

Ekonomia

Optymalizacja w logistyce

tom 1

Modelowanie logistycznych procesów decyzyjnych

Iwona Konarzewska, Maciej Jewczak, Adam Kucharski
redakcja naukowa Iwona Konarzewska



Optymalizacja w logistyce

tom 1

Modelowanie logistycznych procesów decyzyjnych



WYDAWNICTWO
UNIWERSYTETU
ŁÓDZKIEGO

Ekonomia

Optymalizacja w logistyce

tom 1

Modelowanie logistycznych procesów decyzyjnych

Iwona Konarzewska, Maciej Jewczak, Adam Kucharski
redakcja naukowa Iwona Konarzewska



WYDAWNICTWO
UNIwersytetu
ŁÓDZKIEGO

Łódź 2020

Iwona Konarzewska, Maciej Jewczak, Adam Kucharski – Uniwersytet Łódzki
Wydział Ekonomiczno-Socjologiczny, Instytut Logistyki i Informatyki
Katedra Badań Operacyjnych, 90-214 Łódź, ul. Rewolucji 1905 r. nr 37

RECENZENT

Józef Stawicki

REDAKTOR INICJUJĄCY

Beata Koźniewska

REDAKTOR NAUKOWY

Iwona Konarzewska

OPRACOWANIE REDAKCYJNE

Anna Dziadzio

SKŁAD I ŁAMANIE

Munda – Maciej Torz

KOREKTA TECHNICZNA

Leonora Gralka

PROJEKT OKŁADKI

Agencja Reklamowa efectoro.pl

Zdjęcie wykorzystane na okładce: © Depositphotos.com/Peshkova

© Copyright by Authors, Łódź 2020

© Copyright for this edition by Uniwersytet Łódzki, Łódź 2020

Wydane przez Wydawnictwo Uniwersytetu Łódzkiego

Wydanie I. W.09912.20.0.K

Ark. wyd. 13,5; ark. druk. 17,75

ISBN 978-83-8220-112-3

e-ISBN 978-83-8220-113-0

<https://doi.org/10.18778/8220-112-3>

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

Spis treści

Wprowadzenie	9
--------------	---

Rozdział I

Liniowa optymalizacja zagadnień logistycznych	13
1. Poszukiwanie decyzji optymalnej	13
2. Programowanie liniowe	16
3. Algorytm <i>simplex</i> i jego odmiany	29
3.1. Klasyczna wersja algorytmu <i>simplex</i>	29
3.2. Degeneracja rozwiązania zadania PL	40
3.3. Dualizm i analiza wrażliwości	42
3.4. Metoda dwóch faz	56
3.5. Zrewidowana metoda <i>simplex</i>	59
3.6. Dualna metoda <i>simplex</i>	64
4. Parametryczne programowanie liniowe	71
4.1. Parametryzacja współczynników funkcji celu	71
4.2. Parametryzacja prawych stron ograniczeń	77
5. Programowanie liniowe w liczbach całkowitych	82
5.1. Konsekwencje wprowadzenia zmiennych o wartościach całkowitych	82
5.2. Metoda podziału i ograniczeń	85
5.3. Algorytm Gomory'ego	92

Rozdział II

Problemy przydziału	99
1. Natura problemów przydziału	99
2. Algorytm węgierski	101
2.1. Schemat postępowania w algorytmie węgierskim	101
2.2. Algorytm węgierski jako wariant metody <i>simplex</i>	111
3. Rozszerzenia problemu przydziału	113
3.1. Uogólniony problem przydziału GAP	113
3.2. Kwadratowy problem przydziału	114

6 Spis treści

Rozdział III

Zarządzanie projektem w logistyce 117

1. Istota i charakterystyka projektów 117
2. Rola projektów w zarządzaniu przedsiębiorstwem 120
3. Projekty logistyczne jako szczególny rodzaj projektów 124
4. Planowanie sieciowe 126
5. Planowanie przebiegu projektu 133
 - 5.1. Technika CPM 135
 - 5.2. Technika MPM 140
 - 5.3. Technika PERT 144
 - 5.4. Zastosowanie optymalizacji liniowej do wyznaczania ścieżki krytycznej 147
6. Planowanie zasobów projektu 148
 - 6.1. Technika LESS 153
 - 6.2. Zastosowanie optymalizacji liniowej do analizy czasowo-kosztowej projektu 155
 - 6.3. Technika PERT-COST 156
7. Studium przypadku – wdrożenie systemu komputerowego w firmie handlowej 157

Rozdział IV

Teoria masowej obsługi 179

1. Teoretyczne aspekty systemów kolejkowych 179
2. Wybrane modele masowej obsługi 186
 - 2.1. Modele typu $M/M/s/\infty/\infty$ 186
 - 2.2. Modele typu $M/M/s/\infty/\infty$ z uchylonymi założeniami 195
 - 2.3. Model $M/G/1/\infty/\infty$ 203

Rozdział V

Prognozowanie popytu 207

1. Zarys teorii popytu w ekonomii 209
2. Podstawowe pojęcia, mierniki oceny jakości prognoz 217
3. Rodzaje metod prognostycznych wykorzystywanych w prognozowaniu popytu 221
 - 3.1. Kryteria wyboru metod prognostycznych 222
 - 3.2. Dobór odpowiedniej metody prognozowania 223
 - 3.3. Błędy prognoz 226
 - 3.4. Trafność i dopuszczalność prognoz 228
4. Zastosowanie wybranych metod w prognozowaniu popytu i sprzedaży 229
 - 4.1. Prognozowanie szeregów o niskiej zmienności 229
 - 4.2. Prognozowanie szeregów o wysokiej zmienności 232

Bibliografia 243

Dodatek 1. Sposoby konstrukcji prognoz dla wybranych metod 249

Dodatek 2. Mierniki trafności prognoz 253

Dodatek 3. Podsumowanie wyników dla metody wskaźników sezonowości addytywnej i multiplikatywnej	257
Dodatek 4. Wybrane elementy teorii prawdopodobieństwa	265
1. Pojęcia podstawowe	265
2. Prawdopodobieństwo – definicje	265
3. Zmienna losowa i jej wybrane charakterystyki	267
4. Wybrane jednowymiarowe rozkłady prawdopodobieństwa	268
Dodatek 5. Optymalizacja liniowa w środowisku R	271
Spis tabel	279
Spis rysunków	283

Wprowadzenie

Termin „logistyka” obejmuje procesy systemowego planowania i zarządzania łańcuchem dostaw, przepływem produktów oraz informacji. Procesy te mają miejsce zarówno w przypadku wielkich korporacji, jak i małych przedsiębiorstw. Ważną funkcję w logistyce pełni podejmowanie decyzji, które przynoszą danemu podmiotowi maksymalne korzyści, ewentualnie – jak najmniejsze straty. Oczywiście powyższa próba definicji nie obejmuje wszystkich aspektów podejmowania decyzji w procesach logistycznych.

W logistyce wyróżnia się:

- logistykę zaopatrzenia – gdzie lokalizuje się źródła zaopatrzenia, zaopatruje przedsiębiorstwo w niezbędne do produkcji materiały, negocjuje ceny materiałów, kontroluje jakość dostaw, optymalizuje zapasy surowców;
- logistykę dystrybucji – obejmującą planowanie i kontrolę dystrybucji produktów i usług przedsiębiorstwa, utrzymywanie zapasów wyrobów gotowych, lokalizowanie magazynów dystrybucyjnych, organizowanie kanałów dystrybucyjnych, optymalizację dróg transportowych;
- logistykę serwisu – obejmującą zaopatrzenie nabywców w części zamienne oraz obsługę zwrotów towarowych;
- logistykę produkcji – obejmującą planowanie i kontrolę przebiegu produkcji, optymalizację zużycia materiałów oraz odległości przemieszczania się materiałów w procesie produkcji (ten dział logistyki wymaga także przewidywania wielkości popytu na towary i usługi);
- logistykę magazynową – obejmującą magazynowanie części, surowców oraz gotowych produktów;
- logistykę kontraktową – obejmującą operacje prowadzone na powierzchni magazynowej, w tym: przepływy towarów w magazynie, bieżące monitorowanie poziomu stanów magazynowych, zarządzanie towarami, pracą magazynierów i kompletacją dostaw;
- ekologistykę – zajmującą się gromadzeniem, transportem oraz utylizacją odpadów szkodliwych dla środowiska;
- e-logistykę – działania logistyczne związane z wykorzystaniem Internetu oraz systemów informatycznych.

Ze względu na zasięg oddziaływania decyzje logistyczne mogą dotyczyć zarówno pojedynczego przedsiębiorstwa, jak i całego świata. Wyodrębnia się także logistykę społeczną, obejmującą wpływ sieci społecznych oraz przepływów materialnych i informacyjnych na kształtowanie decyzji o charakterze społecznym.

Każdy wymieniony dział logistyki wymaga podejmowania decyzji. Decydenci starają się dokonywać optymalnych i racjonalnych wyborów, poszukując narzędzi wspomagających wszelkie postanowienia. Występujące w logistyce procesy decyzyjne często pokrywają się z tymi, które modelowane są przy pomocy metod badań operacyjnych.

Celem niniejszego tomu jest przybliżenie czytelnikowi metod optymalizacyjnych znajdujących zastosowanie w logistyce (wraz z przykładami) oraz najnowszych wyników badań z tego zakresu.

Rozdział pierwszy obejmuje szeroko rozumiane modelowanie zagadnień logistycznych za pomocą modeli programowania liniowego. Przedstawione zostały w nim metody optymalizacji zagadnień liniowych – zarówno klasyczne (np. metoda *simplex*), jak i ich modyfikacje (metoda dualna czy zrewidowana), często przyspieszające uzyskanie rozwiązania optymalnego. Autorzy starali się zaprezentować te metody w sposób spójny, na podstawie prostych przykładów, a także podkreślić ich ogromny walor, jakim jest możliwość analizy wrażliwości rozwiązań w przypadku niepewności odnośnie parametrów modeli. Parametrami są najczęściej oszacowania czy też prognozy cen lub kosztów, obarczone błędami przewidywań. Ponadto przedstawiono najważniejsze aspekty związane z żądaniem, aby rozwiązania stawianych zadań były całkowitoliczbowe oraz wskazano, jakie komplikacje powodują tak postawione zadania. Zaprezentowano sposób modelowego opisu i rozwiązania problemu parametryzacji funkcji kryterium: uzależnienia wartości współczynników od zadanego, istotnego – z punktu widzenia zastosowań – parametru (np. wielkości marży). Podkreślono również konsekwencje takiego wpływu dla optymalnego rozwiązania postawionego zadania.

