Marcin GODLEWSKI
Katedra Badań Operacyjnych
Akademia Ekonomiczna w Poznaniu

WYKORZYSTANIE ALGORYTMU GENETYCZNEGO DO SYMULACYJNEGO WYZNACZANIA PORTFELA PAPIERÓW WARTOŚCIOWYCH

Wprowadzenie

Algorytmy genetyczne stanowią specyficzną klasę metod optymalizacji, wyprowadzającą swoje korzenie z teorii ewolucji żywych organizmów i nauki o genetyce. Procedury te dokonują przeszukiwania, oceny i poprawy rozwiązań, wzorując się nabiologicznych mechanizmach doboru naturalnego i reprodukcji. Pojedyncze rozwiązania problemu optymalizacyjnego traktowane są jak żywe organizmy, których wiele współistnieje jednocześnie w populacji. Poszczególne rozwiązania różnią się od siebie, co znajduje odzwierciedlenie w ich ocenie według przyjętego kryterium, dokonywanej za pomocą *funkcji przystosowania*, będącej odpowiednikiem funkcji celu. Rozwiązania o wysokim przystosowaniu do narzuconych kryteriów są nagradzane przetrwaniem. Te, które słabo realizują postawione cele, są eliminowane drogą selekcji naturalnej. Cechy zwycięskich rozwiązań są przekazywane drogą reprodukcji do kolejnych pokoleń, jednocześnie ulegając wzajemnemu przemieszaniu (na podobieństwo krzyżowania kodów genetycznych), co sprzyja procesowi poszukiwania nowych, coraz doskonalszych zestawów cech. Dokładny opis funkcjonowania algorytmów genetycznych można znaleźć między innymi w pracach [1], [6], [7].

Polem, na którym algorytmy genetyczne wykazują znaczną przydatność, są zagadnienia optymalizacyjne o dużej złożoności obliczeniowej, na przykład problemy kombinatoryczne. Należy podkreślić, że algorytmy genetyczne nie dają rozwiązań ściśle optymalnych, a nawet nie gwarantują dotarcia w pobliże

rozwiązania optymalnego. Jednakże gdy brak analitycznych algorytmów wyznaczania rozwiązania określonego problemu, lub gdy praktyczne wykorzystanie takiego algorytmu jest niemożliwe ze względu na znaczną złożoność obliczeniową, zastosowanie algorytmu genetycznego może dać dobry skutek.

Algorytmy genetyczne są wykorzystywane jako narzędzie wspomagające różnego rodzaju procesy decyzyjne, również tworzenie portfela papierów wartościowych. Opisy przykładowych implementacji oraz uzyskanych wyników można znaleźć na przykład w pracach [2], [3], [4]. Należy zwrócić jednakże uwagę, że proponowane metody wykorzystywały algorytm genetyczny jedynie jako narzędzie rozwiązywania określonych wariantów klasycznego zadania optymalizacji portfela papierów wartościowych, wywodzącego się z modelu Markowitza.

Niniejszy artykuł opisuje alternatywną, nieklasyczną procedurę wyznaczania składu portfela papierów wartościowych. Ocenie zostanie poddana jakość otrzymywanych rozwiązań, na przykładzie portfeli akcji notowanych na GPW w Warszawie. Zweryfikowana zostanie wiarygodność i stosowalność procedury w realnej działalności inwestycyjnej.

1. Idea procedury

Klasyczne metody konstrukcji portfela papierów wartościowych, oparte na modelu Markowitza, zazwyczaj dokonują oceny portfela na bazie oczekiwanej stopy zwrotu oraz odchylenia standardowego lub wariancji stopy zwrotu. Obie charakterystyki są wyznaczane na podstawie danych historycznych, opisujących zmienność notowań poszczególnych walorów w przeszłości.

Proponowana w niniejszym artykule procedura nie wykorzystuje żadnych analogicznych miar oceny portfela. Zamiast tego, zastosowana jest bardzo dosłowna miara nieznanej przyszłej zyskowności: obliczana jest stopa zwrotu, jaką przyniósłby rozpatrywany portfel, gdyby taka sama inwestycja została dokonana w pewnym momencie w przeszłości. Oczywiście stopa zwrotu w jednym wybranym odcinku czasowym danych historycznych nie może być wystarczającym argumentem dla wyboru portfela. Jednakże porównanie stóp zwrotu dla pewnego wystarczająco licznego zbioru portfeli, powtórzone wielokrotnie w różnych podokresach przeszłości, powinno pozwolić wskazać te portfele, które często wypadały korzystnie na tle konkurencji.

Problemem jest przegląd wystarczająco dużej liczby portfeli. Gdy tworzymy portfel podzielony na p równowartościowych udziałów (czyli o strukturze

określonej z dokładnością do $1/p$) na podstawie zbioru k walorów, wtedy liczba wszystkich możliwych rozwiązań wynosi:

$$l(k,p)=\frac{(p+k-1)!}{(p)!(k-1)!}. \tag{1}$$

Nietrudno obliczyć, że już dla portfela składającego się z akcji 20 spółek o udziałach wartościowych określonych z dokładnością do 0,05, liczba kombinacji sięga 1,76 biliona. Zastosowanie procedur przeglądu zupełnego wiązałoby się z niemożliwą do zaakceptowania złożonością obliczeniową, natomiast specyfika działania algorytmu genetycznego mogłaby wykazać się w tym przypadku dużą skutecznością.

Działanie algorytmu genetycznego opiera się na wielokrotnym ocenianiu rozpatrywanego zbioru rozwiązań pod kątem wartości funkcji przystosowania. Rozwiązania dające zadowalające rezultaty są reprodukowane do następnych pokoleń, gdzie w osobie swoich potomków – są ponownie oceniane. Przyjęcie zasady, że każde pokolenie byłoby oceniane na podstawie stóp zwrotu w innym losowo wybranym przedziale czasowym przeszłości, powodowałoby, iż wielokrotną weryfikację w kolejnych pokoleniach przetrwałyby wyłącznie portfele, które w praktycznie każdych warunkach, niezależnie od okresu weryfikacji, dawały stabilne, wysokie zyski. Inaczej ujmując, przeprowadzenie wielopokoleniowej symulacji oceny portfeli powinno pozwolić wyeliminować takie zestawy walorów, które narażałyby inwestora w pewnych sytuacjach na straty.

