów rośnie szybciej od kapitalizacji wówczas wnioskujemy o wzroście płynności rynku. Wartość obrotów bezpośrednio zależy od liczby zawieranych transakcji oraz zmian cen (notowań). Łatwo więc zrozumieć znaczenie prognoz wolumenu obrotów.

Dane na temat obrotów pochodziły z tego samego okresu co kursy zamknięcia indeksu. Wyodrębniliśmy ponownie trzy segmenty choć tym razem o innych granicach. Pierwszy z nich objął dane od 1.07.2011 do 2.09.2011 r. (45 obserwacji). Kryterium podziału stało się zaobserwowanie serii dużych wahań rosnących wykładniczo. W ciągu owych dwóch miesięcy odnotowaliśmy pojawianie się w nieregularnych odstępach czasu pojedynczych sesji lub ich krótkich serii charakteryzujących się zwyżkami obrotów.

Kolejny segment liczył 33 sesje i sięgał od zakończenia poprzedniego podokresu po 19.10.2011 r. W tym czasie dochodziło do stopniowego zmniejszania amplitudy wahań choć nadal pozostawały one duże. Towarzyszył im stały poziom zmiennej.

W ostatnim, obejmującym 60 sesji segmencie wahania obrotów stają się stopniowo coraz większe. W odróżnieniu jednak od pierwszego segmentu zmiany (tak wzrosty jak i spadki) rosną wykładniczo.

Ze względu na brak trendu i powtarzające się duże wahania losowe dla każdego z wyszczególnionych segmentów przyjęliśmy jako wyjściowy prognostyczny algorytm genetyczny zbudowany w oparciu o średnią ruchomą. Aby zapewnić możliwość porównania z wynikami prognoz notowań indeksu parametry algorytmu pozostały bez zmian. W tabeli 4.4 znalazły się wartości RMSPE dla segmentów, ważone kryterium strategii dla ewolucyjnej oraz zbiory metod, które zapewniły zawarte w niej wyniki dla obu wersji prognozowania kaskadowego oraz $t_m = 10$ i $t_m \in \langle 4, 10 \rangle$.

Pierwszym co zwraca uwagę są dużo wyższe (w porównaniu z prognozami kursu WIG) uzyskane wartości kryteriów. Oscylują one około 17-22%. Nie powinno to jednak dziwić. Po pierwsze, szereg obrotów często przechodził od bardzo niskich do wysokich wielkości i odwrotnie. Po drugie, same obroty są kategorią trudną w modelowaniu. Stanowią bowiem wypadkową decyzji inwestorów składających zlecenia kupna i sprzedaży. Podejmowane decyzje nie zawsze dadzą się racjonalnie uzasadnić. Z tego punktu widzenia uzyskane rezultaty niekoniecznie wyglądają tak źle.

W pracy (Kucharski 2006) można znaleźć porównanie PAG w wersji dla metody naiwnej z tradycyjną metodą naiwną prostą właśnie podczas prognozowania obrotów spółek giełdowych. Algorytm dawał znacznie niższe błędy predykcji dla zawartych w artykule wyników.

Tabela 4.4: RMSPE dla prognoz wolumenu WIG przy $t_m = 10$

	Kaskada 1		Kaskada 2 D	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,2326	0,2040	0,2342	0,1859
Segment 2	0,1609	0,1029	0,1315	0,1436
Segment 3	0,2219	0,2194	0,2189	0,2064
Ważone kryt.	0,2105	0,1863	0,2025	0,1878
Metody	[1 4 1]	[1 1 1]	[1 1 1]	[1 4 1]

Źródło: opracowanie własne

Porównując w tabeli 4.4 odpowiadające sobie wersje prognozowania kaskadowego zauważamy, że trochę różnią się co do wartości ważonego błędu prognoz *ex post*. Najniższe RMSPE udało się uzyskać dla drugiego podokresu. Segmenty pierwszy i trzeci odznaczały prognozami się z wartością kryterium o około 5 punktów wyższą.

Wszystkie przebadane zestawy metod wykazały się umiarkowaną zgodnością podczas ustalania optymalnego zestawu metod dla całego szeregu

Tabela 4.5: RMSPE dla prognoz wolumenu WIG przy $t_m = 5$

	Kaskada 1		Kaskada 2 D	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,2150	0,1908	0,2004	0,2039
Segment 2	0,1717	0,1617	0,1712	0,1241
Segment 3	0,2251	0,2091	0,2368	0,1988
Ważone kryt.	0,2089	0,1918	0,2094	0,1823
Metody	[1 4 1]	[1 4 1]	[1 4 1]	[1 1 1]

Źródło: opracowanie własne

czasowego, który finalnie przyjął odmienną od początkowej postać. Poza drugim segmentem wykluczony został udział metod uwzględniających poprawkę ze względu na trend[3].

Tabela 4.5 zawiera zestawienie wartości pierwiastka procentowego błędu średniokwadratowego dla tych samych co poprzednio metod, ale z parametrem $t_m = 5$ i $t_m \in \langle 3, 5 \rangle$. Wyniki te generalnie są nieco gorsze od swoich odpowiedników z tabeli 4.4. Ważona wartość kryterium oscylowała w okolicach 20%. Co się tyczy doboru metod to był on bardziej jednoznaczny niż przy $t_m = 10$ gdyż tylko kaskada 2 z wyższym prawdopodobieństwem krzyżowania charakteryzowała się odmienną metodą prognozowania drugiego segmentu.

Śledząc znajdujące się w tabeli 4.6 rezultaty trafności prognoz *ex post* dostrzegamy dużą przewagę algorytmu prognostycznego. O ile dla notowań indeksu można było mówić o porównywalności jakości predykcji, o tyle tutaj takie stwierdzenie staje się kompletnie nieuzasadnione. Wyższość PAG nad metodami klasycznymi zapewniła elastyczność algorytmu w doborze składowych tworzących prognozy.

Tabela 4.6: RMSPE dla prognoz wolumenu obrotów WIG metodami klasycznymi

	Metoda	RMSPE
Segment 1	Browna	0,3372
Segment 2	średnia ruch. o $k = 5$	0,2115
Segment 3	Browna	0,3727
Ważone kryt.		0,3253

Źródło: opracowanie własne

[3] Gwoli ścisłości należy dodać, że ze względu na skalę zmian w segmentach 1 i 3 przetestowane zostały wektory metod uwzględniające trend. Jednak uzyskane wyniki okazały się wyraźnie gorsze niż PAG oparty na średniej ruchomej o tym samym t_m.

Zachowanie szeregu wolumenu obrotów zdecydowanie różniło się od zachowania szeregu notowań. Decydującą rolę odegrały bardzo silne, następujące często z sesji na sesję wahania ilości kupowanych i sprzedawanych akcji. Stąd wysokie wartości RMSPE tak dla segmentów jak i dla całego szeregu.

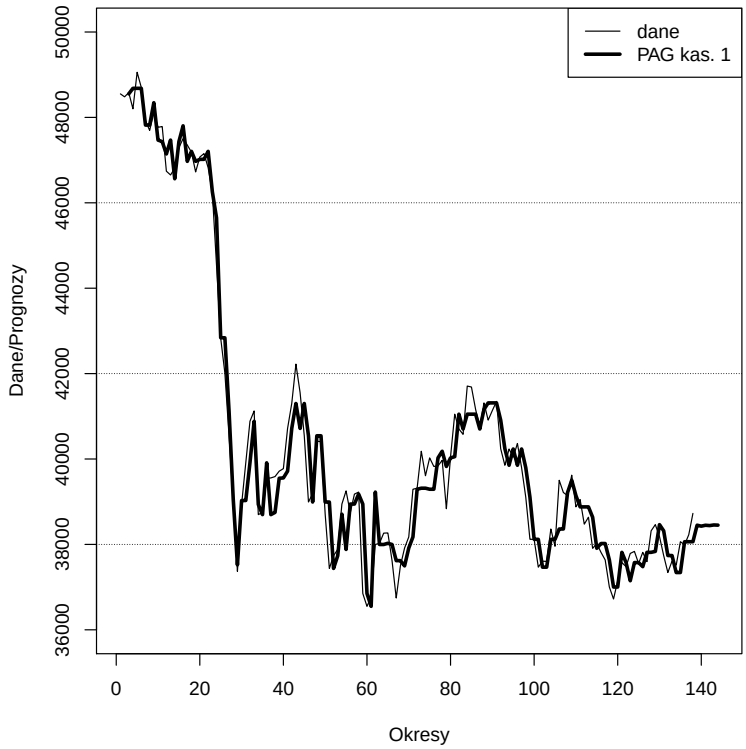
Po omówieniu zachowania prognoz *ex post* przejdźmy do wyznaczenia prognoz *ex ante* wykorzystując analizę bayesowską, o której wspominaliśmy w rozdziale trzecim. Porównaliśmy wyniki kaskady 1 oraz 2 i spośród wszystkich zebranych w tabelach 4.1-4.2 a także 4.4-4.5 wybrane zostały prognozy o najlepszych parametrach.

W każdym z czterech wymienionych powyżej przypadków ostatni segment prognozowany był przy pomocy PAG rozszerzającego metodę naiwną bez poprawki. Brak trendu i sezonowości w tych okresach rzutuje na sposób tworzenia prognoz *ex ante*, które obliczone zostały wg wzoru (3.13). Wykresy na rysunkach 4.1 i 4.2 prezentują prognozy *ex post* oraz bayesowskie *ex ante* notowań a także wolumenu obrotów indeksu WIG. Wykorzystaliśmy parametry, które zapewniły najlepsze rezultaty dla tej metody w tabelach 4.2 i 4.5. Same prognozy *ex ante* sięgały 6 okresów poza zakres danych szeregu czasowego (od 18.01.2012 do 25.01.2012 r.).

Analizując wykres na rysunku 4.2 dostrzegamy, że za wysokie wartości RMSPE odpowiadają nietrafione prognozy kilku punktów zwrotnych, szczególnie w ostatniej części szeregu. Choć we wszystkich przypadkach, które znalazły się w tabelach 4.4 i 4.5 w ostatnim segmencie prognoz wolumenu WIG wybrany został PAG w wersji pierwszej to zachowanie szeregu prognoz na wykresie przypomina średnią ruchomą o niskiej stałej wygładzania. W efekcie uzyskaliśmy wyższe niż w przypadku notowań reszty z prognoz. Trzeba jednak zwrócić uwagę, że część nagłych wzrostów i spadków zmiennej została przez algorytm wychwycona i uwzględniona.

Jakość prognoz *ex ante* weryfikujemy w momencie ich wygaśnięcia. Znamienne jest, że w literaturze rzadko spotykamy się z takim sposobem oceny wyników predykcji. Łatwo to zrozumieć w pracach traktujących o prognozach bazujących na danych o niskiej częstotliwości, zwłaszcza przygotowywanych na średni lub długi horyzont. W momencie złożenia publikacji ich autorzy nie posiadają odpowiedniego materiału porównawczego. Nawet jednak dysponując taką możliwością, zwykle ograniczają się jedynie do przedstawienia trafności prognoz *ex post*. W skrajnym przypadku może dojść do tego, że dobrane w ten czy inny sposób parametry pozwalają metodzie doskonale dopasować się do obserwacji z szeregu czasowego, a po dodaniu kolejnych danych okazuje się, że prognozy *ex ante* są mało trafne[4]. Postanowiliśmy wykorzystać szansę jaką daje częstotliwość danych w omawianych

[4] Z podobnym problemem spotykamy się w procesie uczenia sieci neuronowych. Nosi on tam nazwę „przeuczenia sieci".



Rysunek 4.1: Prognozy *ex post* i *ex ante* kursu WIG

Źródło: opracowanie własne

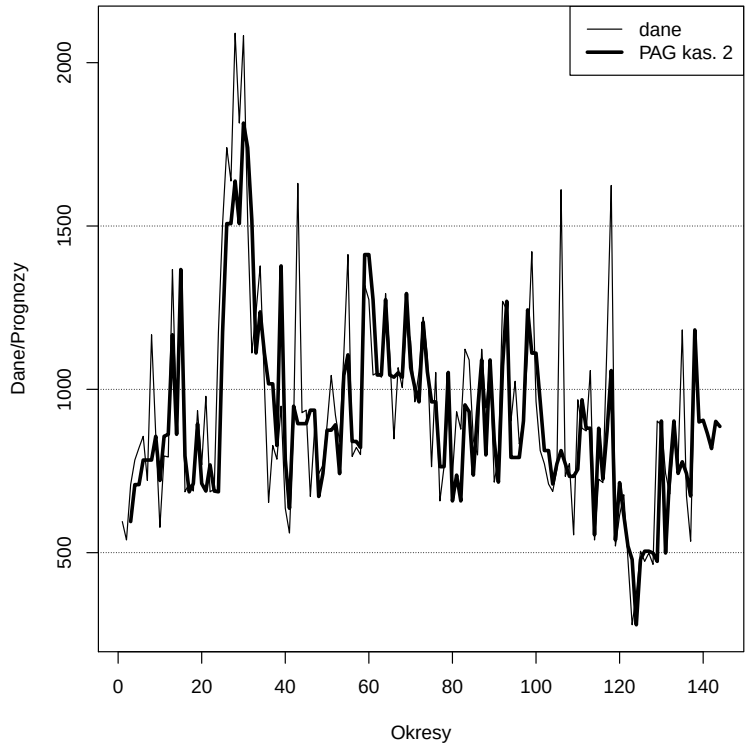
zastosowaniach i przetestować algorytm prognostyczny pod kątem jakości prognoz wygasłych.

Przypomnijmy, że w kaskadę 2 wbudowany został mechanizm krótkookresowych prognoz *ex ante* dla sąsiadujących segmentów. *De facto* korzystamy z prognoz wygasłych już na etapie dopasowywania PAG w kolejnych fragmentach szeregu. W tabeli 4.7 przedstawione zostały wartości RMSPE uzyskane po wygaśnięciu prognoz wykonanych poza zakresem całości dostępnych danych czyli poza szeregiem czasowym.

Tabela 4.7: Dokładność prognoz wygasłych notowań i wolumenu obrotów WIG

Wartość	Notowania		Wolumen	
	Kaskada 1	Kaskada 2	Kaskada 1	Kaskada 2
RMSPE	0,0258	0,0250	0,2808	0,3001

Źródło: opracowanie własne



Rysunek 4.2: Prognozy *ex post* i *ex ante* wolumenu WIG

Źródło: opracowanie własne

Jak pamiętamy prognozy *ex post* notowań i wolumenu obrotów różniły się o cały rząd wielkości. Tendencja ta została zachowana również w okresie *ex ante*. Obserwujemy także wyższe wartości RMSPE prognoz wygasłych niż dla *ex post*, ale tego akurat należało się spodziewać. Ważniejsze jest, że nie następuje drastyczny spadek dokładności predykcji.

Drugi wariant prognozowania kaskadowego charakteryzował się minimalnie większą trafnością podczas predykcji notowań. Kaskada 1 dała lepsze prognozy *ex ante* (ocenione po ich po wygaśnięciu) jeśli chodzi o obroty. Ze względu na większą zmienność obserwowaną w szeregu różnica w porównaniu do stosownej wartości kryterium konkurencyjnej metody okazała się większa.

Na warszawskiej giełdzie łącznie notowanych jest ponad 400 spółek. Taka ilość zdecydowanie wykracza poza ramy książki. Z drugiej strony indeks giełdowy podaje syntetyczną informację na temat sytuacji na parkiecie oraz panujących na nim nastrojów. Zachowanie pojedynczych uczestników rynku może odbiegać od zachowania rynku jako całości. Aby zatem rozszerzyć

analizę wykorzystania PAG do szeregów danych giełdowych postanowiliśmy zaprezentować rezultaty osiągnięte w przypadku dwóch konkretnych spółek: KGHM i TVN.

Wybór padł akurat na te walory z dwóch powodów: interesująca, skomplikowana dekompozycja szeregu czasowego oraz kapitalizacja, z racji której KGHM oraz TVN cieszą się dużym zainteresowaniem m.in. drobnych inwestorów. Nie bez znaczenia był także fakt przynależności wspomnianych firm do dwóch zupełnie różnych branż co pozwoli udowodnić elastyczność algorytmu.

Notowania KGHM (konkretnie były to kursy zamknięcia) obejmowały ten sam przedział czasu co dane wykorzystane do prognozowania WIG czyli od 1.07.2011 do 17.01.2012 r. Szereg podzielony został na cztery segmenty. Pierwszy liczył sobie 39 obserwacji (od początku próby do 25.08.2011 roku włącznie) i początkowo charakteryzuje się trendem malejącym o małych wahaniach losowych. W połowie szeregu dochodzi do gwałtownego spadku kursu zamknięcia, po którym odnotowujemy trwający już do końca segmentu wzrost amplitudy wahań przypadkowych.

Drugi podokres zawiera się między 26.08.2011 a 26.09.2011 r. więc tworzą go 22 obserwacje. Tu również obserwujemy trend malejący, ale wahania są znacznie większe niż w poprzednim segmencie. Pod koniec wspomnianego okresu nastąpił silny spadek kursu KGHM.

Trzeci segment obejmujący sesje między 27.09.2011 a 21.11.2011 r. (38 obserwacji) przynosi zmianę tendencji rozwojowej w porównaniu do poprzednich dwóch segmentów. Trendowi rosnącemu towarzyszą tu nieduże wahana losowe. Jedynie na początku i na końcu tego okresu pojawiają się wahania o wyższej amplitudzie.

W ostatnim, czwartym segmencie znalazły się notowania od 22.11.2011 do końca próby czyli 17.01.2012 roku. W tym czasie obserwujemy generalnie stały poziom zmiennej z tym, że w środkowej części segmentu nastąpił trwający kilka sesji duży spadek kursu, który jednak do końca okresu został zniwelowany. Ponadto mamy również do czynienia z dużymi wahaniami losowymi.

Jako że parametry zastosowane dla indeksu WIG okazały się dawać dobre rezultaty, w prognostycznym algorytmie genetycznym użytym do predykcji notowań KGHM skorzystaliśmy z ich dotychczasowych wartości. Zmianie uległ jedynie początkowy zestaw metod, który uwzględnić musiał wyniki bieżącej dekompozycji.

Strategia ewolucyjna, jak poprzednio liczyła sobie 10 iteracji, w trakcie których wektor metod podlegał mutacji. Segmenty od pierwszego do trzeciego włącznie rozpoczynały obliczenia z PAG w wersji trzeciej (czyli wykorzystującej trend) zaś segmentowi czwartemu przypisaliśmy pierwszą wersję algorytmu (będącą rozwinięciem metody naiwnej prostej).

W tabelach 4.8 i 4.9 znalazły się wartości RMSPE prognoz *ex post* dla poszczególnych segmentów oraz ważona wielkość dla całego szeregu. Po raz kolejny zestawiliśmy ze sobą wyniki osiągnięte przy $t_m = 10$ oraz $t_m = 5$.

Tabela 4.8: RMSPE dla prognoz notowań KGHM przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0184	0,0187	0,0194	0,0176
Segment 2	0,0307	0,0255	0,0297	0,0253
Segment 3	0,0472	0,0368	0,0438	0,0392
Segment 4	0,0337	0,0322	0,0336	0,0288
Ważone kryt.	0,0328	0,0287	0,0320	0,0281
Metody	[3 3 3 1]	[3 3 3 1]	[3 3 3 1]	[3 3 3 1]

Źródło: opracowanie własne

Tabela 4.9: RMSPE dla prognoz notowań KGHM przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0192	0,0166	0,0183	0,0175
Segment 2	0,0345	0,0341	0,0349	0,0341
Segment 3	0,0372	0,0349	0,0394	0,0347
Segment 4	0,0325	0,0294	0,0324	0,0286
Ważone kryt.	0,0305	0,0282	0,0309	0,0280
Metody	[3 3 3 1]	[3 3 3 1]	[3 3 3 1]	[3 3 3 1]

Źródło: opracowanie własne

Porównując uśredniony miernik błędów prognoz *ex post* dla poszczególnych segmentów zauważymy, że najwyższe odchylenia wartości prognozowanych od rzeczywistych wystąpiły w segmentach trzecim i czwartym. Jeśli chodzi o segment trzeci to tutaj przyczyną jest amplituda wahań losowych na początku i na końcu wyróżnionego okresu. Notowania akcji KGHM gwałtownie spadły 4.04.2011 r. i 21.11.2011 r. by po kilku sesjach odzyskać straconą wartość, ale pomiędzy owymi datami kurs zachowywał się już znacznie spokojniej.

Ciekawą sprawą jest zachowanie prognoz w czwartym segmencie. Przypomnijmy, że w połowie podokresu po raz kolejny wystąpiła gwałtowna przecena akcji. W ciągu kilkunastu kolejnych sesji notowania wróciły na wcześniejszy poziom, ale towarzyszyły temu duże wahania w górę i w dół.

Pomimo jednak tak „urozmaiconego” zachowania szeregu błąd średni kształtował się na poziomie około 3% co stanowi bardzo dobrą wartość.

Całościowo patrząc na trafność prognoz *ex post* nawet w najgorzej prezentującym się segmencie uzyskaliśmy prognozy dobrej jakości dzięki elastyczności PAG. W konsekwencji ważone kryterium przyjęło niski, a zatem korzystny z punktu widzenia predykcji poziom. Trafny również okazał się dokonany na początku dobór metod, ponieważ jak można to zaobserwować w tabelach 4.8 i 4.9 strategia ewolucyjna nie dokonała zmian w ich zbiorze.

Konfrontując wartości RMSPE dla różnych prawdopodobieństw krzyżowania zaobserwujemy wyraźną przewagę prognoz uzyskanych przy p_k należącym do przedziału od 0 do 0,6. Z kolei analizując wpływ poziomu parametru t_m na funkcję kryterium możemy zauważyć, że nieco lepiej sprawdzał się niższy czyli równy 5 poziom maksymalnego opóźnienia obserwacji będącej prognozą na dany okres.

W tabeli 4.10 przedstawione zostały wartości procentowego błędu średniokwadratowego osiągnięte po zastosowaniu optymalizacji parametrów dla metody Holta dla poszczególnych segmentów oraz po obliczeniu ważonego RMSPE. Choć błędy również nie okazały się wysokie to jednak PAG dawał bardziej trafne prognozy w porównaniu z każdyą z wymienionych metod.

Tabela 4.10: RMSPE dla prognoz notowań KGHM metodami klasycznymi

	Metoda	RMSPE
Segment 1	Holta	0,0231
Segment 2	Holta	0,0471
Segment 3	Holta	0,0484
Segment 4	Holta	0,0388
Ważone kryt.		0,0383

Źródło: opracowanie własne

Jako kolejne zaprezentowane zostaną prognozy *ex post* wolumenu obrotów KGHM. Szereg ten, podobnie jak notowania, podzieliliśmy na cztery podokresy. Pierwszy sięgał od początku próby (czyli od 1.07.2011 r.) do 5.08.2011 r. licząc tym samym 26 obserwacji. W tym czasie wyodrębniliśmy stały poziom poziom zmiennej, któremu towarzyszyły duże wahania losowe, cecha charakterystyczna szeregów wolumenu obrotów. 11 sierpnia wystąpił punkt zwrotny, po którym ilość akcji będących w obrocie szybko rosła.

Drugi segment rozpoczyna się 8.08.2011 r. a kończy 22.09.2011 r. (33 obserwacje). Ten okres charakteryzował się dużymi wahaniami, które okazały się szczególnie wysokie na początku i na końcu segmentu. Podobną sytuację dało się zaobserwować w trzecim podokresie (23.09.2011-18.11.2011 r. czyli

39 sesji). Różnica w porównaniu z poprzednikiem była taka, że wahania losowe w centralnej części segmentu miały znacznie wyższą amplitudę.

Ostatni segment oprócz tego, że podobnie jak wszystkie wcześniejsze segmenty charakteryzował się brakiem trendu i dużymi wahaniami, wyróżnił się faktem pojawienia się 15.12.2011 r. i 12.01.2011 r. nietypowo wysokich bo sięgających około 4 milionów sztuk obrotów. Wyższe były tylko obroty w dniu wyznaczającym początek segmentu.

W związku z powyższymi uwagami jako punkt startowy dla wszystkich podokresów wybrany został PAG w wersji pierwszej. Parametry algorytmu pozostały takie same jak poprzednio. W tabelach 4.11 oraz 4.12 prezentujemy wartości procentowego błędu średniokwadratowego uzyskanego dla prognoz wolumenu obrotów KGHM algorytmem genetycznym.

Tabela 4.11: RMSPE dla prognoz wolumenu obrotów KGHM przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,2850	0,2311	0,2534	0,2442
Segment 2	0,2230	0,2080	0,2326	0,1940
Segment 3	0,2760	0,2173	0,3105	0,2390
Segment 4	0,3962	0,3705	0,3841	0,3741
Ważone kryt.	0,3001	0,2625	0,3032	0,2687
Metody	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]

Źródło: opracowanie własne

Tabela 4.12: RMSPE dla prognoz wolumenu obrotów KGHM przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,2825	0,2325	0,2659	0,2434
Segment 2	0,2638	0,2384	0,2463	0,2342
Segment 3	0,3224	0,2938	0,3501	0,2980
Segment 4	0,4696	0,4448	0,5140	0,4476
Ważone kryt.	0,3444	0,3144	0,3583	0,3169
Metody	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]

Źródło: opracowanie własne

Zgodnie z wcześniejszymi spostrzeżeniami, prognozy wolumenu obrotów są o cały rząd wielkości gorsze od prognoz notowań dla tych samych walorów. Powód to oczywiście o wiele większa zmienność szeregów danych tego typu.

Najwyższe błędy *ex post* zanotowaliśmy w czwartym segmencie co nie powinno dziwić zważywszy, że występowały w nim punkty zwrotne a jest to zjawisko, z którym nie radzą sobie żadne metody prognozowania. Zresztą nietypowo wysokie wahania pojawiały się w każdym z podokresów co rzutowało na wartość RMSPE. Da się również zauważyć wyraźna przewaga prognoz algorytmu z wyższym prawdopodobieństwem krzyżowania (o wiele wyraźniejsza niż miało to miejsce przy prognozach notowań).

Kiedy porównamy tabele 4.11 z 4.12 dostrzeżemy również, że wyższy pozom t_m pozwolił na znalezienie dokładniejszych prognoz. Dotyczy to zarówno poszczególnych segmentów a w konsekwencji ważonej wartości kryterium. Niezależnie wszakże od parametrów algorytmu i wersji prognozowania kaskadowego optymalny zestaw metod nie różnił się od zbioru wyjściowego.

Różnice za to dostrzeżemy kiedy zestawimy ze sobą jakość prognoz metodami klasycznymi (tabela 4.13) z prognozami kaskadowymi. Spośród metod wygładzania wykładniczego jedynie w trzecim segmencie zastosowana została metoda Browna. Dla segmentów drugiego i trzeciego jedynym wyjściem okazała się metoda naiwna prosta, która dawała najniższe błędy prognoz *ex post*. Algorytm genetyczny osiągnął dwukrotnie lepsze rezultaty. Jego przewaga jest jeszcze bardziej wyraźna niż miało to miejsce podczas prognozowania obrotów WIG.

Tabela 4.13: RMSPE dla prognoz wolumenu obrotów KGHM metodami klasycznymi

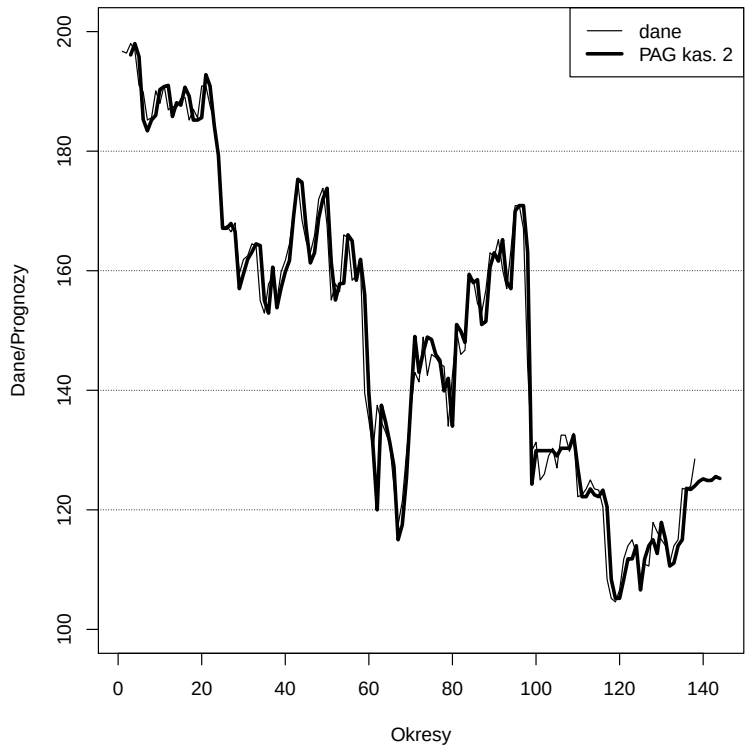
	Metoda	RMSPE
Segment 1	średnia ruch. o $k = 3$	0,7300
Segment 2	naiwna prosta	0,4455
Segment 3	Browna	0,5946
Segment 4	naiwna prosta	0,8787
Ważone kryt.		0,6660

Źródło: opracowanie własne

Zabierając się do wyznaczania prognoz *ex ante* najpierw wybierzemy metodę i jej parametry, które dały najlepsze rezultaty dla prognoz *ex post*. Jeśli chodzi o notowania KGHM to najniższa wartość ważonego RMSPE spośród tych zawartych w tabelach 4.8-4.9 w przypadku kaskady 1 wystąpiła przy $t_m = 5$ i $p_k \in \langle 0, 0{,}6 \rangle$. Przy tych samych parametrach osiągnęliśmy również najlepsze wyniki prognozując przy pomocy kaskady 2. Należy zwrócić uwagę, że ważone kryterium naprawdę niewiele różniło się dla obu wariantów prognozowania kaskadowego.