Rozdział drugi dotyczy problemów przydziału, tj. sytuacji decyzyjnych pojawiających się w przedsiębiorstwach i innych organizacjach, a dotyczących przyporządkowania zadań do pracowników czy też maszyn oraz typów pojazdów do wykonania konkretnych zadań przewozowych. Klasyczny problem przydziału jest problemem liniowym ze zmiennymi zero-jedynkowymi. Stąd w niniejszej pracy przedstawiony został klasyczny algorytm przydziału, znany jako algorytm węgierski, z kryterium minimalizacji kosztów oraz pewnymi modyfikacjami, dotyczącymi zmiany kierunku optymalizacji (maksymalizacja wielkości sprzedaży, maksymalizacja zysku) i blokowania pewnych przyporządkowań. W rozdziale zaprezentowano również propozycje postępowania w przypadku niezbilansowania zadania oraz uogólnienia problemu. Uogólniony problem przydziału uwzględnia, przykładowo, sytuacje, kiedy każde zadanie musi zostać przydzielone dokładnie jednemu wykonawcy, natomiast każdemu wykonawcy można przydzielić więcej zadań niż jedno, w zależności od wielkości zasobu, którym dysponuje (np. czas pracy). Z kolei kwadratowe zadanie

przydziału jako cel stawia dokonanie przydziału w taki sposób, aby zminimalizować sumę iloczynów przepływu towarów i odległości między obiektami.

Rozdział trzeci omawia w sposób dość szczegółowy problemy związane z planowaniem sieciowym i zarządzaniem projektami. Zaprezentowano w nim metody konstrukcji sieci powiązań czynności przy realizacji złożonych projektów, problemy związane z ich harmonogramowaniem oraz planowaniem kosztów wykonania. Przedstawiono metody deterministyczne, jak np. metodę wyznaczania ścieżki krytycznej CPM czy też metodę PERT, uwzględniającą losowość czasów wykonania czynności; wzięto pod uwagę także dodatkowo koszty wykonywania czynności – stąd prezentacja metody LESS oraz PERT-COST. Na przykładach omówiono, w jaki sposób wykorzystać modelowanie liniowe do uzyskiwania optymalnych harmonogramów dla czynności składających się na projekt w obu przypadkach.

Rozdział czwarty poświęcony został teoretycznym aspektom teorii masowej obsługi (teorii kolejek). Sprawność obsługi klienta, np. przez sklep internetowy, jest ważnym narzędziem w walce konkurencyjnej, przyczyniającym się do podjęcia ostatecznej decyzji, m.in. przy wyborze dostawcy. Inne jej zastosowania w logistyce to, chociażby, kolejkovanie zleceń produkcyjnych oczekujących na realizację czy zamówień na części składowane w magazynie, a także pojazdów, które mają zostać załadowane lub rozładowane. Teoria masowej obsługi pomaga określić potencjał stanowiska pracy, poziom jego efektywności; pokazuje, jakie działania należy wdrożyć, aby zwiększyć satysfakcję uczestników systemu obsługi. W rozdziale przedstawione i omówione zostały także najważniejsze stosowane w zagadnieniach logistycznych modele oraz ich własności.

W rozdziale piątym, obok elementów teorii popytu, zaprezentowano wybrane metody prognozowania, w tym metody wygładzania wykładniczego dla danych o niskiej i średniej zmienności. Scharakteryzowano mierniki jakości prognoz, problemy błędów prognozowania oraz kryteria wyboru odpowiednich metod prognozowania.

Tom uzupełniony został o dodatki, w tym dodatek zawierający omówienie doświadczeń Autorów w zakresie modelowania optymalizacyjnego z użyciem pakietów środowiska oprogramowania R oraz wybrane elementy teorii prawdopodobieństwa.

Przedstawiony w niniejszym tomie zestaw metod i przykładów gromadzi w jednej pozycji szerokie spektrum narzędzi pozwalających przeprowadzić optymalizację elementów łańcucha logistycznego. Znajdziemy w nim omówienie modeli rzadko spotykanych w ogólnodostępnej literaturze – łącznie z zaprezentowaniem przykładów ich zastosowań w logistyce. Autorzy dokonali starań, aby język pracy był jednocześnie precyzyjny (matematycznie) i przystępny. W przygotowywanym tomie drugim monografii, zatytułowanym *Optymalizacja w logistyce. Modelowanie przepływów w kanałach dystrybucji*, znajdziemy m.in. omówienie problemów modelowania zapasów, optymalizacji w szeroko rozumianych zagadnieniach transportowych, problemów szeregowania zleceń czy też optymalizacji przepływów produkcyjnych.

Rozdział I

Liniowa optymalizacja zagadnień logistycznych

1. Poszukiwanie decyzji optymalnej

Prowadzenie działalności gospodarczej wymaga podejmowania decyzji. Aby były one racjonalne, muszą opierać się na solidnych podstawach, m.in. na jasno sprecyzowanym celu i znajomości środków koniecznych do jego realizacji. Ponadto funkcjonowanie na konkurencyjnym rynku oznacza konieczność brania pod uwagę wpływu otoczenia na decyzje. Te zaś, w dzisiejszych czasach, podejmuje się coraz szybciej. Menedżera obciąża przy tym odpowiedzialność – gdyż albo nie może on cofnąć swoich działań, albo wiązałoby się to z poniesieniem dużych kosztów. Dlatego tak ważna staje się precyzja i wybór decyzji (spośród wielu innych możliwych w danych warunkach) optymalnej lub chociaż zadowalająco bliskiej optymalności. Tym właśnie zajmują się badania operacyjne. Obszar ich zastosowań obejmuje sporządzanie matematycznych, ekonomicznych i statystycznych opisów (modeli) procesów decyzyjnych, charakteryzujących się dużą złożonością (i często niepewnością). Takie opisy umożliwiają precyzyjne analizowanie złożonych procesów decyzyjnych i ułatwiają podjęcie najlepszej decyzji (Miszczyński i Miszczyńska, 2002).

Logistyka dostarcza wielu danych o ilościowym charakterze, które dają się ująć w ramy metod matematycznych i statystycznych. Może to być koordynacja przepływu zasobów w łańcuchu dostaw, realizacja zadań transportowych czy zarządzanie zapasami. Pojawiające się w logistyce problemy decyzyjne często pokrywają się z tymi, jakie występują w ramach badań operacyjnych. Jednym z nich jest poszukiwanie decyzji optymalnej przy znanym kryterium wyboru oraz otoczeniu opisanym formułami matematycznymi.

Początki badań operacyjnych sięgają II wojny światowej, choć różni autorzy podają odmienne wydarzenia jako moment narodzin tej dziedziny nauki. Wymienia się m.in. budowę statków transportowych typu Liberty, podczas której naukowe podejście do usprawniania produkcji silnie wykazało swoją przydatność. Powstające maszyny charakteryzowały się znacznie skróconym czasem budowy oraz zmniejsz-

szonym zużyciem surowców w porównaniu z początkowym okresem wprowadzania ich do służby – była to zasługa czynionych na bieżąco analiz naukowych.

II wojna światowa okazała się konfliktem, w czasie którego nauka rozwijała się szybko i na wielu płaszczyznach. Naukowcy pracujący na potrzeby projektów wojskowych, realizowanych w USA i Wielkiej Brytanii, mieli za zadanie zadbać o efektywną alokację zasobów na potrzeby operacji wojskowych oraz rozwijać nowe rodzaje broni. Ich badania wiązały się z różnego rodzaju operacjami militarnymi. Stąd wzięła się nazwa, która przetrwała do dziś, mimo że obecnie badania operacyjne znajdują szerokie zastosowanie również (a może przede wszystkim) poza wojskowością.

Boom gospodarczy, jaki pojawił się w USA po zakończeniu wojny, sprawił, że zapotrzebowanie na opracowane dopiero co metody wzrosło. Produkcja przemysłowa stawała się coraz bardziej złożona i wyspecjalizowana, rosła skala działań logistycznych. Klasyczne podejście do zarządzania, opierające się na osobistym doświadczeniu menedżera, przestało wystarczać. Problemy, przed którymi stanęli menedżerowie były podobne do tych z czasu wojny, choć oczywiście zmienił się kontekst. Pojawił się popyt na metody usprawniające proces podejmowania decyzji. Wielu naukowców po zakończeniu pracy dla armii przeszło do biznesu – przybywały nowe rozwiązania i opracowania naukowe na ten temat. Nie bez znaczenia był tu również szybki rozwój komputerów (i ich upowszechnienie).

Można powiedzieć, że badania operacyjne mają zastosowanie do problemów, w których pojawia się potrzeba przeprowadzenia i koordynowania „operacji” w ramach pewnej organizacji (Hillier i Lieberman, 2015, s. 3). Poziom operacyjny oznacza, że efekty podejmowanych decyzji poznamy w nieodległej przyszłości. Implikuje to bliski związek z zarządzaniem a przynajmniej zmusza do przyjęcia punktu widzenia konkretnej organizacji. Z tego powodu badania operacyjne postrzegane są też jako część mikroekonomii. Gdyby przejrzeć katalog dostępnych metod, przekonalibyśmy się o szerokim spektrum zastosowań tego typu badań. Zaliczymy do nich nie tylko produkcję przemysłową, ale również finanse, transport (i ogólnie logistykę), ochronę zdrowia, telekomunikację i wiele innych.

Konieczność podejmowania decyzji wiąże się z wystąpieniem problemu decyzyjnego i wyborem takiej decyzji, która doprowadzi do najlepszego w danych warunkach sposobu działania, zmierzającego do osiągnięcia postawionego celu (Rogalska i inni, 1991, s. 9). Z problemem decyzyjnym mamy do czynienia wtedy, kiedy zachodzą następujące okoliczności:

- pojawia się decydent (osoba lub grupa osób), który musi rozwiązać problem;
- decydent chce osiągnąć jakiś cel;
- istnieją co najmniej dwa sposoby na osiągnięcie zakładanego celu;
- istnieje otoczenie, które wpływa na sposób rozwiązania problemu lub jego wynik.

Sam problem decyzyjny przedstawia się w sposób symboliczny, jako model, który przybiera jedną z trzech postaci:

- 1) ikonyczną (obrazową) – przedstawiającą przedmioty lub zdarzenia w zmiennej skali (np. mapa);
- 2) analogową – przedstawiającą właściwości badanego zjawiska za pomocą własności innych zjawisk;
- 3) symboliczną (matematyczną) – opisującą zjawisko za pomocą zależności matematycznych (równań lub nierówności).