2. Implementacja algorytmu genetycznego

2.1. Kodowanie rozwiązań

Podstawowym elementem każdej implementacji algorytmu genetycznego jest wybór sposobu zapisu rozwiązań (w tym przypadku składów portfeli) w postaci ciągu kodowego, który mógłby pełnić rolę odpowiednika kodu genetycznego, replikowanego i krzyżowanego w procesie ewolucyjnym. W omawianej implementacji przyjęto zapis kodowy oparty na alfabecie o liczbie znaków równej liczbie rozpatrywanych spółek. Każdy znak w ciągu kodowym odpowiada jednostkowemu udziałowi wartościowemu w portfelu, a więc dokładność określenia struktury portfela jest uzależniona od liczby znaków do tego użytych. Udział konkretnego waloru w portfelu jest określony przez liczbę odpowiadających mu znaków w ciągu. Formalnie, dla m-elementowego zbioru spółek

$S=\{s_1, s_2, ..., s_m\}$, portfel podzielony na n równowartościowych pakietów-cegiełek, jest opisywany za pomocą następującego ciągu kodowego:

$$a_1a_2a_3 ... a_n, \quad \text{gdzie} \quad a_i \in S,\ i=1, 2, ..., n. \tag{2}$$

Udział wartościowy waloru k w portfelu jest równy:

$$x_k = \frac{\text{card}(\{i : a_i = s_k\})}{n}. \tag{3}$$

Zaletą takiego sposobu zapisu rozwiązań jest to, że dowolny ciąg znaków o określonej długości i oparty na określonym alfabecie będzie reprezentował dopuszczalne rozwiązanie, czyli portfel, w którym udziały poszczególnych walorów:

- sumują się do 1,
- są nieujemne (nie jest uwzględniana krótka sprzedaż).

2.2. Losowanie składu populacji wyjściowej

Działanie algorytmu genetycznego rozpoczyna się od utworzenia wyjściowej populacji rozwiązań, która będzie później poddawana procesom ewolucyjnym. Zazwyczaj rozwiązania do populacji wyjściowej są dobierane w sposób losowy. W naszym przykładzie oznacza to, że przy użyciu generatora liczb pseudolosowych zostaje utworzonych N ciągów kodowych, z których każdy składa się z n znaków, wylosowanych z równym prawdopodobieństwem z m-elementowego alfabetu, gdzie N jest wielkością populacji, n odpowiada liczbie pakietów, na które podzielony jest każdy portfel, a m oznacza liczbę spółek, których akcje wchodzą potencjalnie w skład portfela.

2.3. Funkcja przystosowania

Wszystkie portfele, wchodzące w skład populacji, muszą zostać ocenione przy użyciu pewnej funkcji przystosowania. Jak zostało to wcześniej wskazane, kryterium oceny ma być zyskowność portfela, przy założeniu że inwestycja dotyczyłaby pewnego odcinka czasu w przeszłości. Aby zachować porównywalność, wszystkie portfele znajdujące się w jednej populacji muszą być oceniane na podstawie tego samego podokresu danych historycznych. Dla każdej kolejnej populacji podokres ma być wybierany losowo. W tym celu ustalane są:

- Długość przedziału czasowego, obliczana według następującego wzoru:

$$s = g + (h - g) rand^2. \tag{4}$$

gdzie parametry g oraz h oznaczają odpowiednio najkrótszy oraz najdłuższy rozpatrywany przedział czasowy weryfikacji (w dniach), a *rand* jest liczbą pseudolosową z przedziału $\langle 0, 1)$. Podniesienie liczby *rand* do kwadratu powoduje, że zwiększone będzie prawdopodobieństwo wyznaczania stosunkowo krótkich okresów utrzymywania portfeli, co powinno sprzyjać wychwytywaniu spółek o silnych krótkookresowych wahaniach notowań.

- Moment początkowy okresu weryfikacji portfeli:

$$t = a + (b - s - a) rand. \tag{5}$$

Zgodnie z powyższym wzorem, odcinek czasowy o oznaczonej długości s jest losowo umiejscawiany w przedziale czasu $\langle a, b \rangle$, dla którego dysponujemy danymi rzeczywistymi co do notowań rozpatrywanych spółek.

Dla wyznaczonego zakresu czasowego okresu weryfikacji portfeli obliczane są odpowiadające im wartości funkcji przystosowania:

$$f_i = 1 + \sum_{j=1}^{m} x_{ij} r_{j,t,t+s}, \tag{6}$$

gdzie:

f_i – wartość funkcji przystosowania portfela i,
m – liczba walorów,
x_{ij} – udział wartościowy waloru j w portfelu i,
$r_{j,t,t+s}$ – rzeczywista stopa zwrotu z waloru j w okresie $\langle t, t+s \rangle$.

2.4. Reprodukcja i krzyżowanie

Nietrudno zauważyć, że wartość funkcji przystosowania jest wartością stopy zwrotu, powiększoną o 1. Jest to zabieg techniczny, który pozwolił uniknąć niebezpieczeństwa przyjmowania przez funkcję przystosowania ujemnych wartości, co jest istotne w procesie reprodukcji rozwiązań. Prawdopodobieństwo bowiem, że portfel o założonym składzie „przetrwa" i znajdzie się w *puli rodzicielskiej*, tworzącej nowe pokolenie portfeli, jest ustalane według następującego wzoru:

$$p_i = \frac{f_i}{\sum_{j=1}^{n} f_j}. \tag{7}$$

Pula rodzicielska składa się z takiej samej liczby rozwiązań jak cała populacja i jest zapełniana w wyniku serii niezależnych losowań, o rozkładach prawdopodobieństw wyznaczonych zgodnie z (7). Każdy portfel, który znajdzie się w puli, będzie miał swojego „potomka" w nowym pokoleniu. Z procedury zapełniania puli wynika, że część portfeli nie znajdzie się w niej w ogóle, natomiast niektóre znajdą się we więcej niż jednej kopii. Można obliczyć, że oczekiwana liczba kopii portfela i w puli rodzicielskiej wynosi:

$$\frac{f_i}{\bar{f}}, \tag{8}$$

gdzie $\bar{f}$ jest przeciętną wartością funkcji przystosowania w populacji. Oznacza to, że portfele, które osiągnęły rzeczywistą stopę zwrotu wyższą niż przeciętna, statystycznie zwiększą swoją reprezentację w nowym pokoleniu, natomiast te, które okazały się gorsze niż przeciętna, zostaną wyparte. Opisana powyżej procedura tworzenia populacji rodzicielskiej, dobrze oddająca istotę doboru naturalnego, jest w literaturze znana pod nazwą *reguły ruletki* (por. [1], s. 32).