Jeśli chodzi o predykcję wolumenu obrotów KGHM to zdecydowanie lepiej wypadły wersje o wyższym t_m i prawdopodobieństwie krzyżowania. Znacznie wyższa zmienność samych danych sprawiła, że różnice między ka-

skadą pierwszą i drugą są większe niż miało to miejsce w przypadku notowań tej spółki. Trzeba zauważyć, że rozpatrując zagadnienie w kategoriach liczb bezwzględnych różnica owa zawiera się w granicach około 0,6 punktu procentowego. To bardzo niewiele i wybrane przez nas prognozy wolumenu możemy uznać za porównywalne. Prognozy *ex ante* dotyczyły 6 kolejnych sesji od 18.01.2012 do 25.01.2012 r. Prawdopodobieństwa *a posteriori*, które wykorzystaliśmy do ich obliczenia posłużyły za wagi średniej ruchomej ponieważ ostatniemu segmentowi obu analizowanych szeregów przyporządkowany został PAG w wersji 1.

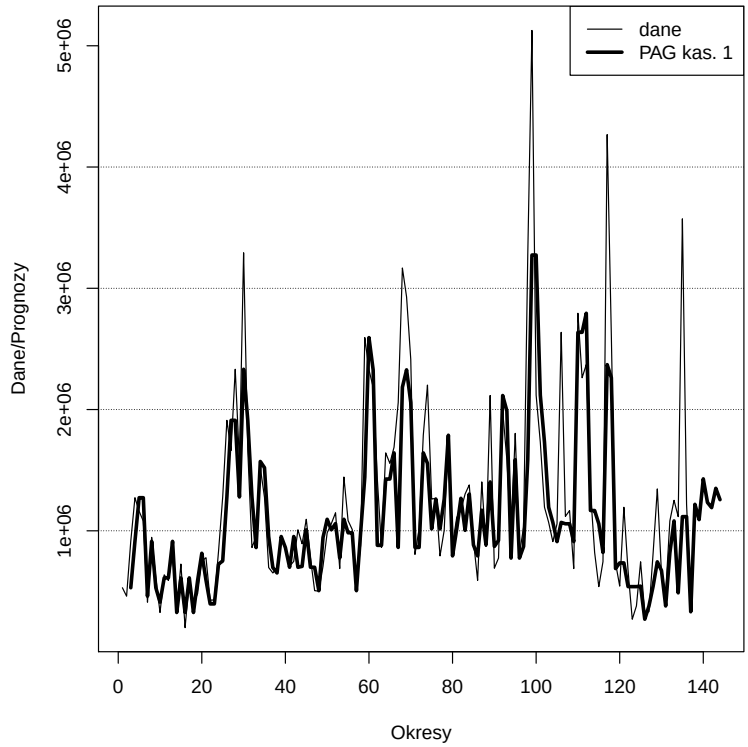


Rysunek 4.3: Prognozy *ex post* i *ex ante* kursu KGHM

Źródło: opracowanie własne

Wykresy na rysunkach 4.3 oraz 4.4 przedstawiają zachowanie prognoz *ex post* a także *ex ante* na podstawie najlepszych wariantów pochodzących z tabel 4.8-4.12.

Prognozy notowań wyróżniają się bardzo dokładnym odwzorowaniem nie tylko małych, ale i dużych wahań. Dla wolumenu obrotów sytuacja była bardziej skomplikowana. Największe błędy algorytm wygenerował dla sesji



Rysunek 4.4: Prognozy *ex post* i *ex ante* wolumenu KGHM

Źródło: opracowanie własne

z ostatniego segmentu charakteryzujących się nieoczekiwanymi wzrostami obrotów. Nie ma zatem co się dziwić, że wzrosło także ważone kryterium.

Analizując dokładność prognoz wygasłych, którą zamieściliśmy w tabeli 4.14 szczególnie mocno rzucają się w oczy bardzo wysokie wartości RMSPE dla wolumenu obrotów. Różnica pomiędzy prognozami *ex post* oraz wygasłymi jest znacznie większa niż miało to miejsce podczas predykcji obrotów WIG. Problem stanowią, wspomniane już kilkakrotnie, nietypowo wysokie wahania wolumenu KGHM. Pojedyncza spółka jest o wiele bardziej podatna na to zjawisko niż rynek jako całość. Dysponując danymi po fakcie jesteśmy w stanie stwierdzić, że na przykład wzrost liczby akcji w obrocie, który miał miejsce 12.01.2012 r. nijak się miał do zachowania podczas sesji 18.01.2012-25.01.2012 r.

Mamy więc szereg czasowy niejako z definicji charakteryzujący się wysokimi wahaniami, w którym pojawiają się obserwacje wywołujące okazjonalne jeszcze większe wzrosty zmienności. Takie zachowanie nie pozostaje bez wpływu na rozkład częstości t_m, który wykorzystujemy podczas tworzenia

Tabela 4.14: Dokładność prognoz wygasłych notowań i wolumenu obrotów KGHM

Wartość	Notowania		Wolumen	
	Kaskada 1	Kaskada 2	Kaskada 1	Kaskada 2
RMSPE	0,0276	0,0370	0,9686	1,2886

Źródło: opracowanie własne

prognoz *ex ante*. Wskutek poszukiwania przez algorytm wartości najbardziej zbliżonych do danych[5] spada zróżnicowanie rozkładu t_m. W prognozach *ex ante* wzrastają również wagi opóźnień wskazujących na obserwacje odstające, które pojawiły się pod koniec próby.

Rozwiązaniem mogłoby być pozbycie się nietypowych danych na przykład wykorzystujące macierz rzutowania (Konarzewska i Karwacki 2005). Niestety przy tak dużej zmienności szeregu czasowego miejsce wykluczonych obserwacji zajmą inne, które na tle pozostałych danych będą różnić się co do wielkości. Nie mamy zatem stuprocentowej gwarancji istotnej poprawy trafności prognoz *ex ante*.

Kolejnym szeregiem, którego prognozami się zajmiemy będzie kurs zamknięcia TVN – innej ze spółek notowanych na warszawskiej giełdzie. Wykorzystane notowania obejmowały ten sam przedział czasu co w wypadku WIG i KGHM.

Pierwszy wyodrębniony podokres liczył sobie 25 obserwacji i sięgał od początku próby czyli 1.07.2011 do 4.08.2011 r. W tym czasie kurs spółki charakteryzował się stałym poziomem i słabymi wahaniami losowymi.

Kolejny segment jest jednym z najkrótszych spośród wszystkich poddanych badaniu. Rozpoczyna się 5.08.2011 r. a kończy 23.08.2011 r. czyli składa się na niego 12 sesji. Dekompozycja tego szeregu wykazała obecność trendu liniowego, któremu towarzyszą małe wahania przypadkowe.

Początek trzeciego segmentu przypada na 24.08.2011 r. zaś koniec na 9.11.2011 r. (55 obserwacji). Tak długi przedział wynika z faktu, że w tym czasie dane podlegały podobnemu zachowaniu. Oprócz braku trendu i sezonowości notowania odznaczały się znacznie wyższymi niż w poprzednich segmentach wahaniami losowymi.

Czwarty segment przypadł na sesje między 10.11.2011 a 17.01.2012 roku (46 obserwacji). Tym razem obok stałego poziomu wyodrębniamy początkowo małe wahania przypadkowe, które w około jednej trzeciej szeregu zyskują dużo wyższą amplitudę. Ponadto około połowy grudnia kurs TVN zaliczył

[5] Zwróćmy uwagę, że przy wolumenach obrotów strategia ewolucyjna preferuje PAG w wersji opartej na metodzie naiwnej prostej.

minimum notowań w podlegającej badaniu próbie, po którym nastąpił trwających zaledwie kilka sesji szybki wzrost.

Dla segmentów pierwszego, trzeciego i czwartego jako metodę startową wybraliśmy PAG w wersji 1. Prognozy w segmencie drugim rozpoczynały się przy założeniu PAG w wersji 3.

W tabelach 4.15 i 4.16 znalazły się wartości RMSPE dla segmentów oraz całego szeregu podzielone ze względu na t_m, rodzaj prognozowania kaskadowego i prawdopodobieństwo krzyżowania. Wynika z nich, że wyższe wahania w dwóch ostatnich segmentach przełożyły się na wyższe błędy prognoz *ex post*. Nadal jednak są to prognozy dobrej jakości z RMSPE poniżej 5%.

Tabela 4.15: RMSPE dla prognoz notowań TVN przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0179	0,0154	0,0164	0,0162
Segment 2	0,0529	0,0170	0,0232	0,0209
Segment 3	0,0378	0,0498	0,0378	0,0369
Segment 4	0,0,0355	0,0313	0,0342	0,0255
Ważone kryt.	0,0350	0,0348	0,0317	0,0281
Metody	[3 1 3 1]	[3 3 2 1]	[1 3 4 1]	[1 3 4 1]

Źródło: opracowanie własne

Tabela 4.16: RMSPE dla prognoz notowań TVN przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0124	0,0169	0,0183	0,0175
Segment 2	0,0174	0,0170	0,0173	0,0209
Segment 3	0,0467	0,0272	0,0438	0,0291
Segment 4	0,0314	0,0190	0,0288	0,0240
Ważone kryt.	0,0331	0,0218	0,0321	0,0247
Metody	[4 3 2 1]	[1 3 3 3]	[1 3 2 3]	[1 3 3 3]

Źródło: opracowanie własne

Po raz kolejny zaznaczyła się przewaga wyższego prawdopodobieństwa krzyżowania. Ważone wartości procentowego błędu średniokwadratowego przy $p_k \in \langle 0,0{,}6 \rangle$ okazały się niższe w obu tabelach. Porównując prognozy z punktu widzenia wartości t_m zauważamy lepsze wyniki przy $t_m = 5$.

Ciekawie przedstawia się sprawa przypisania metod prognozowania do poszczególnych segmentów. Żaden z zawartych w tabelach 4.15-4.16 warian-

tów nie przypomina wyjściowego wektora metod. W dotychczas zaprezentowanych szeregach obserwowaliśmy niewielkie zmiany jakich dokonywała strategia ewolucyjna. Można by pomyśleć, że w prognozach WIG czy KGHM żadne zmiany nie były brane pod uwagę. Jednak przykład kursu TVN pokazuje, że strategia ewolucyjna działa i koryguje pierwotny zbiór metod.

Przyglądając się uważnie ostatecznym wersjom PAG a następnie porównując je z wynikami dekompozycji dostrzeżemy, że w segmentach o stałym poziomie zmiennej pojawiają się warianty algorytmu bazujące na trendzie. Wskażemy dwa powody tego stanu rzeczy. Po pierwsze prognozowanie kaskadowe zakłada związek między krańcowymi prognozami sąsiadujących segmentów. Jest to jednak przyczyna o mniejszym znaczeniu. W końcu to samo prognozowanie kaskadowe odpowiadało za predykcję dla poprzednich szeregów.

Drugim powodem stało się zachowanie szeregów czasowych składających się na segmenty. Choć niewątpliwie mieliśmy do czynienia z brakiem trendu to wykresy przypominają swoim zachowaniem nieregularną sinusoidę. Kursy akcji TVN powoli rosły (malały), osiągały lokalne maksimum (minimum) po czym łagodnie przechodziły do odwrotnej tendencji. Z uwagi na płynny charakter zmian algorytm mógł wskazać jako korzystne prognozy korygowane przyrostem tym bardziej, że przeważa wśród nich PAG w wersji 3. Przypomnijmy, że ma on tę własność iż w szczególnych warunkach może zachowywać się podobnie do wariantu 1. Świadczy o tym częste pojawianie się niskich (rzędu 1 i 2) wartości t_m w chromosomie.

Tabela 4.17: RMSPE dla prognoz notowań TVN metodami klasycznymi

	Metoda	RMSPE
Segment 1	Browna	0,0275
Segment 2	Holta	0,0590
Segment 3	naiwna prosta	0,0334
Segment 4	naiwna prosta	0,0254
Ważone kryt.		0,0317

Źródło: opracowanie własne

Tabela 4.17 zawiera wartości przyjętego kryterium obliczone dla prognoz uzyskanych metodami klasycznymi. Uwagę zwracają w niej dwie rzeczy. Po pierwsze metoda Holta w drugim segmencie wypadła najsłabiej i to pomimo optymalizacji parametrów wygładzania pod kątem RMSPE. Oczywiście PAG dał w tym wypadku o wiele lepsze wyniki.

Drugą cechą charakterystyczną jest wybór metody naiwnej prostej dla dwóch ostatnich segmentów pomimo występujących w nich sporych wahań

losowych. Ani metoda Browna, ani średnia ruchoma nie dały lepszych rezultatów w dopasowaniu prognoz do danych rzeczywistych.

Jako następnymi zajmiemy się prognozami wolumenu obrotów TVN. Cały szereg czasowy został tym razem podzielony na trzy segmenty. Pierwszy z nich obejmował notowania od 1.07.2011 do 29.08.2011 r. czyli składa się na niego 41 obserwacji. W tym czasie dekompozycja szeregu wskazała na stały poziom zmiennej z dużymi, a niekiedy nawet bardzo dużymi wahaniami losowymi. Największe wzrosty obrotów występowały między 7 a 11 lipca i były to przy okazji maksima dla tego okresu. Jednak nawet podczas pozostałych sesji, szczególnie pod koniec segmentu dochodziło do wysokich zmian w wolumenie obrotów.

Drugi segment rozpoczynający się 30.08.2011 a kończący 27.10.2011 roku (43 obserwacje) również charakteryzował się brakiem trendu i pokaźnymi zmianami obrotów choć nie były one już tak duże jak w segmencie poprzednim. Wyjątek stanowiły tylko dwie sesje: 28.09.2011 r. i ostatnia kiedy to wystąpiły nietypowo wysokie jak na ten szereg obroty akcjami.

Ostatni podokres przypadł na sesje między 28.10.2011 a 17.01.2012 roku licząc sobie 54 obserwacje. Po raz trzeci mamy do czynienia z brakiem trendu oraz bardzo dużymi wahaniami losowymi. Szczególnie częste zmiany wolumenu obrotów przypadały na pierwszą połowę wspomnianego przedziału czasowego.

Ze względu na obecność stałego poziomu zmiennej we wszystkich trzech segmentach jako startowy zestaw metod wybraliśmy PAG w wersji 1. Tabele 4.18 oraz 4.19 prezentują dokładność prognoz *ex post* mierzoną RMSPE dla poszczególnych segmentów a także jako średnią ważoną, w której wagę stanowił udział długości danego segmentu w długości całego szeregu.

Tabela 4.18: RMSPE dla prognoz wolumenu obrotów TVN przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,4792	0,4406	0,4567	0,4175
Segment 2	0,3371	0,3004	0,3279	0,3136
Segment 3	0,0,3641	0,3500	0,3419	0,3410
Ważone kryt.	0,3886	0,3603	0,3704	0,3543
Metody	[1 1 1]	[1 1 1]	[1 1 1]	[1 1 1]

Źródło: opracowanie własne

W każdym z segmentów wystąpiły odizolowane obserwacje o nietypowej wysokości. Nie pozostało to bez wpływu na wartość RMSPE, który waha się od 30 do nawet 50%. Najwyższe błędy *ex post* obserwujemy w pierw-

Tabela 4.19: RMSPE dla prognoz wolumenu obrotów TVN przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,4943	0,4604	0,5081	0,4499
Segment 2	0,3680	0,3236	0,3442	0,3164
Segment 3	0,4022	0,3592	0,4075	0,3682
Ważone kryt.	0,4178	0,3770	0,4163	0,3753
Metody	[1 1 1]	[1 1 1]	[1 1 1]	[1 1 1]

Źródło: opracowanie własne

szym segmencie, w którym przez kilka kolejnych sesji obroty były jednymi z najwyższych w całym szeregu.

Po raz kolejny okazało się, że podwyższenie prawdopodobieństwa krzyżowania zaowocowało poprawą jakości prognoz również tą mierzoną ważonym kryterium. Duża zmienność towarzysząca zwykle szeregom czasowym obrotów giełdowych sprawia, że lepiej sprawdza się wyższe t_m. W połączeniu z PAG opartym na metodzie naiwnej prostej algorytm łatwiej znajduje okresy z wartością zbliżoną do tej z momentu prognozy.

Oczywiście jeśli w zasięgu wyznaczonym przez t_m nie ma obserwacji choćby o podobnym rzędzie wielkości pojawiają się wyższe błędy prognoz *ex post*. Tak właśnie stało się w segmentach szeregu obrotów TVN.

Szczególne zachowanie obrotów TVN spowodowało ogromne kłopoty na etapie doboru klasycznych metod prognozowania dla poszczególnych segmentów. Najgorzej prezentuje się pod tym względem pierwszy segment jak można przekonać się na podstawie wartości procentowego błędu średniokwadratowego w tabeli 4.20. Zastosowaliśmy w tym segmencie model trendu jako wielomianu trzeciego stopnia wraz ze zmiennymi zero-jedynkowymi. Pomimo użycia dodatkowych zmiennych RMSPE przyjęło niezwykle wysoką wartość, która nawiasem mówiąc i tak była najniższa spośród wszystkich użytych dla tego segmentu metod.

Prognozy w pozostałych segmentach są obarczone o wiele niższymi błędami, które wszakże nadal trudno uznać za akceptowalne. Ważone kryterium okazało się dwukrotnie wyższe od najgorszych wartości RMSPE, które możemy znaleźć w tabelach 4.18-4.19.

Spośród wariantów prognoz znajdujących się w tabelach 4.15-4.19 wybraliśmy najlepsze, które posłużyły do wyznaczenia prognoz *ex ante* z pomocą analizy bayesowskiej. W przypadku notowań TVN najniższy ważony RMSPE wystąpił w kaskadzie 1 przy $t_m = 5$ i $p_k \in \langle 0,0{,}6\rangle$. W ostatnim segmencie strategia ewolucyjna wskazała na PAG w wersji 3. Z kolei dla wolumenu obrotów tej spółki najbardziej korzystne wartości dała kaskada

Tabela 4.20: RMSPE dla prognoz notowań TVN metodami klasycznymi

	Metoda	RMSPE
Segment 1	trend t^3	1,0516
Segment 2	naiwna prosta	0,7702
Segment 3	średnia ruch. o $k = 2$	0,7785
Ważone kryt.		0,8589

Źródło: opracowanie własne

drugiego typu przy $t_m = 10$ i $p_k \in \langle 0, 0{,}6 \rangle$. W jej wypadku we wszystkich segmentach obowiązywał PAG w wersji 1.

Wykresy na rysunkach 4.5 oraz 4.6 ilustrują wzajemne położenie szeregów danych rzeczywistych i prognoz. Prognozy szeregu notowań odznaczają się na nich lepszym dopasowaniem z powodów, które opisaliśmy wcześniej. Drugi z wymienionych wykresów pozwala zrozumieć dlaczego doświadczyliśmy o wiele wyższych błędów, szczególnie po zastosowaniu metod klasycznych. Wyraźnie widać pojedyncze sesje o ponadprzeciętnie wysokich obrotach w porównaniu do reszty szeregu.

Tabela 4.21: Dokładność prognoz wygasłych notowań i wolumenu obrotów TVN

Wartość	Notowania		Wolumen	
	Kaskada 1	Kaskada 2	Kaskada 1	Kaskada 2
RMSPE	0,0098	0,0127	0,3296	0,3533

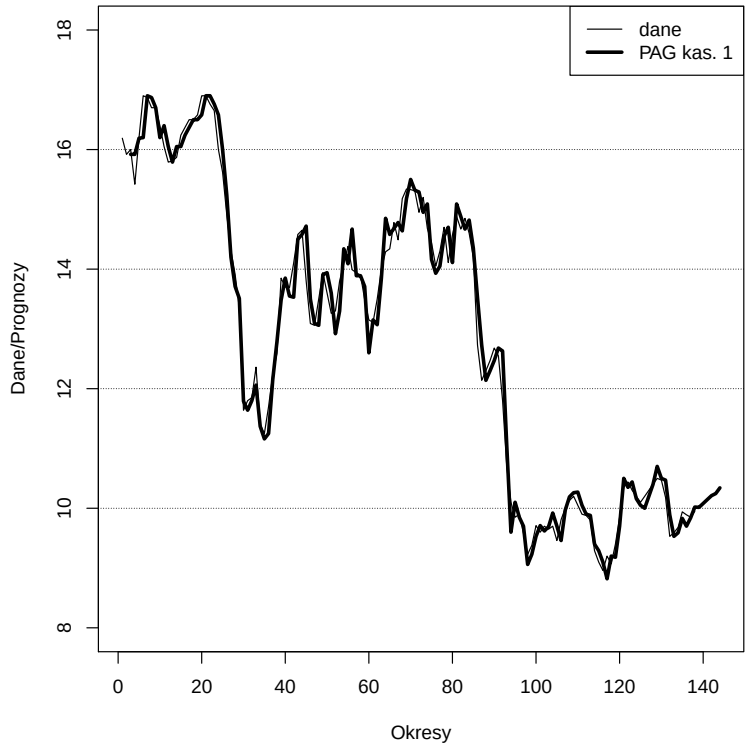
Źródło: opracowanie własne

Ponownie oceniliśmy trafność prognoz wygasłych, tak jak poprzednio na 6 sesji naprzód. Wartości RMSPE ze stosownym podziałem znalazły się w tabeli 4.21. Zarówno prognozy notowań jak i wolumenu są warte omówienia. Procentowy błąd średniokwadratowy w przypadku notowań okazał się niższy od swojego odpowiednika *ex post*.

Duże zaskoczenie stanowi przede wszystkim poziom tej różnicy. Również prognozy wygasłe obrotów uzyskaliśmy nieco lepsze od prognoz *ex post*. Choć różnica nie jest aż tak widoczna to należy ją docenić mając na uwadze zachowanie szeregu w poprzednich okresach i problemy z nim związane.

4.1.2. Prognozy notowań kursów walut

Kolejnym obszarem poddanym działaniu prognostycznego algorytmu genetycznego będą notowania walut. Wybraliśmy do tego celu euro, franka szwajcarskiego i dolara amerykańskiego. Wybór ten nie jest przypadkowy



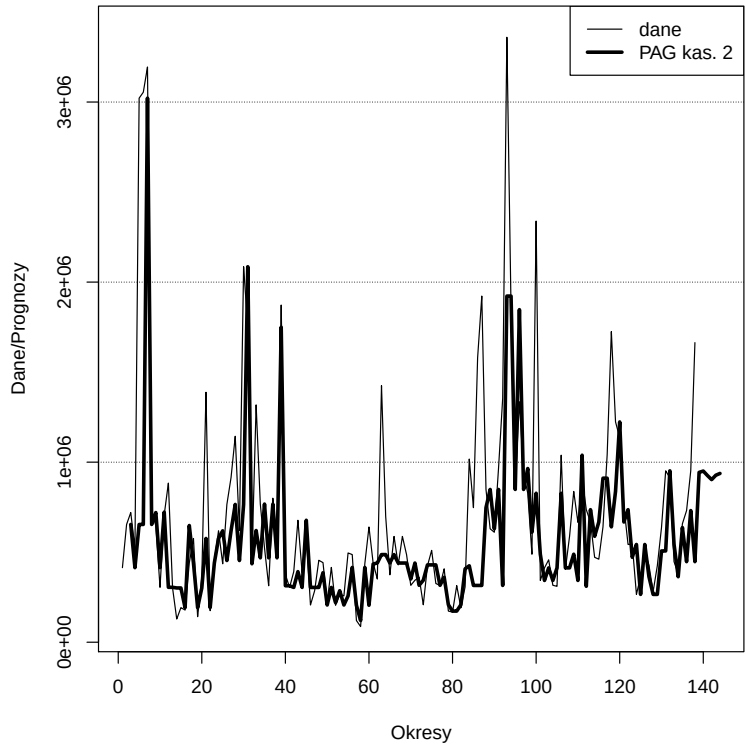
Rysunek 4.5: Prognozy *ex post* i *ex ante* kursu TVN

Źródło: opracowanie własne

gdyż są to waluty o dużym znaczeniu międzynarodowym a przez to mające wpływ również i na polską gospodarkę. Ponadto kryzys finansowy mocno uderzył między innymi w rynki walutowe.

W przypadku Europy szczególnie mocno uwidoczniła się groźba bankructw całych krajów jak Grecji czy Hiszpanii. Groźba ta powracała pomimo, a może na przekór podejmowanym przez instytucje unijne próbom łagodzenia spodziewanych negatywnych skutków ekonomicznych biorących się m.in. z zadłużenia rządów części państw wspólnoty.

Na początek zajmiemy się wyznaczaniem prognoz *ex post* notowań euro. W tym celu sięgniemy po publikowane przez NBP dzienne kursy zamknięcia z okresu od 1.07.2011 do 17.01.2012 r. a więc tego samego z jakiego pochodziły notowania WIG oraz spółek. Same notowania w oparciu o wyniki dekompozycji zostały podzielone na trzy podokresy. W pierwszym znalazły się notowania między 1.07.2011 a 9.09.2011 r. (50 obserwacji). W tym czasie dał się zaobserwować wyraźny trend rosnący z wahaniami losowymi mającymi zróżnicowaną siłę.



Rysunek 4.6: Prognozy *ex post* i *ex ante* wolumenu TVN

Źródło: opracowanie własne

Kolejny przedział zaczynał się 12.09.2011 a kończył 4.11.2011 roku. Podczas liczącego 39 obserwacji okresu notowania początkowo rosły osiągając swoje apogeum 23 września. Potem nastąpiło odwrócenie trendu, który po raz kolejny zmienił kierunek po 16 sesjach. Pod koniec segmentu wystąpiła seria kilkunastu dużych wahań losowych. Patrząc jednak na ów zestaw danych jak na pewną całość możemy potraktować go jako szereg o stałym poziomie i dużych wahaniach przypadkowych.

Trzeci segment rozciągający się od 7.11.2011 r. do końca próby (49 obserwacji) prezentował się najciekawiej z punktu widzenia dekompozycji. Mamy tu bowiem do czynienia z małymi wahaniami przypadkowymi występującymi pomiędzy punktami wyznaczającymi zmiany kierunku trendu. Gdyby nie zróżnicowana długość okresu wahań mówilibyśmy o sezonowości gdyż wykres tego odcinka danych przypomina swoim zachowaniem ten element dekompozycji.

Parametry algorytmu nie uległy zmianie. Prognozy *ex post* wykonaliśmy zatem dla dwóch różnych wartości t_m czyli 5 i 10 oraz dwóch wariantów

prawdopodobieństwa krzyżowania: $\langle 0,0{,}3\rangle$ i $\langle 0,0{,}6\rangle$. Początkowy zestaw metod składał się z PAG opartego na metodzie naiwnej z trendem w pierwszym wariancie (pierwszy segment) oraz dwóch wywołań algorytmu wykorzystującego średnią ruchomą (segmenty 2 i 3).

Tabela 4.22: RMSPE dla prognoz kursu euro przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,0079	0,0048	0,0049	0,0045
Segment 2	0,0095	0,0071	0,0072	0,0083
Segment 3	0,0062	0,0081	0,0067	0,0079
Ważone kryt.	0,0077	0,0066	0,0062	0,0068
Metody	[2 2 3]	[3 3 2]	[3 1 1]	[3 2 2]

Źródło: opracowanie własne

Tabela 4.23: RMSPE dla prognoz kursu euro przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,0047	0,0047	0,0049	0,0059
Segment 2	0,0087	0,0066	0,0078	0,0068
Segment 3	0,0073	0,0072	0,0049	0,0061
Ważone kryt.	0,0068	0,0061	0,0057	0,0062
Metody	[3 2 2]	[4 4 2]	[3 2 4]	[1 3 1]

Źródło: opracowanie własne

W tabelach 4.22 i 4.23 zaprezentowane zostały wartości funkcji kryterium dla poszczególnych podokresów oraz wartość ważona dla całego chromosomu. Ponadto ostatni wiersz zawiera ostateczne wektory metod użytych przez algorytm w kolejnych segmentach.

Ważona wartość kryterium w obu tabelach pokazuje, że między kaskadami pierwszego i drugiego typu występują minimalne różnice w dopasowaniu prognoz *ex post* do danych. Oba podejścia do PAG najwyższe (co nie znaczy wysokie) błędy odnotowały w drugim segmencie.

Wzrost wartości prawdopodobieństwa krzyżowania wpływa na poprawę jakości prognoz w poszczególnych segmentach. Na uwagę zasługują jednak odmienne zestawy metod prognostycznych, które posłużyły do uzyskania prognoz tak podobnej jakości.

Biorąc pod uwagę niewielkie różnice ważonego kryterium dopasowania prognoz do danych, zachowanie kursu w trzecim segmencie (nieregularnie

rozłożone w czasie okresy względnej stabilności notowań) oznaczało, że częściej wybierane były te wersje PAG, które nie opierają się na trendzie.

Prognozy wykonane algorytmem z $t_m = 10$ odznaczają się mniejszą dokładnością w porównaniu z $t_m = 5$. Odpowiada za to z jednej strony większy obszar potencjalnych rozwiązań, który należy przeszukać, a z drugiej serie nakładających się na trend wahań losowych, które dodatkowo utrudniają poszukiwania jak najlepszego rozwiązania. Jednakże obiektywnie rzecz biorąc, miary jakości prognoz z tabeli 4.22 nadal mają zadowalające wartości.

Tabela 4.24: RMSPE dla prognoz notowań euro metodami klasycznymi

	Metoda	RMSPE
Segment 1	Holta	0,0064
Segment 2	Browna	0,0078
Segment 3	naiwna prosta	0,0050
Ważone kryt.		0,0063

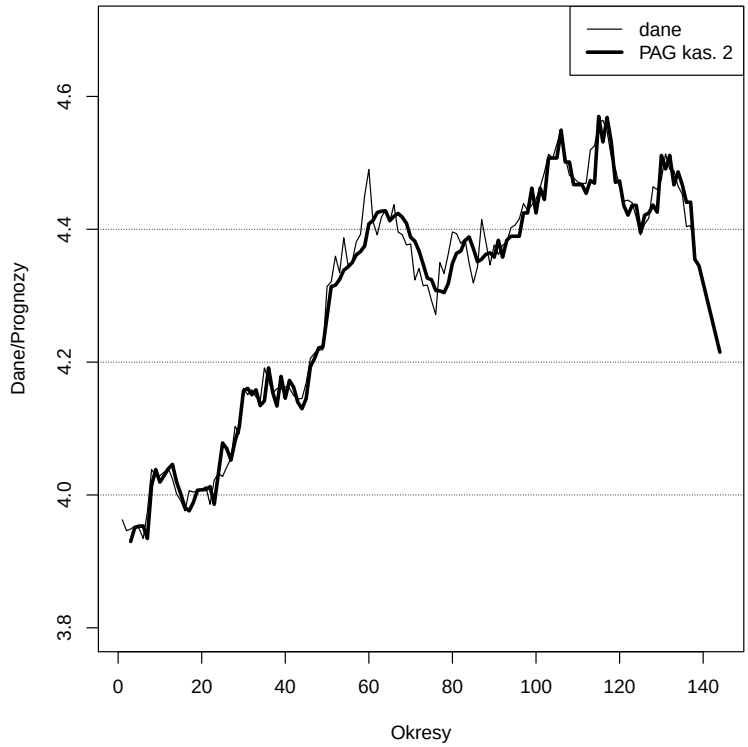
Źródło: opracowanie własne

Warto zauważyć, że prognozy kaskadowe drugiego typu okazały się średnio rzecz biorąc lepsze od kaskady 1. Traktujemy to jako dowód słuszności włączania do mechanizmu tworzącego prognozy dla całego szeregu prognoz *ex ante* poszczególnych segmentów. Po raz pierwszy zdarzyło się jednak, że najniższa wartość ważonego RMSPE ze wszystkich obliczonych przypadła na $p_k \in \langle 0, 0{,}3 \rangle$. Do tej pory w szeregach „notowań" lepiej sprawdzało się wyższe prawdopodobieństwo krzyżowania.