Biorąc pod uwagę ostatnią z wymienionych postaci, wyróżnimy następujące etapy poszukiwania rozwiązania: sformułowanie problemu decyzyjnego, wybór postaci modelu, zebranie potrzebnych danych, budowa modelu i jego rozwiązanie, weryfikacja wyników, wdrożenie decyzji wskazanych przez model. Proces podejmowania decyzji może zostać wsparty przez odpowiednio skonstruowane modele matematyczne, zwane modelami decyzyjnymi. Zazwyczaj znajdujemy się w sytuacji, w której należy dokonać wyboru pomiędzy wieloma możliwymi decyzjami, zwanymi decyzjami dopuszczalnymi.

W przeciwieństwie do podręcznikowych przykładów, większość praktycznych problemów jest przez decydentów początkowo opisywana w niezbyt precyzyjny, słabo sformalizowany sposób. Rzadko bowiem menedżer osobiście zajmuje się budową modelu. Dlatego to analityk (lub inna osoba odpowiedzialna za tę pracę) musi uściślić elementy składowe. Na początku ustala on przesłanki, na podstawie których oceniane będą decyzje. Zakładamy bowiem, że decydent posługuje się pewnym kryterium, pozwalającym mu odróżnić decyzje lepsze od gorszych, i że kryterium to da się przedstawić w postaci funkcji, zwanej funkcją celu. Zupełnie normalne jest więc to, że dwóch różnych decydentów będzie mieć odmienne kryteria oceny tych samych działań.

Kiedy już znamy kryterium wyboru, przechodzimy do ustalenia listy zmiennych sterowanych (decyzyjnych) oraz niesterowanych (parametrów) i określenia warunków, jakie powinna spełniać podjęta decyzja. Na koniec analityk musi przełożyć zebrane informacje na matematyczną wersję modelu. Pomocne w tym będzie programowanie matematyczne zajmujące się tworzeniem algorytmów rozwiązywania określonych klas problemów optymalizacyjnych (Sikora i inni, 2008, s. 10). Słowo „programowanie” nie odnosi się do programowania komputerowego – jest raczej synonimem planowania.

Metody ilościowe wymagają danych do przeprowadzania obliczeń. W przypadku modeli optymalizacyjnych gromadzenie danych liczbowych sprowadza się do poznania wartości określonych parametrów. Na ich podstawie wyznacza się poszukiwane wartości zmiennych decyzyjnych.

Same decyzje podejmowane mogą być w różnych warunkach. W sytuacji pewności każdej decyzji odpowiada tylko jeden wynik z prawdopodobieństwem równym 1. Powiemy wtedy, że proces jest zdeterminowany a parametry modelu są znane i stałe. W warunkach niepewności każdej decyzji odpowiada więcej niż jeden wynik (mówimy, że jest to element stochastyczny). Nie znamy jednak prawdopodobieństw, z jakim dany wynik może wystąpić. Parametry przyjmują bowiem różne wartości, w zależności od zachowania otoczenia. Z kolei podczas podejmowania decyzji

w warunkach ryzyka znamy prawdopodobieństwa powyższych wyników. Wyróżnia się tu ponadto podejmowanie decyzji w warunkach częściowej informacji – każdej decyzji nadal odpowiada więcej niż jeden wynik. Nie znamy co prawda prawdopodobieństwa jego wystąpienia, ale możemy je oszacować dzięki znajomości niektórych charakterystyk nieznanego rozkładu prawdopodobieństwa.

W ramach naszych rozważań posługiwać się będziemy określoną terminologią o matematycznym charakterze. I tak, zbiór (mierzalnych) decyzji do podjęcia reprezentowany będzie przez zmienne decyzyjne. To właśnie ich wartości poszukujemy. Efekty podjętych działań wyraża funkcja kryterium (funkcja celu) określona na wspomnianych zmiennych. Wszelkie ograniczenia wartości, które można przypisać zmiennym decyzyjnym, są również wyrażane matematycznie – zwykle za pomocą nierówności lub równań zwanych warunkami ograniczającymi lub po prostu ograniczeniami. Stałe liczbowe, występujące w funkcji celu oraz w ograniczeniach, będą parametrami modelu – ich wartości ustala się podczas jego tworzenia. Stajemy więc przed problemem wyboru takich wartości zmiennych decyzyjnych, które zapewnią maksymalną (minimalną) wartość funkcji celu¹, pod warunkiem spełnienia narzuconych ograniczeń.

Klasyfikując modele w najbardziej ogólny sposób, podzielimy je na liniowe i nieliniowe. W naszym opracowaniu skupimy się na modelach liniowych, gdyż posiadają one wiele zalet. Po pierwsze dają szerokie możliwości interpretacji. Po drugie są dobrze zbadane, ponieważ naukowcy zajmują się nimi od kilkudziesięciu lat. Po trzecie część modeli nieliniowych da się sprowadzić (i sprowadza się je) do postaci liniowej. Mimo swojej prostoty, modele liniowe sprawdzają się w zaskakująco wielu obszarach.

Zapisany w formie matematycznej liniowy model decyzyjny przyjmuje postać modelu programowania liniowego. Zarówno funkcja celu, jak i wszystkie ograniczenia wyrażone są przy pomocy funkcji liniowych. Zatem programowanie liniowe (PL) obejmuje planowanie działań w celu uzyskania optymalnego rezultatu, tj. takiego, który zapewnia osiągnięcie założonego celu w najlepszy możliwy sposób (zgodnie z modelem matematycznym), spośród wszystkich alternatyw.

2. Programowanie liniowe

Przeanalizujemy następujący przykład problemu, przed którym może stanąć decydent. Jest to celowo uproszczona sytuacja, ograniczona do najbardziej istotnych – z punktu widzenia procesu decyzyjnego – elementów. Posłuży ona do zilustrowania koncepcji budowy modelu programowania liniowego w dalszej części rozdziału.

¹ Domyślnie w tym rozdziale przyjmujemy, że decydent posługuje się pojedynczym kryterium.

Przykład 1

Firma Paralel S.A. wytwarza części do sprzętu gospodarstwa domowego na zamówienie zewnętrznych kontrahentów. Zarząd firmy rozważa podjęcie produkcji filtrów montowanych w dwóch nowych modelach odkurzaczy tego samego producenta. Odbłyły się już negocjacje między Paralel S.A. a producentem sprzętu. Wiadomo więc, że filtry do pierwszego modelu sprzedawane będą po siedemdziesiąt złotych za sztukę, a do drugiego – po czterdzieści. Biorąc pod uwagę stan realizacji bieżących zamówień, zarząd Paralel S.A. jest zdecydowany na podjęcie produkcji. Musi jednak określić jej docelową wielkość i strukturę w taki sposób, aby osiągnąć jak największy przychód ze sprzedaży. Kontrahent gwarantuje odbiór każdej liczby wyprodukowanych filtrów pod warunkiem, że powstaną one najpóźniej w ciągu najbliższych sześciu miesięcy. Wtedy bowiem zaplanowano wprowadzenie odkurzaczy do sprzedaży.

Opierając się na doświadczeniach z realizacji wcześniejszych zamówień oraz dysponując dokumentacją techniczną, pracownicy Paralel S.A. oszacowali, że jeden filtr do modelu pierwszego wymaga zużycia dwóch dekagramów tworzywa sztucznego, zaś specjalistyczna maszyna potrzebuje jednej minuty, aby uformować wymagany kształt. Filtr do drugiego modelu powstawać ma z jednego dekagrama tego samego tworzywa; na jego uformowanie potrzeba dwóch minut. Specyficzny proces technologiczny sprawia, że filtry opuszczają linię produkcyjną pojedynczo. Na potrzeby realizacji kontraktu zgromadzono dwieście pięćdziesiąt kilogramów tworzywa – i wielkość ta nie zostanie zwiększona. Dostępny limit czasu pracy maszyn wynosi dwadzieścia tysięcy minut – ze względu na to, że Paralel S.A. ma zobowiązania wynikające z wcześniej podpisanych umów. Każdy z filtrów, niezależnie od modelu, zawiera wkład wykonany z włókniny o masie trzech dekagramów. Aby zapewnić dokładność filtrowania zanieczyszczeń, wymaganą przez zamawiającego, należy zużyć łącznie co najmniej sto pięćdziesiąt kilogramów włókniny.

Rozłóżmy powyższy przykład na elementy składowe problemu decyzyjnego. Decydentem w nim jest zarząd firmy Paralel S.A. Jego celem jest osiągnięcie jak najwyższych przychodów wynikających z podjętej decyzji o produkcji obu rodzajów filtrów. Nie nakładamy przy tym ograniczeń na wielkość produkcji. Zakładamy jedynie, że będzie to liczba nieujemna. Możliwe więc, że najbardziej korzystne okaże się wytwarzanie tylko jednego rodzaju filtra, a być może najlepsze wyniki finansowe zapewni rozłożenie produkcji na oba wyroby.

Jak już jednak wspomnieliśmy, niekoniecznie to sam zarząd układa model. Analityk (lub analitycy) z pewnością zwróciłby (zwróciliby) uwagę na otoczenie wpływające na podejmowanie decyzji. W naszym przykładzie sprowadza się ono do procesu technologicznego, czyli zużycia ograniczonej ilości surowców i wykorzystania dostępnego czasu pracy. W ten sposób wyodrębniliśmy elementy składowe i możemy przejść do budowy modelu matematycznego, a następnie znalezienia jego rozwiązania.

Problem decyzyjny z przykładu 1 posłuży do zilustrowania procesu tworzenia typowego (choć niedużych rozmiarów) zadania programowania liniowego. Miejmy jednak na uwadze, że jest to narzędzie zbyt wszechstronne, aby można je było w pełni opisać za pomocą jednego przykładu. Dlatego kolejne elementy i własności będą sukcesywnie dodawane w dalszej części rozdziału.

Każdy model PL spełnia następujące założenia:

- wszystkie parametry modelu, tj. zasoby, ceny, technologie, są znane i z góry ustalone (deterministyczne);
- technologia jest stała, a zatem wymagania produkcyjne, jeżeli problem dotyczy optymalizacji planu produkcji, są ustalone w czasie planowania i nie ulegną zmianie w analizowanym okresie;
- zmienne decyzyjne są zmiennymi ciągłymi i nieujemnymi;
- funkcja kryterium i funkcje definiujące ograniczenia są liniowe – oznacza to, że ten sam, co do wielkości, przyrost zmiennej, bez względu na początkowy poziom, powoduje zawsze taki sam przyrost wartości funkcji (*constant returns to scale*);
- liniowość funkcji kryterium i funkcji definiujących ograniczenia oznacza proporcjonalność i addytywność.

Z matematycznego punktu widzenia wystarczą założenia mówiące, że model musi mieć liniową funkcję celu, podlegającą również liniowym ograniczeniom. Jednak, z punktu widzenia modelowania problemu decyzyjnego, warto sformułować dodatkowe założenia, które są w zasadzie implikowane przez założenia o liniowości funkcji kryterium i ograniczeń, a dotyczą pewnych konsekwencji dla modelowanego problemu.