Dla dobra skuteczności działania procedury, dobór naturalny musi jednak być w pewnej mierze sterowany. Należy bowiem pamiętać, że wysoka wartość funkcji przystosowania danego portfela świadczy jedynie o wysokiej stopie zwrotu w pojedynczym, przykładowym podokresie historii, co nie powinno być przesłanką do nadmiernej replikacji składu takiego portfela w następnych pokoleniach. Dopiero pomyślne przetrwanie weryfikacji w wielu pokoleniach świadczy o jakości portfela. Rozwiązania o ponadprzeciętnej wartości funkcji przystosowania powinny umacniać swoją pozycję w populacji, jednakże w ograniczonym tempie. Regulacja tempa może być dokonywana za pomocą skalowania wartości funkcji przystosowania. W rozpatrywanej implementacji zastosowano liniową funkcję skalowania wokół średniej:

$$f_i' = \bar{f} + (f_i - \bar{f})M, \tag{9}$$

gdzie M jest wartością mnożnika, daną poniższą zależnością:

$$M = \min\left\{\left(\frac{\bar{f}}{\max_j\{f_j\} - \bar{f}}(q-1)\right); \left(\frac{\bar{f}}{\bar{f} - \min_j\{f_j\}}\right)\right\}. \tag{10}$$

Powyższa metoda skalowania wartości funkcji przystosowania zapewnia spełnienie poniższych warunków:

- wartość przeciętna funkcji przystosowania pozostaje na tym samym poziomie,

- przeskalowane wartości funkcji przystosowania są nieujemne.
- oczekiwana liczba potomków najlepszego rozwiązania w populacji jest równa dokładnie q lub jest jak najbardziej do tej wartości zbliżona, jeśli dokładna równość byłaby sprzeczna z drugim warunkiem.

Rozwiązania, które znalazły się w puli rodzicielskiej, nie są bezpośrednio przenoszone do nowego pokolenia, gdyż nie stwarzałoby to możliwości postępu i poprawy rozwiązań. Algorytm replikujący bez zmian dobre rozwiązania szybko zbiegłby się do najlepszego z rozwiązań początkowych i nie miałby szans na ocenę innych wariantów, które nie znalazły się w wygenerowanej losowo populacji wyjściowej. Aby temu zapobiec, na parach rozwiązań z puli rodzicielskiej wykonywana jest operacja krzyżowania, czyli wzajemnej zamiany fragmentów ciągów kodowych, opisujących te rozwiązania. Dwa ciągi są „rozcinane" w losowo wybranym miejscu, a następnie łączone „na krzyż", dając dwa zupełnie nowe rozwiązania, łączące wybrane cechy obu rozwiązań macierzystych. Na przykład, dwa ciągi:

$$X_1X_2X_3X_4X_5X_6X_7X_8X_9$$
$$Y_1Y_2Y_3Y_4Y_5Y_6Y_7Y_8Y_9$$

rozcięte pomiędzy 6. a 7. znakiem i skrzyżowane, dadzą następujące „potomstwo":

$$X_1X_2X_3X_4X_5X_6Y_7Y_8Y_9$$
$$Y_1Y_2Y_3Y_4Y_5Y_6X_7X_8X_9$$

Należy jednakże wziąć pod uwagę niebezpieczeństwo, że w wyniku skrzyżowania rozwiązania bardzo dobrego z rozwiązanie słabym, mogą powstać dwa również słabe rozwiązania. Pomimo wielkiej wagi działania tego operatora dla całości procesu ewolucji rozwiązań, wskazana może być ochrona najlepszych portfeli przed ewentualną utratą w wyniku krzyżowania. Może to zostać zrealizowane poprzez określanie dla każdego z rozwiązań indywidualnego parametru krzyżowania:

$$\pi_c^i = \begin{cases} \dfrac{z+(1-z)(\max\limits_j\{f_j\}-f_i)}{(\max\limits_j\{f_j\}-\bar{f}} & \text{dla } f_i \geqslant \bar{f}, \\ 1 & \text{dla } f_i < \bar{f}. \end{cases} \tag{11}$$

Powyższa formuła jest niewielką modyfikacją wzoru zaproponowanego przez M. Srinivasa i L. Patnaika ([6], s. 152). Wyznaczana wartość parametru,

będąca podstawą do określenia prawdopodobieństwa krzyżowania danego portfela, jest równa 1 dla rozwiązań słabszych niż przeciętne, natomiast dla portfeli ponadprzeciętnych maleje, aż do $\pi_c = z$ dla portfela o najwyższej wartości funkcji przystosowania w całej populacji.

Prawdopodobieństwo, że dane dwa rozwiązania i_1 oraz i_2 zostaną skrzyżowane, jest przyjmowane na poziomie:

$$p_c(i_1, i_2) = \sqrt{\pi_c^{i_1} \pi_c^{i_2}}. \tag{12}$$

Gdyby okazało się, że dane dwa portfele nie miałyby zostać skrzyżowane, wtedy odpowiadające im ciągi kodowe byłyby bezpośrednio przenoszone do nowej populacji (następuje reprodukcja prosta).