Porównując wyniki pochodzące z prognozowania kaskadowego do zawartych w tabeli 4.24 rezultatów dla metod klasycznych dostrzeżemy bardzo podobne wartości RMSPE w poszczególnych segmentach oraz w wartości ważonej. Nieco słabiej wypadła metoda Holta w pierwszym segmencie, za to lepiej metoda Browna w drugim. Warto zwrócić uwagę na dobry wynik metody naiwnej prostej.

Wykres na rysunku 4.7 prezentuje szereg prognoz *ex post* i *ex ante* w porównaniu z danymi rzeczywistymi. Do jego wykonania posłużyły najlepsze wyniki dla prognozowania kaskadowego 2. Prognozy bayesowskie wykonaliśmy według drugiego z opisanych w rozdziale 3 wariantów zakładającego trend. Warto zwrócić uwagę na dobre dopasowanie w ostatnim segmencie, w którym szereg czasowy podlega quasi-sezonowemu zachowaniu.

Trafność prognoz wygasłych obejmowała notowania kursu euro w okresu 18-25.01.2012 r. (6 sesji następujących po ostatniej prognozie *ex post*). Oceniliśmy je dla obu typów prognozowania kaskadowego a stosowne wyniki zamieściliśmy w tabeli 4.25. Jak można się na jej podstawie przekonać,



Rysunek 4.7: Prognozy *ex post* i *ex ante* kursu euro

Źródło: opracowanie własne

kaskada 2 ma dokładność zbliżoną do tej z prognoz *ex post*. W przypadku kaskady 1 wzrost błędów okazał się wyraźnie wyższy.

Tabela 4.25: Dokładność prognoz wygasłych kursu euro

Wartość	Kaskada 1	Kaskada 2
RMSPE	0,0202	0,0098

Źródło: opracowanie własne

Jako kolejne przedstawimy prognozy franka szwajcarskiego. Waluta ta zawsze miała w Polsce szczególny status ze względu na popularność denominowanych w niej kredytów (mieszkaniowych itp.). Wprawdzie w ostatnim czasie zachowywała większą stabilność notowań niż pozostałe liczące się na świecie środki płatnicze, nie znaczy to jednak, że okazała się odporna na ataki spekulacyjne. Doszło nawet do interwencji szwajcarskiego banku centralnego na rynku międzynarodowym w celu ochrony narodowej waluty.

Wpływ franka na polską gospodarkę wiąże się, jako się rzekło, z dużą ilością udzielonych kredytów, których raty gwałtownie wzrosły w momencie wzrostu notowań samej waluty. Spowodowało to dodatkowe zamieszanie na rynku[6] i doprowadziło w końcu do zmian legislacyjnych, które ograniczyły dostęp do kredytów w obcej walucie. Właśnie z tych powodów frank szwajcarski znalazł się w polu naszych zainteresowań.

Zachowanie notowań franka w okresie od 1.07.2011 do 17.01.2012 roku (podobnie jak w przypadku kursu euro) ulegało bardzo istotnym zmianom. W związku z tym wyodrębniliśmy cztery podokresy opierając się na dekompozycji szeregu czasowego.

W pierwszym segmencie, trwającym od 1.07.2011 do 2.08.2011 roku (23 obserwacje) dostrzegamy wyraźny trend rosnący. Towarzyszą mu niezbyt duże wahania losowe. W środkowej części, przez kilka sesji, trend znajduje się na wyższej trajektorii, po czym wraca do wcześniejszego rzędu wielkości.

Z formalnego punktu widzenia o drugim podokresie (od 3.08.2011 do 9.09.2011 r. – 27 sesji) można powiedzieć, że występuje w nim stały poziom zmiennej. Należy jednak zwrócić uwagę na obserwowane w kilku okresach bardzo duże wahania będące w istocie punktami zwrotnymi. Następują one w odstępie zaledwie kilku sesji. Powiemy więc, że drugi segment został zdominowany przez duże wahania przypadkowe.

Trzeci segment zawarty pomiędzy 12.09.2011 a 21.11.2011 r. (49 obserwacji) ma najbardziej podręcznikowy charakter jeśli chodzi o jego dekompozycję. Obserwujemy stały poziom zmiennej oraz małe i duże wahania losowe.

Zachowanie w ostatnim, czwartym segmencie jest bardzo podobne do tego co obserwowaliśmy dla notowań euro. Po pierwsze pokrywają się daty wyznaczające trwanie tego podokresu – od 22.11.2011 do 17.01.2012 r. (39 obserwacji). Po drugie (i ważniejsze) mamy do czynienia z niemal kalką zachowania notowań. A zatem w przypadku franka również możemy mówić o wahaniach quasi-sezonowych. Wykres tego fragmentu danych przypomina nieregularną sinusoidę.

Aby zachować szansę porównania wyników prognoz obu walut, poza liczbą segmentów, nie zmienialiśmy parametrów algorytmu. Tak więc ponownie sprawdziliśmy zachowanie dla dwóch znacząco różnych wartości t_m oraz prawdopodobieństw krzyżowania.

Wyjściowy wektor metod poddany działaniu strategii ewolucyjnej zakładał przypisanie pierwszemu segmentowi PAG w wersji 3 (trend) zaś pozostałym trzem segmentom w wersji 1 (metoda naiwna prosta). W tabeli 4.27 zawarte zostały wartości RMSPE dla poszczególnych segmentów i całego szeregu (średnia ważona) przy $t_m = 5$. Analiza tych danych pokazuje, że

[6] Kredytobiorcy decydowali się na przewalutowanie kredytów. Pojawiły się problemy ze spłatami wyższych rat.

obie wersje prognozowania kaskadowego najlepiej poradziły sobie w ostatnim segmencie, najsłabiej zaś w drugim. Jednak, podobnie jak miało to miejsce przy okazji prognoz kursu euro, „najgorzej" to względne pojęcie. Najwyższe RMSPE okazało się bowiem nieco wyższe niż 2%.

Porównując ważoną wartość kryterium stwierdzamy, że zarówno kaskada 1 jak i kaskada 2 charakteryzowały się bardzo zbliżonymi wynikami. Metody te okazały się też bardzo zgodne co do wyboru metod prognozowania dla poszczególnych segmentów. Co się zaś tyczy efektów działania strategii ewolucyjnej to na podstawie tabel 4.26 i 4.27 można zauważyć pewne zmiany w stosunku do zbioru wyjściowego w przypadku kaskady 1. Kaskada 2 dla obu rozpatrywanych wariantów t_m pozostała przy metodach ustalonych na początku postępowania.

Tabela 4.26: RMSPE dla prognoz kursu franka szwajcarskiego przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0126	0,0108	0,0109	0,0109
Segment 2	0,0147	0,0139	0,0165	0,0145
Segment 3	0,0096	0,0073	0,0093	0,0072
Segment 4	0,0064	0,0065	0,0070	0,0062
Ważone kryt.	0,0102	0,0089	0,0103	0,0089
Metody	[3 1 1 1]	[3 4 1 1]	[3 1 1 1]	[3 1 1 1]

Źródło: opracowanie własne

Tabela 4.27: RMSPE dla prognoz kursu franka szwajcarskiego przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0109	0,0108	0,0120	0,0119
Segment 2	0,0157	0,0142	0,0147	0,0141
Segment 3	0,0071	0,0067	0,0076	0,0068
Segment 4	0,0060	0,0053	0,0064	0,0059
Ważone kryt.	0,0091	0,0086	0,0093	0,0088
Metody	[4 1 1 4]	[3 1 1 4]	[3 1 1 1]	[3 1 1 1]

Źródło: opracowanie własne

W tabeli 4.28 zaprezentowaliśmy wartości RMSPE obliczone dla klasycznych metod prognozowania. W segmentach od 1 do 3 najlepiej sprawdzały się metody wygładzania wykładniczego, które jednak wypadły wyraźnie gorzej od prognoz otrzymanych za pomocą PAG i to niezależnie od

wersji prognozowania kaskadowego. W ostatnim segmencie zastosowana została metoda naiwna prosta, której trafność okazała się na bardzo podobnym poziomie do algorytmu genetycznego. Choć prognozy wyznaczone klasycznie zasługują na miano dobrych to również z punktu widzenia kryterium ważonego są mniej dokładne niż ich odpowiedniki z tabel 4.26 oraz 4.27.

Tabela 4.28: RMSPE dla prognoz notowań franka szwajcarskiego metodami klasycznymi

	Metoda	RMSPE
Segment 1	Holta	0,0179
Segment 2	Holta	0,0245
Segment 3	Browna	0,0092
Segment 4	naiwna prosta	0,0061
Ważone kryt.		0,0127

Źródło: opracowanie własne

Jak pamiętamy (por. tabele 4.22 i 4.23) prognozy *ex post* euro charakteryzowały się mniejszą trafnością od prognoz wygasłych. W tabeli 4.25 jest to szczególnie widoczne dla kaskady 2. Jak pokazują wartości z tabeli 4.29 błąd prognoz wygasłych obliczony dla 6 kolejnych sesji notowań franka szwajcarskiego (między 18 a 25 stycznia 2012 r.) był tym razem wyższy zarówno dla kaskady 1 jak i kaskady 2. Prognozy *ax ante* wyznaczyliśmy w oparciu o najlepsze parametry spośród zawartych w tabelach 4.26 i 4.27. Tak się składa, że przypadły one na $t_m = 5$ oraz $p_k \in \langle 0, 0{,}6 \rangle$. Trzeba jednak podkreślić, że RMSPE utrzymał się poniżej 2%.

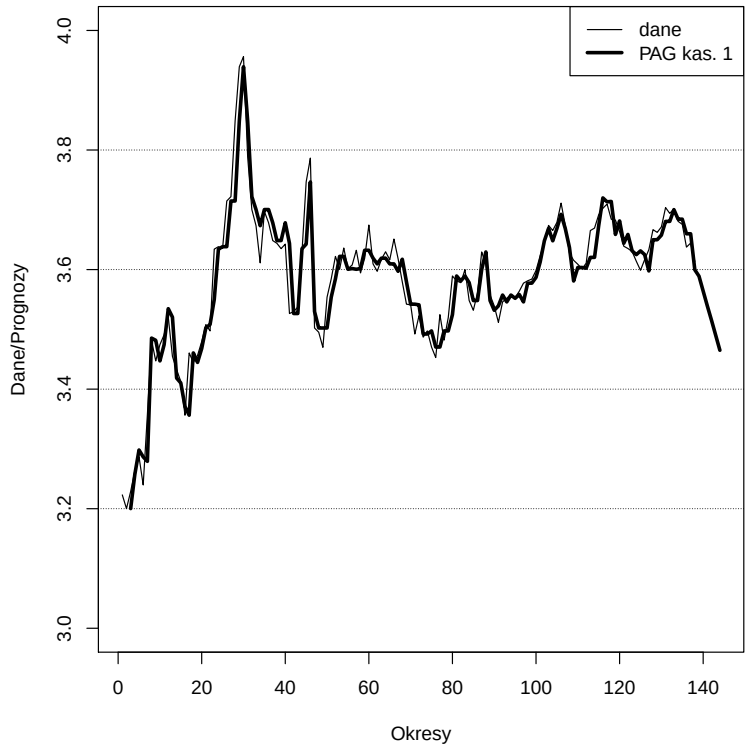
Tabela 4.29: Dokładność prognoz wygasłych kursu franka szwajcarskiego

Wartość	Kaskada 1	Kaskada 2
RMSPE	0,0147	0,0156

Źródło: opracowanie własne

Prognozy, których jakość oceniamy w tabeli 4.29 różniły się jednak wariantem użytego prognozowania bayesowskiego. Kaskada 1 w ostatnim segmencie zawierała PAG w wersji 4 czyli zakładający trend oparty na przyrostach dwóch sąsiednich wartości. Z kolei kaskada 2 pozostała przy metodzie przyjmującej za punkt wyjścia stały poziom zmiennej i metodę naiwną prostą. Jak widać bardziej trafne było założenie wystąpienia trendu w kolejnych notowaniach. Zachowanie prognoz *ex ante* można prześledzić na wykresie na rysunku 4.8.

Sytuacja rynkowa sprawiła, że zarówno notowania euro jak i franka szwajcarskiego podlegały zjawisku, które nazwaliśmy wcześniej quasi-sezonowością. Seria niedużych, ale systematycznych wzrostów (spadków) osiąga swój punkt kulminacyjny, po którym następuje odwrócenie tendencji. Punkty zwrotne wypadają przy (sięgając po terminologię rodem z analizy technicznej) stałych liniach oporu i wsparcia, lecz w niejednakowych odstępach mierzonych liczbą sesji. Na wykresie obserwujemy wówczas kształt przypominający nieregularną sinusoidę.



Rysunek 4.8: Prognozy *ex post* i *ex ante* kursu franka szwajcarskiego

Źródło: opracowanie własne

W poprzednim rozdziale opisaliśmy zachowanie prognostycznego algorytmu genetycznego w razie napotkania sezonowości. Stwierdziliśmy wtedy, że parametr t_m powinien być co do swojej wartości co najmniej równy okresowi wahań sezonowych. W sytuacji z jaką zetknęliśmy się w ostatnich segmentach szeregów notowań euro i franka okres wahań sezonowych możemy określić co najwyżej w przybliżeniu. Dlatego wyższa wartość t_m włącza do prognoz efekty quasi-sezonowe. Prognozy *ex post* przy $t_m = 5$ wypadały

dobrze gdyż wahania losowe w obrębie ostatnich pięciu okresów przeważnie (poza okolicami punktów zwrotnych) nie były duże.

Jeśli chodzi o błędy prognoz wygasłych opisanych dotąd walut to warto zauważyć, że zwycięski pod tym względem wariant prognoz kaskadowych dla ostatniego segmentu dobierał PAG oparty na trendzie w szeregu. Kursy euro i franka w okresach *ex ante* spadały niewiele odchylając się od trendu. Korekta ważonym przyrostem, który był ujemny, trafnie prognozowała kierunek, w którym podążą notowania. Biorąc jednak pod uwagę częstotliwość notowań, ich zmienność oraz skłonność do zmiany trajektorii z sesji na sesję wskazanym wydaje się, aby prognozy kaskadowe powtarzać dodając nowe obserwacje, a także korygując na bieżąco granice ostatnich segmentów. Spowoduje to jeszcze bardziej elastyczną reakcję algorytmu na zachowanie zmiennej.

W następnej części tego rozdziału zajmiemy się prognozami dolara amerykańskiego. Waluta ta także ma duże znaczenie dla polskiej i światowej gospodarki. Potoczne powiedzenie mówi, że „kiedy USA ma katar, reszta świata łapie grypę". Dolar ciągle ogrywa istotną rolę w rozliczeniach międzynarodowych wciąż pełniąc funkcję światowej waluty. Nie mogło go więc zabraknąć w naszych rozważaniach.

Okres objęty działaniem prognostycznego algorytmu genetycznego przypadł na ten sam przedział czasu co w przypadku spółek giełdowych i dwóch omówionych wcześniej walut a więc od 1.07.2011 do 17.01.2012 r.

Cały szereg podzieliliśmy tym razem na 4 segmenty. Na pierwszy z nich składają się 42 pierwsze obserwacje co oznacza, że kończy się on 30.08.2011 roku. W tym okresie stwierdziliśmy obecność trendu rosnącego, któremu towarzyszą w pierwszej połowie duże wahania losowe. W dalszej części segmentu ich amplituda bardzo spada.

Drugi segment trwa od 31.08.2011 do 17.10.2011 r. (34 obserwacje). Na początku tego przedziału nastąpił szybki wzrost kursu dolara przy małych wahaniach przypadkowych. Dwukrotnie następowało chwilowe odwrócenie trendu, które po kilku sesjach było korygowane. Dopiero 3.10.2011 r. trend na trwałe zmienił nachylenie na ujemne i w tej formie utrzymał się do końca segmentu.

Trzeci wyodrębniony segment rozpoczął się 18.10.2011 r. i trwał do 21.11.2011 r. rozciągając się tym samym na 23 obserwacje. W porównaniu z segmentem poprzednim nastąpiła kolejna zmiana kierunku trendu, znowu na rosnący tyle, że tempo wzrostu nie jest duże. Przeważająca większość wahań losowych jest mała, tylko na początku pojawiło się kilka większych.

Ostatni, czwarty segment liczy sobie 39 notowań wypadających między 22.11.2011 a 17.01.2012 r. Na trend rosnący nakładają się wahania losowe oraz quasi-sezonowe, które jednak mają mniej regularny przebieg w porównaniu z szeregami euro i franka szwajcarskiego.

W każdym segmencie stwierdziliśmy obecność trendu. Dlatego w początkowym zestawie metod zdecydowaliśmy się na PAG w wersji 3 czyli uwzględniający przyrost między obserwacjami odległymi od siebie o nie więcej niż t_m okresów.

Analizując wartości RMSPE dla segmentów, które umieściliśmy w tabelach 4.30 i 4.31 zauważamy, że prognozy wyznaczone dla $t_m = 10$ są z punktu widzenia przyjętego kryterium wyraźnie gorsze od swoich odpowiedników dla $t_m = 5$. Szczególnie wyraźnie widać to w segmentach drugim i trzecim. Za taki stan rzeczy w drugim segmencie odpowiadają chwilowe odwrócenia kierunku trendu, o których wspominaliśmy przy okazji dekompozycji.

W początkowej części następnego, trzeciego segmentu mieliśmy do czynienia z kilkoma wyższymi wahania losowymi. Algorytm o $t_m = 10$ generując prognozy mocniej sięga w przeszłość. Proces ma charakter stochastyczny więc zdarza się, że do chromosomu włączony zostanie gen mający wartość pochodzącą z okresu charakteryzującego się podwyższoną amplitudą wahań.

Ponadto należy podkreślić, że w tabeli 4.30 dochodziło do zmian początkowego zestawu metod podczas gdy w tabeli 4.31 takie zmiany nie występują. Generalnie jednak wszystkie otrzymane prognozy *ex post* dolara okazały się bardzo dokładne przy czym te dla $p_k \in \langle 0,0{,}6 \rangle$ odznaczały się wyższą trafnością.

Tabela 4.30: RMSPE dla prognoz kursu dolara amerykańskiego przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0110	0,0091	0,0093	0,0090
Segment 2	0,0120	0,0084	0,0127	0,0089
Segment 3	0,0118	0,0107	0,0123	0,0111
Segment 4	0,0101	0,0099	0,0097	0,0083
Ważone kryt.	0,0111	0,0094	0,0108	0,0091
Metody	[3 3 3 4]	[3 3 3 3]	[4 3 3 1]	[3 3 3 3]

Źródło: opracowanie własne

Ciekawie wypadł dobór metod klasycznych podczas prognozowania kursu dolara w tabeli 4.32. W segmentach 2 i 3, które jak pamiętamy tak mocno różniły się między algorytmem o $t_m = 10$ i $t_m = 5$ najkorzystniej wypadł model trendu wielomianowego stopnia 2 przy czym RMSPE w segmencie trzecim przyjęło niższe wartości niż miało to miejsce w przypadku PAG.

Jeśli chodzi o metodę Wintersa, którą wykorzystaliśmy w segmencie czwartym to bazowała ona na sezonowości addytywnej z okresem wahań sezonowych wynoszącym 12. Wahania w tym segmencie, jak to już napisaliśmy omawiając dekompozycję, nie podlegały czystej, regularnej sezono-

Tabela 4.31: RMSPE dla prognoz kursu dolara amerykańskiego przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0098	0,0094	0,0103	0,0090
Segment 2	0,0087	0,0080	0,0086	0,0079
Segment 3	0,0102	0,0100	0,0109	0,0098
Segment 4	0,0093	0,0083	0,0086	0,0082
Ważone kryt.	0,0094	0,0088	0,0095	0,0086
Metody	[3 3 3 3]	[3 3 3 3]	[3 3 3 3]	[3 3 3 3]

Źródło: opracowanie własne

wości. Jednak przyjęcie wysokiego jak na szereg liczący sobie 39 obserwacji okresu wahań sezonowych dało mimo wszystko dobre rezultaty. Zestawiając wartość ważonego kryterium z trzech opisanych tabel stwierdzamy, że metodom klasycznym bliżej do prognoz z tabeli 4.30 niż 4.31.

Tabela 4.32: RMSPE dla prognoz notowań dolara amerykańskiego metodami klasycznymi

	Metoda	RMSPE
Segment 1	Holta	0,0121
Segment 2	trendu t^2	0,0112
Segment 3	trendu t^2	0,0077
Segment 4	Wintersa	0,0131
Ważone kryt.		0,0111

Źródło: opracowanie własne

Prognozy *ex ante* obliczyliśmy przy niższym t_m oraz wyższym prawdopodobieństwie krzyżowania, ponieważ w tabeli 4.31 ważone RMSPE tu właśnie okazało się najniższe. Prognozy bayesowskie zakładające obecność trendu w kolejnych 6 sesjach zostały ponownie ocenione już jako prognozy wygasłe a rezultaty tej oceny znalazły sie w tabeli 4.33.

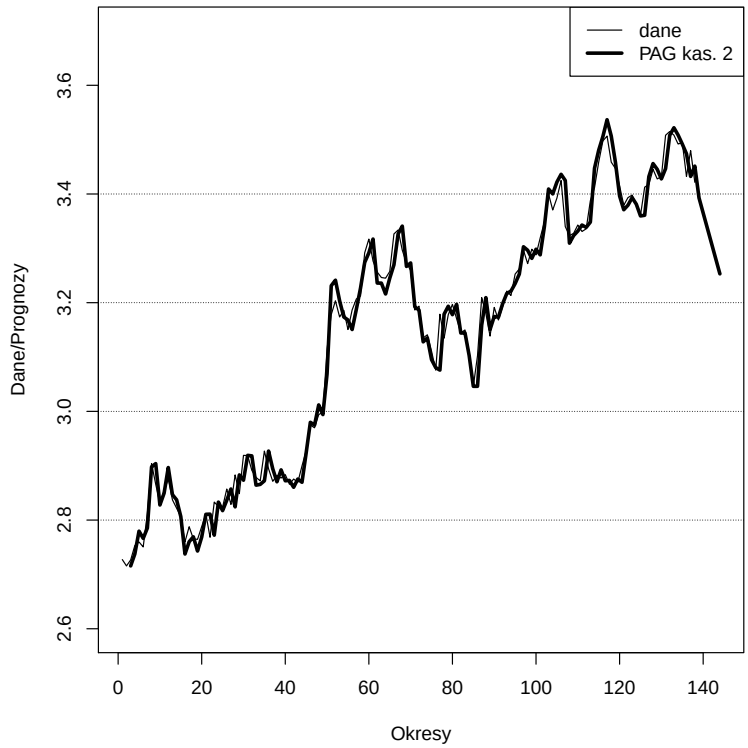
Pełen obraz zachowania prognoz *ex post* i *ex ante* kursu dolara daje wykres na rysunku 4.9. Szczególnie zwracają na nim uwagę dobrze wyprognozowane wahania w środkowej części szeregu oraz w ostatnim segmencie (quasi-sezonowość). Sam wykres powstał przy wykorzystaniu najlepszych wyników z tabeli 4.31. Po raz drugi kaskada 2 pozwoliła na uzyskanie średnio rzecz biorąc niższych błędów. Predykcja przy użyciu kaskady 1, choć wciąż dobra nie zbliżyła się zbytnio do wyniku drugiego wariantu prognozowania kaskadowego. Jak się okazuje włączenie prognoz wygasłych do mechanizmu

wyznaczania prognoz *ex post* szczególnie dobrze sprawdza się gdy pojawią się wahania quasi-sezonowe.

Tabela 4.33: Dokładność prognoz wygasłych kursu dolara amerykańskiego

Wartość	Kaskada 1	Kaskada 2
RMSPE	0,0210	0,0071

Źródło: opracowanie własne



Rysunek 4.9: Prognozy *ex post* i *ex ante* kursu dolara amerykańskiego

Źródło: opracowanie własne

4.1.3. Prognozy IRDN oraz inflacji

Do tej pory testowaliśmy zachowanie prognostycznego algorytmu genetycznego na szeregach o charakterze finansowym – notowaniach giełdowych i kursów walut. Sprawdzimy teraz jak wypadną prognozy dwóch szeregów o odmiennej naturze. Pierwszym będzie notowany na Towarowej Giełdzie

Energii Indeks Rynku Dnia Następnego (IRDN) zaś drugim indeks cen towarów i usług konsumpcyjnych (CPI) liczony jako indeks łańcuchowy miesiąc bieżący do miesiąca poprzedniego.

Towarowa Giełda Energii funkcjonuje od 1999 roku[7]. Od 2003 roku TGE prowadzi giełdę towarową i od tego czasu sukcesywnie rozszerzała działalność związaną z obrotem energią w Polsce. W chwili obecnej odbywa się na niej handel na czterech rynkach, na których Członkowie Giełdy handlują energią elektryczną, prawami majątkowymi i uprawnieniami do emisji CO2. Jednym z nich jest Rynek Dnia Następnego.

Cytując stronę internetową TGE: „jednym z podstawowych jego celów jest kreowanie cen energii elektrycznej dla pozostałych kontraktów zawieranych na hurtowym rynku energii elektrycznej w Polsce. Jego rolą jest również:

— umożliwianie uczestnikom rynku wstępne zbilansowanie swoich pozycji kontraktowych;
— umożliwianie pośredniej wyceny wartości przedsiębiorstw energetycznych (przede wszystkim wytwórców) poprzez wycenę produkowanego przez nich towaru, czyli energii elektrycznej;
— generowanie istotnych sygnałów inwestycyjnych w zakresie budowy nowych mocy wytwórczych."

Ciekawostką związaną z organizacją obrotu na tym rynku jest odbywanie się notowań codziennie, również w dni świąteczne. Na giełdzie papierów wartościowych obowiązuje pięciodniowy tydzień pracy. Na skutek zawartych na Rynku Dnia Następnego transakcji następuje fizyczny przepływ energii elektrycznej od jej wytwórców, przez sprzedawców po firmy zajmujące się hurtowym obrotem energią elektryczną.

Notowany na tym rynku indeks IRDN wyraża cenę średnią ważoną wolumenem ze wszystkich transakcji na sesji giełdowej, liczoną po dacie dostawy dla całej doby. Oprócz niego na giełdzie towarowej notowane są inne indeksy m.in. sIRDN określający cenę liczoną po dacie dostawy dla godz. 8-22 czy IRDN24 dla ceny liczonej po dacie dostawy dla całej doby.

W naszym wypadku wykorzystaliśmy notowania IRDN z tego samego przedziału czasu co poprzednie szeregi jednak specyfika obrotów czyli notowania odbywające się 7 dni w tygodniu spowodowała, że cały szereg liczył 201 obserwacji wobec 138 sesji na GPW. Podzieliliśmy go na pięć segmentów. Pierwszy obejmował 52 obserwacje od początku szeregu do 21.08.2011 roku włącznie. W tym czasie wystąpiły małe i duże wahania losowe oraz nieliniowy trend. Ponadto pojawiła się sezonowość o okresie równym 7 dni.

W drugim segmencie przypadającym między 22.08.2011 a 20.10.2011 roku (60 obserwacji) dekompozycja wskazuje na brak trendu i duże waha-

[7] Opracowanie na podst. strony internetowej tge.pl

nia przypadkowe. Trudno jednoznacznie stwierdzić czy mają one sezonowy charakter jak miało to miejsce w poprzednim podokresie.

Trzeci segment tworzyły 24 notowania od 1.10.2011 do 13.11.2011 r. Wyodrębniliśmy w nim stały poziom zmiennej i niebyt duże wahania losowe za wyjątkiem środkowej części szeregu. Dokładniej 1 listopada zanotowany został duży spadek wartości indeksu, który jednakże został zniwelowany już następnego dnia.

Czwarty segment obejmował sesje od 14.11.2011 do 25.12.2011 r. (42 obserwacje). Charakteryzował się on trendem malejącym, wyraźną sezonowością i dużymi (szczególnie na początku) wahaniami.

Ostatni z wyróżnionych segmentów przypadający na notowania między 26.12.2011 a 17.01.2012 r. (23 obserwacje) odznaczał się z kolei trendem rosnącym z małymi wahaniami losowymi i jedną nietypową niską wartością przypadającą na 1.01.2012 r.

Indeks IRDN notowany jest codziennie przez cały tydzień. Ponadto w części segmentów stwierdziliśmy obecność sezonowości z okresem 7 dni. Zmianie zatem uległy niektóre parametry algorytmu. Wartości t_m przyjęliśmy na poziomie 7 i 14 dni[8]. Prawdopodobieństwa krzyżowania, mutacji a także liczebności populacji pozostały na dotychczasowym poziomie. Startowy zbiór metod zakładał wykorzystanie PAG w wersji 1 dla wszystkich pięciu segmentów.

W tabelach 4.34-4.35 prezentujemy wartości RMSPE prognoz *ex post* notowań indeksu. Jak widać algorytm pozostał przy wyjściowych metodach w każdym z rozpatrywanych zestawów parametrów. Analizując zaś trafność predykcji w rozbiciu na poszczególne segmenty zauważymy, że najwyższy błąd średni wystąpił w czwartym podokresie. Odpowiadają za ten stan rzeczy bardzo duże wahania notowań w pierwszych dwóch tygodniach tego segmentu. Z kolei wpływ na wartość kryterium w trzecim i piątym segmencie miały pojedyncze większe spadki kursu indeksu odpowiednio 1.11.2011 oraz 1.01.2012 roku. W rezultacie średnia wartość funkcji w tabelach 4.34-4.35 oscylowała około 4-5% co wciąż stanowi bardzo dobry rezultat.