Wymieniona w ostatnim założeniu proporcjonalność dotyczy zarówno funkcji celu, jak i jej ograniczeń. Efekt wpływu każdej ze zmiennych decyzyjnych na łączną wartość funkcji celu jest proporcjonalny do poziomu tej zmiennej. Podobnie rzecz się ma z wartością lewej strony każdego z ograniczeń. Założenie to wyklucza wykładnik inny niż równy 1 dla dowolnej zmiennej w modelu, niezależnie od tego, czy rozważamy funkcję celu, czy funkcję po lewej stronie ograniczenia. Dodatkowo zakładamy, że każda funkcja w modelu programowania liniowego jest sumą udziałów poszczególnych zmiennych. Nazywamy to warunkiem addytywności.

Przybliżenia i uproszczenia, obecne w powyżej sformułowanych założeniach, są potrzebne, aby w ogóle dało się model zastosować. Dodanie zbyt wielu szczegółów może sprawić, że model będzie zbyt skomplikowany, aby przeprowadzić przydatną analizę problemu. To, czego naprawdę potrzebujemy, to istnienie odpowiednio wysokiej korelacji między wynikami uzyskanymi na podstawie modelu a tym, co faktycznie wydarzy się w rzeczywistości. W realnych zastosowaniach powszechną jest sytuacja, w której niektóre z założeń nie są całkowicie zachowane i musimy pogodzić się z niedużymi odchyleniami. Poważne naruszenie któregoś z założeń oznacza konieczność sięgnięcia po inne, zwykle bardziej skomplikowa-

ne metody. Niektóre z nich nie dają takich możliwości analitycznych i interpretacyjnych, jak modele liniowe – choć badacze cały czas pracują nad zmniejszeniem tej różnicy.

Na potrzeby przykładu 1 zdefiniujemy następujące zmienne decyzyjne:

- x_1 – wielkość produkcji filtrów modelu 1 [szt.],
- x_2 – wielkość produkcji filtrów modelu 2 [szt.].

Znając ceny sprzedaży, konstruujemy funkcję przychodu (cel decydena), której maksymalnej wartości poszukujemy:

$$f(x_1, x_2) = 70x_1 + 40x_2 \rightarrow \max \quad (1)$$

Ograniczenia wynikające z procesu technologicznego są opisywane przez nierówności (2)–(4). Dołączamy do nich warunki brzegowe (5), gwarantujące, że żadna ze zmiennych decyzyjnych nie przyjmie wartości ujemnej.

$$\text{(czas [min])} \quad x_1 + 2x_2 \leq 20\,000 \quad (2)$$

$$\text{(tworzywo [dag])} \quad 2x_1 + x_2 \leq 25\,000 \quad (3)$$

$$\text{(wkład [dag])} \quad 3x_1 + 3x_2 \geq 15\,000 \quad (4)$$

$$x_1 \geq 0, x_2 \geq 0 \quad (5)$$

Zapiszmy ogólną postać modelu programowania liniowego.

$$f(\mathbf{x}) = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max/\min \quad (6)$$

$$\begin{cases} a_{11}x_{11} + a_{12}x_{12} + \dots + a_{1n}x_{1n} \leq b_1 \\ \vdots \\ a_{k1}x_{k1} + a_{k2}x_{k2} + \dots + a_{kn}x_{kn} = b_k \\ \vdots \\ a_{m1}x_{m1} + a_{m2}x_{m2} + \dots + a_{mn}x_{mn} \geq b_m \end{cases} \quad (7)$$

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0 \quad (8)$$

Symbolem c_j ($j = 1, \dots, n$) oznaczać będziemy parametry funkcji celu, x_j ($j = 1, \dots, n$) – zmienne decyzyjne, a_{ij} ($i = 1, \dots, m; j = 1, \dots, n$) – parametry lewych stron ograniczeń, zaś b_i ($i = 1, \dots, m$) – wartości prawych stron ograniczeń.

Interpretacja modeli PL nie ogranicza się tylko do podania wartości zmiennych decyzyjnych, dla których funkcja celu osiąga wartość największą lub najmniejszą. Znając rozwiązanie optymalne (oczywiście, o ile takie istnieje), jesteśmy w stanie odpowiedzieć na następujące pytania:

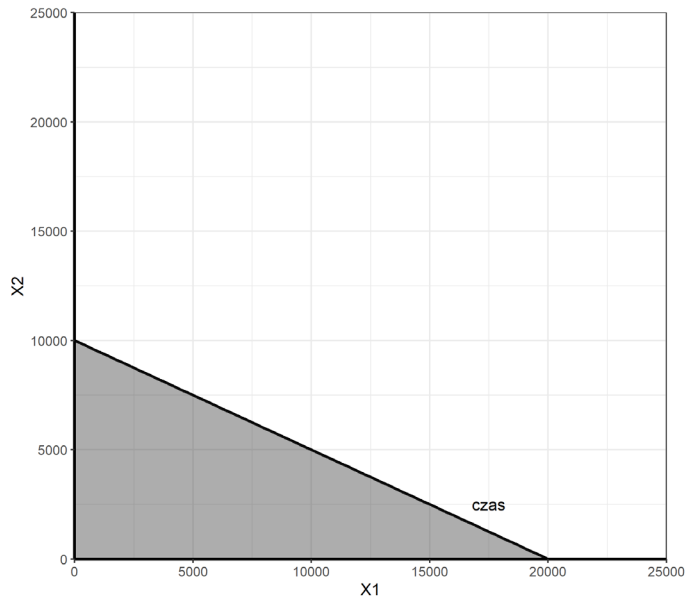
1. Jak zareaguje optymalna wartość funkcji celu na marginalne zmiany wybranego środka (limitu) reprezentowanego przez określony wyraz wolny w ograniczeniach?
2. Jakie są granice zmian przedziałów dopuszczalnych zmian w zasobach środków (limitów), dla których siła i kierunek reakcji, wymienionej w poprzednim punkcie, pozostają bez zmian?
3. Jakie są granice przedziałów dopuszczalnych zmian dla współczynników funkcji celu, które nie powodują zmiany rozwiązania optymalnego?
4. Jak zareaguje rozwiązanie optymalne na dołączenie lub usunięcie zmiennej decyzyjnej?
5. Jak zareaguje rozwiązanie optymalne na dołączenie lub usunięcie ograniczenia?

Jeżeli problem można opisać przy pomocy tylko dwóch zmiennych decyzyjnych, wówczas da się go też przedstawić, a nawet rozwiązać graficznie, w prostokątnym układzie współrzędnych. Warunki brzegowe narzucają konieczność ograniczenia się do jego pierwszej ćwiartki. Wyznaczamy część wspólną nierówności i równań tworzących ograniczenia modelu. Będzie to (matematycznie rzecz ujmując) iloczyn wszystkich półpłaszczyzn i prostych odpowiadających nierównościom i równaniom tworzącym ograniczenia zadania PL. Jeśli ograniczenia występują wyłącznie jako nierówności (jak to ma miejsce w przykładzie 1), zbiór rozwiązań dopuszczalnych jest wielobokiem wypukłym.

Chcąc naszkicować dane ograniczenie w pierwszej ćwiartce układu współrzędnych, zazwyczaj zaczynamy od wyznaczenia punktów przecięcia prostej odcinającej część płaszczyzny w ramach danego ograniczenia z osiami. Dla ograniczenia (2) punkty przecięcia z osiami mają współrzędne odpowiednio: (0, 10000) i (20000, 0). Na rysunku 1 umieszczono to ograniczenie, zaznaczając obszar, którego punkty spełniają obowiązujący limit.

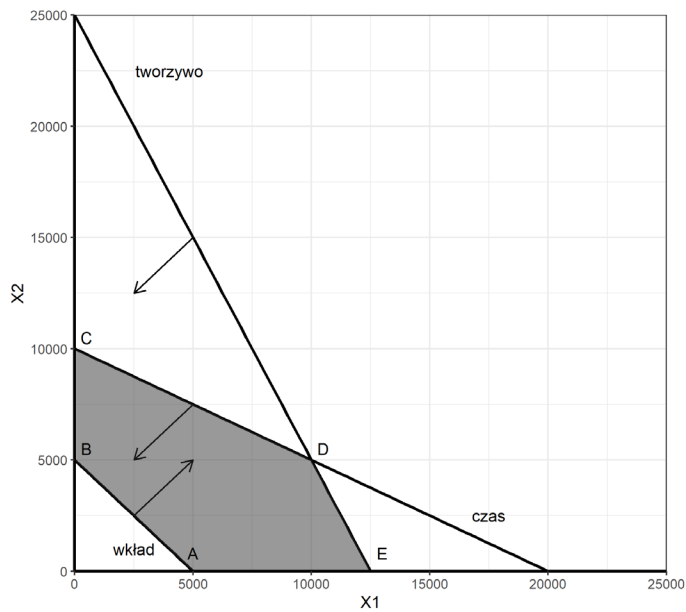
W ten sam sposób dodajemy do wykresu następne ograniczenia. Każda nowa półpłaszczyzna odcina odpowiedni fragment pierwszej ćwiartki układu, przy czym – obszary te częściowo lub całkowicie pokrywają się ze sobą. W efekcie tworzy się zbiór punktów spełniających wszystkie ograniczenia jednocześnie. Na rysunku 2 reprezentuje go wielobok *ABCDE*. Strzałki pokazują, gdzie znajdują się punkty spełniające dane ograniczenie. Zbiór punktów spełniających wszystkie ograniczenia modelu nosi nazwę zbioru rozwiązań dopuszczalnych – i oznaczmy go jako *X*. Co istotne, do zbioru tego należą także jego krawędzie – w modelach optymalizacyjnych praktycznie nie występują ostre nierówności. Ma to niebagatelne znaczenie przy wyznaczaniu rozwiązania optymalnego. Poszukiwana przez nas optymalna decyzja odnośnie wielkości i struktury produkcji musi być dopuszczalna. Z pewnością należy więc do zbioru *X*.

Jeżeli istnieje choć jedna decyzja dopuszczalna, tj. jeżeli zbiór decyzji dopuszczalnych jest niepusty, wówczas zadanie PL okazuje się zadaniem niesprzecznym.



Rys. 1. Zbiór punktów spełniających ograniczenie (2)

Źródło: opracowanie własne.



Rys. 2. Zbiór rozwiązań dopuszczalnych dla przykładu 1

Źródło: opracowanie własne.