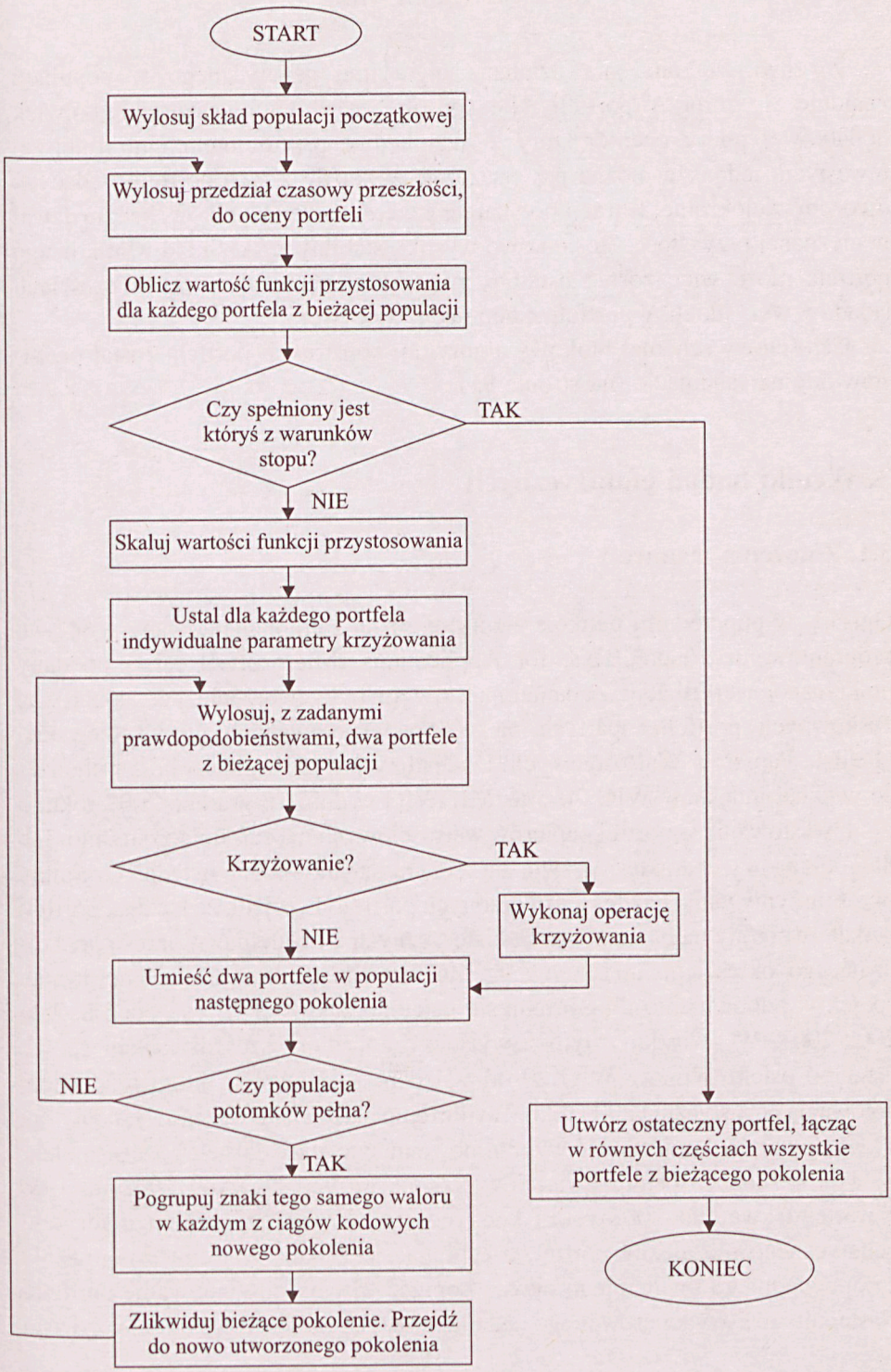
Ostatnim krokiem pojedynczej iteracji algorytmu jest porządkowanie zapisów kodowych nowego pokolenia w taki sposób, aby w każdym ciągu znaki odpowiadające temu samemu walorowi zajmowały sąsiednie pozycje. Taka zmiana kolejności znaków w ciągu nie wpływa na skład portfeli, gdyż ze sposobu kodowania wynika, że dowolna permutacja ciągu daje w rezultacie ten sam portfel. Umieszczenie znaków tego samego rodzaju na sąsiadujących pozycjach ma służyć zmniejszeniu prawdopodobieństwa rozdzielenia tych znaków w procesie krzyżowania w następnym pokoleniu, co mogłoby doprowadzić do „rozmycia" informacji na temat wielkości udziału konkretnego waloru w portfelu. Znajduje to również swoje uzasadnienie w fundamentalnym dla algorytmów genetycznych twierdzeniu o schematach i w hipotezie bloków budujących (por. [7], s. 159).

2.5. Zatrzymanie algorytmu

Po zakończeniu procesu reprodukcji i krzyżowania, nowe pokolenie zastępuje populację macierzystą i wykonywana jest kolejna iteracja algorytmu. Portfele wchodzące w skład nowo powstałego pokolenia są oceniane, tworzona jest pula rodzicielska, która daje początek kolejnemu pokoleniu. Cały proces jest powtarzany aż do momentu zajścia któregoś z dwóch alternatywnych warunków stopu:

- upłynięcie określonej liczby pokoleń G,
- zmniejszenie się rozstępu stóp zwrotu w ocenianym okresie pomiędzy najlepszym a najgorszym portfelem w populacji, poniżej ustalonego poziomu ε, pod warunkiem, że upłynęła co najmniej połowa założonej liczby pokoleń.

Decyzja o ewentualnym zatrzymaniu algorytmu następuje po wyznaczeniu wartości funkcji przystosowania dla portfeli z bieżącej populacji.



Schemat blokowy algorytmu

W chwili zakończenia działania algorytmu genetycznego w populacji znajduje się zbiór N portfeli. Nie sposób bez przyjmowania jakiejkolwiek dodatkowej miary ocenić, który z nich będzie przedmiotem najtrafniejszej inwestycji, jednakże można przypuszczać, że każdy z tych portfeli, będących owocem wielokrotnej filtracji pod kątem generowanych wyników, będzie dawał w nieznanej przyszłości stosunkowo wysoki, stabilny zysk. Skład wynikowego portfela może więc zostać ustalony poprzez złożenie w równych częściach składów wszystkich N portfeli z populacji końcowej.

Całościowy schemat blokowy algorytmu konstrukcji portfela został przedstawiony na schemacie (na stronie 91).

3. Wyniki badań empirycznych

3.1. Założenia testowe

Opisany w poprzednim punkcie algorytm został zaimplementowany w języku programowania Visual Basic for Applications (Microsoft Excel) i poddany praktycznej weryfikacji. Zadaniem testowym było tworzenie jak najbardziej zyskownych portfeli, opartych na walorach notowanych na warszawskiej Giełdzie Papierów Wartościowych. Do analizy wytypowano spółki[1], wchodzące w skład indeksów WIG 20 oraz MIDWIG w dniu 10 grudnia 2003 roku.

Inwestowanie w portfel papierów wartościowych ma zasadniczo średnio- lub długookresowy charakter, w związku z czym uznano za odpowiedni kwartalny okres utrzymywania każdego z tworzonych portfeli. Przyjęto, że każdy z portfeli będzie tworzony na bazie notowań historycznych z dwuletniego okresu poprzedzającego okres jego utrzymywania. Ze względu na konieczność notowania spółek w odpowiednio długim okresie ustalono zakres czasowy badań na lata 2002-2003. W związku z tym z wyjściowego zbioru sześćdziesięciu spółek usunięto osiem (jedną z WIG 20 oraz siedem z MIDWIG), które rozpoczęły notowania po 1 stycznia 2000 roku. Wydłużenie okresu badawczego o kolejny rok spowodowałoby konieczność wyeliminowania z analizy dalszych sześć spółek.