Porównując wyniki z punktu widzenia poziomów t_m lepsze (czyli obarczone niższymi błędami) prognozy dał wariant o niższym parametrze. Działo się tak nie tylko dla średniej ważonej RMSPE, ale również w przypadku pojedynczych segmentów.

Metody klasyczne (por. tabela 4.36) dały dość zbliżone do algorytmu prognostycznego wyniki. Jedynie metoda Browna w trzecim segmencie odstaje na niekorzyść od swoich odpowiedników z prognozowania kaskadowego. W czterech pierwszych segmentach dobrze sprawdziły się metody wygładzania wykładniczego zaś w ostatnim model trendu liniowego ze zmienną

[8] Przypomnijmy sobie co na temat sezonowości i PAG pisaliśmy w rozdziale trzecim.

Tabela 4.34: RMSPE dla prognoz notowań IRDN przy $t_m = 14$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,0291	0,0279	0,0297	0,0263
Segment 2	0,0327	0,0311	0,0425	0,0304
Segment 3	0,0516	0,0425	0,0520	0,0471
Segment 4	0,0913	0,0827	0,0972	0,0817
Segment 5	0,0479	0,0386	0,0489	0,0568
Ważone kryt.	0,0480	0,0432	0,0525	0,0451
Metody	[1 1 1 1 1]	[1 1 1 1 1]	[1 1 1 1 1]	[1 1 1 1 1]

Źródło: opracowanie własne

Tabela 4.35: RMSPE dla prognoz notowań IRDN przy $t_m = 7$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,0292	0,0254	0,0271	0,0244
Segment 2	0,0290	0,0284	0,0347	0,0309
Segment 3	0,0527	0,0518	0,0546	0,0524
Segment 4	0,0819	0,0702	0,0613	0,0765
Segment 5	0,0397	0,0362	0,0376	0,0500
Ważone kryt.	0,0442	0,0401	0,0454	0,0436
Metody	[1 1 1 1 1]	[1 1 1 1 1]	[1 1 1 1 1]	[1 1 1 1 1]

Źródło: opracowanie własne

zero-jedynkową. Powtórzyła się sytuacja najwyższe błędy odnotowaliśmy dla czwartego segmentu. Ważone kryterium przekroczyło nieco 5%.

Na podstawie najlepszych wartości ważonego kryterium zostały wyznaczone prognozy *ex ante* notowań IRDN, których trafność następnie oceniliśmy w kategoriach prognoz wygasłych. Punktem wyjścia analizy bayesowskiej stał się PAG w wersji 1, który obowiązywał w ostatnim segmencie kiedy zakończyły się obliczenia.

Obecność sezonowości w interesujący sposób wpłynęła na poziom wartości RMSPE. Prognozy bayesowskie charakteryzowały się trzykrotnie wyższym uśrednionym błędem w porównaniu z prognozami *ex post*. Winę ponosi chwilowy zanik wahań sezonowych w ostatnim segmencie, a które ponownie pojawiły się w kolejnych okresach. Z tego powodu postanowiliśmy wypróbować inne rozwiązanie. Otóż wyznaczyliśmy prognozy traktując najlepszy chromosom z populacji metodą naiwną z sezonowością. Zatem na 7 okresów

Tabela 4.36: RMSPE dla prognoz notowań IRDN metodami klasycznymi

	Metoda	RMSPE
Segment 1	Wintersa	0,0288
Segment 2	Browna	0,0379
Segment 3	Browna	0,0878
Segment 4	Wintersa	0,0917
Segment 5	trendu	0,0542
Ważone kryt.		0,0541

Źródło: opracowanie własne

poza próbą prognozami stało się 7 ostatnich genów rozkodowanego, najlepszego osobnika wyznaczonego zarówno przy pomocy kaskady 1 jak i kaskady 2. Zestawienie wyników zawiera tabela 4.37.

Okazało się, że udało się nam osiągnąć wyraźną poprawę trafności predykcji przy czym kaskada w pierwszej wersji charakteryzowała się nieco niższą wartością RMSPE. Kryterium podziału na segmenty jest jednorodność zachowania z punktu widzenia dekompozycji. Analizując trajektorie danych we wszystkich pięciu segmentach zauważamy, że poza drugim i ostatnim zawsze pojawiały się wahania sezonowe. Zastosowane rozwiązanie sprawdziło się więc choć trudno je uznać za odpowiadające w pełni wymogom analizy bayesowskiej[9], warto mimo wszystko mieć je na uwadze i stosować w podobnych sytuacjach.

Tabela 4.37: Dokładność prognoz wygasłych notowań IRDN

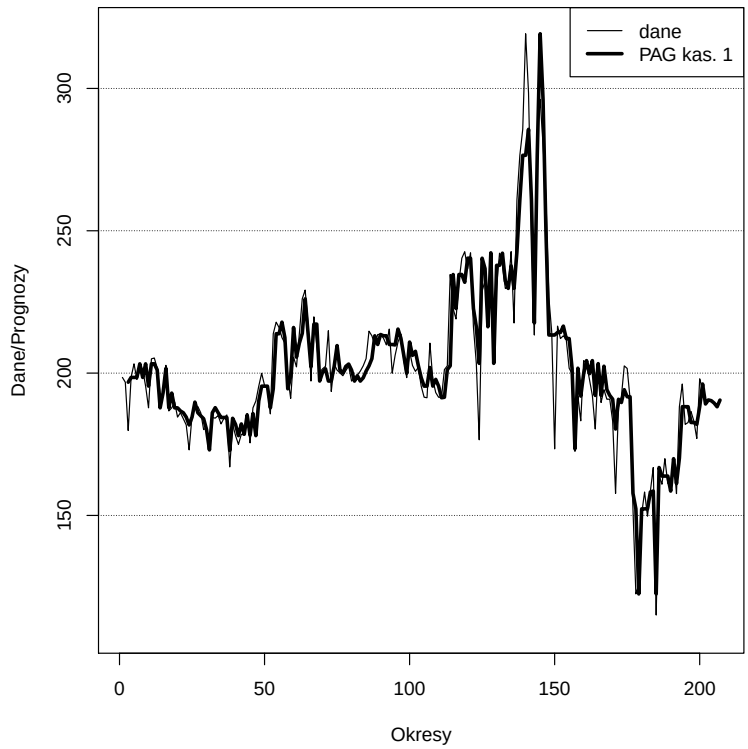
RMSPE	Kaskada 1	Kaskada 2
średnia ważona	0,1330	0,1347
naiwna z sezonowością	0,1076	0,1146

Źródło: opracowanie własne

Podsumowanie efektów osiągniętych dzięki prognostycznemu algorytmowi genetycznemu stanowi wykres na rysunku 4.10. Jak to miało miejsce do tej pory w szeregach notowań różnego rodzaju lepsze prognozy zapewniała niższa wartość t_m oraz wyższe prawdopodobieństwo krzyżowania. Zamieszczone na rysunku 4.10 prognozy powstały dzięki zastosowaniu kaskady 1 przy $t_m = 7$ i p_k z przedziału od 0 do 0,6. Pomimo segmentów charakteryzujących się sezonowością oraz okresów z nagłymi wzrostami i spadkami

[9] Nadal punktem wyjścia jest ostatni segment, ale nie dochodzi do wyznaczenia prawdopodobieństw *a posteriori*

notowań indeksu udało się wiernie odwzorować mechanizmy odpowiadające za kształtowanie wartości szeregu.



Rysunek 4.10: Prognozy *ex post* i *ex ante* notowań IRDN

Źródło: opracowanie własne

Kolejną kategorią, której prognozy zdecydowaliśmy się wykonać przy pomocy prognostycznego algorytmu genetycznego jest indeks cen towarów i usług konsumpcyjnych (CPI) zwany potocznie inflacją. Choć jest to tylko jeden ze znanych ekonomii indeksów cen trudno przecenić jego rolę. Inflacja ma nie tylko bezpośredni wpływ na stan gospodarki wpływając na przykład na oprocentowanie kredytów wszelkiego rodzaju. Służy również do urealniania wartości pieniądza w czasie. Deflatory jednak wyraża się jako indeksy jednopodstawowe zaś nas bardziej interesuje dynamika zmian poziomu cen. Wykonamy więc prognozy CPI jako indeksu łańcuchowego podającego zmianę w danym miesiącu do miesiąca poprzedniego. Tego typu prognozy przydają się przy planowaniu inwestycji czy w sytuacji kiedy chcemy ocenić przyszłe nastroje konsumentów.

Punktem wyjścia stały się dane na temat wskaźnika cen towarów i usług konsumpcyjnych publikowane przez GUS w okresie od stycznia 2009 r. do

lutego 2012 r. (38 obserwacji). Były to indeksy łańcuchowe mierzące zmianę cen w stosunku do miesiąca poprzedniego. W obrębie tego szeregu wyodrębniliśmy dwa segmenty. Pierwszy obejmował 20 miesięcy od początku szeregu do sierpnia 2010 r. włącznie. Do sierpnia 2009 r. inflacja charakteryzowała się trendem malejącym by następnie do końca wydzielonego podokresu przyjąć stały poziom. Zmianom tym towarzyszyły małe wahania losowe.

Do drugiego segmentu zaliczyliśmy pozostałe 18 obserwacji. W tym okresie czasu dekompozycja szeregu wskazała na brak trendu i sezonowości oraz małe wahania przypadkowe.

Parametry algorytmu prognostycznego przyjęliśmy na tym samym poziomie co dla szeregów z rynków finansowych a zatem rozpatrywaliśmy t_m jako równe 5 oraz 10. Populacja liczyła sobie 100 osobników poddanych działaniu algorytmu przez 150 pokoleń. Początkowy zestaw metod zakładał PAG w wersji 1 (bazujący na metodzie naiwnej prostej) dla obu segmentów.

W tabelach 4.38 i 4.39 znalazły się wartości RMSPE dla prognoz wygasłych z podziałem na dwa poziomy t_m. Nieduże wahania losowe w obu segmentach pozwoliły uzyskać bardzo niskie wartości błędów predykcji. Dlatego właśnie ważone kryteria są bardzo zbliżone dla $t_m = 10$ oraz $t_m = 5$ przy lekkiej przewadze pierwszej z wymienionych wartości.

Jeśli chodzi o efekty działania strategii ewolucyjnej to zmiany obserwujemy tylko dwukrotnie dla pierwszego segmentu w tabeli 4.39. Wybrany został PAG zakładający obecność trendu za co odpowiada systematyczny spadek inflacji w kilku początkowych miesiącach tego segmentu.

Tabela 4.38: RMSPE dla prognoz inflacji przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,0018	0,0018	0,0020	0,0018
Segment 2	0,0029	0,0027	0,0022	0,0031
Ważone kryt.	0,0023	0,0022	0,0021	0,0024
Metody	[1 1]	[1 1]	[1 1]	[1 1]

Źródło: opracowanie własne

Niskie wahania losowe i stały poziom zmiennej predestynują do zastosowania (spośród metod klasycznych) metody Browna. Tak właśnie postąpiliśmy zaś rezultaty mierzone poziomem RMSPE prognoz wygasłych zamieściliśmy w tabeli 4.40. Niewątpliwie są to dobre prognozy, ale wypadają słabiej od tych uzyskanych za pomocą algorytmu genetycznego.

Prognozy *ex ante* zostały wyznaczone na sześć kolejnych miesięcy. Sięgnęliśmy po najlepsze wyniki spośród obu wersji prognozowania kaskadowego i wyznaczyliśmy prognozy bayesowskie jako średnie ważone gdzie wagi

Tabela 4.39: RMSPE dla prognoz inflacji przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0, 0{,}3 \rangle$	$p_k \in \langle 0, 0{,}6 \rangle$	$p_k \in \langle 0, 0{,}3 \rangle$	$p_k \in \langle 0, 0{,}6 \rangle$
Segment 1	0,0021	0,0020	0,0020	0,0014
Segment 2	0,0034	0,0031	0,0034	0,0040
Ważone kryt.	0,0028	0,0025	0,0027	0,0027
Metody	[3 1]	[1 1]	[1 1]	[4 1]

Źródło: opracowanie własne

Tabela 4.40: RMSPE dla prognoz inflacji metodami klasycznymi

	Metoda	RMSPE
Segment 1	Browna	0,0029
Segment 2	Browna	0,0043
Ważone kryt.		0,0036

Źródło: opracowanie własne

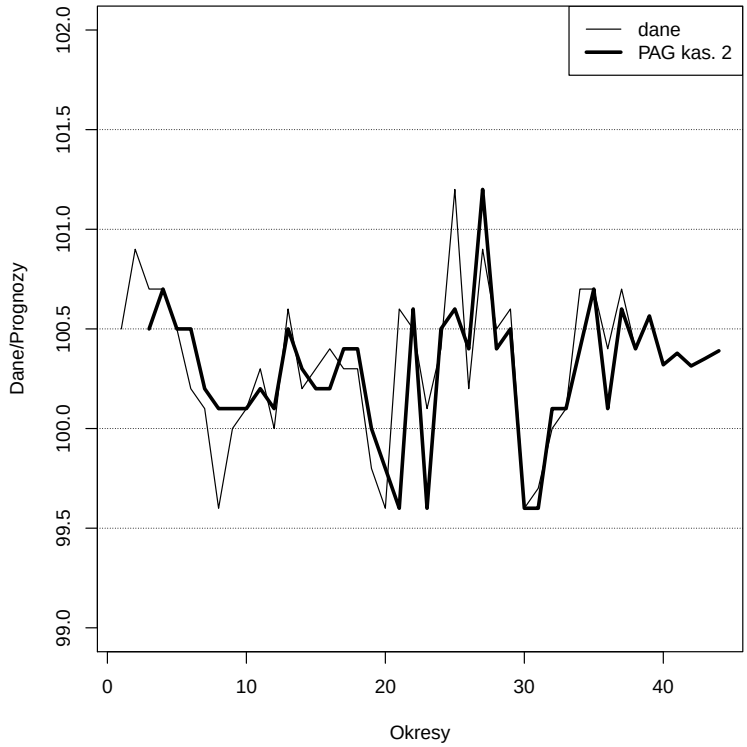
były prawdopodobieństwami *a posteriori*. Następnie obliczyliśmy RMSPE tych prognoz po ich wygaśnięciu. Wyniki znalazły się w tabeli 4.41. Kaskada 1 jak i kaskada 2 osiągnęły identyczny wynik. W porównaniu z okresem *ex post* kryterium przyjęło niemal dwukrotnie wyższą wartość, ale wyniosła ona zaledwie 0,5%.

Tabela 4.41: Dokładność prognoz wygasłych inflacji

Wartość	Kaskada 1	Kaskada 2
RMSPE	0,0050	0,0050

Źródło: opracowanie własne

Najlepsze prognozy *ex post* zapewniło użycie kaskady 2 przy $t_m = 10$ oraz $p_k \in \langle 0, 0{,}3 \rangle$ (por. tab. 4.38). Na podstawie tych wyników oraz odpowiadających im prognoz *ex ante* powstał wykres na rysunku 4.11. Tytułem komentarza warto wspomnieć, że największe rozbieżności między danymi rzeczywistymi a prognozami pojawiły się w sierpniu 2009 r. oraz w marcu 2011 r. W pozostałych okresach wykres potwierdza dobrą trafność prognoz kaskadowych, które dobrze naśladowały zmiany zachodzące w czasie.



Rysunek 4.11: Prognozy *ex post* i *ex ante* kursu inflacji

Źródło: opracowanie własne

4.1.4. PAG a zmienne losowe

Dotychczasowe rozważania dotyczyły rzeczywistych szeregów czasowych, w których odzwierciedlenie znajdują faktyczne mechanizmy stojące za kształtowaniem się zmiennych takie jak oczekiwania inwestorów czy związki miedzy obserwacjami z różnych okresów. Postanowiliśmy sprawdzić jak PAG zachowa się w przypadku zmiennych losowych *sensu stricte* – o konkretnych, założonych z góry rozkładach. W tego typu szeregu wyeliminowany zostanie wpływ zewnętrznych czynników na jego zachowanie a my będziemy mogli sprawdzić słuszność założeń metody.

Do analizy wybraliśmy dwa rozkłady ciągłe: beta i normalny. Dla każdego z nich wygenerowaliśmy po 120 wartości, które zostały podzielone na cztery segmenty po 30 elementów w każdym.

Zachowanie zmiennych różniło się między sobą m.in. dlatego, że miały one różne rozkłady. Pamiętając o wcześniejszych doświadczeniach z rzeczywistymi szeregami czasowymi skupiliśmy się jednak na przypadku pojawiania się większych wahań losowych. Szereg zmiennej o rozkładzie normalnym

(z wartością oczekiwaną 3 oraz odchyleniem standardowym 0,5) charakteryzował się częstymi zmianami wartości z okresu na okres[10] dającymi całkiem spory bo wynoszący około 15% współczynnik zmienności. Z kolei w rozkładzie beta parametry zostały dobrane w taki sposób, aby występowała w nim duża asymetria lewostronna. W efekcie wygenerowany szereg odznaczał się o wiele większą zmiennością w porównaniu do rozkładu normalnego. Duże amplitudy wahań oraz częste przejścia od niskich do wysokich poziomów obserwujemy choćby w zachowaniu obrotów spółek.

Dla zachowania porównywalności z wcześniejszymi wynikami podajemy wyniki obliczeń algorytmu, w którym liczba pokoleń pozostała równa 150. Nie zmieniły się też poziomy t_m i prawdopodobieństwa krzyżowania oraz mutacji. W tabelach 4.42 i 4.43 zaprezentowane zostały wartości RMSPE dla szeregu o rozkładzie normalnym różniące się wielkością t_m. Każdy segment reprezentował stały poziom zmiennej bez sezonowości dlatego początkowy zestaw metod składał się z czterokrotnego użycia PAG w wersji pierwszej (metoda naiwna prosta).

Tabela 4.42: RMSPE dla szeregu o rozkładzie normalnym przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,0777	0,0742	0,0988	0,0774
Segment 2	0,0906	0,0776	0,0779	0,0723
Segment 3	0,0688	0,0608	0,0691	0,0618
Segment 4	0,1580	0,1434	0,1525	0,1288
Ważone kryt.	0,0991	0,0893	0,0996	0,0852
Metody	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]

Źródło: opracowanie własne

Ważona wartość kryterium okazała się wyższa gdy prawdopodobieństwo krzyżowania było niższe. Różnice te wyraźniej widać w tabeli 4.43. Porównanie obu wersji prognozowania kaskadowego wskazuje, że przy $t_m = 5$ wyniki były porównywalne zaś przy $t_m = 10$ korzystniej wypadła kaskada 2. Strategia ewolucyjna pozostawiła początkowy zestaw metod. Same wartości RMSPE są dość wysokie, ale biorąc pod uwagę zmienność szeregu danych należało się tego spodziewać.

Wyraźnie wyższe zróżnicowanie wartości wygenerowanych na podstawie rozkładu beta przełożyło się na wyższe błędy prognoz. Wszystkie znajdujące się w tabelach 4.45 oraz 4.44 warianty obliczeń były zgodne co do jakości

[10] Dla zachowania zgodności z używanym dotąd nazewnictwem nadal będziemy mówić o okresach i obserwacjach.

predykcji z uwzględnieniem zachowania danych w poszczególnych podokresach. Najniższe błędy odnotowaliśmy w pierwszym, najwyższe w czwartym segmencie co szczególnie widać w prognozach przy niższym t_m.

Tabela 4.43: RMSPE dla szeregu o rozkładzie normalnym przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,1143	0,0914	0,1021	0,0942
Segment 2	0,0831	0,0863	0,0990	0,0860
Segment 3	0,0688	0,0616	0,0655	0,0581
Segment 4	0,1603	0,1486	0,1596	0,1518
Ważone kryt.	0,1065	0,0971	0,1066	0,0976
Metody	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]

Źródło: opracowanie własne

Tabela 4.44: RMSPE dla szeregu o rozkładzie beta przy $t_m = 10$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$	$p_k \in \langle 0,0{,}3\rangle$	$p_k \in \langle 0,0{,}6\rangle$
Segment 1	0,1425	0,1411	0,1462	0,1393
Segment 2	0,2342	0,2288	0,2375	0,2300
Segment 3	0,2808	0,2710	0,2904	0,2638
Segment 4	0,4098	0,4053	0,4128	0,4048
Ważone kryt.	0,2688	0,2636	0,2747	0,2615
Metody	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]	[1 1 1 1]

Źródło: opracowanie własne

Zgodność obserwujemy także przy wyborze metod metod prognozowania dla kolejnych podokresów, które objęte zostały PAG pierwszego typu.

Ważone wartości RMSPE okazały się wyraźnie niższe dla $t_m = 10$ przy czym w tabeli 4.45 najlepiej wypadła pod tym względem kaskada 1 o prawdopodobieństwie krzyżowania z przedziału od 0 do 0,3. Był to najlepszy wynik w przypadku rozkładu beta w ogóle choć najlepsze prognozy otrzymane przez kaskadę 2 były tylko o około 1 punkt procentowy gorsze.

Zmienne wygenerowane według zadanego z góry procesu losowego okazały się nie lada problemem dla metod klasycznych o czym możemy przekonać się przeglądając wartości RMSPE w tabeli 4.46. Już samo dobranie metod dla poszczególnych segmentów okazało się wyzwaniem. Pomimo, że podlegały one takiej samej dekompozycji różnią się co do wyboru ostatecznych metod predykcji. Szczególnie duże błędy prognoz *ex post* wystąpiły

Tabela 4.45: RMSPE dla szeregu o rozkładzie beta przy $t_m = 5$

	Kaskada 1		Kaskada 2	
	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$	$p_k \in \langle 0,0{,}3 \rangle$	$p_k \in \langle 0,0{,}6 \rangle$
Segment 1	0,1388	0,1339	0,1419	0,1342
Segment 2	0,1895	0,1449	0,2356	0,2299
Segment 3	0,1971	0,1574	0,2295	0,1626
Segment 4	0,3136	0,2958	0,2637	0,2427
Ważone kryt.	0,2110	0,2636	0,2190	0,1933
Metody	[1 4 1 1]	[1 4 1 1]	[1 1 1 1]	[1 1 1 1]

Źródło: opracowanie własne

Tabela 4.46: RMSPE dla prognoz zmiennych losowych metodami klasycznymi

	Rozkł. beta		Rozkł. normalny	
	Metoda	RMSPE	Metoda	RMSPE
Segment 1	Browna	0,2949	Browna	0,1725
Segment 2	śr. ruchoma $k = 3$	0,4520	Browna	0,1510
Segment 3	Browna	0,5314	śr. ruchoma $k = 3$	0,1576
Segment 4	naiwna prosta	0,6470	Browna	0,2650
Ważone kryt.		0,4813		0,1865

Źródło: opracowanie własne

dla zmiennej o rozkładzie beta. Prognostyczny algorytm genetyczny dzięki wrodzonej elastyczności zdecydowanie góruje w tym aspekcie.

Generując zmienne o założonych rozkładach możemy pozwolić sobie na wyznaczenie dodatkowych elementów szeregu, aby sprawdzić jak wypadną prognozy wygasłe. Tabela 4.47 zawiera wartości RMSPE prognoz wygasłych dla pięciu okresów następujących po ostatnim segmencie. Powstały one na podstawie parametrów, które zapewniły najlepsze wyniki w tabelach 4.42-4.44. Strategia ewolucyjna preferowała PAG 1 typu we wszystkich segmentach, dlatego prawdopodobieństwa *a posteriori* zostały włączone do prognoz *ex ante* jako wagi średniej ważonej odpowiednich danych z szeregu.

Prognozy wygasłe dla rozkładu normalnego odznaczają się minimalnie mniejszą trafnością od prognoz *ex post*, przy czym kaskada 2 nieco lepiej przewidywała przyszłe zachowanie zmiennej od kaskady 1. RMSPE prognoz wygasłych z rozkładu beta wyraźniej różnią się od ważonej wielkości tej miary dla prognoz *ex post* niż miało to miejsce przy okazji rozkładu normalnego. W ich przypadku prognozy wygasłe przewyższyły trafnością prognozy *ex post* dla niektórych wariantów nawet o około 10 punktów procentowych.

Tabela 4.47: Dokładność prognoz wygasłych szeregu o rozkładzie normalnym

Rozkł. beta		
Wartość	Kaskada 1	Kaskada 2
	0,1718	0,1713
Rozkł. normalny		
Wartość	Kaskada 1	Kaskada 2
	0,1508	0,1015

Źródło: opracowanie własne

Porównując zaś odpowiadające sobie wyniki prognozowania kaskadowego da się zauważyć sporą przewagę kaskady drugiego rodzaju.

Wykresy na rysunkach 4.12 i 4.13 przedstawiają prognozy *ex post* oraz *ex ante* w odniesieniu do danych wygenerowanych na podstawie rozkładu normalnego i beta. Szereg o rozkładzie normalnym zestawiony został z prognozami pochodzącymi z kaskady drugiego typu o $t_m = 10$ i $p_k \in \langle 0, 0{,}6 \rangle$ zaś szereg o rozkładzie beta z predykcją kaskady drugiego typu o $t_m = 10$ i $p_k \in \langle 0, 0{,}6 \rangle$. Były to parametry, które dały najlepsze prognozy *ex post* spośród przedstawionych w tabelach 4.42-4.44.

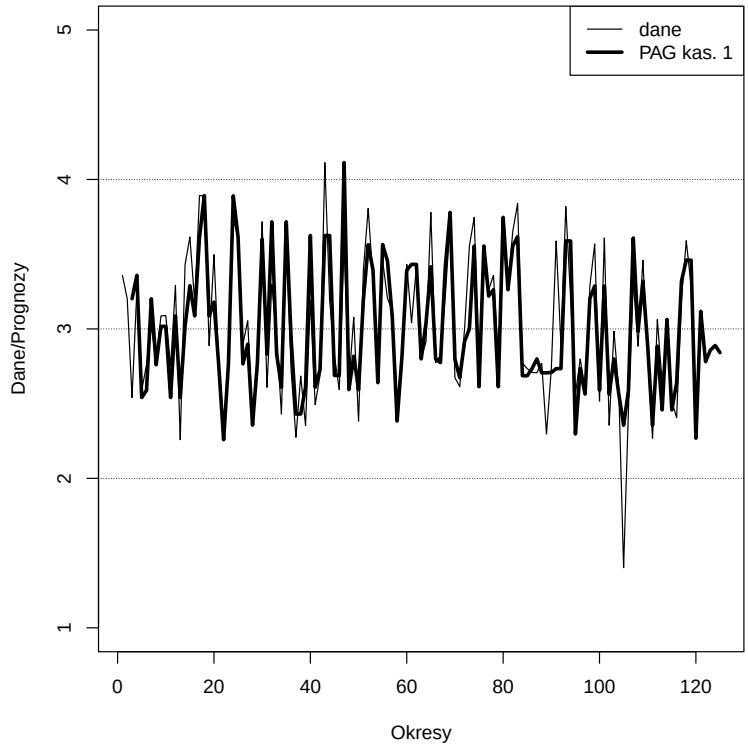
Na wykresach wyraźnie widać różnicę w zmienności między obydwoma rozkładami. Duże zróżnicowanie rozkładów spowodowało, że lepiej sprawdzał się PAG o wyższym parametrze t_m i prawdopodobieństwie krzyżowania.

Warto zapamiętać, że choć prognozy *ex post* zmiennej o rozkładzie beta bazujące na tych parametrach wypadły słabiej, to prognozy wygasłe znalazły się znacznie bliżej rzeczywistych wartości wygenerowanego szeregu, szczególnie gdy weźmiemy pod uwagę prognozy bayesowskie.

Wykorzystaliśmy dwa szeregi o losowym charakterze (choć z odgórnie zadanymi parametrami rozkładu). Algorytm poradził sobie z nimi w podobny sposób jak uczynił to z rzeczywistymi szeregami czasowymi wyrażającymi pewną sytuację rynkową i związane z nią mechanizmy. To dobra wiadomość, gdyż oznacza, iż dysponujemy narzędziem nadającym się do wykorzystania nawet wówczas gdy nie jesteśmy w stanie wskazać czynników o istotnym znaczeniu wpływających na kształtowanie się zmiennej.

4.2. Efektywność PAG podczas przeszukiwania zbioru wariantów prognostycznych

Wyznaczając prognozy dla opisanych wcześniej szeregów mieliśmy okazję przekonać się, że zachowanie (w sensie dekompozycji) danych w poszczególnych segmentach wpływa na efekt końcowy pracy algorytmu. Szczególnie



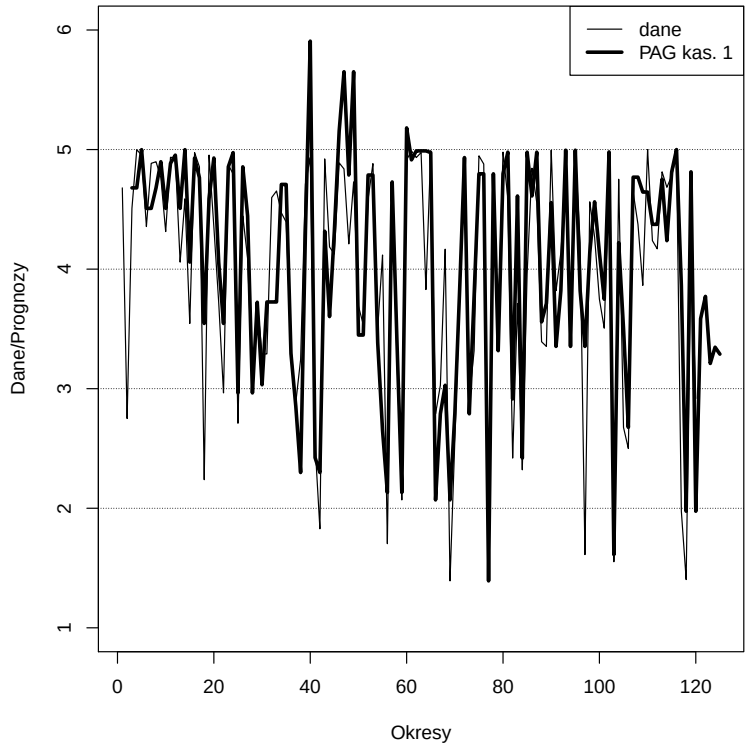
Rysunek 4.12: Prognozy *ex post* i *ex ante* zmiennej o rozkładzie normalnym

Źródło: opracowanie własne

wyraźnie widać to kiedy porównamy trafność predykcji *ex post* instrumentów notowanych na giełdzie. Prognozy notowań charakteryzowały się bardzo dobrymi (niskimi) wartościami RMSPE. Jednak odpowiadające im wolumeny obrotów osiągały kilkakrotnie gorsze wyniki.