Rozwiązań dopuszczalnych, którym odpowiadają punkty na zacięniowanym obszarze rysunku 2, jest nieskończenie wiele. Należy znaleźć w tym zbiorze punkt będący rozwiązaniem optymalnym. Pomocne będą dwa poniższe twierdzenia (Miszczuński i Miszczuńska, 2002, s. 38):

Twierdzenie 1

Zbiór rozwiązań dopuszczalnych zadania PL jest zbiorem domkniętym, wypukłym, o skończonej liczbie wierzchołków.

Twierdzenie 2 (Weierstrassa)

Funkcja liniowa $f(\mathbf{x})$ określona na domkniętym zbiorze wypukłym X , o skończonej liczbie wierzchołków, osiąga swoją wartość największą (najmniejszą) w wierzchołku tego zbioru. Jeżeli wartość taką osiąga w więcej niż jednym wierzchołku, to osiąga ją również w każdym punkcie $\mathbf{x} \in X$, będącym wypukłą kombinacją liniową takich wierzchołków.

Drugie z przytoczonych twierdzeń znacząco ogranicza liczbę punktów, jakie będziemy brali pod uwagę, poszukując maksimum funkcji celu w przykładzie 1. Naturalnym podejściem wydaje się wyznaczenie współrzędnych wszystkich wierzchołków, podstawienie do funkcji przychodu i sprawdzenie, dla którego z nich osiąga ona wartość największą. Jest to tzw. przegląd zupełny wierzchołków. Tabela 1 ilustruje takie właśnie podejście do poszukiwania wierzchołka optymalnego.

Tab. 1. Współrzędne wierzchołków i wartości przychodu dla przykładu 1

Wierzchołek	Współrzędna x_1	Współrzędna x_2	Wartość funkcji celu
A	5000	0	350 000
B	0	5000	200 000
C	0	10 000	400 000
D	10 000	5000	900 000
E	12 500	0	875 000

Źródło: opracowanie własne.

Łatwo zauważyć, że w wierzchołku D występuje największy przychód, a zatem to on jest rozwiązaniem optymalnym zadania. Decydent powinien zatem wyprodukować dziesięć tysięcy sztuk filtrów modelu pierwszego i pięć tysięcy sztuk modelu drugiego, osiągając przy tym przychód na poziomie dziewięciuset tysięcy złotych.

Sprawdzanie wartości funkcji celu, w każdym z wierzchołków zbioru rozwiązań dopuszczalnych, nie jest metodą zbyt wydajną. Szczególnie, gdy liczba takich wierzchołków jest duża, a tak będzie w problemach o większej liczbie ograniczeń.

Istnieje jednak inna metoda poszukiwania rozwiązania optymalnego, opierająca się na tzw. gradiencie funkcji celu (Sikora i inni, 2008, s. 27–29).

Gradientem funkcji celu jest wektor, którego składowymi są pochodne cząstkowe $f(\mathbf{x})$ względem zmiennych decyzyjnych. Pokazuje lokalnie kierunek najszybszego wzrostu wartości $f(\mathbf{x})$. Formalnie zapiszemy go następująco:

$$\nabla_{f(\mathbf{x})} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \quad (9)$$

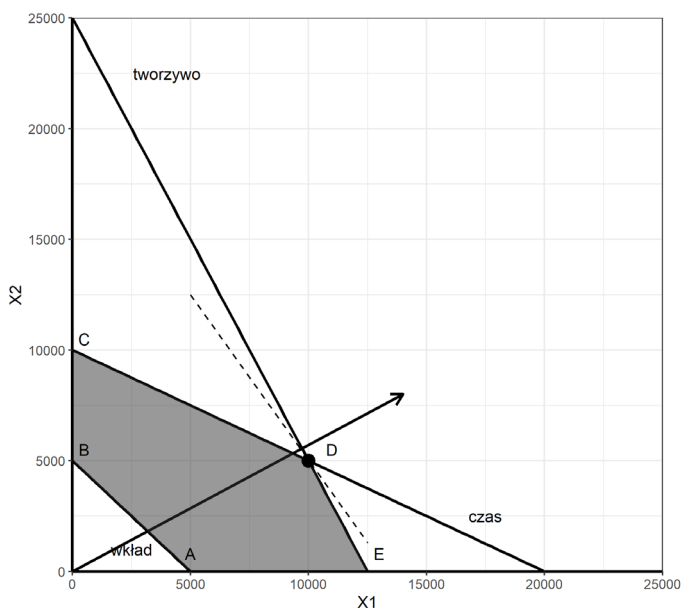
Kiedy funkcja ma postać liniową, gradient jest wektorem złożonym ze współczynników funkcji celu, a wskazujący przez niego kierunek najszybszego wzrostu wartości $f(\mathbf{x})$ okazuje się niezależny od wartości zmiennych. Gradient funkcji celu dla przykładu 1 jest zatem następujący:

$$\nabla_{f(\mathbf{x})} = \begin{bmatrix} 70 \\ 40 \end{bmatrix} \quad (10)$$

Wektor gradientu nanosi się na wykres, zaczepiając go w początku układu współrzędnych. Istotny dla rozwiązania jest kierunek wskazywany przez ten wektor – to kierunek najszybszego wzrostu wartości funkcji kryterium. Biorąc pod uwagę skalę na rysunku 2, rozsądnie będzie przeskalować ten wektor, tj. pomnożyć go przez stałą. Poprawi to precyzję analizy graficznej, a nie wpłynie na rozwiązanie.

Sam gradient nie wystarczy do znalezienia poszukiwanego wierzchołka. Dlatego na wykresie umieszczamy dodatkowo warstwicę (izokwantę funkcji kryterium), czyli prostą prostopadłą do półprostej wyznaczonej przez gradient. Oczywiście takich warstwic możemy wykreślić dowolnie wiele, ale nas interesują tylko te z nich, które na rysunku przechodzą przez zbiór rozwiązań dopuszczalnych. Poszukując największej wartości $f(\mathbf{x})$ (izokwanta odpowiadająca maksymalnej wartości funkcji kryterium), przesuwamy warstwicę równolegle, zgodnie ze zwrotem wektora gradientu (czyli oddalając się od początku układu współrzędnych), tak długo, dopóki nie dotrzemy do najdalszego punktu dopuszczalnego (wierzchołka) przed opuszczeniem zbioru X – to właśnie wierzchołek będący rozwiązaniem optymalnym. Poszukując wartości najmniejszej, postępujemy podobnie, z tą różnicą, że warstwicę przesuwamy w kierunku przeciwnym do zwrotu wektora (poszukujemy izokwanta odpowiadającej minimalnej wartości funkcji kryterium).

Na rysunku 3 warstwicę, odpowiadającą maksymalnej wartości funkcji kryterium zadania, zaznaczono przerywaną linią. Jest to prosta najbardziej oddalona od początku układu współrzędnych, mająca jednocześnie punkt wspólny ze zbiorem rozwiązań dopuszczalnych. Ten punkt wspólny to wierzchołek D – i właśnie on wyznacza maksymalną wartość funkcji celu, a jego współrzędne – rozwiązanie optymalne.



Rys. 3. Gradient i warstwica dla przykładu 1

Źródło: opracowanie własne.

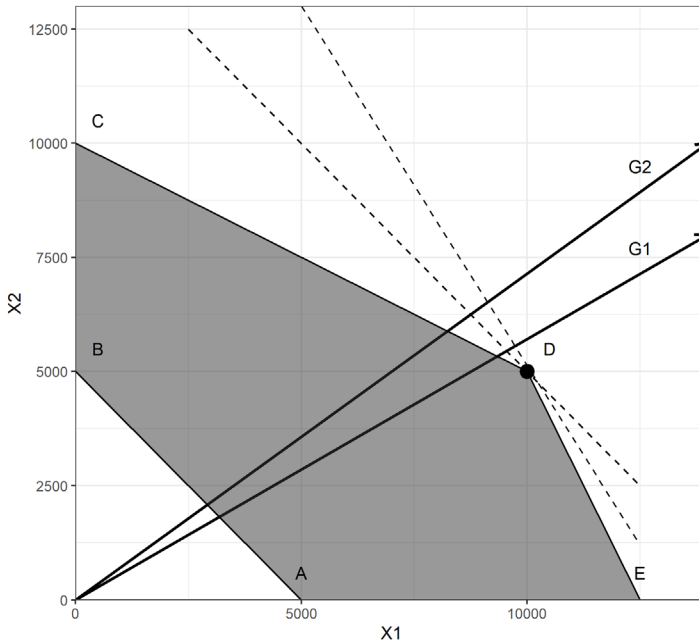
Wyzaczyliśmy interesujące nas optymalne wartości zmiennych decyzyjnych i odpowiadającą im wartość kryterium. Na tym nie kończą się możliwości analizy. W tabeli 2 umieszczone zostały wartości lewych stron ograniczeń oraz moduły różnic między prawą a lewą stroną każdego z nich. Interesuje nas ostatnia kolumna. Dowiadujemy się z niej, że limit czasu pracy oraz aktualnie posiadany zapas tworzywa zostały wykorzystane w całości. Lewe strony tych ograniczeń równają się prawym. Inną sytuację obserwujemy w ograniczeniu opisującym wykorzystanie wkładu z włókniny. Zakładaliśmy w jego przypadku wykorzystanie co najmniej stu pięćdziesięciu kilogramów. Tymczasem wartość modułu różnicy wskazuje, że wielkość ta została przekroczona o trzysta kilogramów i faktyczne zużycie włókniny wyniosło czterysta pięćdziesiąt kilogramów.

Tab. 2. Wartości ograniczeń po podstawieniu rozwiązania optymalnego

Ograniczenie	Lewa strona	Kierunek	Prawa strona	Moduł różnicy
Czas	20 000	\leq	20 000	0
Tworzywo	25 000	\leq	25 000	0
Wkład	45 000	\geq	15 000	30 000

Źródło: opracowanie własne.

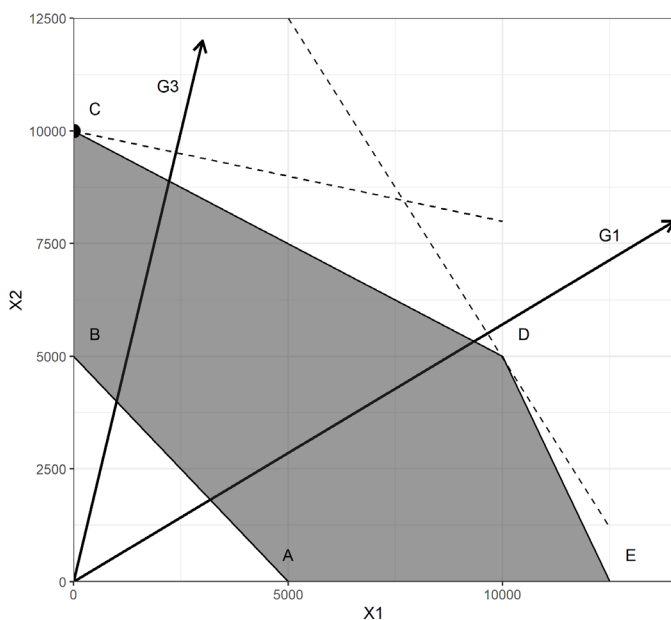
W tym miejscu rodzi się pytanie, co musiałoby się stać, żeby decyzja wskazana przez model uległa zmianie. Załóżmy więc, że cena za jedną sztukę filtra modelu drugiego wzrosła do pięćdziesięciu złotych. Wykres na rysunku 4 zawiera dotychczasowy gradient ($G1$) oraz nowy ($G2$), wynikający ze zmiany ceny drugiego modelu filtra.