Dla każdego z ośmiu kwartałów (od pierwszego kwartału 2002 roku do czwartego kwartału 2003 roku) konstruowano z wykorzystaniem algorytmu genetycznego niezależny portfel, przyjmując jako dane wejściowe notowania z odpowiedniego dwuletniego okresu poprzedzającego utrzymywanie portfela. Następnie dla wygenerowanego składu portfela obliczana była rzeczywista

[1] Źródło danych: serwis internetowy www.parkiet.com.pl.

stopa zwrotu w trzymiesięcznym okresie, dla którego był konstruowany. Wyjątkiem był portfel utworzony na czwarty kwartał 2003 roku, dla którego rzeczywista stopa zwrotu została obliczona w krótszym niż kwartalny okresie (od 1 października do 10 grudnia 2003 roku), ze względu na dostępność danych w czasie przeprowadzania obliczeń.

Szczegółowe parametry techniczne działania algorytmu zostały dobrane metodą „prób i błędów", wspartą pewną dozą intuicji, wynikającej z uwzględnienia specyfiki rozpatrywanego problemu. Ustalone wartości parametrów to:

- liczba portfeli rozpatrywanych w pojedynczym pokoleniu, czyli wielkość populacji, wynosi 100,
- każdy portfel jest opisany ciągiem kodowym składającym się ze 100 znaków, czyli struktura udziałów jest określona z dokładnością[2] do 0,01,
- współczynnik q funkcji skalującej wartości przystosowania, interpretowany jako oczekiwana liczba potomków najlepszego portfela, został ustalony na 1,30,
- długość okresu danych historycznych, dla którego obliczana jest stopa procentowa, będąca podstawą funkcji przystosowania, jest określana w przedziale od dwóch do dziewięciu miesięcy,
- wartość indywidualnego parametru krzyżowania dla najlepszego portfela w populacji wynosi $z = 0{,}30$,
- liczbę pokoleń G, po których zatrzymywany jest algorytm genetyczny, ustalono na1000,
- opcjonalnie algorytm zatrzymuje się po nie mniej niż 500 pokoleniach, jeżeli rozstęp stopy zwrotu między najlepszym a najgorszym portfelem spadnie poniżej $\varepsilon = 1{,}5$ p.p.

Działanie algorytmu badano dla dwóch wariantów zestawu spółek – w wariancie pierwszym przyjęto, że portfel będzie tworzony na bazie spółek wchodzących w skład indeksów WIG 20 oraz MIDWIG, natomiast w wariancie drugim zbiór ten został ograniczony wyłącznie do spółek notowanych w ramach indeksu WIG 20.

3.2. Pomiar rzeczywistej stopy zwrotu generowanych portfeli

Charakterystyczną cechą algorytmu genetycznego jest to, że przy każdym wywołaniu może dawać inne rozwiązanie. Kilkakrotne wygenerowanie portfela dla tego samego kwartału zazwyczaj da składy portfeli bardziej lub mniej

---

[2] Podana dokładność dotyczy portfeli rozpatrywanych w trakcie działania algorytmu genetycznego. Po zakończeniu obliczeń, portfel wynikowy jest generowany na podstawie średniej z całej populacji, więc dokładność z jaką jest określony, wzrasta 100 razy, do 0,0001.

różniące się od siebie. Jest to naturalna konsekwencja przyjęcia stochastycznych procedur działania algorytmu. W przypadku rozpatrywanego zagadnienia, zróżnicowanie jakości otrzymywanych wyników będzie jeszcze dodatkowo powiększane ze względu na to, że składy portfeli są wyznaczane na bazie analizy danych historycznych, a ocena portfeli następuje na podstawie nieznanych w momencie konstruowania portfela rzeczywistych zmian cen walorów. Wtedy dwa portfele, które mogły się wydawać równoważne na etapie konstruowania i przez to być zamiennie wskazywane przez algorytm, mogą zdecydowanie różnić się efektywnością w okresie utrzymywania portfela. W świetle powyższych spostrzeżeń niecelowe byłoby ograniczenie analizy do portfeli uzyskanych w jednorazowym wywołaniu algorytmu genetycznego. W związku z tym wygenerowano – oddzielnie dla dwóch wariantów zestawów spółek po 200 „ciągów portfeli"[3]. Kryterium oceny jakości otrzymanych rozwiązań będzie łączna 24-miesięczna rzeczywista stopa zwrotu z ciągu portfeli, obliczona przy założeniu, że pełna kwota uzyskana ze sprzedaży każdego kwartalnego portfela jest reinwestowana w następnym kwartale (nie są uwzględniane koszty transakcyjne).

Poniższe tabele (oddzielne dla dwóch wariantów zestawu spółek) zawierają:

- stopy zwrotu dla – przykładowo – pierwszego, pięćdziesiątego i setnego ciągu portfeli (*AG#1, AG#50, AG#100*),
- stopy zwrotu dla najbardziej i najmniej zyskownego spośród dwustu ciągów portfeli (*AG-Min*, *AG-Max*),
- stopy zwrotu dla ciągu portfeli, utworzonego z połączenia w równych częściach wszystkich dwustu ciągów (*AG-Średnia*).

Pierwszym wnioskiem, który nasuwa się na podstawie analizy stóp zwrotu z wybranych ciągów portfeli (kolumny *AG#1, AG#50* oraz *AG#100* w obu tabelach) jest znaczne zróżnicowanie wyników. Stopa zwrotu, jaką osiągnąłby inwestor, decydując się na portfele akcji wskazywane przez pojedyncze wywołania algorytmu genetycznego dla każdego z kwartałów, byłaby obarczona dużą dozą zróżnicowania. W przypadku portfela tworzonego na podstawie walorów z indeksów WIG 20 oraz MIDWIG, wynikiem dwuletniej inwestycji mogłaby być zarówno utrata niespełna 11% kapitału, jak i jego powiększenie o prawie 520% (kolumny *AG-Min* oraz *AG-Max* w tabeli 1). Dla portfeli tworzonych z ograniczonego zestawu walorów (tabela 2) zmienność wyników jest mniejsza – spośród dwustu wygenerowanych ciągów portfeli, w najgorszym przypadku uzyskano prawie 30% zysk, a najwyższa stopa zwrotu nie przekroczyła 219%.