Z punktu widzenia dekompozycji te dwie grupy szeregów różniły się zmiennością. Cechą wspólną wolumenów obrotów analizowanych walorów jest gwałtowne przechodzenie od bardzo niskich do bardzo wysokich wartości i na odwrót. Często dochodziło do pojawiania się punktów zwrotnych, które z uwagi na swoją nieprzewidywalność nie dają się uwzględnić w procedurach prognostycznych.

Rodzi się w tym miejscu pytanie o to, jak efektywny jest prognostyczny algorytm genetyczny podczas przeszukiwania zbioru potencjalnych rozwiązań. Musimy też zastanowić się czy i jak mocno na efektywność poszukiwań wpływa zachowanie w czasie objętego działaniem algorytmu szeregu. Postaramy się w tej części pracy odpowiedzieć na te pytania a następnie wy-



Rysunek 4.13: Prognozy *ex post* i *ex ante* zmiennej o rozkładzie beta

Źródło: opracowanie własne

ciągnąć wnioski pomagające w ustaleniu wartości parametrów sterujących algorytmem pod kątem przewidywania mechanizmów jego funkcjonowania.

Ocenimy efektywność działania algorytmu prognostycznego z kilku punktów widzenia, ale wspólnym mianownikiem będzie śledzenie drogi prowadzącej do końcowego rozwiązania. Nie staramy się jedynie szczegółowo analizować czasu pracy algorytmu, ponieważ różniłby się on w zależności od konfiguracji komputera, na którym przeprowadza się obliczenia.

4.2.1. Zmiany przystosowania najlepszego chromosomu populacji

Część prognozowanych szeregów dotyczyła notowań różnego rodzaju zmiennych. Ich cechą wspólną była mała zmienność wartości w sąsiadujących ze sobą okresach. Do tej grupy zaliczymy również szereg danyvh na temat inflacji. Oczywiście zdarzały się obserwacje odstające w niektórych segmentach, ale na tle całego szeregu stanowiły one nieduży odsetek danych i jak pokazały przeprowadzone obliczenia nie dochodziło z ich powodu do

znaczącego pogorszenia trafności prognoz *ex post*. Zmiany trendu udało się wyeliminować dzieląc szereg na odpowiednie segmenty.

W pierwszej kolejności zajmiemy się analizą zmian przystosowania najlepszego chromosomu populacji, które zachodziły z pokolenia na pokolenie. Ma to dla nas istotne znaczenie ze względu na zaimplementowany w algorytmie mechanizm selekcji metodą elitarną, w którym najlepszy osobnik pochodzący z bieżącego pokolenia jest przekazywany do następnej generacji chromosomów (gdzie zastępuje chromosom o najgorszym przystosowaniu).

Wykresy na rysunku 4.14 przedstawiają wartości RMSPE najlepszych chromosomów pokolenie po pokoleniu dla prognoz notowań WIG z podziałem na segmenty. Dane do jego utworzenia pochodziły z kaskady 1 o $t_m = 5$ i prawdopodobieństwie krzyżowania od 0 do 0,6 czyli tych parametrów algorytmu, które zapewniły prognozy *ex post* obarczone najmniejszym średnim błędem. Są one wystarczająco reprezentatywne również dla reszty wariantów prognoz tego szeregu. Wykresy te ilustrują wpływ metody elitarnej na poprawę wyników w kolejnych pokoleniach. Przedstawione wartości funkcji przystosowania dotyczą najlepszych chromosomów w danym pokoleniu dla ostatniej iteracji strategii ewolucyjnej.

Dla pierwszego segmentu, który charakteryzował się najniższą cząstkową wartością RMSPE algorytm szybko i sukcesywnie poprawiał rezultaty do około 100 pokolenia. Ostatnie 50 pokoleń charakteryzuje się wyraźnie wolniejszymi zmianami wartości przystosowania najlepszego chromosomu. Wiąże się to z występowaniem w większej ilości lepiej przystosowanych osobników, które dominowały w populacji. Ponadto średni błąd prognoz znalazł się poniżej 1% a zatem trudno oczekiwać pojawienia się radykalnie lepszej predykcji. Współczynnik zmienności szeregu wartości funkcji przystosowania najlepszych chromosomów okazał się najwyższy i wyniósł 34%. Algorytm w ciągu wspomnianych 100 pokoleń znalazł się w okolicach optimum. RMSPE dla reszty generacji niewiele się zmieniało czyli mocno różniło się od wcześniejszych wartości. Stąd wysoki poziom współczynnika.

Drugi segment charakteryzował się najwyższymi błędami prognoz *ex post* w całym szeregu. Przez pierwsze 50 pokoleń dominowało w nim kilka osobników o zbliżonym przystosowaniu. Następnie doszło do szybkiej poprawy wartości kryterium, która wyhamowała około 30 pokoleń przed końcem postępowania.

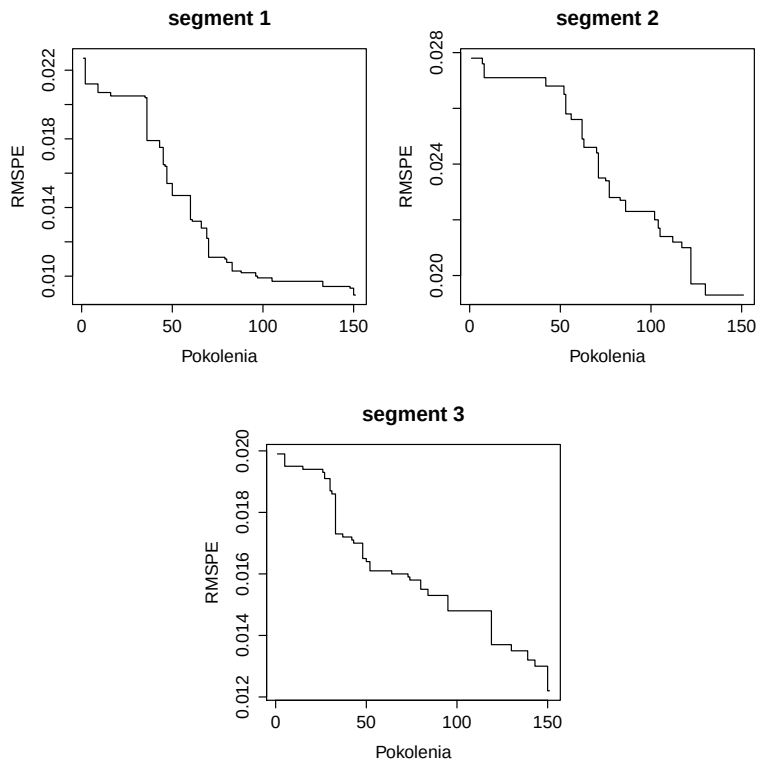
Przypomnijmy, że w segmencie tym wyodrębniliśmy duże wahania losowe wraz z kilkoma lokalnymi szczytami notowań WIG i właśnie te elementy wpłynęły na sposób przeszukiwania zbioru potencjalnych prognoz.

Segment numer 3 sprawił algorytmowi więcej kłopotów. Co kilka pokoleń dotychczas najlepszy chromosom był zastępowany osobnikiem lepszym. Sytuacja taka trwała od trzydziestego do niemal ostatniego pokolenia. Co ciekawe RMSPE nie było wcale w tym podokresie największe. Znaczenie

jednak miało zachowanie szeregu czyli trend zmieniający co kilkanaście dni kierunek czemu towarzyszyły niekiedy duże wahania losowe. Jak widać na ostatnim z trzech wykresów na rysunku 4.14 poprawa przystosowania najlepszego chromosomu trwała praktycznie do ostatniego pokolenia. Zarówno w tym jak i w poprzednim segmencie współczynnik zmienności wartości RMSPE najlepszych osobników wyniósł 13%.

Szczególne zachowanie łamanych na wykresach wynika z faktu, że poprawa wyników nie dokonuje się natychmiastowo z pokolenia na pokolenie. Dany chromosom może pozostawać najlepszy przez kilka przebiegów algorytmu.

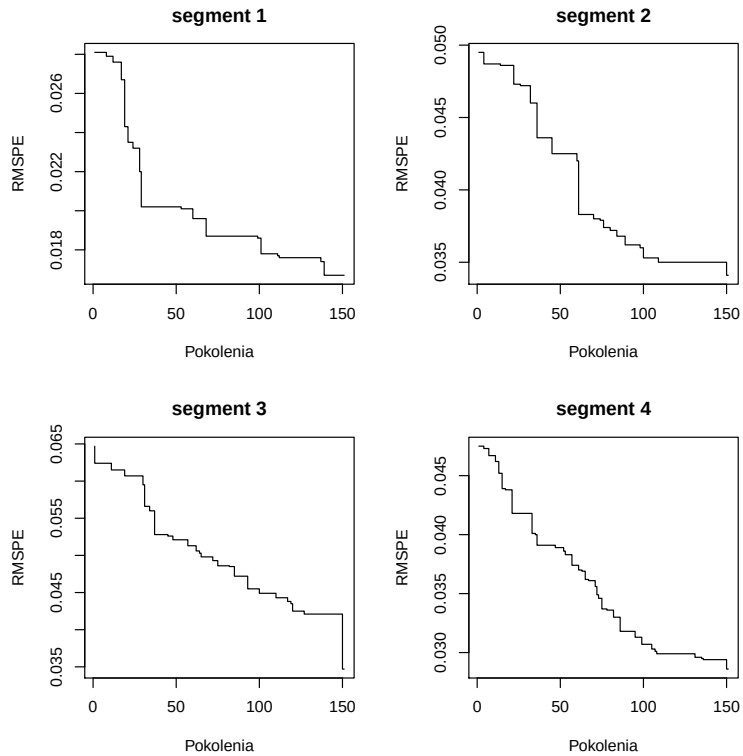
Badania przeprowadzone do tej pory (Kucharski 2008*b*) wykazały, że w przypadku notowań dziennych PAG dość szybko zbiega (mierząc to liczbą pokoleń) w pobliże rozwiązania optymalnego pod warunkiem stabilnego zachowania danych w czasie. Wówczas w następnych pokoleniach ma już miejsce nieduża poprawa wyników.



Rysunek 4.14: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy notowań WIG

Źródło: opracowanie własne

Jako ilustrację zmian przystosowania najlepszych chromosomów w kolejnych populacjach dla notowań KGHM wybraliśmy prognozy otrzymane przy pomocy kaskady 2 o $t_m = 5$ i $p_k \in \langle 0, 0{,}6 \rangle$. W tym szeregu wyodrębniliśmy 4 segmenty. Wykresy na rysunku 4.15 pokazują w jaki sposób zmieniały się wartości RMSPE najlepszego osobnika w populacji przez 150 generacji.



Rysunek 4.15: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy notowań KGHM

Źródło: opracowanie własne

Szczególną uwagę chcielibyśmy zwrócić na segmenty 1 i 3. Pierwszy doświadczył w połowie okresu mocnej korekty kursu. Przed i po tym fakcie dane zachowywały się jednak stabilnie tj. podlegały małym wahaniom losowym. Okazało się, że algorytm w tej sytuacji bardzo szybko (w ciągu 30 pierwszych pokoleń) poprawił znalezione rozwiązanie po czym już do końca obliczeń dokonywał jedynie pomniejszych korekt. Warto zauważyć, że dany chromosom pozostawał najlepszy zwykle przez kilkadziesiąt generacji. Właśnie w tym segmencie zanotowaliśmy najniższe średnie błędy predykcji.

Segment 3 szeregu notowań kursu KGHM wyróżnił się dwukrotną zmianą kierunku trendu. Dodatkowo zaobserwowaliśmy w nim duże wahania przypadkowe. Analiza stosownego wykresu na rysunku 4.15 pozwala dostrzec, że wartość funkcji przystosowania najlepszego chromosomu jest bezustannie poprawiana. Dopiero pod sam koniec obliczeń algorytm znalazł znacznie lepsze rozwiązanie w porównaniu z dotychczasowymi. Trochę podobną sytuację mamy w segmencie 4 z tym, że duże zmiany wartości RMSPE zakończyły się około 40 pokoleń przed zakończeniem postępowania.

Zmienność przystosowania najlepszych osobników była we wszystkich czterech segmentach dość podobna i zawierała się w przedziale od 13 do nieco ponad 16 procent.

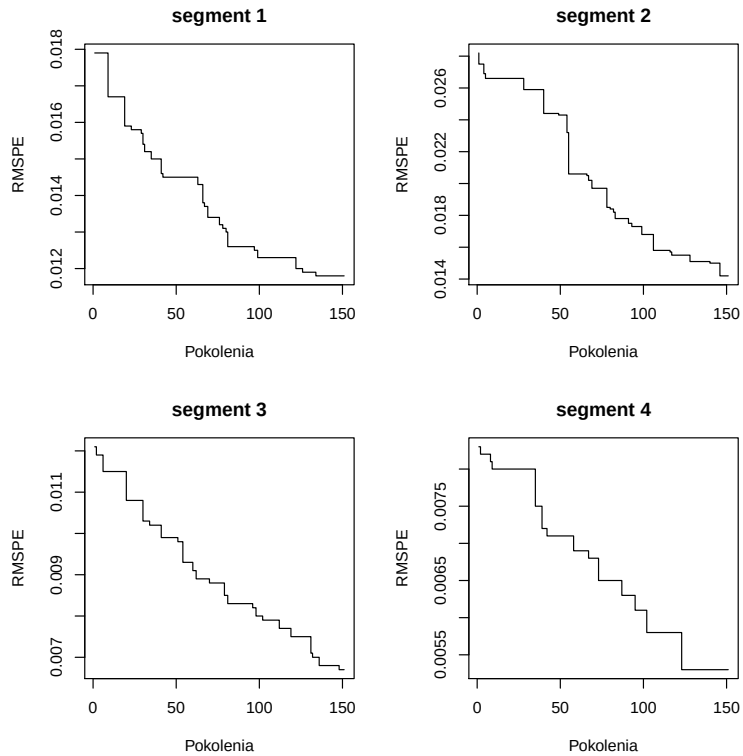
Zachowanie algorytmu podczas poprawy przystosowania najlepszego chromosomu prognoz notowań TVN nie odbiegało od opisanego dla kursu WIG czy KGHM. Na nieco więcej uwagi zasługuje jedynie drugi segment. Jest on jednym z najkrótszych ze wszystkich wyróżnionych i opisanych segmentów w tej pracy. Tworzące go obserwacje podlegały małym wahaniom losowym oraz trendowi malejącemu. Chromosom otrzymany w setnym pokoleniu nie ulega zmianie aż do końca postępowania. Właśnie w tym segmencie współczynnik zmienności dla szeregu najlepszych osobników z kolejnych pokoleń okazał się najwyższy i równał się 44%. Drugi co do zmienności, segment czwarty charakteryzował się dwukrotnie mniejszą dyspersją.

W podokresach 1, 3 i 4 w danych pojawiały się wyższe wahania więc stabilizacja rozwiązania następowała 20-30 pokoleń później niż w segmencie drugim. Poza wymienionym już ostatnim segmentem współczynniki zmienności wyniosły odpowiednio 17% (segment 1) i 19% (segment 3).

Omówienie poprawy wartości RMSPE dla notowań kursów walut rozpoczniemy od franka szwajcarskiego. Wykresy na rysunku 4.16 prezentują zmianę przystosowania najlepszych chromosomów we wszystkich czterech segmentach. Do ich utworzenia posłużyły dane z prognozy kaskadowej pierwszego typu o $t_m = 5$ i $p_k \in \langle 0,0{,}6 \rangle$, a więc parametrach, które zapewniły najlepszą jakość prognoz *ex post*. Trzeba jednak pamiętać, że w stosunku do kaskady 2 różnica była naprawdę minimalna zaś prognozy *ex post* tego szeregu okazały się obarczone bardzo małymi błędami. RMSPE dla segmentów 3 i 4 znalazło się wyraźnie poniżej 1%.

Wartości owe algorytm osiągnął pomimo występujących w danych punktów zwrotnych czy obserwacji nietypowych. Jednak analizując rysunek 4.16 zauważamy, że wspomniane elementy dekompozycji zaznaczyły swój wpływ na przebieg procesu poszukiwania jak najbardziej trafnych prognoz.

We wszystkich segmentach najlepszy znaleziony chromosom pozostawał takim przez zaledwie przez kilka pokoleń po czym był zastępowany jeszcze lepszym osobnikiem. Proces poprawy rozwiązania trwał praktycznie przez całe 150 pokoleń. Najrzadziej do zmian dochodziło w czwartym



Rysunek 4.16: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy kursu franka szwajcarskiego

Źródło: opracowanie własne

segmencie, który charakteryzował się wahaniami określonymi w pracy mianem quasi-sezonowych. Pod względem zróżnicowania wyróżnił sie jedynie segment drugi, w którym współczynnik zmienności osiągnął 23%. W pozostałych podokresach był on o około 10 punktów niższy.

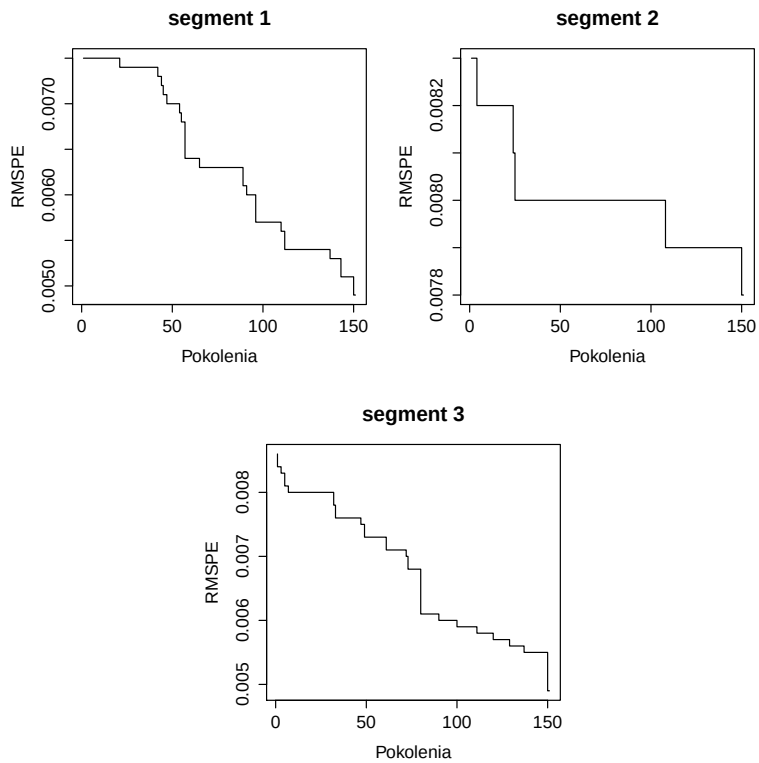
Całkiem podobną sytuację obserwujemy w prognozach kursu euro gdzie najbardziej dokładne prognozy powstały dzięki zastosowaniu kaskady 2 dla niższych poziomów parametrów sterujących algorytmem (por. tabela 4.23). Wartości RMSPE, podobnie jak miało to miejsce przy okazji prognozowania kursu franka znalazły się poniżej 1%. Ponadto w ostatnim segmencie również wystąpiły wahania quasi-sezonowe.

Uznaliśmy wykresy na rysunku 4.17 za warte zamieszczenia ze względu na ilustrację zachowania algorytmu w segmentach drugim a także trzecim. W pierwszym z wymienionych dwukrotnie doszło do odwrócenia trendu. Pomiędzy wyznaczającymi zmiany punktami dane wykazywały się obecnością tendencji rozwojowej z małymi wahaniami. Dopiero pod koniec drugiego

segmentu wzrosła ich amplituda. Algorytm tylko kilka razy poprawiał znalezione rozwiązanie co przypomina sytuację opisaną przy okazji notowań TVN. Odmienność polega jednak na tym, że segment dla kursu euro liczył sobie 39 a nie 12 obserwacji.

Znalazło to odzwierciedlenie również w zróżnicowaniu wartości przystosowania najlepszych chromosomów, która okazała się bardzo niska i równa zaledwie 1,3%. Dla porównania w segmencie 1 współczynnik osiągnął 13 zaś w trzecim 15%.

Jak zapewne pamiętamy, w segmencie trzecim (podobnie jak w czwartym kursu franka) odnotowaliśmy obecność wahań quasi-sezonowych. Porównując odpowiadające sobie wykresy na rysunkach 4.16 oraz 4.17 na pewno dostrzeżemy, że poprawa przystosowania najlepszego chromosomu dokonuje się z podobną częstotliwością wyrażoną liczbą okresów, w ciągu których dany osobnik funkcjonuje jako najlepiej przystosowany. Podobne okazuje się też rozproszenie wartości funkcji przystosowania.



Rysunek 4.17: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy kursu euro

Źródło: opracowanie własne

Zachowanie najlepszych chromosomów zbliżone do opisanego wcześniej obserwujemy również w prognozach *ex post* kursu dolara. W drugim segmencie tego szeregu trend kilka razy zmieniał kierunek. Tak jak miało to już miejsce na przykład przy okazji prognoz kursu KGHM, algorytm często poprawiał dotychczasowe rozwiązanie co trwało do samego końca postępowania. Ponownie w takiej sytuacji współczynnik zmienności okazał się najwyższy (26%).

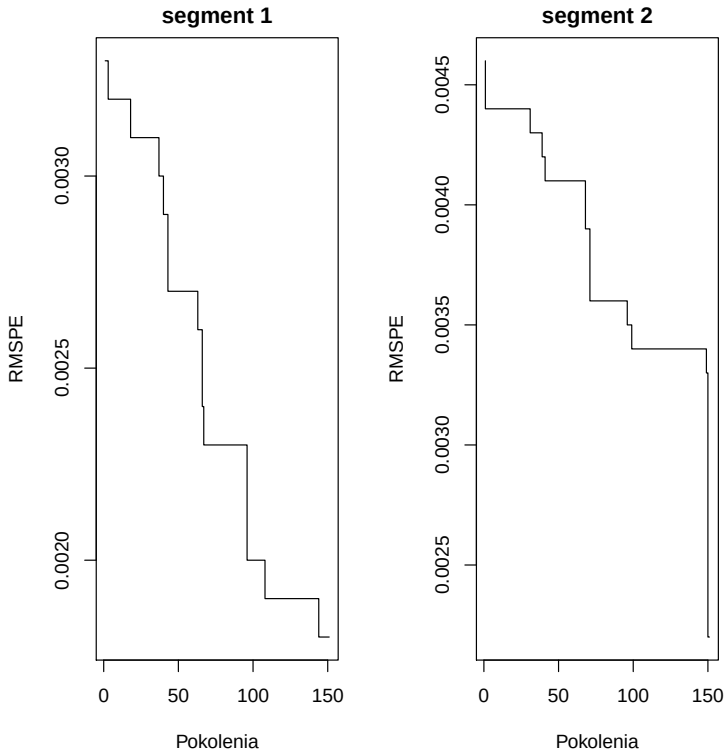
Zauważmy, że pojawianie się punktów zwrotnych powoduje utrudnia PAG poszukiwanie globalnego optimum. Trzeba jednak wspomnieć przy tej okazji, że RMSPE najlepszego chromosomu pochodzącego z kaskady 2 pomimo problemów okazało się jednak bardzo niskie – poniżej 1%. Dobrze to świadczy o elastyczności algorytmu genetycznego.

Ostatni, czwarty segment wyodrębniony spośród danych tworzących szereg kursu dolara charakteryzował się wahaniami quasi-sezonowymi, które nałożyły się na trend liniowy. W podobnie zachowujących się segmentach franka i euro nie mieliśmy do czynienia z obecnością tendencji rozwojowej. Jak się okazało nie wpłynęła ona na przebieg przeszukiwania zbioru rozwiązań. W porównaniu do pozostałych segmentów poprawa wartości kryterium najlepszego chromosomu w pokoleniu następuje rzadziej i jest wyższa. Ostateczne rozwiązanie zostaje znalezione około 120 pokolenia. Współczynnik zmienności osiągnął wartość 15% czyli bardzo podobną do tej z odpowiednich segmentów pozostałych walut.

Jeśli chodzi o segmenty 1 i 3 to zachowanie algorytmu było typowe dla danych charakteryzujących się trendem i wahaniami losowymi o zmiennej amplitudzie. Znacząca poprawa rozwiązania zajmowała około 100 generacji po czym zmiany wartości przystosowania najlepszego osobnika stabilizowały się i wartość RMSPE pozostawała na stałym poziomie.

W następnej kolejności zajmiemy się poprawą przystosowania najlepszego chromosomu podczas wyznaczania prognoz wskaźnika cen towarów i usług konsumpcyjnych. Cały szereg podzielony został na dwa segmenty zaś najlepsze prognozy algorytm wyznaczył dla $t_m = 10$ i $p_k \in \langle 0, 0{,}3 \rangle$. Na rysunku 4.18 zamieściliśmy wykresy poprawy RMSPE chromosomu zawierającego najlepsze prognozy w danym pokoleniu. Szereg danych dotyczących inflacji był znacznie krótszy od pozostałych poddanych działaniu PAG zestawów danych co wpłynęło na krótszy czas obliczeń.

Małe wahania w obu segmentach sprawiły, że osobnik o najlepszym przystosowaniu potrafił przetrwać nawet kilkadziesiąt pokoleń zanim został zastąpiony co odbiło się na współczynniku zmienności, który dla pierwszego segmentu równał się 21% a dla drugiego 11%. Na uwagę zasługuje nietypowe zachowanie algorytmu w drugim segmencie, w którym w ostatnim pokoleniu dochodzi do dużego zmniejszenia wartości RMSPE. To jedyny taki przypadek spośród wszystkich przeanalizowanych szeregów.



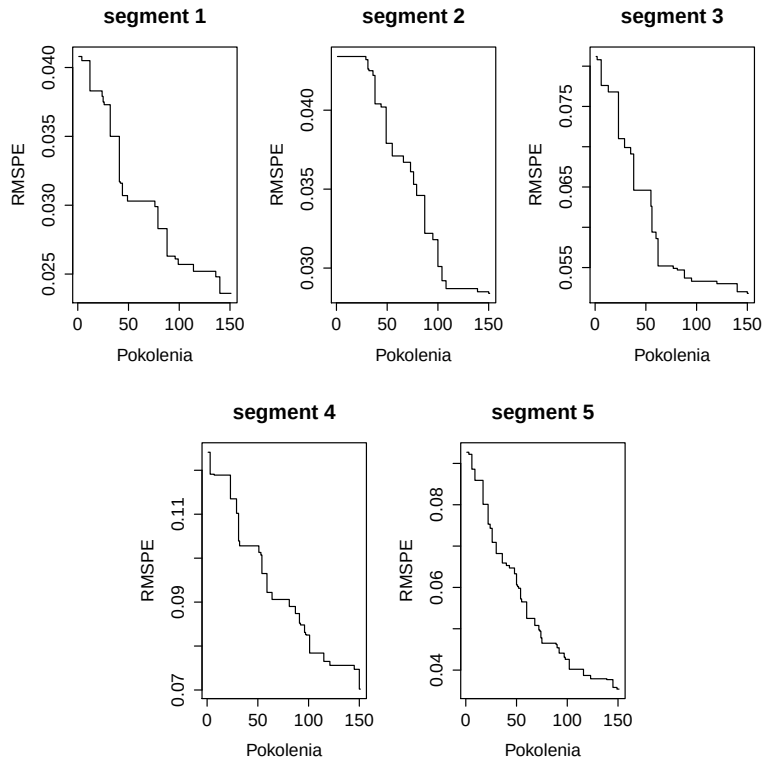
Rysunek 4.18: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy inflacji

Źródło: opracowanie własne

Przyjrzyjmy się teraz najlepszym chromosomom wyznaczonym podczas prognozowania notowań indeksu IRDN. Cechami charakterystycznymi tego szeregu były: sezonowość o okresie równym 7 oraz duże wahania losowe. Wyjątek stanowił drugi segment, w którym dekompozycja nie wykazała wahań sezonowych. Wykresy na rysunku 4.19 ilustrują zmiany wartości funkcji przystosowania z podziałem na pięć segmentów. Wykorzystaliśmy w tym celu najbardziej korzystne wyniki kaskady 1 rodzaju z tabeli 4.35.

Generalnie algorytm wielokrotnie poprawiał znalezione rozwiązanie we wszystkich wyróżnionych podokresach przy czym zmiany zachodziły najczęściej w ostatnim, piątym segmencie. W nim również współczynnik zmienności był najwyższy (31%). Dla reszty segmentów zróżnicowanie wartości funkcji przystosowania w ciągu 150 pokoleń zawierało się między 16 a 18%.

W segmentach drugim oraz trzecim po upływie około 100 pokoleń nie obserwujemy już dużych zmian wartości RMSPE. Sezonowość w segmencie trzecim nie była aż tak wyraźna co wobec braku trendu zaowocowało



Rysunek 4.19: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy notowań IRDN

Źródło: opracowanie własne

radykalnym obniżeniem się średniego błędu prognoz *ex post* do poziomu bliskiego ostatecznemu rozwiązaniu w ciągu pierwszych 50 pokoleń. Nieliniowy trend w segmencie pierwszym nie przeszkodził algorytmowi osiągnąć RMSPE poniżej 3%.

Uprzedzając nieco fakty możemy powiedzieć, że wysokie wahania losowe w głównej mierze odpowiadają za większy „wysiłek” jaki algorytm musi włożyć, aby dotrzeć do satysfakcjonującego rozwiązania. O tym, że potrafi tego dokonać przekonaliśmy się wcześniej przy okazji omawiania trafności prognoz *ex post*.