**Rys. 4.** Zmiana gradientu po zmianie współczynnika c_2

Źródło: opracowanie własne.

Zmieniło się oczywiście nachylenie warstwy, ale wierzchołek D nadal pozostaje decyzją optymalną. Nie znaczy to jednak, że nie zaszły żadne zmiany w rozwiązaniu, ponieważ wartość przychodu maksymalnego wzrosła do dziewięćset pięćdziesięciu złotych. Wartości lewych stron ograniczeń pozostają bez zmian.

Założmy teraz, że kontrahent zamawiający filtry wycofuje część modeli odkurzaczy. W związku z tym spada popyt na filtry modelu pierwszego. Zarząd zdecydował więc o obniżeniu ceny za taki filtr do dziesięciu złotych za sztukę. Na rysunku 5 nowej sytuacji odpowiada gradient oznaczony jako $G3$. Tym razem zmiany są dużo większe. Przede wszystkim warstwica dla gradientu $G3$ wskazała nowe rozwiązanie optymalne, które znajduje się w wierzchołku $C(0; 10000)$. Dochodzi do radykalnej zmiany decyzji, gdyż – z punktu widzenia maksymalizacji przychodu – należy zrezygnować z produkcji jednego modelu filtrów. Pociąga to za sobą zmianę nie tylko wartości funkcji celu, ale także lewych stron ograniczeń.



Rys. 5. Zmiana gradientu po zmianie współczynnika c_1

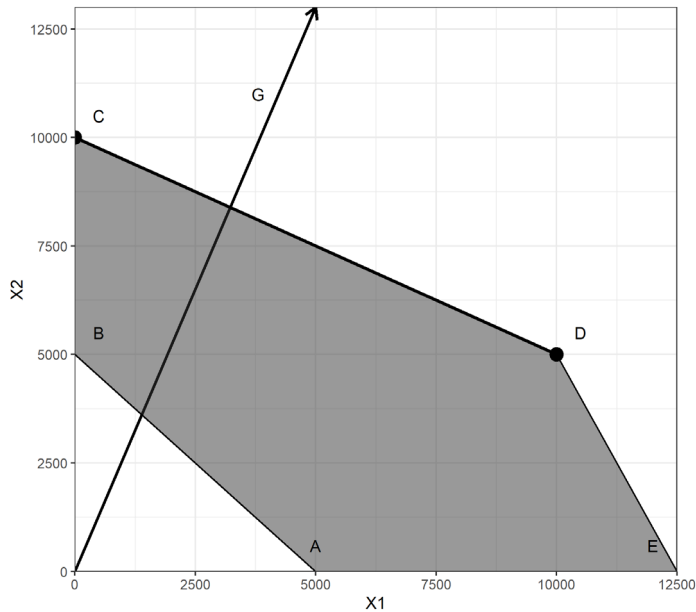
Źródło: opracowanie własne.

Powyższy przykład udowadnia, że nie każda zmiana współczynnika funkcji celu prowadzi do zmiany rozwiązania optymalnego. Musi ona być odpowiednio duża, a raczej wykraczać poza pewien konkretny przedział wartości, o czym przekonamy się w następnym podrozdziale.

Możemy też zastanowić się nad wpływem zmiany prawej strony ograniczenia na rozwiązanie optymalne. Spójrzmy na rysunek 2 i wyobraźmy sobie, że prawa strona ograniczenia („wkład”) wzrosła do trzystu kilogramów. Nowy zbiór rozwiązań dopuszczalnych straci wówczas jeden wierzchołek, gdyż wierzchołek B pokryje się z wierzchołkiem C . Zmieni się też położenie wierzchołka A . Mimo

wszystko nie dojdzie do zmiany rozwiązania optymalnego, pozostanie nim wierzchołek D . Z kolei spadek zapasu tworzywa do poziomu dwustu dwudziestu kilogramów spowoduje przesunięcie wierzchołka D do punktu $(8000; 6000)$. Nadal pozostanie on rozwiązaniem optymalnym, ale z maksymalną wartością funkcji celu równą ośmiuset tysiącom złotych. Przypomnijmy sobie w tym miejscu, że ograniczenie dotyczące wkładu z włókniny zostało przekroczone aż o trzysta kilogramów, tymczasem nasza zmiana wynosi „zaledwie” sto pięćdziesiąt kilogramów. Ograniczenie „tworzywo” nie posiadało niewykorzystanego limitu, a zatem – każda zmiana w jego przypadku skutkować będzie nową decyzją, nową wartością kryterium i nowymi wartościami lewych stron ograniczeń.

Interesującą cechą modeli optymalizacyjnych jest to, że ich rozwiązania końcowe mogą przybrać jedną z kilku postaci. W przykładzie 1, kiedy poznaliśmy konkretny wierzchołek wskazujący najlepszą w danych warunkach decyzję, było to tzw. jednoznaczne rozwiązanie optymalne. Ale już w twierdzeniu Weierstrassa pojawiła się wzmianka o tym, że funkcja celu swoją największą lub najmniejszą wartość może osiągać w więcej niż jednym wierzchołku.

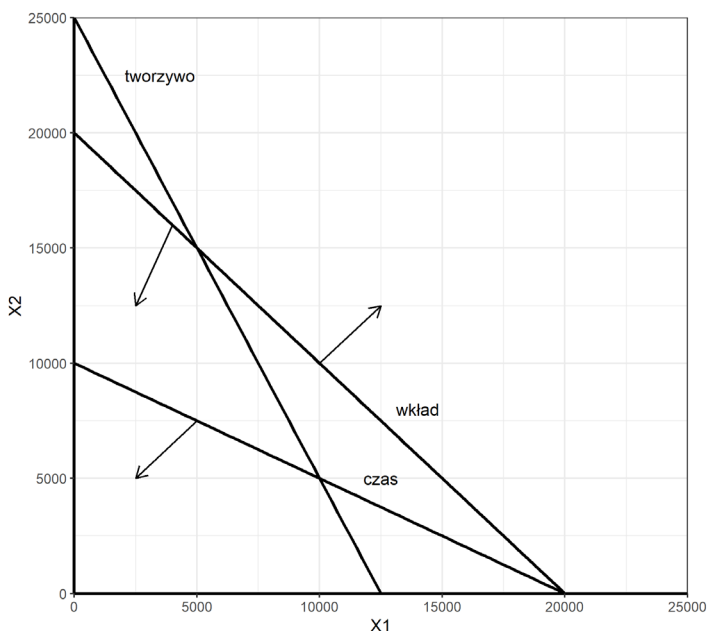


Rys. 6. Rozwiązanie niejednoznaczne

Źródło: opracowanie własne.

Ilustracją takiej sytuacji jest wykres na rysunku 6. Warstwica odpowiadająca maksymalnej wartości funkcji kryterium pokryła się z krawędzią CD zbioru

rozwiązań dopuszczalnych. Funkcja kryterium osiąga w wierzchołkach C i D tę samą, maksymalną wartość. Co więcej, osiąga ją w każdym punkcie na krawędzi CD . Mogą być one wyznaczone jako kombinacje liniowe tych wierzchołków. Punktów na krawędzi CD jest nieskończenie wiele. Z tego powodu ten typ rozwiązania nazywa się niejednoznacznym.



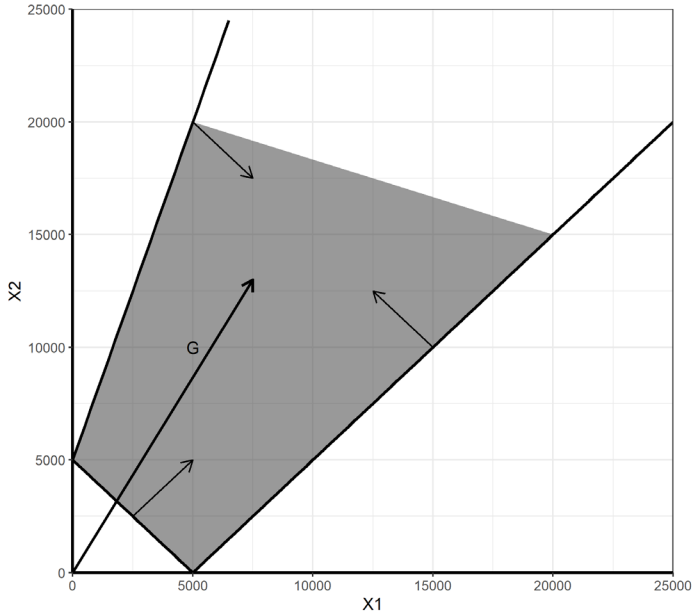
Rys. 7. Rozwiązanie nie istnieje

Źródło: opracowanie własne.

Podczas budowy modelu decyzyjnego zdarza się czasem sytuacja, kiedy decydynt wymaga spełnienia dwóch (lub więcej) wzajemnie wykluczających się warunków. W przytaczanym dotąd przykładzie odpowiadałoby jej przyjęcie, że należy zużyć co najmniej sześćset kilogramów włókny (ostatnie z ograniczeń). Ograniczenia naniesione na dwuwymiarowy wykres wyglądałyby tak, jak na rysunku 7. W pierwszej ćwiartce układu współrzędnych nie istnieje część wspólna wszystkich ograniczeń, a zatem – mamy do czynienia z zadaniem sprzecznym.

Przyjrzyjmy się teraz sytuacji na rysunku 8. Jest to problem maksymalizacji funkcji celu. Istnieje w nim część wspólna wszystkich trzech ograniczeń, ale nie jest ona ograniczona od góry. Powiemy, że zbiór rozwiązań dopuszczalnych nie został domknięty. Tym samym, maksymalizując funkcję kryterium, nie uda się znaleźć warstwy, która utraci kontakt ze zbiorem rozwiązań dopuszczalnych. Sytuację tę

nazywamy brakiem skończonego rozwiązania optymalnego – traktuje się ją jako symptom źle skonstruowanego modelu. Najprawdopodobniej decydent pominął jakieś istotne ograniczenie podczas budowy modelu.