[3] Przez ciąg potfeli będziemy rozumieć zbiór ośmiu, wygenerowanych niezależnie kolejnych portfeli, po jednym na każdy kwartał inwestycji (od początku 2002 do końca 2003 roku).

Tabela. 1. Rzeczywiste stopy zwrotu z portfeli uzyskanych metodą genetyczną spółki z WIG 20 i MIDWIG (w %)

Stopy zwrotu	*AG#1*		*AG#50*		*AG#100*		*AG-Min*		*AG-Max*		*AG-Średnia*	
	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie
2002												
I	21,10	21,10	3,16	3,16	10,81	10,81	8,22	8,22	12,62	12,62	12,37	12,37
II	4,46	26,50	−4,52	−1,50	−7,47	2,54	−0,79	7,37	−15,98	−5,37	−0,62	11,67
III	−8,17	16,17	1,23	−0,29	−8,02	−5,69	−2,99	4,16	0,53	−4,86	−5,52	5,51
IV	7,47	24,85	6,34	6,03	7,20	1,10	9,35	13,90	20,45	14,59	10,31	16,38
2003												
I	−9,12	13,47	−4,98	0,75	−7,90	−6,88	−12,28	−0,09	31,76	50,99	8,25	25,99
II	19,02	35,05	62,85	64,08	38,40	28,87	4,40	4,30	63,35	146,64	30,82	64,82
III	43,20	93,39	16,60	91,32	−2,42	25,76	−12,57	−8,81	46,03	260,17	14,38	88,51
IV	6,18	105,35	53,66	193,98	−2,95	22,05	−2,00	−10,63	71,88	519,06	21,69	129,40

Tabela 2. Rzeczywiste stopy zwrotu z portfeli uzyskanych metodą genetyczną spółki z WIG 20 (w %)

Stopy zwrotu	*AG#1*		*AG#50*		*AG#100*		*AG-Min*		*AG-Max*		*AG-Średnia*	
	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie
2002												
I	22,39	22,39	3,65	3,65	23,44	23,44	9,92	9,92	22,19	22,19	17,12	17,12
II	−4,99	16,28	−6,77	−3,37	−7,64	14,01	−8,99	0,04	−8,12	12,27	−6,66	9,33
III	−4,13	11,48	−3,51	−6,76	−3,34	10,20	−9,12	−9,08	1,57	14,04	−5,70	3,10
IV	19,91	33,69	12,60	4,99	14,98	26,71	11,80	1,65	17,24	33,70	14,60	18,15
2003												
I	14,64	53,26	−0,99	3,94	26,35	60,10	−10,67	−9,20	31,16	75,37	12,66	33,11
II	18,54	81,68	15,55	20,11	12,62	80,30	8,20	−1,75	28,22	124,85	11,96	49,03
III	33,27	142,12	32,89	59,61	20,06	116,47	23,49	21,33	32,141	97,12	29,70	93,30
IV	7,45	160,17	7,72	71,93	8,22	134,26	6,85	29,64	7,31	218,85	8,18	109,11

Tabela 3. Rozkład empiryczny częstości rzeczywistych 24-miesięcznych stóp zwrotu

Stopa zwrotu (w %)	Wariant I (WIG 20 + MIDWIG)		Wariant II (WIG 20)	
	częstość	częstość skumulowana	częstość	częstość skumulowana
Poniżej 0	3	3	0	0
⟨0, 10)	2	5	0	0
⟨10, 20)	8	13	0	0
⟨20, 30)	7	20	1	1
⟨30, 40)	7	27	2	3
⟨40, 50)	8	35	3	6
⟨50, 60)	13	48	7	13
⟨60, 70)	7	55	16	29
⟨70, 80)	9	64	16	45
⟨80, 90)	11	75	17	62
⟨90, 100)	11	86	20	82
⟨100, 110)	13	99	17	99
⟨110, 120)	13	112	19	118
⟨120, 130)	10	122	27	145
⟨130, 140)	10	132	22	167
⟨140, 150)	4	136	10	177
⟨150, 160)	6	142	11	188
⟨160, 170)	7	149	5	193
⟨170, 180)	7	156	2	195
⟨180, 190)	6	162	3	198
⟨190, 200)	6	168	1	199
⟨200, 300)	22	190	1	200
⟨300, 400)	5	195	0	200
⟨400, 500)	3	198	0	200
500 i więcej	2	200	0	200

Dokładniejszy pogląd na charakterystykę zmienności wyników w kolejnych wywołaniach algorytmu genetycznego daje tabela 3, zawierająca informacje o tym, ile spośród dwustu ciągów portfeli odnotowało stopy zwrotu w odpowiednich przedziałach. Warto zauważyć, że nawet w pierwszym wariancie jedynie trzy ciągi portfeli spośród dwustu przyniosły stratę na koniec okresu

inwestycji, a połowa dała zysk przekraczający 110%. Większe skupienie wartości stopy zwrotu portfeli opartych na spółkach z indeksu WIG 20 jest zrozumiałe, jeśli weźmie się pod uwagę fakt, że indeks ten skupia spółki największe, a przez to – można domniemywać – względnie stabilne, wiarygodne i przewidywalne.

Mimo wszystko tak znaczne zróżnicowanie stóp zwrotu z konstruowanych portfeli jest dla inwestora poważnym problemem i może budzić brak zaufania do proponowanej metodyki. Znając rzeczywiste wartości notowań, możemy wskazywać, który spośród wygenerowanych portfeli był lepszą inwestycją, a który gorszą. Prawdziwy inwestor dowiedziałby się tego dopiero po tym, jak ulokowałby już swoje pieniądze. Niepewność wyniku inwestycji wynikającą ze zróżnicowania rzeczywistych stóp zwrotu portfeli wygenerowanych w niezależnych wywołaniach algorytmu genetycznego możemy określić **ryzykiem procedury**.

Ryzyko procedury może być jednakże w bardzo prosty sposób wyeliminowane. Wystarczy, aby inwestor zamiast kupić akcje tworzące portfel wskazany przez pojedyncze wywołanie algorytmu genetycznego, wygenerował na przykład dwieście takich portfeli i połączył je w równych udziałach w jeden portfel. Wtedy stabilność wyniku (przynajmniej w zakresie ryzyka procedury) zostanie praktycznie zapewniona. Łączna stopa zwrotu uzyskana w przypadku stosowania takiej strategii inwestycyjnej byłaby w naszym przypadku równa odpowiednio 129% lub 109%, zależnie od rozpatrywanego wariantu zestawu spółek (kolumny *AG-Średnia* w obu tabelach powyżej).