Szeregi czasowe opisane do tej pory łączył fakt, że pojawiające się duże wahania losowe odchylały się od trendu w granicach, które można było uznać za rozsądne (rzędu kilkunastu procent). Poza tym zwykle występowały w pojedynkę lub w małych grupach i dawały się modelować z dobrym skutkiem. Nie dało się jednak nie zauważyć, że ich obecność wpływała na

przebieg obliczeń algorytmu, a jeszcze większy wpływ miały towarzyszące dużym wahaniom zmiany kierunku tendencji rozwojowej szeregu.

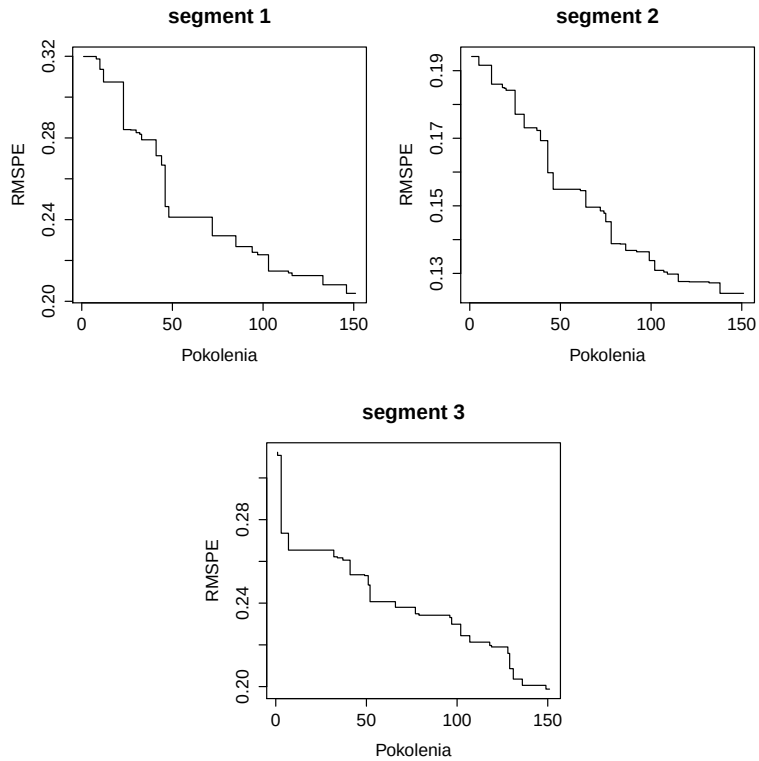
W drugiej grupie wykorzystanych danych czyli wolumenach obrotów WIG, KGHM i TVN oraz zmiennych losowych wysoka zmienność jest normą. Z okresu na okres dochodzi w nich do przejścia od bardzo niskich do bardzo wysokich wartości. Przewidywanie siły i kierunku zmian staje się niezmiernie trudnym zagadnieniem. Błędy prognoz *ex post* otrzymanych przy pomocy prognozowania kaskadowego okazywały się nawet kilkanaście razy wyższe niż ich odpowiedniki z szeregów notowań.

Weźmy na przykład wolumen obrotów WIG. Środkowy segment odróżniał się swoim zachowaniem, ponieważ duże wahania losowe oscylowały przez cały wyróżniony okres w stałych granicach. Zresztą na tej podstawie wyodrębniliśmy ten segment. Na wykresie na rysunku 4.20 widzimy jednak, że poprawa przystosowania najlepszego chromosomu następowała najdalej co kilkanaście pokoleń a przeważnie częściej. Właśnie dla tego segmentu ważona wartość RMSPE okazała się najniższa. Wyciągamy więc wniosek, że wysokie wahania przypadkowe przyczyniają się do spowolnienia (mierzonego częstością zastępowania najlepszego chromosomu) procesu poszukiwania rozwiązania przez algorytm genetyczny. Z drugiej strony ograniczenie amplitudy wahań przyczynia się do tego, że spowolnienie jest mniejsze.

Potwierdzeniem powyższej tezy niech będą pozostałe dwa wykresy na rysunku 4.20. I w pierwszym, i w trzecim segmencie wahania losowe zwiększają swoją amplitudę z upływem czasu. Ponadto niektóre notowania odznaczały się nagłymi, niezwykle wysokimi wzrostami by już w następnej sesji spaść w okolice minimum dla całego szeregu czasowego. Skutkuje to szybkim polepszeniem RMSPE w początkowych pokoleniach i wzrostem częstotliwości zastępowania najlepszych chromosomów później co szczególnie dobrze ilustruje sytuacja w segmencie trzecim. Co ciekawe we wszystkich segmentach współczynnik zmienności RMSPE najlepszych osobników ze 150 pokoleń okazał się dość podobny zamykając się w granicach 10-15%

Analogiczny stan rzeczy obserwujemy w przypadku najlepszych chromosomów dla prognoz obrotów akcjami KGHM (por. rys. 4.21). Pojedyncze wahania na początku i na końcu segmentu drugiego odstawały od pozostałych obserwacji, które dawały się zamknąć w jednym kanale trendowym. Podobnie jak drugim segmencie obrotów WIG algorytm poprawia rozwiązanie niemal do samego końca postępowania przy czym właśnie w tym podokresie zanotowaliśmy najniższy RMSPE.

Spośród pozostałych wykresów wyróżnia się ten dla trzeciego segmentu, w którym szczególnie często dochodziło do przejść od niskich do wysokich obrotów KGHM. Na wykresie obserwujemy, że co około 20-30 pokoleń dochodziło do skokowej zmiany wartości przystosowania najlepszego chromosomu. Zmienność obrotów w segmentach 1 i 4 sprawiała, że algorytm znalazł



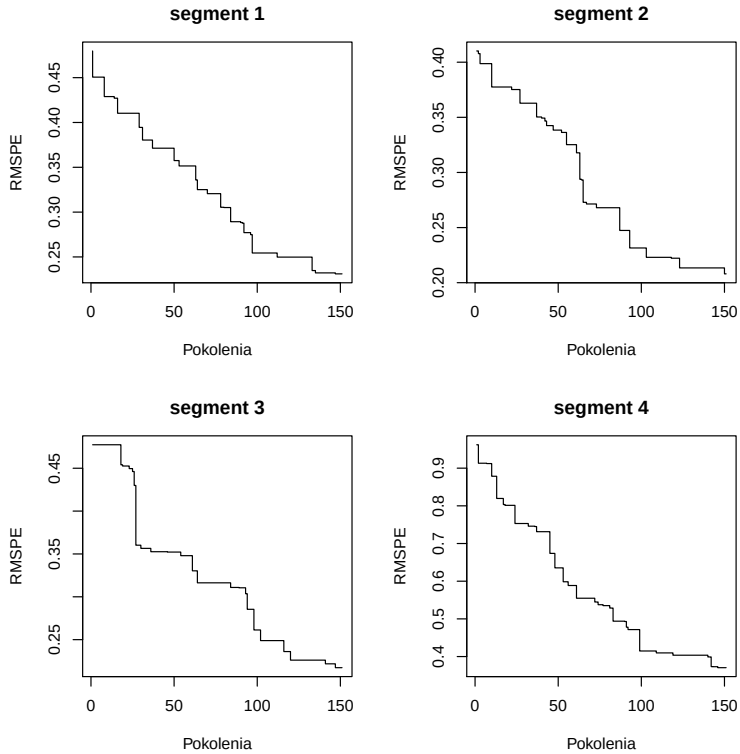
Rysunek 4.20: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy wolumenu WIG

Źródło: opracowanie własne

się w pobliżu rozwiązania po około 100 pokoleniach. Po dotarciu do tej granicy przez kolejne 50 generacji nie dochodzi już do szybkiej poprawy wartości funkcji przystosowania. Dobrze to widać jeśli porównamy je z wartościami funkcji przystosowania 50 generacji wcześniej. Najniższe zróżnicowanie rozwiązań wystąpiło w segmencie pierwszym (22%) zaś najwyższe w czwartym (współczynnik zmienności równy 30%).

Bardzo duże wahania losowe wystąpiły również na początku pierwszego segmentu szeregu wolumenu obrotów TVN. Na podstawie wykresu na rysunku 4.22 możemy przekonać się, że przez około 70 pokoleń algorytm poprawił wartość kryterium o połowę. Potem nie dochodziło już do dużych zmian aż do końca obliczeń.

O specyfice tego segmentu zadecydowały sesje z 8 i 11 lipca 2011 roku. Choć w pozostałych okresach segmentu obroty nadal podlegały dużej zmienności, osiągnęły one co najwyżej połowę wielkości z owych dwóch sesji. Współczynnik zmienności oczywiście okazał się najwyższy i wyniósł 27%.

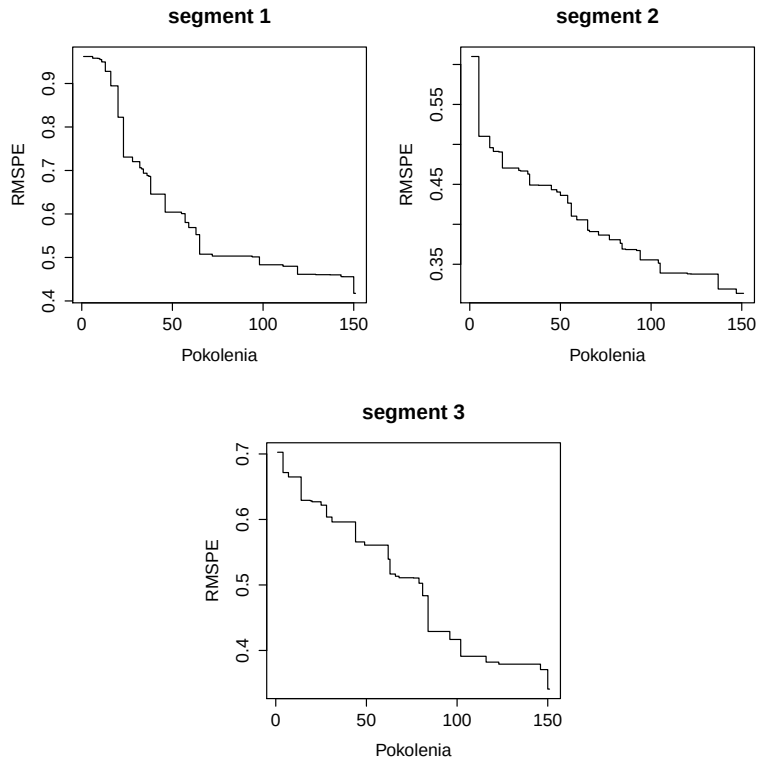


Rysunek 4.21: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy wolumenu KGHM

Źródło: opracowanie własne

W segmencie 2 również mieliśmy do czynienia z dwoma nietypowo wysokimi obserwacjami, po których doszło do natychmiastowego spadku. Stosowny wykres na rysunku 4.22 wskazuje, że algorytm dość szybko zneutralizował wpływ tych sesji po czym przeszedł do tradycyjnego przy dużym zróżnicowaniu częstego poprawiania rozwiązania co kilka pokoleń aż do setnej generacji kiedy proces poprawy rozwiązania wyhamował. W segmencie trzecim różnica polegała tylko na tym, że obserwacji odstających było więcej w związku z czym zróżnicowanie zaprezentowanych na wykresie wartości było wyższe i osiągnęło poziom 21% wobec 17% dla poprzedniego segmentu.

Ostatnim typem danych o dużej zmienności, które ocenimy pod kątem poprawy rozwiązania w kolejnych generacjach są szeregi zmiennych losowych o rozkładach beta i normalnym. Różnią się one od szeregów wolumenów obrotów tym, że poziom zmienności został w nich z góry założony i jest sterowany określonymi parametrami. W rozkładzie beta założyliśmy większe



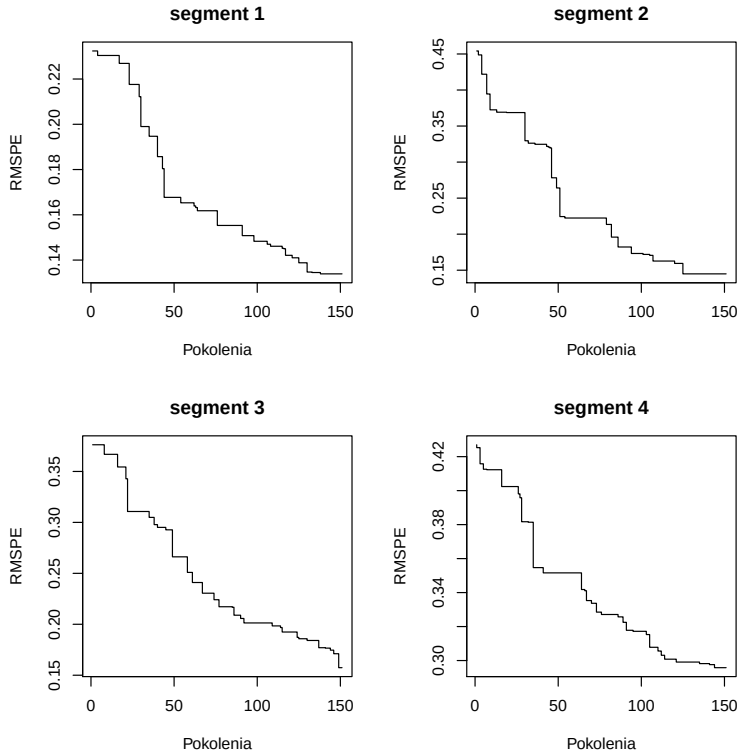
Rysunek 4.22: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy wolumenu TVN

Źródło: opracowanie własne

zróżnicowanie oraz silną asymetrię. Oba szeregi podzielone zostały na cztery jednakowej długości segmenty po 30 obserwacji w każdym.

Wykres na rysunku 4.23 zawiera przystosowanie najlepszych chromosomów dla rozkładu beta. Prognozy zmiennej o rozkładzie normalnym zachowywały się podobnie czyli najniższe wyznaczone RMSPE zmienia się w odstępie kilku-kilkunastu pokoleń. W przypadku rozkładu beta na uwagę zasługuje co innego. W trzech z czterech segmentów po około 50 pokoleniach spada poziom zmian wartości kryterium. Same zmiany są równie częste jak dotychczas, ale najlepsze chromosomy z sąsiednich pokoleń różnią się mniej niż miało to miejsce na początku obliczeń.

Dysponując sporym zestawem wyników możemy wypowiedzieć się na temat związku między liczbą pokoleń a skalą zmian wartości funkcji kryterium. Jak już kilka razy nadmieniliśmy decydują dwa czynniki: zróżnicowanie samych danych oraz ewentualne zmiany trendu. Dla części segmentów w ciągu mniej więcej 50 pokoleń dochodzi do najbardziej istotnej popra-



Rysunek 4.23: Poprawa RMSPE najlepszych chromosomów w kolejnych pokoleniach, prognozy zmiennej o rozkładzie beta

Źródło: opracowanie własne

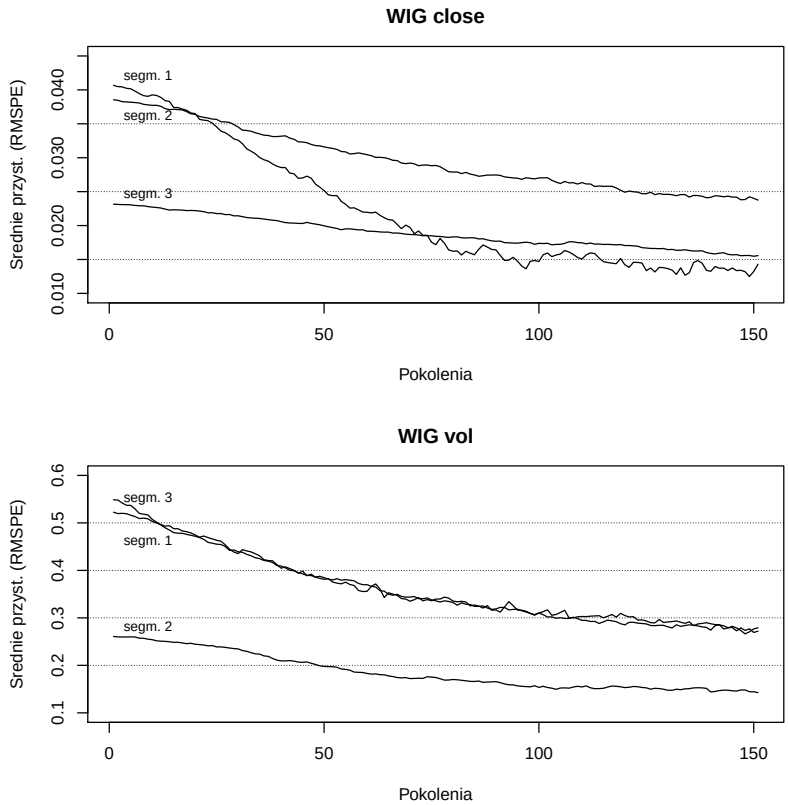
wy przystosowania chromosomów mierzonego ważonym RMSPE prognoz *ex post*. Rząd poprawy zależał od tego czy były to szeregi notowań gdzie rzadko spotyka się duże zmiany wartości z okresu na okres czy ich zachowanie przypominało wolumen obrotów.

4.2.2. Zmiany średniego przystosowania chromosomów populacji

Wyznaczyć rozwiązanie w algorytmie genetycznym oznacza wskazać chromosom o najbardziej korzystnej wartości funkcji przystosowania. Dlatego w poprzednim podrozdziale często mówiliśmy o poprawie rozwiązania gdyż skupialiśmy się na jednym konkretnym osobniku z całej populacji. Skoncentrowaliśmy się przy tym na wpływie dekompozycji szeregu na szybkość poszukiwania rozwiązania mierzoną liczbą pokoleń. Nie analizowaliśmy dokładnie tempa zmian samej wartości kryterium. Skokowe, często duże zmiany wartości zmiennej zaburzają możliwe do wykorzystania mierniki.

Nie wyczerpaliśmy jednak tematu oceny efektywności pracy algorytmu. Zajmiemy się bowiem badaniem zmian średniego przystosowania chromosomów całej populacji w kolejnych pokoleniach.

Na jednym rysunku (por. rys. 4.24) umieściliśmy wykresy średniego przystosowania chromosomów populacji w segmentach notowań WIG (WIG close) oraz wolumenu obrotów tego indeksu (WIG vol). Punktem wyjścia był algorytm o parametrach zapewniających najniższe ważone RMSPE.



Rysunek 4.24: Średnie wartości funkcji przystosowania w kolejnych pokoleniach populacji prognoz WIG

Źródło: opracowanie własne

Średnie przystosowanie populacji chromosomów zawierających prognozy notowań WIG mierzone wartością RMSPE najszybciej spadało w pierwszym segmencie zaś wolniej w drugim i trzecim. Zachowanie segmentu numer 1 wyróżnia się tym, że do połowy postępowania średnia szybko się zmniejsza przy niewielkich odchyleniach po czym spada tempo zmian zaś wzrasta zmienność. Co ciekawe, właśnie dla tego segmentu uzyskaliśmy najbardziej

trafne prognozy. Przeciętne tempo poprawy wyników mierzone średnią geometryczną indeksów łańcuchowych średnich w kolejnych pokoleniach wyniosło dla segmentu pierwszego 0,69%, dla pozostałych dwóch odpowiednio 0,3% i 0,26%. Segmenty drugi i trzeci charakteryzowały sie też mniejszym zróżnicowaniem średniego przystosowania populacji. Współczynniki zmienności wyniosły odpowiednio 15 i 12% wobec aż 41% w segmencie 1.

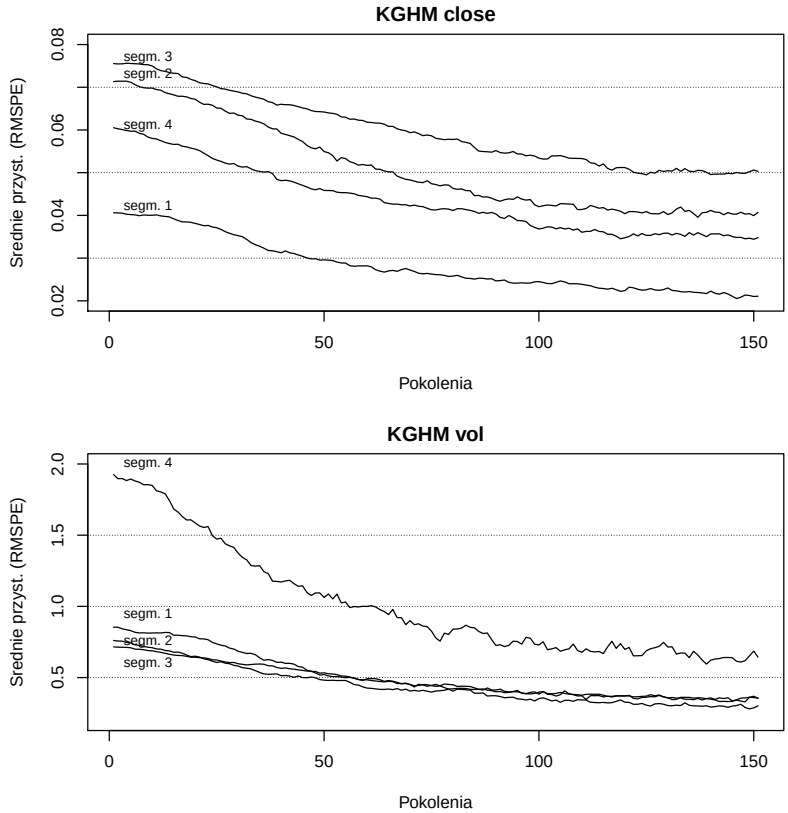
Nieco inaczej przedstawia się, zilustrowane dolnym wykresem na rysunku 4.24 średnie przystosowanie chromosomów wolumenu obrotów WIG. Segment drugi, który zawierał prognozy obarczone najniższymi błędami odstaje od pozostałych dwóch podokresów co do średniego przystosowania populacji. Tempo poprawy wyników z pokolenia na pokolenie wynosiło przeciętnie 0,4%. Segmenty 1 i 3 miały podobne średnie RMSPE, ale poprawa wyników następowała tylko nieco szybciej.

Średnia geometryczna indeksów liczonych pokolenie do pokolenia pozwoliła ustalić tempo spadku wartości kryterium odpowiednio na: 0,43% oraz 0,45%. Zmienność średniego przystosowania jest we wszystkich trzech segmentach bardzo podobna i plasuje się około 20%. Wyłączając pierwszy segment notowań WIG, przeciętne tempo spadku wartości RMSPE jest wyższe w szeregu obrotów za co odpowiada większe zróżnicowanie danych.

Same zmiany z generacji na generację na pierwszy rzut oka nie wydają się duże, ale należy wziąć pod uwagę, że RMSPE jest miarą względną czyli mniej podatną na wahania wartości niż miary bezwzględne. Poza tym analizujemy poziom tego miernika uśredniony dla całej, licznej populacji więc nawet nieduże zmiany trwające przez dostatecznie wiele pokoleń potrafią doprowadzić do pożądanych rezultatów.

Średnie przystosowanie w populacji chromosomów zawierających prognozy kursu KGHM (por. rys. 4.25) w porównaniu z notowaniami indeksu giełdowego odpowiadało zachowaniu prognoz *ex post*. Najniższe przeciętne przystosowanie wystąpiło dla segmentu pierwszego zaś najwyższe dla trzeciego przy czym to właśnie w pierwszym odnotowaliśmy największą zmienność. Współczynnik zmienności kształtował się na poziomie 22%. Przeciętne tempo zmian z pokolenia na pokolenie było równe 0,44%. Najwolniej średnia wartość funkcji kryterium spadała w trzecim segmencie z tempem poprawy wynoszącym średnio 0,27%. Segmenty 2 i 4 plasowały się pod tym względem pomiędzy segmentami 1 i 3.

Zgoła odmienną sytuację obserwujemy w szeregu wolumenu obrotów akcjami KGHM. Segment 4 odstaje od pozostałych trzech pod każdym względem. Przede wszystkim już jego wykres sugeruje znaczne zróżnicowanie, które zmierzone współczynnikiem zmienności sięgnęło 40%. Średnie tempo zmian wyznaczone na podstawie średniej geometrycznej indeksów łańcuchowych również okazało się najwyższe i równe 0,77%. W tym czasie obroty akcjami KGHM podlegały stopniowo coraz większym wahaniom. Dla



Rysunek 4.25: Średnie wartości funkcji przystosowania w kolejnych pokoleniach populacji prognoz KGHM

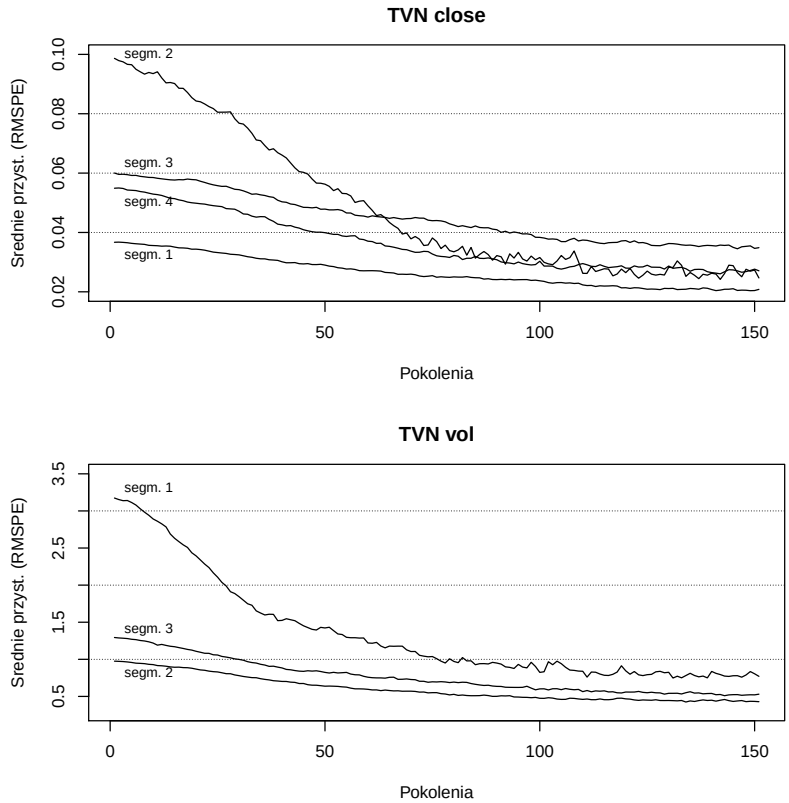
Źródło: opracowanie własne

pozostałych segmentów średnie tempo spadku RMSPE obliczonego dla całej populacji zmieniało się od 0,49% (segment 3) przez 0,59% (segment 1) do 0,63% (segment 2). Zróżnicowanie między pierwszym a ostatnim pokoleniem oscylowało około 30%.

Na pierwszym z wykresów zamieszczonych na rysunku 4.26 uwagę przyciąga segment 2. Był to jeden z najkrótszych spośród wszystkich segmentów charakteryzujący się wyraźnym trendem malejącym. Do około 75 pokolenia następowała szybka poprawa średniego przystosowania chromosomów po czym tempo zmian wyhamowuje. Wzrastają za to wahania średniej wartości funkcji przystosowania. Współczynnik zmienności dla całego segmentu był duży gdyż wyniósł 50%.

Wartość średniego RMSPE dla populacji z pokolenia na pokolenie poprawiała się przeciętnie o 0,92%. Na tle dotychczas przedstawionych seg-

mentów jest to wysokie tempo zmian. Dla pozostałych segmentów wykresy mają znacznie łagodniejszy przebieg a przeciętne tempo zmian wskazywało na poprawę rzędu 0,4% z pokolenia na pokolenie.



Rysunek 4.26: Średnie wartości funkcji przystosowania w kolejnych pokoleniach populacji prognoz TVN

Źródło: opracowanie własne

W przypadku obrotów akcjami TVN wyróżnia się pierwszy segment. Wahania losowe osiągnęły w nim maksimum w całym szeregu zaś z tego powodu błędy prognoz *ex post* były tu najwyższe. Podokres ten również odznaczał się zróżnicowaniem średniego przystosowania osobników na poziomie 51% oraz przeciętnym tempem zmian z poprawą rzędu 0,94% z pokolenia na pokolenie. W porównaniu do segmentu 2 kursu tej spółki najszybsza poprawa zachodziła przez pierwsze 40 pokoleń. Segmenty 2 i 3 pod względem średniego przystosowania chromosomów zachowywały się bardzo podobnie tak pod względem zróżnicowania (30%) jak i tempa zmniejszania RMSPE (odpowiednio 0,55% i 0,59%).

Spośród trzech walut, których prognozy wykonaliśmy zachowanie średniego przystosowania chromosomów populacji najbardziej przewidywalne było dla franka szwajcarskiego. W segmentach pierwszym i drugim mieliśmy do czynienia z wyższymi błędami prognoz *ex post* z powodu obecności wysokich wahań lub zmian trendu. Z tego powodu charakteryzowały się one wyższą zmiennością w obrębie 150 pokoleń i przeciętnym tempem spadku średniego RMSPE w populacji na poziomie 0,44%. Segmenty trzeci i czwarty okazały się do siebie podobne. Poprawa przeciętnej wartości kryterium chromosomów następowała w tempie 0,36-0,38% średnio na pokolenie.