Rys. 8. Brak skończonego rozwiązania optymalnego

Źródło: opracowanie własne.

Co ciekawe, problem dotyczy tylko przypadku maksymalizacji funkcji celu. Obecność warunków brzegowych powoduje, że zawsze uda się znaleźć punkt (punkty), będący (będące) rozwiązaniem zadania minimalizacji.

3. Algorytm *simplex* i jego odmiany

3.1. Klasyczna wersja algorytmu *simplex*

Metoda graficzna ma tę zaletę, że dobrze pokazuje, o co chodzi w poszukiwaniu rozwiązania optymalnego. Tworzymy zbiór rozwiązań dopuszczalnych jako część wspólną ograniczeń. Jest on wielobokiem, a więc ma wierzchołki. Sprawdzamy,

który z tych wierzchołków okaże się rozwiązaniem optymalnym, a gdy już go poznamy – wyznaczamy samo rozwiązanie.

Niestety metoda ta ma również swoje ograniczenia. Najpoważniejsze stanowi liczba zmiennych decyzyjnych. Skoro wykres sporządzamy na dwuwymiarowej płaszczyźnie, siłą rzeczy mogą się na niej znaleźć tylko dwie zmienne. Teoretycznie dałoby się sporządzić rysunek dla trzech zmiennych, wykorzystując rzut izometryczny, ale nie będzie to łatwe zadanie. W modelu o czterech zmiennych można w tym celu wykorzystać sztuczne zmienne. Każda z nich przechowywać będzie informacje o dwóch zmiennych decyzyjnych. Po odpowiednich przekształceniach model sprowadzamy do postaci dwuwymiarowej.

Nawet jednak w przypadku tylko dwóch zmiennych decyzyjnych, możemy mieć problem ze sporządzeniem wykresu. Jeśli pojawi się duża liczba ograniczeń, znalezienie części wspólnej staje się bardzo trudne, o ile w ogóle możliwe. Dlatego potrzebna jest inna metoda, wolna od powyższych ograniczeń. Taką metodę, nazwaną algorytmem *simplex*², zaproponował w 1947 roku George Dantzig (Dantzig, 1963). Okazała się ona niezwykle skuteczna i dziś jest standardem stosowanym do rozwiązywania małych i dużych problemów optymalizacyjnych. Algorytm *simplex* to procedura algebraiczna – i dlatego model (6)–(8) przedstawimy w postaci macierzowej. Wprowadźmy następujące oznaczenia:

- $\mathbf{A} = [\mathbf{a}_{ij}]$ – macierz prostokątna o wymiarach $m \times n$ ($i = 1, 2, \dots, m; j = 1, 2, \dots, n$), której wiersze są liniowo niezależne;
- $\mathbf{b} = [\mathbf{b}_i]$ – m -wymiarowy wektor kolumnowy ($i = 1, 2, \dots, m$);
- $\mathbf{c} = [\mathbf{c}_j]$, $\mathbf{x} = [\mathbf{x}_j]$ – n -wymiarowe wektory kolumnowe ($j = 1, 2, \dots, n$).

Zapiszmy model algebraiczny w jego najbardziej ogólnej, mieszanej postaci, w której ograniczenia mogą być zarówno nierównościami, jak i równaniami.

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \rightarrow \max/\min \quad (11)$$

$$\mathbf{A}_1 \mathbf{x} \leq \mathbf{b}_1 \quad (12)$$

$$\mathbf{A}_2 \mathbf{x} = \mathbf{b}_2 \quad (13)$$

$$\mathbf{A}_3 \mathbf{x} \geq \mathbf{b}_3 \quad (14)$$

$$\mathbf{x} \geq 0 \quad (15)$$

Macierz \mathbf{A} powyższego modelu jest macierzą blokową o postaci:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}_1 \\ \mathbf{A}_2 \\ \mathbf{A}_3 \end{bmatrix} \quad (16)$$

² W polskiej literaturze spotyka się ją również pod następującymi nazwami: „simpleks” i „sympleks”.

Przyjmijmy, że m_1 jest liczbą warunków ograniczających typu „ \leq ”, m_2 to liczba warunków równościowych a m_3 – warunków typu „ \geq ”. Macierze \mathbf{A}_1 , \mathbf{A}_2 i \mathbf{A}_3 , stanowiące bloki macierzy \mathbf{A} , mają liczbę wierszy odpowiednio równą m_1 , m_2 , m_3 ; $m_1 + m_2 + m_3 = m$. Wektor \mathbf{b} , w odpowiadającej postaci blokowej, przedstawia wzór (17). Liczba elementów poszczególnych bloków \mathbf{b}_1 , \mathbf{b}_2 , \mathbf{b}_3 wynosi odpowiednio m_1 , m_2 , m_3 .

$$\mathbf{b} = \begin{bmatrix} \mathbf{b}_1 \\ \mathbf{b}_2 \\ \mathbf{b}_3 \end{bmatrix} \quad (17)$$

Simplex jest procedurą iteracyjną, która wymaga uprzedniego przekształcenia zbioru ograniczeń (często nierówności) do postaci układu równań liniowych. Następnie w określony sposób dokonujemy przeglądu rozwiązań bazowych. Rozwiązania bazowe dopuszczalne możemy utożsamiać z wierzchołkami zbioru rozwiązań dopuszczalnych X . Istnieją także rozwiązania bazowe niedopuszczalne, odpowiadające punktom przecięcia prostych związanych z ograniczeniami leżącymi poza zbiorem X . Celem przeglądania jest znalezienie punktu, w którym funkcja celu osiąga wartość największą lub najmniejszą. Metoda nie potrzebuje przy tym sprawdzać wszystkich rozwiązań bazowych. Algorytm wędruje od bieżącego do sąsiedniego rozwiązania bazowego w przestrzeni m -wymiarowej, gwarantując przy tym, że w nowym rozwiązaniu bazowym wartość funkcji celu jest nie gorsza niż w poprzednim rozwiązaniu.

Ciekawy jest mechanizm sprawiający, że tak właśnie się dzieje. Na podstawie twierdzenia Weierstrassa wiemy, dlaczego szczególnie ważne są wierzchołki zbioru rozwiązań dopuszczalnych, a także, że wśród nich znajduje się rozwiązanie optymalne. Rozwiązania bazowe podzielimy na sąsiadujące i niesąsiadujące (Hillier i Lieberman, 2015, s. 94). W przypadku dowolnego problemu programowania liniowego z n zmiennymi decyzyjnymi, dwa rozwiązania bazowe sąsiadują ze sobą, jeśli mają wspólne $n-1$ ograniczeń. W modelu z dwiema zmiennymi są one połączone segmentem prostej odpowiadającej ograniczeniu. Wierzchołek sąsiedni w zbiorze rozwiązań dopuszczalnych znajdziemy, przesuwając się wzdłuż krawędzi zbioru, nie napotykając po drodze innego wierzchołka. Chociaż w przykładzie 1 każdy wierzchołek bierze się z przecięcia dwóch prostych związanych z ograniczeniami, to generalnie w danym punkcie przecinać się może ich więcej. Najczęściej jednak z każdego wierzchołka wychodzą po dwie krawędzie, prowadzące do sąsiadujących dopuszczalnych rozwiązań wierzchołkowych. Algorytm *simplex* potrafi wystartować z dowolnego rozwiązania bazowego (także z rozwiązania bazowego niedopuszczalnego) i „odwiedzić” jedynie wierzchołki znajdujące się w jego sąsiedztwie. Jeśli dla danego wierzchołka nie uda się znaleźć sąsiada

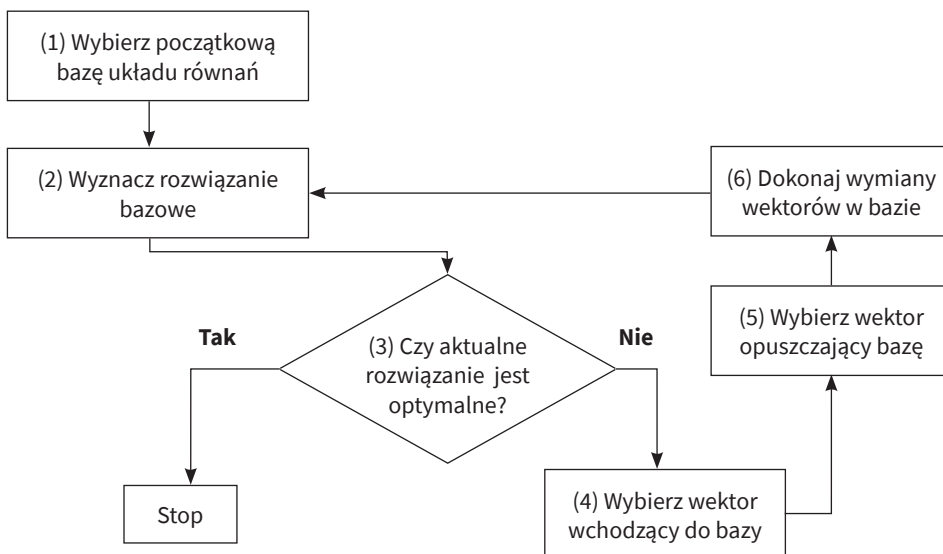
dającego lepszy wynik, to znaczy, że znalezione zostało rozwiązanie optymalne. Rysunek 9 opisuje sposób postępowania w algorytmie *simplex*. Jego szczegóły omówimy, wykorzystując do tego przykład 2.

Przykład 2

Importer planuje wprowadzenie na rynek herbaty powstającej z mieszanki trzech różnych gatunków tego krzewu. W przeliczeniu na jedną tonę sprowadzenie herbaty pierwszego gatunku kosztuje dwieście pięćdziesiąt złotych, drugiego dwieście dziesięć, a trzeciego – trzysta; przy czym łączny koszt zakupu nie powinien przekroczyć ośmiu tysięcy. Herbata musi przejść obróbkę w specjalnych komorach oraz charakteryzować się określoną zawartością garbników. Czas obróbki i zawartość garbników w zależności od gatunku przedstawić można następująco:

	Herbata gat. 1	Herbata gat. 2	Herbata gat. 3
Czas przebywania w komorze [min]	30	60	60
Zawartość garbnika [mg/100 g]	18	14	20

Dostępny czas pracy komory wynosi dwieście godzin. Walory smakowe wymagają, aby garbnika w mieszance znalazło się co najmniej 350 mg/100 g. Opracować należy plan zakupu poszczególnych gatunków, zapewniający poniesienie minimum kosztów.



Rys. 9. Schemat postępowania w algorytmie *simplex*

Źródło: Miszczyński i Miszczyńska, 2002, s. 44.