Można przyjąć, że inwestor, który chciałby posłużyć się prezentowaną postacią algorytmu genetycznego do wyznaczania składu portfela, powinien opierać się na uśrednionych wskazaniach co najmniej kilku wywołań procedury. Obranie takiej strategii powinno skutkować stabilną, wysoką stopą zwrotu z inwestycji. Takimi też wynikami, uzyskanymi na podstawie dwustukrotnego generowania składów portfeli, będziemy się posługiwać w dalszej analizie.

Dla zobiektywizowania oceny jakości portfeli generowanych przez algorytm genetyczny ich wyniki zostaną przedstawione na tle:

- ogólnych tendencji rynkowych, znajdujących swoje odzwierciedlenie w zmianach wartości indeksów giełdowych,
- stóp zwrotu z portfeli otrzymywanych za pomocą klasycznego zadania maksymalizacji użyteczności inwestora, opartego na modelu Markowitza.

Do porównania wykorzystano wartości indeksów giełdowych WIG, WIG 20 oraz MIDWIG. Składy portfeli Markowitza dla kolejnych kwartałów zostały wygenerowane przy wykorzystaniu modułu Solver (Microsoft Excel) jako rozwiązania poniższego zadania optymalizacji (por. [5], s. 185-186):

Tabela 4. Zestawienie porównawcze stóp zwrotu z portfeli (w %)

Stopy zwrotu	WIG		WIG 20		MIDWIG		Markowitz		AG I		AG II	
	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie	kwartalnie	łącznie
2002												
I	7,93	7,93	10,81	10,81	1,08	1,08	10,38	10,38	12,37	12,37	17,12	17,12
II	−5,54	1,95	−9,16	0,66	−4,88	−3,85	0,41	10,83	−0,62	11,67	−6,66	9,32
III	−9,68	−7,92	−14,07	−13,50	−11,42	−14,83	−5,98	4,20	−5,52	5,51	−5,70	3,09
IV	12,76	3,83	13,13	−2,15	10,31	−6,05	8,76	13,33	10,31	16,39	14,60	18,14
2003												
I	−4,22	−0,55	−6,99	−8,99	−5,60	−11,31	2,75	16,45	8,25	25,99	12,66	33,10
II	16,14	15,50	14,63	4,33	10,41	−2,08	28,55	49,70	30,82	64,82	11,96	49,01
III	22,72	41,74	20,31	25,52	20,78	18,27	6,75	59,80	14,38	88,52	29,70	93,27
IV	4,96	48,77	4,00	30,54	3,54	22,45	33,02	112,57	21,69	129,41	8,18	109,08

$$
\begin{aligned}
&(0)\ E(R_p)-\frac{1}{2}\lambda D^2(R_p)\to \max,\\
&(1)\ \sum_{i=1}^{m} x_i=1,\\
&(2)\ x_i\geqslant 0,\ i=1,2,\ldots,m.
\end{aligned}
\tag{13}
$$

Oczekiwane stopy zwrotu poszczególnych walorów oraz ich wariancje były szacowane na podstawie danych historycznych. Przyjmowaną oddzielnie dla każdego z portfeli próbą, był zbiór siedmiodniowych stóp zwrotu każdego z 52 walorów (z WIG 20 oraz MIDWIG), w odpowiednim dwuletnim okresie poprzedzającym moment, dla którego konstruowany był portfel.

Wartość współczynnika ostrożności inwestora $\lambda=6$ została przyjęta na podstawie sugesti zawartej w pracy [5], s. 196.

Naturalnie istnieje wiele, alternatywnych do powyższego, klasycznych wariantów konstrukcji portfela papierów wartościowych. Dla zapewnienia pełnej porównywalności wyników należałoby przedstawić szersze ich spektrum, jednakże celem autora było jedynie ukazanie pewnego tła dla oceny portfeli generowanych przy użyciu AG, stąd ograniczenie analizy do jednej wybranej metody.

3.3. Ocena ryzyka inwestycji

W przypadku portfela papierów wartościowych skonstruowanego za pomocą omawianej procedury opartej na algorytmie genetycznym, można wyróżnić dwa główne źródła niepewności inwestora odnośnie do przyszłej wartości zainwestowanego kapitału:

- **ryzyko procedury** – związane ze stochastycznym charakterem procedur algorytmu genetycznego i w konsekwencji różnicami pomiędzy portfelami wskazywanymi przez kolejne wywołania algorytmu,
- **ryzyko notowań** – przejawiające się w częściowej nieprzewidywalności zmian cen walorów, wynikającej z zaburzeń w kontynuowaniu tendencji historycznych, na których opiera się konstrukcja portfela.

Pierwszy rodzaj ryzyka został omówiony w poprzednim punkcie. Wskazano tam prostą metodę jego eliminacji, pozwalającą inwestorowi przede wszystkim na uniknięcie sporadycznych błędnych wskazań algorytmu genetycznego, które mogły powstać w wyniku podążenia „ślepą ścieżką ewolucji".

Aktualną groźbą pozostaje jednak ryzyko notowań. Ryzyko to dotyczy wszystkich metod konstrukcji portfela papierów wartościowych. Obroną przed nieprzewidywalnością zmian cen walorów jest dywersyfikacja portfela.

Tabela 5. Porównanie wartości odchylenia standardowego stóp zwrotu jako miary dywersyfikacji

	Odchylenie standardowe dziennych rzeczywistych stóp zwrotu			Odchylenie standardowe siedmiodniowych rzeczywistych stóp zwrotu		
	Markowitz	AG I	AG II	Markowitz	AG I	AG II
2002						
I	0,0094	0,0089	0,0162	0,0172	0,0175	0,0309
II	0,0096	0,0084	0,0145	0,0177	0,0161	0,0316
III	0,0139	0,0120	0,0159	0,0342	0,0305	0,0380
IV	0,0068	0,0058	0,0121	0,0142	0,0139	0,0279
2003						
I	0,0099	0,0089	0,0126	0,0210	0,0190	0,0306
II	0,0115	0,0089	0,0101	0,0313	0,0233	0,0290
III	0,0225	0,0190	0,0186	0,0479	0,0365	0,0470
IV	0,0214	0,0174	0,0156	0,0598	0,0482	0,0401
I 2002 - IV 2003	0,0141	0,0119	0,0147	0,0342	0,0282	0,0357
Średnia kwartalna	0,0131	0,0112	0,0145	0,0304	0,0256	0,0344

W klasycznym modelu Markowitza, dywersyfikacja jest przeprowadzana drogą dążenia do możliwie jak najniższej wariancji (czy też odchylenia standardowego) stopy zwrotu z portfela. Wykorzystywana przy tym jest miara współzmienności stóp zwrotu poszczególnych walorów, jaką jest kowariancja.