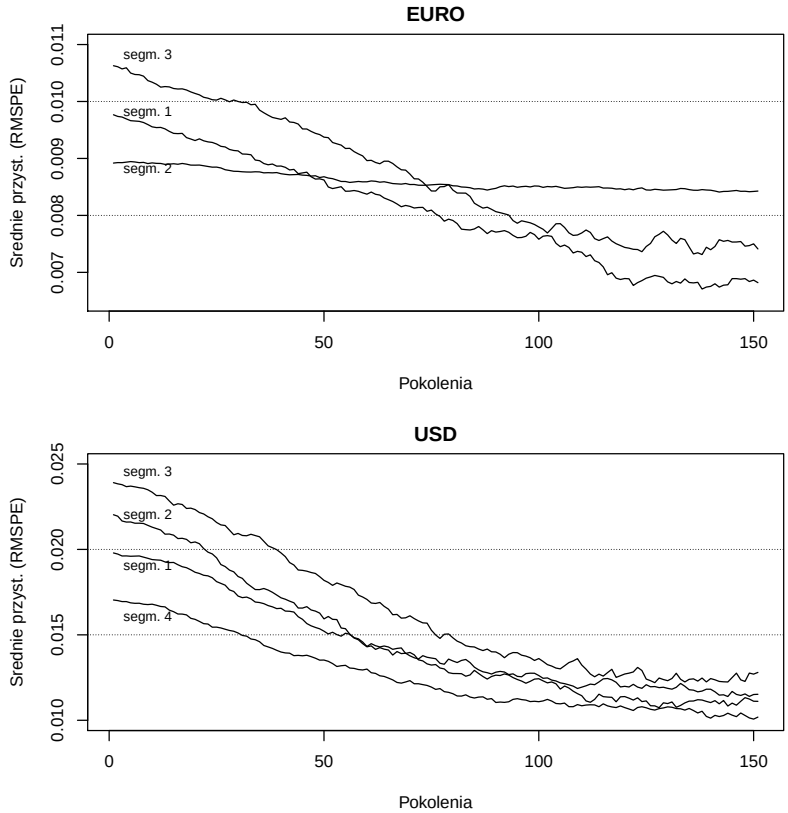
Rzeczą wartą wzmianki jest to, że w czwartym segmencie występowały wahania quasi-sezonowe. Ich względna regularność (w odróżnieniu od segmentu drugiego gdzie kurs franka po nagłym wzroście szybko wracał na wcześniejszą trajektorię) sprawiła, że algorytm tworzył jednorodne populacje w kolejnych pokoleniach. Współczynnik zmienności znalazł dla tego segmentu się na poziomie 19%.

Jeśli chodzi o średnie przystosowanie chromosomów składających się na populację prognoz kursu euro to zdecydowanie wyróżnia się drugi segment tego szeregu (por. górny wykres na rys. 4.27). Charakteryzuje się on minimalną zmiennością na poziomie 2%. To najniższa wartość ze wszystkich segmentów, których prognozy wykonaliśmy. Średnia geometryczna indeksów łańcuchowych również wskazała na minimalne tempo spadku RMSPE z pokolenia na pokolenie przyjmując wartość 0,04%.

Algorytm już na początku znalazł bliskie optymalnemu dla danego segmentu rozwiązanie a zachodzące zmiany wiązały się z faktem obliczeń wykonywanych dla pozostałych segmentów bo sama poprawa wyników jest jak wspomnieliśmy minimalna. W końcu prognozowanie kaskadowe obejmuje cały szereg czasowy. Już analiza wartości funkcji przystosowania najlepszego chromosomu populacji w poprzednim podrozdziale wskazywała, że od połowy liczby generacji algorytm nie dokonuje większych zmian u najlepszego osobnika. W drugim segmencie współczynnik kierunkowy trendu zmieniał się dwa razy, ale pomiędzy tymi zmianami notowania nie podlegały większym wahaniom.

W pozostałych dwóch segmentach kursu euro obserwujemy typowe zachowywanie czyli do około 100 pokolenia następuje wyraźny spadek średniej wartości funkcji kryterium po czym wzrasta zróżnicowanie populacji między kolejnymi pokoleniami. Zmienność średniego RMSPE uplasowała się dla obu segmentów na poziomie 12%. Podobne również okazało się przeciętne tempo spadku tej miary wynoszące 0,24%.

Z perspektywy dekompozycji sytuacja w tych segmentach zmieniała się znacznie bardziej dynamicznie. Pierwszy segment podlegał silnemu trendowi rosnącemu w całym swoim zakresie zaś w trzecim wystąpiły wyraźne wahania quasi-sezonowe.



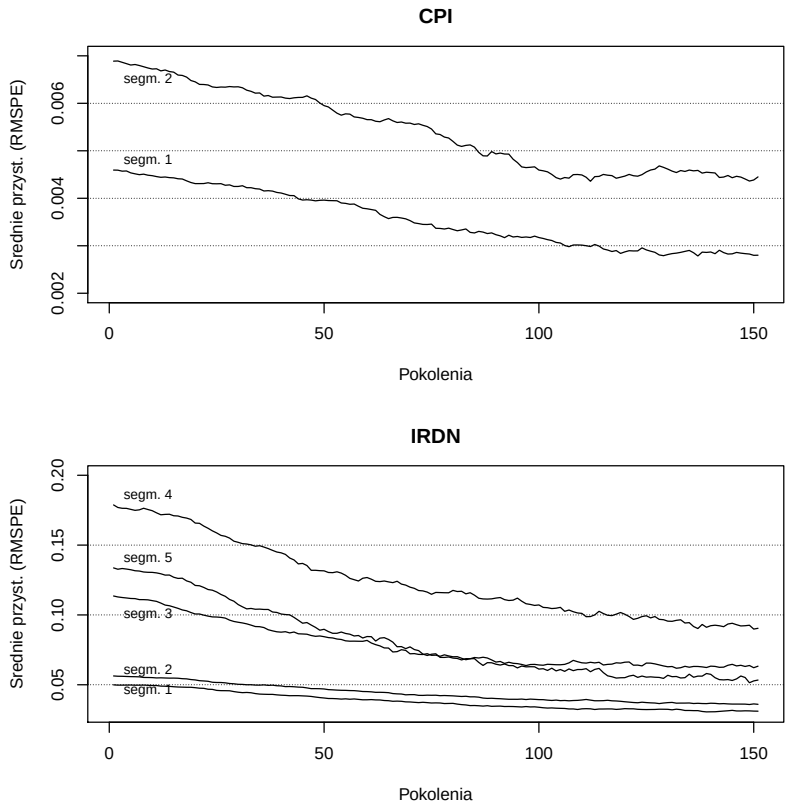
Rysunek 4.27: Średnie wartości funkcji przystosowania w kolejnych pokoleniach populacji prognoz EURO i USD

Źródło: opracowanie własne

Zachowanie średniego przystosowania chromosomów w segmentach prognoz kursu dolara amerykańskiego jest podobne. Różni je jedynie rząd wielkości. A zatem przez nieco ponad 100 pokoleń obserwujemy szybszy spadek przeciętnej wartości RMSPE po czym trajektoria na wykresie wyhamowuje, wzrasta za to zróżnicowanie. Współczynniki zmienności dla wszystkich 150 generacji wahały się od 17% (segment 4) do 24% (segmenty 2 i 3). Średnie tempo zmian najwyższe okazało się w segmencie drugim (0,46%) a najniższe w czwartym (0,34%).

Jeżeli przypomnimy sobie, że RMSPE prognoz *ex post* oscylował w okolicach 1% we wszystkich wyróżnionych podokresach, takie zachowanie populacji jako całości nie powinno dziwić. Już wcześniej stwierdziliśmy wpływ zmian trendu (segmenty 2 i 3) oraz wahań quasi-sezonowych (segment 4) na średnią wartość funkcji przystosowania osobników tworzących populację.

Do podobnych wniosków dojdziemy analizując zachowanie średniego przystosowania chromosomów w obu segmentach prognoz indeksu łańcuchowego cen towarów i usług konsumpcyjnych – pierwszy wykres na rysunku 4.28. Poprawa przeciętnego RMSPE populacji następowała w tempie 0,33% oraz 0,29% odpowiednio w segmentach pierwszym i drugim co pod względem skali zmian czyni je podobnymi do prognoz USD dla segmentu 4.



Rysunek 4.28: Średnie wartości funkcji przystosowania w kolejnych pokoleniach populacji prognoz inflacji i IRDN

Źródło: opracowanie własne

Zwróćmy teraz uwagę na drugi wykres na rzeczonym rysunku a konkretnie na wyniki uzyskane w segmentach pierwszym i drugim. Zachowanie szeregu czasowego w okresach je obejmujących można chyba określić jako najbliższe podręcznikowemu. Wyraźna siedmiodniowa sezonowość, wahania losowe nieprzejawiające skłonności do nietypowych wartości (w odróżnieniu od dalszych segmentów szeregu) – wszystko to spowodowało, że przystosowanie chromosomów podlegało bardzo podobnym zmianom. Średnie

tempo zmian kształtowało się na poziomie spadku rzędu 0,3% z pokolenia na pokolenie. Dla porównania, w najbardziej dynamicznie zmieniającym się segmencie piątym miernik ów wyniósł 0,61%. W pierwszym oraz w drugim segmencie nie obserwujemy wzrostu zmienności wartości funkcji przystosowania pod koniec obliczeń w odróżnieniu od pozostałych trzech segmentów.

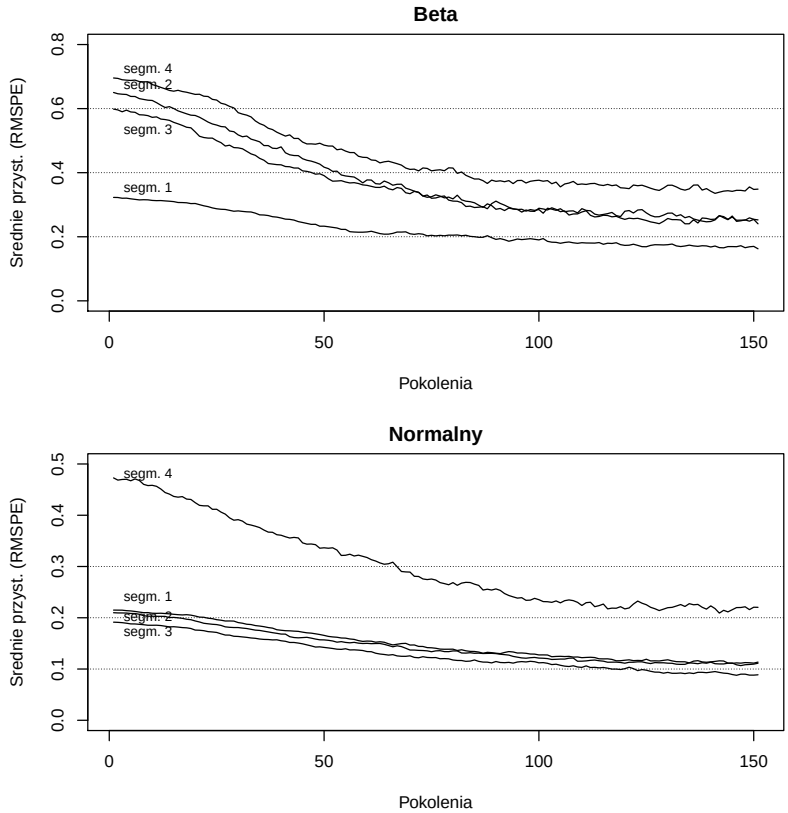
Na zakończenie rozważań na temat średniego przystosowania chromosomów w kolejnych generacjach zajmiemy się wynikami osiągniętymi dla szeregów wygenerowanych na podstawie założonych rozkładów zmiennych losowych. Wykresy na rysunku 4.29 przedstawiają zmiany jakie zachodziły w ramach całych populacji zachowując podział na segmenty. Różnice przystosowania chromosomów wynikają ze zmienności wylosowanych wartości, które składały się na segmenty. Należy przez to rozumieć, że ponowne użycie generatora liczb pseudolosowych dałoby inne rezultaty. Jednakże po zmienne losowe sięgnęliśmy, aby przetestować działanie algorytmu w kontrolowanym środowisku.

Rozkład beta charakteryzował się znacznie wyższą zmiennością niż rozkład normalny. Pomijając rząd wielkości błędów prognoz *ex post*, RMSPE spadało z pokolenia na pokolenie w średnim tempie od 0,46% dla segmentów pierwszego i czwartego do 0,66% dla segmentu drugiego. Współczynnik zmienności w tym segmencie wyniósł 34% wobec 22% dla segmentu 1. Same różnice między danymi a prognozami nie wpłynęły na szybkość poprawy średniego przystosowania chromosomów skoro segment pierwszy (posiadający najniższy RMSPE) i czwarty (o najwyższym) zachowywały się pod tym względem podobnie.

W przypadku rozkładu normalnego segmenty od 1 do 3 charakteryzowały się zbliżoną jakością prognoz *ex post* zaś ostatni segment odstawał pod tym względem na minus. Jednak jeśli porównamy na przykład chromosomy w czwartym i w trzecim segmencie wówczas okaże się, że średnie tempo spadku przeciętnego przystosowania było w ich wypadku takie samo i równało się 0,51%. W segmentach pierwszym i drugim wyniosło ono odpowiednio 0,43% oraz 0,42%.

Nasuwa się zatem wniosek, że o ile zmienność szeregów rozkładów wpływała na szybkość i sposób poszukiwania najlepszego chromosomu w populacji to w przypadku średniego przystosowania chromosomu populacji takiego wpływu trudno się dopatrzyć.

Uważne przeanalizowanie wykresów przedstawiających zmiany przeciętnego przystosowania chromosomów pozwala zauważyć, że w przeważającej części przypadków pod koniec postępowania zwiększa się zróżnicowanie średniego kryterium całej populacji przy jednoczesnym spadku tempa poprawy wyników. Tak algorytm reaguje na zmodyfikowane mechanizmy krzyżowania i mutacji, w których w miarę tworzenia kolejnych generacji zmienia się nacisk selektywny. W początkowych pokoleniach, kiedy to zróżnicowanie osobników



Rysunek 4.29: Średnie wartości funkcji przystosowania w kolejnych pokoleniach populacji prognoz rozkładów: beta i normalnego

Źródło: opracowanie własne

jest większe prawdopodobieństwa krzyżowania i mutacji są niższe. W dalszych generacjach przybywa jednak podobnych osobników tym bardziej, że w nasz algorytm wbudowana została selekcja elitarna, bardzo przydatna na początku, ale później mogąca doprowadzić do uporczywego trzymania się tego samego obszaru zbioru rozwiązań. Dlatego rzeczone prawdopodobieństwa stopniowo są coraz wyższe co zapobiega zdominowaniu populacji przez dużą grupę identycznych chromosomów. Uboczny efekt podwyższenia tych parametrów stanowi pojawienie się osobników odbiegających od średniej populacji. Stąd większe zróżnicowanie.

Aspektem wiążącym średnie przystosowanie chromosomów w ramach populacji ze zmianą przystosowania najlepszego znalezionego osobnika jest kwestia polepszenia wyników pomiędzy punktem startowym a końcowym mierzona różnicą dopasowania chromosomów populacji końcowej i począt-

kowej. Wartości genów chromosomów wyjściowej populacji generowane są wprawdzie losowo, ale przyjęty mechanizm prognozowania powoduje, że startowa populacja składa się z osobników dopuszczalnych. Istnieje zatem powód, aby porównać punkt wyjścia i efekt końcowy pracy algorytmu tym bardziej, że wyposażony on został w mechanizm preselekcji. Jak wykazane zostało w pracy (Kucharski 2008*b*) mechanizm ten w istotny sposób wpływa na poprawę jakości prognoz *ex post* otrzymywanych za pomocą PAG.

Jako ilustracji zmian zachodzących między populacją początkową a finalną użyjemy danych uzyskanych dla dwóch szeregów: inflacji oraz wolumenu obrotów akcji TVN. Wybraliśmy akurat te szeregi ponieważ w pierwszym z nich wystąpiły najniższe zaś w drugim najwyższe błędy prognoz *ex post* mierzone ważoną wartością RMSPE. Pozostałe zmienne znalazły się pomiędzy granicami wytyczonymi przez wymienione dwa przypadki.

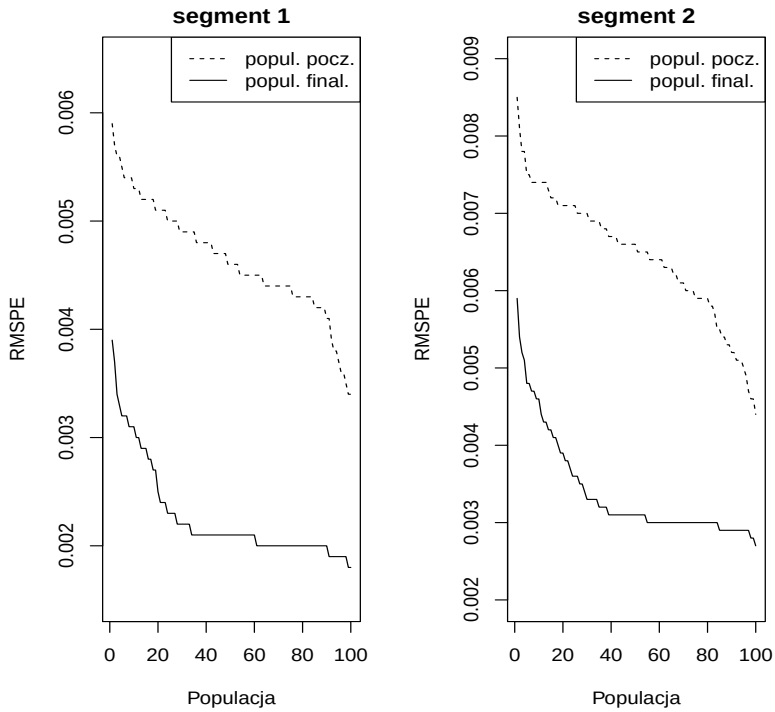
Rysunek 4.30 ilustruje wspomnianą różnicę dla obu segmentów prognozowania kaskadowego 1, ale sytuacja na nim przedstawiona jest typowa dla postępowania jako takiego. Na wykresach znalazły się wartości funkcji przystosowania wszystkich chromosomów obu populacji, które zostały uporządkowane malejąco oddzielnie dla każdej z populacji.

Dolna łamana zawiera niższe, a zatem lepsze wartości kryterium. Udało się osiągnąć znaczącą poprawę rezultatów w porównaniu z punktem wyjścia. Średnia różnic funkcji przystosowania pomiędzy populacjami pierwszą i ostatnią osiągnęła w przypadku pierwszego segmentu 0,0024 zaś drugiego 0,0031. Biorąc pod uwagę rząd wielkości błędu prognozy występujący w tabeli 4.38 różnice te należy uznać za istotne. Za skalą poprawy wyników stoi właśnie mechanizm preselekcji Cavichio na co zresztą wskazywały badania opisane we wspomnianym wcześniej artykule.

Prognozy *ex post* szeregu czasowego inflacji zaliczały się do obarczonych najmniejszymi błędami. Na drugim końcu skali znalazł się wolumen obrotów TVN, których predykcja ze względu na bardzo dużą dynamikę zmian danych charakteryzowała się o wiele wyższymi błędami.

Zachowanie obserwacji odbiło się na różnicach między populacją początkową a końcową. Średnia tych różnic wahała się od 0,546 dla segmentu 2 do aż 2,4029 w segmencie 1. Jeśli zestawić te wyniki z najlepszym ważonym RMSPE z tabeli 4.18 wówczas zmiana, która dokonuje się między początkową a finalną populacją staje się jeszcze bardziej wyraźna niż miało to miejsce w przypadku wskaźnika cen.

Porównanie pierwszej i ostatniej populacji stanowi klamrę spinającą spostrzeżenia, które poczyniliśmy analizując zmiany zachodzące w wartości przystosowania najlepszego chromosomu z uwagami poczynionymi dla średniego przystosowania populacji. Dolne łamane na wykresach z rysunku 4.30 wskazują na obecność większej ilości osobników o podobnym przystosowaniu, które wyróżniają się ze względu na lepsze przystosowanie. Różnice



Rysunek 4.30: Różnice przystosowania populacji końcowej i początkowej prognoz CPI

Źródło: opracowanie własne

między początkiem a końcem pracy algorytmu są z kolei dobrym uzupełnieniem średniego tempa zmian przeciętnego przystosowania chromosomów populacji gdyż pozwalają ocenić jak duży obszar zbioru dopuszczalnych rozwiązań został przeszukany.

4.3. Dobór parametrów PAG

Po zapoznaniu się z wynikami działania prognostycznego algorytmu genetycznego we wcześniejszych częściach tego rozdziału czas przejść do sformułowania bardziej ogólnych wniosków na temat wpływu warunków narzuconych przez dekompozycję szeregu na funkcjonowanie PAG. Pomoże to określić parametry metody w jej potencjalnych zastosowaniach. Wprawdzie nasz algorytm posiada w tym względzie pewną autonomię, ale właściwe określenie punktu startowego ułatwi i przyspieszy znalezienie rozwiązania najbardziej korzystnego z punktu widzenia przyjętego kryterium.

Parametry sterujące algorytmem funkcjonują na różnych jego poziomach. Najniżej działają te, które wpływają na pojedynczy chromosom. Za-

liczymy do nich operatory selekcji, krzyżowania i mutacji. Same operatory są niezmienne w ramach PAG, ale przed rozpoczęciem obliczeń decydujemy o wartościach prawdopodobieństw krzyżowania i mutacji.

Mutacja powoduje zmianę pojedynczego genu chromosomu na inny zawierający się w przedziale $\langle 1, t_m \rangle$ z pewnym ustalonym prawdopodobieństwem. Przyjęliśmy, że prawdopodobieństwo uruchomienia tego operatora zależy od numeru pokolenia i w związku z tym zawiera się w pewnym wąskim przedziale, którego krańce ustalamy przed rozpoczęciem obliczeń. Jak dowiodły wcześniejsze badania (Kucharski 2007), poważne zmiany prawdopodobieństwa mutacji wpływają wprawdzie na wyniki końcowe, ale nie do tego stopnia by radykalnie je zmienić. Trzeba jednak przyznać, że maksymalna wartość 0,1 (taka obowiązywała w trakcie obliczeń) jak na operator mutacji jest dość wysoka.

Inaczej sprawa miała się z prawdopodobieństwem krzyżowania. Operator krzyżowania bardzo istotnie wpływa na populację poddaną działaniu prognostycznego algorytmu genetycznego. Razem z selekcją decydują o kierunku i szybkości poszukiwania najbardziej korzystnych wartości funkcji przystosowania. Niskie prawdopodobieństwo krzyżowania zmniejsza liczbę chromosomów poddanych działaniu operatora i pomaga przyspieszyć obliczenia. Wysokie z kolei zwiększa różnorodność osobników tworzących populację potomków choć czas obliczeń przy tej okazji ulega wydłużeniu.

Przeprowadzając obliczenia korzystaliśmy z dwóch wyraźnie różniących się poziomów prawdopodobieństwa krzyżowania co pozwoliło zauważyć zależność między zachowaniem szeregu a rezultatami odpowiadającymi danej wartości parametru. W segmentach szeregów o małej zmienności bardziej opłaca się korzystać z niższych prawdopodobieństw krzyżowania. Po pierwsze RMSPE dla $p_k \in \langle 0, 0{,}3 \rangle$ zwykle okazywało się niższe. Po drugie czas obliczeń był wyraźnie krótszy co wynikało z mniejszej ilości operacji krzyżowania chromosomów.

Sytuacja ulegała zmianie gdy zróżnicowanie szeregu rosło, na przykład poprzez pojawianie się grup obserwacji nietypowych lub częste przechodzenie danych od niskich do wysokich wartości. Wyższe prawdopodobieństwo krzyżowania (zmieniające się w przedziale $\langle 0, 0{,}6 \rangle$) sprawiało, że błąd prognoz *ex post* stawał się niższy niż uzyskiwany przy $p_k \in \langle 0, 0{,}3 \rangle$ dla tych samych szeregów. Więcej chromosomów potomków zwiększa szansę znalezienia lepiej przystosowanych osobników wśród bardziej różnorodnej populacji. W szeregu o dużej zmienności rodzice i potomkowie mogą silniej różnić się wartością kryterium. Prawdopodobne zatem jest, że potomek, który mocno różni się od rodzica będzie dawał lepsze prognozy.

Większa różnorodność populacji zapewniona przez wyższe p_k podnosi w PAG prawdopodobieństwo znalezienia się bliżej rozwiązania optymalnego ponieważ w trakcie obliczeń sprawdzana jest większa liczba potencjalnych

zestawów prognoz. Oczywiście drugą stroną medalu jest wzrost czasu potrzebnego na obliczenia.

Kolejnym parametrem wpływającym na funkcjonowanie PAG jest maksymalny dopuszczalny poziom opóźnienia obserwacji będącej prognozą na dany okres, który w pracy oznaczyliśmy symbolem „t_m". Działa on na wyższym poziomie funkcjonowania algorytmu niż prawdopodobieństwa krzyżowania i mutacji ponieważ obejmuje swym zasięgiem całą populację, a nie tylko wylosowane chromosomy lub geny.

Z punktu widzenia samego algorytmu genetycznego t_m determinuje skład chromosomu i zakres jego zmian podczas działania operatorów krzyżowania i mutacji. Od strony merytorycznej t_m odpowiada (choć nie jako jedyny czynnik) za jakość prognoz *ex post* gdyż wyznacza granice przedziału okresów, z którego korzysta algorytm. Z tego powodu jest ściśle powiązany z dekompozycją szeregu czasowego.

Części szeregów składające się na segmenty opisane wcześniej w rozdziale charakteryzowały się bardzo różnorodnym zachowaniem w czasie. Możemy jednak podjąć próbę pogrupowania ich w celu wyciągnięcia ogólnych wniosków co do doboru wartości t_m.

Obecność jedynie małych wahań losowych w szeregach z trendem lub o stałym poziomie zmiennej ułatwia algorytmowi znalezienie rozwiązania. Przykład takiego zachowania obserwujemy w pierwszym segmencie szeregu notowań WIG lub drugim segmencie kursu TVN. Wiemy już, że prawdopodobieństwo krzyżowania nie musi być wtedy wysokie, ale niskie wartości może przyjmować również parametr t_m. Skraca się wówczas czas obliczeń. Ponadto błędy prognoz *ex post* są bardzo niskie (poniżej 5%), obserwujemy też rzadsze zmiany funkcji przystosowania najlepszych chromosomów pochodzących z kolejnych populacji.

Poprawa przystosowania najlepszego chromosomu następuje rzadko. Kolejne etapy dzieli przynajmniej kilku pokoleń (z uwagi na selekcję elitarną najlepszy osobnik z pokolenia poprzedniego jest kopiowany do bieżącego co pozwala prześledzić proces polepszania rozwiązania). Świadczy to o tym, że algorytm szybko dociera w pobliże optimum globalnego. Przyjęcie $t_m = 5$ wystarczało w tej sytuacji w zupełności.

Pojawienie się większych wahań przypadkowych, takich jak na przykład w pierwszym segmencie notowań kursu euro, nie zwiększa błędów *ex post*, ale powoduje podniesienie zmienności przystosowania najlepszych chromosomów z kolejnych pokoleń o około 2-3 p% (mierzone współczynnikiem zmienności). Średnie przystosowanie osobników w całej populacji odznacza się zróżnicowaniem podobnym do szeregów z małymi wahaniami.

Postęp w przystosowaniu (na podstawie wartości funkcji dla najlepszego chromosomu w danym pokoleniu) następuje stopniowo, nie rzadziej niż co dwa pokolenia. Pod koniec obliczeń (rozumiemy przez to przedział od 100

do 150 pokolenia) wartość RMSPE stabilizuje się i nie ulega już dużym zmianom. Przy takiej dekompozycji zastosowanie wciąż znajdzie algorytm z niskim t_m (u nas nie przekraczającym 5).

Oczywiście zwrot „niski t_m" można różnie rozumieć. Prognozując kurs indeksu rynku dnia następnego na Towarowej Giełdzie Energii musieliśmy wziąć pod uwagę obecność wahań sezonowych, szczególnie dobrze widoczną w pierwszych dwóch segmentach. Nawiązując do uwag z paragrafu 3.1.2. okres wahań sezonowych musiał zawierać się w granicach przedziału wyznaczonego przez wartość maksymalnego opóźnienia obserwacji będącej prognozą *ex post* na daną sesję. Dlatego niski t_m w tym wypadku równał się 7 a pomimo tego ważony błąd predykcji oscylował w okolicach zaledwie 5%.

Warto wspomnieć o zaobserwowanej w pierwszym segmencie notowań franka szwajcarskiego a trwającej kilka okresów zmianie trajektorii wartości zmiennej. Przez kilka kolejnych sesji kurs tej waluty był wyraźnie wyższy od kursu w reszcie segmentu. Zaowocowało to wzrostem zróżnicowania średniego przystosowania populacji do 23% co dowodzi wyższej różnorodności składających się na nie osobników. W ten sposób algorytm zasygnalizował dostosowanie do nowej sytuacji – część chromosomów uwzględniła szczególne zachowanie niektórych obserwacji podczas gdy część tego nie uczyniła. Poprawa wartości funkcji przystosowania następowała najczęściej w środkowych pokoleniach aż do momentu znalezienia chromosomu najlepiej dostosowanego do aktywności szeregu czasowego. Również w tym przypadku t_m można przyjąć na niższym poziomie.

Pojawienie się dużych wahań losowych należy rozpatrywać w dwóch aspektach: gdy są to pojedyncze obserwacje oddzielone od siebie co najmniej kilkoma okresami o wyraźnie niższych wartościach oraz gdy grupują się one w serie. Z pierwszą z wymienionych sytuacji zetknęliśmy się na przykład w drugim segmencie notowań franka szwajcarskiego. Pomimo obecności wysokich co do wartości kursów błąd prognoz *ex post* wciąż pozostawał na niewysokim poziomie choć wzrosła zmienność średniego przystosowania chromosomów populacji.

Specyfika PAG polega na typowaniu w szeregu obserwacji zapewniającej niski błąd predykcji. Dlatego algorytm preferuje geny, które do prognoz na okresy o niższych wahaniach przypadkowych zawierają (w granicach przedziału określonego parametrem t_m) dane o zbliżonym zachowaniu. Innymi słowy pojedyncza odstająca obserwacja zostanie pominięta w procesie tworzenia prognoz dla okresów mocno różniących się od niej wartością.

Odstająca obserwacja zostaje objęta działaniem PAG, który także dla tego okresu wyznaczyć musi prognozę. Generalnie parametr t_m powinien zawierać tyle okresów, aby dać algorytmowi możliwość wyboru spośród danych o odpowiednim zakresie wahań. Jako ilustracja niech posłuży nam

sytuacja w drugim segmencie notowań franka szwajcarskiego[11]. Występujące w nim duże wahania przypadkowe (mogące *de facto* uchodzić za punkty zwrotne) rozdzielone są kilkoma sesjami o mniejszych wahaniach. Przedział czasu określony parametrem t_m powinien objąć swoim zasięgiem wahania o różnej amplitudzie. W tym konkretnym przypadku wystarczyła wartość równa 5 choć i dla $t_m = 10$ prognozy *ex post* też miały dobrą jakość. Należy zwrócić uwagę na to, że wyższy poziom t_m lepiej będzie sprawdzać się przy dłuższych seriach dużych wahań losowych. Szerzej aspekt ten zostanie omówiony poniżej.

Zmiana wartości funkcji przystosowania najlepszego chromosomu zależy od częstotliwości pojawiania się dużych wahań. Jeżeli nie powtarzają się one często w obrębie segmentu, dotychczasowe rozwiązanie poprawiane jest rzadziej. Gdy zaczynają grupować się w serie dotychczasowy najlepiej przystosowany osobnik będzie zastępowany znacznie częściej.

Sprawy zmienności danych i jej wpływu na jakość prognoz otrzymywanych za pomocą PAG dotyka również zachowanie obserwowane w szeregach wolumenu obrotów wybranych przez nas walorów notowanych na warszawskiej giełdzie oraz obu zmiennych losowych. Na przykład praktycznie każdy z trzech wyodrębnionych segmentów szeregu dla WIG składa się z dużych wahań losowych. Niskie i wysokie wolumeny obrotów występują w następujących po sobie sesjach. Do tego amplituda wahań potrafi wykładniczo rosnąć lub maleć w kolejnych okresach.

Na podstawie porównania uzyskanych wyników dla możemy wysnuć wniosek, że przy dużej i bardzo dużej zmienności w połączeniu z wysoką częstotliwością zmian lepsze rezultaty prognozowania uda się osiągnąć przy wyższym t_m – w tym konkretnym przypadku równym 10.

Zmieniające się w czasie[12] zachowanie danych, na które nakładają się skokowe zmiany amplitudy wahań bardzo utrudniają prognozowanie nawet z wysoką (w porównaniu do liczby obserwacji składających się na segment) wartością t_m. Algorytm ma wówczas tendencję do wybierania PAG w wersji 1 czyli uogólnienia metody naiwnej prostej.

Ciekawostką jest fakt, że pomimo wysokiego względnego błędu średniokwadratowego zmienności najlepszych chromosomów w pokoleniach oraz średniego przystosowania nie przekraczają zwykle 10 punktów procentowych choć różnice rozproszenia samych danych są kilkakrotnie wyższe. Poprawa przystosowania najlepszego chromosomu dokonuje się jednak częściej i trwa nawet do ostatniego pokolenia. Aby radykalnie poprawić wyniki nie wystarczy wysokie t_m, ale należy kilkukrotnie zwiększyć liczbę generacji. Ogólny

[11] Podobną analizę Czytelnik może przeprowadzić również w czwartym segmencie szeregu notowań IRDN.

[12] Na przykład przechodzenie od trendu do stałego poziomu.

wzrost ilości koniecznych operacji i czasu obliczeń czyni takie wyjście mało przydatnym z praktycznego punktu widzenia.

Kolejnym elementem dekompozycji związanym z parametrem t_m są wahania sezonowe i quasi-sezonowe. Te ostatnie pojawiły się w ostatnim segmencie szeregów kursów walu). Z technicznego punktu widzenia wahania sezonowe nie są niczym innym jak dużymi wahaniami losowymi, które powtarzają się w regularnych odstępach czasu. Tak też traktuje je PAG co wzięliśmy pod uwagę dobierając wartość t_m w segmentach szeregu IRDN. Okres wahań powinien zawierać się w przedziale czasu objętym parametrem o czym wspominaliśmy już w rozdziale trzecim.

Jeżeli szereg czasowy zachowuje się w sposób nazwany przez nas quasi-sezonowym ustalenie poziomu t_m staje się trudniejsze. Okres wahań nie jest bowiem stały. Mogą też pojawić się krótkie podokresy z przejściowo występującym trendem jak w szeregu kursu euro w trzecim segmencie. Jeśli cały nieregularny cykl obejmuje większą część długości szeregu istnieje jeszcze inne wyjście. Parametr t_m należy przyjąć tak, żeby zawarł w sobie co najmniej dwie kolejne fazy całego cyklu na przykład fazę wzrostu do momentu przesilenia i następujący po nim spadek. Być może błędy prognoz nieco wzrosną, ale pozostaną wciąż na przyzwoitym poziomie.

W trakcie badań przyjmowaliśmy jedną wartość t_m dla wszystkich segmentów. Nic nie stoi na przeszkodzie ustalenia jej oddzielnie dla każdego podokresu, ale nie jest to wymagane. PAG wybiera wartość z przedziału $\langle 1, t_m \rangle$ co zapewnia mu wystarczającą elastyczność we wszystkich segmentach[13]. To korzystne z punktu widzenia nadzorowania pracy algorytmu gdyż spada liczba parametrów, które należy ustalić na początku postępowania.

Przenosząc się na wyższy poziom sterowania stajemy przed wyborem wersji PAG (jednej z czterech) dla każdego z wyodrębnionych segmentów. Algorytm posiada mechanizm doboru najlepszego zestawu metod i właściwie ten etap można by pominąć przyjmując jako startowy dowolny wektor wersji prognostycznego algorytmu genetycznego. Korzystniej jednak będzie ustalić na początku postępowania metody adekwatne do dekompozycji podokresów składających się na szereg czasowy. W ten sposób algorytm znajdzie się w pobliżu optymalnego zbioru metod prognostycznych. Zaimplementowanie strategii ewolucyjnej pomaga poprawić początkowy wybór jeśli nie był on do końca trafiony. Ma to szczególnie istotne znaczenie przy wyższych poziomach t_m, kiedy zbiór rozwiązań gwałtowanie rośnie. Zawężamy wówczas poszukiwania do najbardziej obiecującego podzbioru możliwych wariantów prognostycznych.

Na najwyższym poziomie sterowania algorytmem znajduje się wybór rodzaju prognozowania kaskadowego. Jeśli przyjrzymy się wynikom prognoz

[13] Wariant PAG ze średnią ruchomą wymaga ustalenia dolnej i górnej wartości t_m

ex post pod tym kątem trudno będzie jednoznacznie wskazać zwycięzcę. Przewaga kaskady 2 polega na tym, że dopasowując metody do segmentów bierze pod uwagę jakość tworzonych przez owe metody prognoz wygasłych dla luk między segmentami. Tym samym prognoza *ex post* dobrej jakości otrzymana dzięki prognozowaniu kaskadowemu drugiego typu daje większe szanse, że prognozy *ex ante* również okażą się trafne.

Prognozowanie kaskadowe 2 dobrze sprawdza się tam, gdzie do głosu dochodzą serie dużych wahań losowych. O dokładności prognoz wygasłych w takiej sytuacji była mowa wcześniej. Kaskada 2 wskazuje przy okazji wyraźniej, które wartości t_m z zakładanego zakresu są silniej preferowane przez algorytm. W przypadku kaskady 1 częstości występowania poszczególnych wartości t_m są bardziej wyrównane.

Kaskadę pierwszą od drugiej odróżnia fakt, że luki między segmentami wypełniają rzeczywiste elementy szeregu czasowego zamiast prognoz. Z tego względu kaskada 1 jest w stanie posłużyć za punkt odniesienia dla drugiego wariantu. Innymi słowy, można ją użyć do kalibrowania parametrów prognozowania kaskadowego 2.

Prognozując zmienną najpierw sięgamy po kaskadę 1. Następnie, przy tych samych parametrach początkowych uruchamiamy kaskadę 2 i porównujemy wyniki. Dążymy do tego, aby efekty jednej i drugiej metody predykcji charakteryzowały się podobną trafnością. Wówczas prognozy *ex ante* wykonujemy na podstawie kaskady drugiego typu[14].

Wszelako analiza wartości RMSPE prognoz wygasłych wskazuje, że nie należy rezygnować z prognozowania kaskadowego pierwszego typu ponieważ dla części szeregów wypadło ono lepiej. Sęk w tym, że opinię tą wygłaszamy gdyż dysponowaliśmy danymi, które pozwoliły prognozy wygasłe wyznaczyć i dokonać porównania. W praktyce nie wiemy z góry, która z metod okaże się lepsza. W wypadku wątpliwości co do oczekiwanej trafności prognoz sugerujemy zatem, aby prognozy *ex ante* wyznaczyć obiema metodami po czym wyniki uśrednić.

[14] Oczywiście prognozowanie kaskadowe drugiego typu od początku może sprawdzać się lepiej. Odnotowaliśmy takie przypadki.

Zakończenie

Do celów prognozowania szeregów czasowych opracowano wiele metod. Z upływem czasu stają się one jednak coraz bardziej wyspecjalizowane. Pojawienie się anomalii naruszającej któreś z założeń potrafi w skrajnym wypadku uniemożliwić wykorzystanie danego narzędzia. Oczywiście możemy użyć kolejnej metody lub zmodyfikować już istniejącą tak, aby odpowiadała nowym warunkom czyniąc ją bardziej wyrafinowaną. Coraz trudniej osobie nie będącej ekspertem poruszać się w obrębie metodologii prognozowania ze względu na prawdziwy gąszcz proponowanych przez autorów publikacji pomysłów oraz modyfikacji. Dlatego chcieliśmy przynajmniej częściowo uprościć proces tworzenia prognoz zdejmując z barków prognosty niektóre obowiązki. Użytkownik PAG musi jedynie podać kilka parametrów nawet bez znajomości mechaniki obliczeniowej stojącej za wyznaczaniem prognoz.

Połączenie algorytmu genetycznego (to skrót myślowy bo *de facto* chodzi o hybrydę łączącą cechy algorytmu genetycznego i ewolucyjnego) z prognozowaniem jest wykonalne czego dowodzi nie tylko niniejsza książka, ale cytowane wcześniej prace. Udowodniliśmy ponadto, że prognozowanie niestrukturalne wciąż posiada obszary warte zbadania i dające obiecujące wyniki. Choć przykłady obliczeń skupiały się na szeregach rodem z rynków finansowych, prognostyczny algorytm genetyczny znalazłby zastosowanie podczas predykcji innych kategorii ekonomicznych. Po prostu zmienne wywodzące się z giełdy i rynku walut reprezentowały zachowanie w czasie, na którym najbardziej nam zależało.

Punktem wyjścia był proces decyzyjny, który miał odpowiedzieć na pytanie: którą prognozę wybrać? Sięgnięcie po optymalizację heurystyczną zapewniło elastyczność charakterystyczną dla tej grupy metod oraz nieco niezależności od subiektywizmu prognosty.

Algorytm genetyczny powstał jako jeden ze sposobów naśladowania procesów obserwowanych w środowisku naturalnym. Zachowanie konsumentów lub inwestorów na rynku także podlega pewnym ogólnym prawom. Algorytm próbuje owe zachowania modelować i wspomagać podejmowanie decyzji.

W prognostycznym algorytmie genetycznym kluczowe staje się podejście do dekompozycji szeregu czasowego. PAG pozwala ograniczyć ją do oceny wahań losowych (duże czy małe, grupujące się w serie czy nie) oraz

podzielić szereg na jednorodne pod względem zachowania segmenty. Dzięki temu powstają prognozy nie wymagające żmudnego poszukiwania czynników, które miałyby wywierać wpływ na zmienną. Ma to tym większe znaczenie, że PAG dedykowany jest szeregom źle uwarunkowanym, które wymykają się predykcji klasycznymi metodami. Poza tym wprowadzenie podziału na segmenty pozwala włączyć do prognozowania niestrukturalnego długookresowe zależności dzięki prognozowaniu kaskadowemu i statystyce bayesowskiej. Wprowadzenie analizy bayesowskiej uwypukliło wpływ mechanizmów odpowiadających za kształtowanie się danych. Korzyści z tego postępowania najlepiej widać w jakości prognoz wygasłych.

We wstępie wspomnieliśmy o próbie przerzucenia pomostu między prognozowaniem niestrukturalnym a optymalizacją heurystyczną. Żywimy przekonanie, że takie połączenie zostało osiągnięte z sukcesem. Algorytm dowiódł swojej elastyczności i przydatności szczególnie tam, gdzie zawodzą inne metody.

Biorąc pod uwagę ciągle niezbyt obfitą literaturę na temat wykorzystania algorytmów genetycznych do predykcji należy zauważyć, że praca na tym polu nie jest jeszcze ukończona. Wciąż pozostaje miejsce na dalsze badania gdyż trzymana w ręku książka z pewnością nie wyczerpała tematu.

Bibliografia

Aczel A. D. (2000), *Statystyka w zarządzaniu*, PWN, Warszawa.

Adya M., Collopy F., Armstrong J. S., Kennedy M. (2012), *Automatic identification of time series features for rule-based forecasting*, *Omega 40*, 748–757.

Agustín-Blas L. E., Salcedo-Sanz S., Jiménez-Fernández S., Carro-Calvo L., Del Ser J., Portilla-Figueras J. A. (2012), *A new grouping genetic algorithm for clustering problems*, *Expert Systems with Applications 39*, 9695–9703.

Aiello G., La Scalia G., Enea M. (2012), *A multi objective genetic algorithm for the facility layout problem based upon slicing structure encoding*, *Expert Systems with Applications 39*, 10352–10358.

Allen F., Karjalainen R. (1999), *Using genetic algorithms to find technical trading rules*, *Journal of Financial Economics 13*, 245–271.

Arabas J. (2004), *Wykłady z algorytmów genetycznych*, wyd. 2, WNT, Warszawa.

Arroyo J., Maté C. (2009), *Forecasting histogram time series with k-nearest neighbours method*, *International Journal of Forecasting 25*, 192–207.

Bai J., Ng S. (2008), *Forecasting economic time series using targeted predictors*, *Journal of Econometrics 146*, 304–317.

Błaszczuk D. (2006), *Wstęp do prognozowania i symulacji*, wyd. 2, PWN, Warszawa.

Beenstock M., Szpiro G. (2001), *Specification search in nonlinear time-series models using the genetic algorithm*, *Journal of Economic Dynamic & Control 26*, 811–835.

Bérard J. (2005), *Genetic algorithms in random environments: two examples*, *Probab. Theory Relat. Fields 133*, 123–140.

Biecek P. (2008), *Przewodnik po pakiecie R*, wyd. 1, Oficyna Wydawnica GiS, Wrocław.

Borgulya I. (2005), *A multi-objective evolutionary algorithm with a separative archive*, *Central European Journal of Operations Research 13*, 233–253.

Bunn D. W., Vassilopoulos A. I. (1999), *Comparison of seasonal estimation methods in multi-item short-term forecasting*, *International Journal of Forecasting 15*, 431–443.

Buszkowska E. (2008), *Porównywanie zdolności prognostycznej modeli zmienności indeksu wig20 za pomocą testu spa*, w *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, nr 10 w *Rynek kapitałowy, skuteczne inwestowanie*, Uniwersytet Szczeciński, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin, 306–319.

Cieślak M. (1997), *Prognozowanie gospodarcze. Metody i zastosowania*, PWN, Warszawa.

Cleveland R. B., Cleveland W. S., McRae J. E., Terpenning I. (1990), *Stl: A seasonal-trend decomposition procedure based on loess*, *Journal of Official Statistics 6*(1), 3–73.

Corberán-Vallet A., Bermúdez J. D., Vercher E. (2011), *Forecasting correlated time series with exponential smoothing models*, *International Journal of Forecasting 27*, 252–265.

Cytowski J. (1996), *Algorytmy genetyczne. Podstawy i zastosowania*, Akademicka Oficyna Wydawnicza PLJ, Warszawa.

Dittman P. (2004), *Prognozowanie w przedsiębiorstwie*, wyd. 2, Oficyna Ekonomiczna, Kraków.

Dong R., Pedrycz W. (2008), *A granular time series approach to long-term forecasting and trend forecasting*, *Physica A 387*, 3253–3270.

Ducange P., Lazzerini B., Marcelloni F. (2010), *Multi-objective genetic fuzzy classifiers for imbalanced and cost-sensitive datasets*, *Soft Comput 14*, 713–728.

Easton F., Mansour N. (1999), *A distributed genetic algorithm for deterministic and stochastic labor scheduling problems*, *European Journal of Operational Research 118*, 505–523.

Falkenauer E. (1992), *The grouping genetic algorithm – widening the scope of the gas*, *Proceedings of the Belgian Journal of operations research, statistics and computer science 33*, 79–102.

Falkenauer E. (1998), *Genetic algorithms for grouping problems*, Wiley, New York.

Fok D., van Dijk D., Franses P. H. (2005), *Forecasing aggregates using panels of nonlinear time series*, *International Journal of Forecasting 21*, 785–794.

Fonseca C. M., Fleming P. J. (1995), *An overview of evolutionary algorithms in multiobjective optimizations*, *Evolutionary Computation 3*, 165–180.

Friedman D. (1998), *On economic applications of evolutionary game theory*, *Journal of Evolutionary Economics 8*, 15–43.

Gajda J. B. (2001), *Prognozowanie i symulacja a decyzje gospodarcze*, Wydawnictwo C.H. Beck, Warszawa.

Gajda J. B. (2004), *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa.

Gnat S. (2008), *Badanie wpływu szerokości aproksymanty segmentowej na błędy prognoz ex post w prognozach z wykorzystaniem trendu pełazjącego z wagami harmonicznymi*, w J. B. Gajda i A. Kucharski, (red.), *Badania operacyjne. Metody i zastosowania*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 107–115.

Goldberg D. E. (2003), *Algorytmy genetyczne i ich zastosowania*, wyd. 3, WNT, Warszawa.

Gong D., Yamazaki G., Gen M., Xu W. (1999), *A genetic algorithm method for one-dimensional machine location problems*, *International Journal Production Economics 60–61*, 337–342.

Goryl A., Jędrzejczyk Z., Kukuła K., Osiewalski J. (1996), *Wprowadzenie do ekonometrii w przykładach i zadaniach*, PWN, Warszawa.

Goud P. G., Koehler A. B., Ord J. K., Snyder R. D., Hyndman R. J., Vahig-Araghi F. (2008), *Forecasting time series with multiple seasonal patterns*, *European Journal of Operational Research 191*, 207–222.

Gran J., Aral M. M. (1999), *Progresive genetic algorithm for solution of optimization problems with nonlinear equality and inequality constraints*, *Applied Mathematical Modelling 23*, 329–343.

Granger C. W. J., Jeon Y. (2003), *A time-distance for evaluating forecasting models*, *International Journal of Forecasting 19*, 199–215.

Gwiazda T. (1995), *Algorytmy genetyczne. Wstęp do teorii*, Biblioteka Sztucznej Inteligencji, Warszawa.

Hanne T. (1999), *On the convergence of multiobjective evolutionary algorithms*, *European Journal of Operational Research 117*, 553–564.

Hong T.-P., Huang K.-Y., Lin W.-Y. (2002), *Applying genetic algorithms to game search tree*, *Soft Computing 6*, 277–283.

Ignasiak E. (2001), *Badania operacjne*, wyd. 3, PWE, Warszawa.

Jajuga K., Jajuga T. (2001), *Inwestycje. Instrumenty finansowe, ryzyko finansowe, inżynieria finansowa*, PWN, Warszawa.

Jędrzejczyk Z., Kukuła K., Skrzypek J., Walkosz A. (2002), *Badania operacyjne w przykładach i zadaniach*, PWN, Warszawa.

Joines J. A., Houck C. R. (1994), *On the use of non-stationary penalty functions to solbe nonlinear constrained optimization problems with gas*, w (Michalewicz 2003), IEEE Service Center, Piscataway NJ.

Kang I.-B. (2003), *Multi-period forecasting using different models for different horizons: an application to us economic time series data*, *International Journal of Forecasting 19*, 387–400.

Koehler A. B., Snyder R. D., Ord J. K., Beaumont A. (2012), *A study of outliers in the exponential smoothing approach to forecasting*, *International Journal of Forecasting 28*, 477–484.

Koloch G. (2008), *O przeszukiwaniu przestrzeni modeli prognostycznych*, w J. B. Gajda i A. Kucharski, (red.), *Badania operacyjne. Metody i zastosowania*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź, 195–206.

Konarzewska I., Karwacki Z. (2005), *Planowanie i kontrola kosztów – wybrane problemy statystyczne*, w *Zarządzanie organizacjami w świetle wyzwań XXI wieku – od teorii do praktyki*, Wydawnictwo Naukowe Wyższej Szkoły Kupieckiej, Łódź, 445–459.

Kopczewska K., Kopczewski T., Wójcik P. (2009), *Metody ilościowe w R. Aplikacje ekonomiczne i finansowe*, CeDeWu, Warszawa.

Kucharski A. (2006), *Wykorzystanie algorytmów genetycznych do krótkookresowych prognoz na giełdzie papierów wartościowych*, w *Zeszyty Naukowe Uniwersytetu Szczecińskiego*, Vol. 462, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin, 135–145.

Kucharski A. (2007), *O pewnym zastosowaniu algorytmów genetycznych do prognozowania szeregów czasowych*, w *Prace Naukowe AE we Wrocławiu*, Vol. 1167, Wydawnictwo AE im. O. Langego we Wrocławiu, Wrocław, 143–153.

Kucharski A. (2008*a*), *Algorytmy genetyczne w prognozowaniu danych giełdowych - usuwanie obserwacji nietypowych*, *Badania Operacyjne i Decyzje 1*, 35–45.

Kucharski A. (2008*b*), *Efektywność algorytmu genetycznego jako narzędzia prognozowania szeregów czasowych*, *Badania Operacyjne i Decyzje 3*, 43–52.

Kucharski A. (2008*c*), *Krótkookresowe prognozy notowań - implementacja średniej ruchomej ze zmiennym efektem wygładzania w algorytmie genetycznym*, w *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, Vol. 9, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin, 663–670.

Kucharski A. (2008*d*), *Wpływ sezonowości w szeregu czasowym na prognozy ex post otrzymywane przy pomocy algorytmu genetycznego*, w H. Sobocka-Szczapa, (red.), *Przedsiębiorczość i Zarządzanie*, t. IX, Wydawnictwo SWSPiZ w Łodzi, Łódź, 99–107.

Kucharski A. (2008*e*), *Wykorzystanie zjawiska nisz i gatunków w prognozach na podstawie algorytmu genetycznego*, w D. Kopańska-Bródka, (red.), *Metody i zastosowania badań operacyjnych '07*, Prace Naukowe AE im. K. Adamieckiego w Katowicach, Katowice, 99–107.

Lawton R. (1998), *How should additive holt-winters estimates be corrected?*, *International Journal of Forecasting 14*, 393–403.

Lipowski A., Lipowska D. (2012), *Roullette-wheel selection via stochastic acceptance*, *Physica A 391*, 2193–2196.

Liu H.-T., Wei M.-L. (2010), *An improved fuzzy fuzzy forecasting method for seasonal time series*, *Expert Systems and Applications 37*, 6310–6318.

Locatelli M., Schoen F. (2003), *Efficient algorithms for large scale global optimization: Lennard-jones clusters*, *Computational Optimization and Applications 26*, 173–190.

Lukoseviciute K., Ragulskis M. (2010), *Evolutionary algorithms for the selection of time lags for time series forecasting by fuzzy inference systems*, *Neurocomputing 73*, 2077–2088.

Maddala G. S. (2008), *Ekonometria*, PWN, Warszawa.

Majcherczak J. (2004), *Analityczne metody krótkookresowego prognozowania wielkości popytu przedsiębiorstwa produkcyjnego*, w T. Trzaskalik, (red.), *Prace Naukowe AE im. K. Adamieckiego*, *Metody i zastosowania badań operacyjnych'04*, Akademia Ekonomicza im. K. Adamieckiego w Katowicach, Katowice, 223–230.

Mascheck M. K. (2010), *Intelligent mutation rate control in an economic application of genetic algorithms*, *Comput Econ 35*, 25–49.

Michalewicz Z. (2003), *Algorytmy genetyczne + struktury danych = programy ewolucyjne*, wyd. 3, WNT, Warszawa.

Miszczyńska D., Miszczyński M. (2002), *Wybrane metody badań operacyjnych*, Wyższa Szkoła Ekonomiczno-Humanistyczna w Skierniewicach, Skierniewice.

Montero E., Riff M.-C. (2011), *On-the fly calibrating strategies for evolutionary algorithms*, *Information Sciences 181*, 552–566.

Ong Bun T., Fukushima M. (2011), *Genetic algorithm with automatic termination and search space rotation*, *Memetic Comp. 3*, 111–127.

Ostasiewicz S., Rusnak Z., Siedlecka U. (2001), *Statystyka. Elementy teorii i zadania*, Wydawnictwo AE im. O. Langego we Wrocławiu, Wrocław.

Pasley A., Austin J. (2004), *Distribution forecasting of high frequency time series*, *Decision Support Systems 37*, 501–513.

Powell D., Skolnick M. M. (1993), *Using genetic algorithms in engineering design optimization with non-linear constraints*, w (Michalewicz 2003), Morgan Kaufman, San Mateo CA.

Ragulskis M., Lukoseviciute A., Navickas Z., Palivonaite R. (2011), *Short-term time series forecasting based on the identification of skeleton algebraic sequences*, *Neurocomputing 74*, 1735–1744.

Rechenberger I. (1973), *Evolutionsstrategie: Optimization technischer systeme nach prinzipien der biologischen evolution*, w (Michalewicz 2003), Frohmann-Holzberg, Stuttgart.

Remus W., O'Connor M., Griggs K. (1998), *The impact of information of unknown correctness on the judgmental forecasting process*, *International Journal of Forecasting 14*, 313–322.

Ren L., Glasure Y. (2009), *Applicability of the revised mean absolute percentage errors (mape) approach to some popular normal and non-normal independent time series*, *Int. Adv. Econ. Res. 15*, 409–420.

Riechmann T. (2001), *Genetic algorithm learning and evolutionary games*, *Journal of Economic Dynamic & Control 25*, 1019–1037.

Rościzak P. (2010), *Wykorzystanie algorytmu genetycznego w generowaniu decyzji inwestycyjnych na rynku walut*, w J. Siedlecki i P. Peternek, (red.), *Prace naukowe Uniwersytetu Ekonomicznego we Wrocławiu*, Vol. 108 *Współczesne tendencje rozwojowe badań operacyjnych*, Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław, 231–241.

Rutkowska D., Piliński M., Rutkowski L. (1999), *Sieci neuronowe, algoytmy genetyczne i systemy rozmyte*, PWN, Warszawa.

Schaffer J. D. (1984), *Some experiments in machine learning using vector evaluated genetic algorithms*, w (Michalewicz 2003), Vanderbilt University, Nashville.

Schoenauer M., Xanthakis S. (1993), *Constrained GA optimization*, w (Michalewicz 2003), Morgan Kaufman, San Mateo CA.

Schwefel H.-P. (1981), *Numerical optimization of computer models*, w (Michalewicz 2003), Wiley, Chichester.

Shang H. L., Hyudman R. J. (2011), *Nonparametric time series forecasting with dynamic updating*, *Mathematics and Computers in Simulation 81*, 1310–1324.

Sikora R., Piramuthu S. (2005), *Efficient genetic algorithm based data mining using feature selection with Hausdorff distance*, *Information Technology and Management 6*, 315–331.

Szapiro T. (2000), *Decyzje menedżerskie z Excelem*, PWE, Waraszawa.

Takahashi Y. (1998), *Convergence of simple genetic algorithms for the two-bit problem*, *BioSystems 46*, 235–282.

Tarczyński W. (1997*a*), *Rynki kapitałowe*, Vol. I, Agencja Wydawnicza Placet, Warszawa.

Tarczyński W. (1997*b*), *Rynki kapitałowe*, Vol. II, Agencja Wydawnicza Placet, Warszawa.

Tashman L. J. (2000), *Out-of-sample tests of forecasting accuracy: an analysis and review*, *International Journal of Forecasting 16*, 437–450.

Taylor J. W. (2010), *Exponentialy weighted methods for forecasting intraday time series with multiple seasonal cycles*, *International Journal of Forecasting 26*, 627–646.

Taylor J. W., Snyder, R. D. (2001), *Forecasting intraday time series with multiple seasonal cycles using parsimonious seasonal exponential smoothing*, *International Journal of Forecasting 17*, 143–157.

Theodosiou M. (2011), *Forecasting monthly and quarterly time series using stl decomposition*, *International Journal of Forecasting* .

Trzaskalik T. (2008), *Wprowadzenie do badań operacyjnych z komputerem*, PWE, Warszawa.

Łuniewska M. (2008), *Ekonometria finansowa. Analiza rynku kapitałowego*, PWN, Warszawa.

van Bellegem S., von Sachs R. (2004), *Forecasting economic time series with unconditional time-varying variance*, *International Journal of Forecasting 20*, 611–627.

Venkatesan R., Kumar V. (2002), *A genetic algorithms approach to growth phase forecasting of wireless subscribers*, *International Journal of Forecasting 18*, 625–646.

Wagner J. (1999), *A genetic algorithm solution for one-dimensional boundled stock cutting*, *European Journal of Operational Research 117*, 368–381.

Wang J.-J., Wang J.-Z., Zhang Z.-G., Guo S.-P. (2012), *Stock index forecasting based on a hybrid model*, *Omega 40*, 758–766.

Welfe A. (2003), *Ekonometria*, PWE, Warszawa.

Weron R., Misiorek A. (2008), *Forecasting spot electricity prices: a comparison of parametric and semiparametric time series models*, *International Journal of Forecasting 24*, 744–763.

Wichard J. D. (2011), *Forecasting the nn5 time series with hybrid models*, *International Journal of Forecasting 27*, 700–707.

Winston W. L. (2005), *Microsoft Excel. Analiza i modelowanie danych*, APN PROMISE, Microsoft Press, Warszawa.

Wójcik M., Szewieczek D., Kiedos L. (2008), *Prognozy wartości aktywów tfi i depozytów bankowych gospodarstw domowych*, w *Studia i Prace Wydziału Nauk Ekonomicznych i Zarządzania*, Vol. 10 *Rynek kapitałowy, skuteczne inwestowanie*, Uniwersytet Szczeciński, Wydawnictwo Uniwersytetu Szczecińskiego, Szczecin, 306–319.

Zeliaś A. (1997), *Teoria prognozy*, PWE, Warszawa.

Zheng Z., Dong X., Zheng S. (2001), *Assigning task by parallel genetic algorithm based on pvm*, *Wuhan Univeristy Journal of Natural Sciences 6*(1–2), 579–584.

Zou H., Yang Y. (2004), *Combining time series models for forecasting*, *International Journal of Forecasting 20*, 69–84.

Zou X., Kong, L., Li Y. Chen, Y. (2002), *Finding global minima with a new dynamical evolutionary algorithm*, *Wuhan Univeristy Journal of Natural Sciences 7*(2), 157–160.