Algorytm genetyczny nie posługuje się żadną analogiczną miarą powiązania stóp zwrotu akcji, jednakże zróżnicowanie składu portfela owocuje większą stabilnością jego wartości i jest nagradzane w procesie ewolucji portfeli. Można więc przypuszczać, że algorytm genetyczny generuje portfele w pewnej mierze zdywersyfikowane.

Aby móc porównywać jakość dywersyfikacji portfeli utworzonych za pomocą algorytmu genetycznego oraz portfeli konstruowanych na podstawie modelu Markowitza, należy przyjąć możliwą do zastosowania w obu sytuacjach miarę. Taką miarą może być odchylenie standardowe rzeczywistej stopy zwrotu *wewnątrz* okresu utrzymywania portfela. Wielkość taka będzie wyznacznikiem stabilności zmian wartości portfela, a więc również oceną związanego z portfelem ryzyka.

Typowym problemem w przypadku obliczania odchylenia standardowego (lub wariancji) stopy zwrotu, jest przyjęcie określonego okresu stopy zwrotu. Rozpatrywanie odchylenia standardowego z miesięcznych stóp zwrotu może prowadzić do innych wniosków niż w przypadku stóp zwrotu, na przykład tygodniowych. Biorąc pod uwagę, że będziemy obliczać odchylenia standardowe w ramach kwartału, przyjęto dwa warianty okresu stopy zwrotu dzienną oraz tygodniową.

Z obliczeń, których wyniki znajdują się w tabeli 5, wynika, że portfele AG-I są praktycznie zawsze lepiej zdywersyfikowane niż odpowiadające im portfele otrzymane w modelu Markowitza, niezależnie od przyjętego wariantu okresu stopy zwrotu. Natomiast wysokie czasami wartości odchylenia standardowego stopy zwrotu z portfeli AG-II są przypuszczalnie konsekwencją konstruowania tych portfeli na podstawie mniejszego zbioru walorów (tylko WIG 20, bez MIDWIG), co w naturalny sposób ogranicza możliwość skutecznej dywersyfikacji.

4. Wnioski

Przeprowadzona analiza wyników portfeli generowanych za pomocą algorytmu genetycznego wskazuje na wysoką skuteczność prezentowanego narzędzia. W całym dwuletnim okresie utrzymywania portfeli rzeczywiste stopy zwrotu, jakie osiągałby inwestor, przewyższały wyraźnie tendencje ogólnorynkowe,

wyrażane zmianami wartości indeksów giełdowych. Uzyskiwana stopa zwrotu była zazwyczaj wyższa niż w przypadku przykładowego zbioru portfeli, utworzonego na podstawie klasycznego modelu Markowitza.

Algorytm genetyczny wykazał się dużą sprawnością techniczną – porównanie wyników uzyskiwanych przez portfele tworzone na bazie różnych zbiorów spółek wykazało brak potrzeby wstępnej selekcji walorów – automatyczna selekcja na poziomie działania algorytmu okazała się wystarczająca. Co więcej, portfele wygenerowane przez algorytm genetyczny odznaczały się większą, niż w przypadku portfeli Markowitza, stabilnością zmian wartości. Świadczyć to może o dobrym poziomie dywersyfikacji ich składu.

Należy jednocześnie pamiętać, że omawiana procedura opiera się wyłącznie na analizie historycznych danych co do zmian notowań poszczególnych walorów. Powiązanie wskazań proponowanej procedury wyznaczania portfela z dodatkową analizą rozpatrywanych spółek za pomocą na przykład analizy fundamentalnej, mogłoby pozwolić wykorzystać szanse czy też uniknąć niebezpieczeństw, które nie zawsze dają się wyodrębnić z obserwacji historycznych.

Autor nadał proponowanej procedurze miano *symulacyjnej procedury wyznaczania portfela papierów wartościowych*, aby wprowadzić odróżnienie w stosunku do klasycznych metod portfelowych. Użycie określenia *symulacyjne* wydaje się uzasadnione ze względu na opieranie wskazań algorytmu na wynikach wielokrotnych „eksperymentów inwestycyjnych" w zróżnicowanych warunkach, a nie na określaniu jakiejkolwiek syntetycznej miary oceny portfela.

Literatura

[1] D. Goldberg, *Algorytmy genetyczne i ich zastosowania*, WNT, Warszawa 1998

[2] T. Gwiazda, *Dywersyfikacja portfela akcji algorytmem genetycznym*, Wydawnictwo Naukowe Wydziału Zarządzania UW, Warszawa 1997

[3] S. Jarek, *Zastosowanie algorytmów genetycznych do tworzenia portfela papierów warościowych*, w: T. Trzaskalik (red.), *Metody i zastosowania badań operacyjnych*, cz. 1, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice 1998

[4] S. Jarek, *Algorytmy genetyczne w dywersyfikacji ilościowej*, w: T. Trzaskalik (red.), *Modelowanie preferencji a ryzyko '99*, cz. 1, Wydawnictwo Akademii Ekonomicznej w Katowicach, Katowice 1999

[5] W. Jurek, *Konstrukcja i analiza portfela papierów wartościowych o zmiennym dochodzie*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2001

[6] Z. Michalewicz, *Algorytmy genetyczne + struktury danych = programy ewolucyjne*, WNT, Warszawa 1996
[7] D. Rutkowska, M. Piliński, L. Rutkowski, *Sieci neuronowe, algorytmy genetyczne i systemy rozmyte*, Wydawnictwo Naukowe PWN, Warszawa 1999

ADAPTATION OF A GENETIC ALGORITHM FOR PORTFOLIO SELECTION PROBLEM

Summary

This paper presents a procedure of building a portfolio of stocks using a genetic algorithm. Selection of stocks is based on historical rates of return in numerous, random periods. Resulting portfolios consist of combination of stocks that would have given relatively high and stable rate of return, if the investment was made in any past period. Empirical analysis is performed, considering long-term sequential portfolio selection on Warsaw Stock Exchange Market. Investment outcomes are compared to market indices and classical portfolio methods results in corresponding timeframe.