Model decyzyjny w przykładzie 2 wygląda następująco. Lista zmiennych decyzyjnych:

- x_1 – planowany zakup herbaty gatunku 1 [t],
- x_2 – planowany zakup herbaty gatunku 2 [t],
- x_3 – planowany zakup herbaty gatunku 3 [t].

Funkcja celu – łączny koszt zakupu [zł]:

$$f(\mathbf{x}) = 250x_1 + 210x_2 + 300x_3 \rightarrow \min \quad (18)$$

Ograniczenia i warunki brzegowe:

$$(\text{maks. koszt [zł]}) \quad 250x_1 + 210x_2 + 300x_3 \leq 8000 \quad (19)$$

$$(\text{czas [h]}) \quad 0,5x_1 + x_2 + x_3 \leq 200 \quad (20)$$

$$(\text{garbnik [mg/100 g]}) \quad 15x_1 + 14x_2 + 20x_3 \geq 350 \quad (21)$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \quad (22)$$

W metodzie *simplex*, aby w ogóle myśleć o rozwiązaniu modelu (18)–(22), musimy sprowadzić go do postaci kanonicznej. Jest to postać, w której wszystkie ograniczenia przedstawione są jako równania. Postępujemy według następujących reguł:

- jeżeli k -te ograniczenie jest typu „ \leq ”, wówczas do lewej strony ograniczenia dodajemy zmienną swobodną s_k , która jest zmienną bilansującą;
- jeżeli k -te ograniczenie jest typu „ \geq ”, wówczas od lewej strony ograniczenia odejmujemy zmienną swobodną s_k , która jest zmienną bilansującą oraz dodajemy zmienną sztuczną t_k ;
- jeżeli k -te ograniczenie jest typu „ $=$ ”, wówczas do lewej strony ograniczenia dodajemy zmienną sztuczną t_k .

Dodatkowe zmienne są, podobnie jak zmienne decyzyjne, nieujemne. Pojawiają się w funkcji celu w postaci kanonicznej modelu z ustalonymi parametrami. Parametry przy zmiennych decyzyjnych pozostają bez zmian. Zmienne swobodne wprowadzane są do funkcji celu z parametrem 0, zmienne sztuczne z parametrem $-M$ (gdy funkcja celu jest maksymalizowana) lub $+M$ (gdy funkcja celu jest minimalizowana), gdzie M to pewna bardzo duża liczba dodatnia.

Jeśli zmienna swobodna jest równa 0 w bieżącym rozwiązaniu dopuszczalnym, wówczas rozwiązanie leży na krawędzi zbioru rozwiązań dopuszczalnych, wyznaczonego przez dane ograniczenie. Wartość większa od 0 oznacza, że rozwiązanie znajduje się po dopuszczalnej stronie, a wartość mniejsza – że ograniczenie nie jest spełnione. Jeżeli zmienna sztuczna w bieżącym rozwiązaniu przyjmuje wartość większą od 0, oznacza to, że bieżące rozwiązanie bazowe jest niedopuszczalne.

Przyjmijmy ogólny zapis postaci kanonicznej (po zamianie ograniczeń wyrażonych jako układ nierówności na układ równań) modelu (11)–(15):

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \rightarrow \max/\min \quad (23)$$

$$\mathbf{Ax} = \mathbf{b} \quad (24)$$

$$\mathbf{x} \geq 0 \quad (25)$$

Postać kanoniczna modelu w rozważanym przykładzie 2 jest następująca:

$$f(\mathbf{x}) = 250x_1 + 210x_2 + 300x_3 + 0s_1 + 0s_2 + 0s_3 + Mt_3 \rightarrow \min \quad (26)$$

$$250x_1 + 210x_2 + 300x_3 + s_1 = 8000 \quad (27)$$

$$0,5x_1 + x_2 + x_3 + s_2 = 200 \quad (28)$$

$$15x_1 + 14x_2 + 20x_3 - s_3 + t_3 = 350 \quad (29)$$

$$x_1 \geq 0, x_2 \geq 0, s_1 \geq 0, s_2 \geq 0, s_3 \geq 0, t_3 \geq 0 \quad (30)$$

Kroki algorytmu *simplex* realizuje się, korzystając ze specjalnych tablic. W tym celu zdefiniujemy macierz \mathbf{P} , zbudowaną z kolumn \mathbf{p}_j w taki sposób, że pierwsze n_1 kolumn zawiera składowe odpowiadające zmiennym decyzyjnym, następne n_2 kolumn odpowiada zmiennym swobodnym, a pozostałe n_3 kolumn – zmiennym sztucznym. Na potrzeby dalszych rozważań założymy, że $n = n_1 + n_2 + n_3$ oznacza łączną liczbę zmiennych w modelu (decyzyjnych, swobodnych i sztucznych), zaś n_1 to liczba zmiennych decyzyjnych. Rząd macierzy parametrów lewych stron ograniczeń jest równy liczbie ograniczeń m i jest to mniejszy wymiar tej macierzy.

$$\mathbf{P} = [\mathbf{p}_1 \dots \mathbf{p}_{n_1} \mathbf{p}_{n_1+1} \dots \mathbf{p}_{n_2} \mathbf{p}_{n_2+1} \dots \mathbf{p}_n] \quad (31)$$

Równania ograniczeń można teraz zapisać w postaci kombinacji liniowej wektorów \mathbf{p}_i ($i = 1, \dots, n$), w której zmienne modelu (23)–(27) pełnią funkcję nieznanych współczynników tej kombinacji:

$$\mathbf{p}_1 x_1 + \mathbf{p}_2 x_2 + \dots + \mathbf{p}_n x_n = \mathbf{b} \quad (32)$$

Prześledzimy teraz kolejne kroki algorytmu *simplex*, dodając do nich stosowny komentarz. Metoda *simplex* bazuje na przeglądaniu rozwiązań bazowych układu równań liniowych (24).


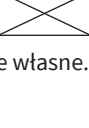
Krok pierwszy. Wybór początkowej bazy układu równań

Ważnym i często powtarzającym się pojęciem jest baza układu równań. Tworzy ją układ wektorów, które są liniowo niezależne. Warunek ten spełnia, przykładowo, macierz jednostkowa. Początkową bazę w algorytmie *simplex* tworzą wektory

stojące przy zmiennych swobodnych i/lub sztucznych, ponieważ zmiennym tym zawsze towarzyszą wektory jednostkowe. Łatwo zauważyć, że do bazy należy tyle wektorów, ile występuje ograniczeń. Odpowiadające im zmienne nazywamy bazowymi, pozostałe zmienne – niebazowymi. Powstaje macierz jednostkowa, której wymiar równy jest liczbie ograniczeń. W tym momencie dostrzegamy sens wprowadzenia w ograniczeniu (29) zmiennej sztucznej. Gdyby nie ona, nie zdołalibyśmy utworzyć podmacierzy jednostkowej w macierzy parametrów lewych stron ograniczeń. W naszym przykładzie tworzą ją wektory: \mathbf{p}_4 , \mathbf{p}_5 i \mathbf{p}_7 :

$$\mathbf{B}^{pocz} = [\mathbf{p}_4 \quad \mathbf{p}_5 \quad \mathbf{p}_7] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (33)$$

Tab. 3. Iteracje algorytmu *simplex* dla przykładu 2

	\mathbf{c}^B	Baza	$[c_j]$	250	210	300	0	0	0	+M	Wart. zm. baz.	Ilorazy wyj.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	\mathbf{p}_7		
				x_1	x_2	x_3	s_1	s_2	s_3	t_3		
1.	0	\mathbf{p}_4	s_1	250	210	300	1	0	0	0	8000	26,67
	0	\mathbf{p}_5	s_2	1/2	1	1	0	1	0	0	200	200
	M	\mathbf{p}_7	t_3	15	14	20	0	0	-1	1	350	17,5
	$\Delta_j = c_j - z_j$			250 -15M	210 -14M	300 -20M	0	0	M	0	350M	
2.	0	\mathbf{p}_4	s_1	25	0	0	1	0	15	-15	2750	
	0	\mathbf{p}_5	s_2	-1/4	3/10	0	0	1	1/20	-1/20	182,5	
	300	\mathbf{p}_3	x_3	3/4	7/10	1	0	0	-1/20	1/20	17,5	
	$\Delta_j = c_j - z_j$			25	0	0	0	0	15	M-15	5250	

Źródło: opracowanie własne.

Taką samą postać ma w tej chwili, kluczowa w późniejszych rozważaniach, macierz odwrotna do macierzy bazy, \mathbf{B}^{-1} (ta informacja będzie miała znaczenie przy analizie wrażliwości).

Wektor wartości zmiennych bazowych w każdej iteracji algorytmu można znać, korzystając z następującej formuły algebraicznej (i oznacza numer iteracji):

$$\mathbf{x}^{B^{(i)}} = (\mathbf{B}^{(i)})^{-1} \mathbf{b} \quad (34)$$

Krok drugi. Wyznaczenie rozwiązania bazowego

Aktualne rozwiązanie bazowe odczytujemy z przedostatniej kolumny tablicy simpleksowej (zob. tabela 3). W pierwszej iteracji tworzą je po prostu przepisane prawe strony ograniczeń. W kolejnych jest ono związane z aktualną bazą i powstaje na drodze przekształceń tablicy z iteracji poprzedniej (patrz: krok szósty). Jeżeli zmienna jest niebazowa, jej wartość wynosi 0; w pozostałych przypadkach jest równa liczbie z odpowiedniego wiersza kolumny \mathbf{x}^B . Dlatego w pierwszej iteracji w tabeli 3 zmienna swobodna s_1 równa się 8000. Wartość funkcji celu obliczamy, mnożąc transponowany wektor \mathbf{c}^B przez wektor wartości zmiennych bazowych \mathbf{x}^B , osiągając (dla przykładu 2) 350M.

Krok trzeci. Sprawdzenie optymalności rozwiązania

Sprawdzenia optymalności rozwiązania dokonuje się na podstawie tzw. wskaźników optymalności. Oblicza się je dla wszystkich zmiennych znajdujących się w tablicy simpleksowej, przy czym dla zmiennych bazowych zawsze równają się one 0. Formuła ich wyznaczania jest następująca³:

$$\Delta_j = c_j - z_j = c_j - (\mathbf{c}^B)^T \mathbf{p}_j \quad (35)$$

gdzie:

- c_j – współczynnik w funkcji celu przy j -tej zmiennej, dla której obliczamy wskaźnik optymalności;
- \mathbf{c}^B – wektor współczynników funkcji celu przy zmiennych bazowych;
- \mathbf{p}_j – j -ta kolumna centralnej części tablicy simpleksowej.

Przykładowo, wskaźnik optymalności dla zmiennej x_1 w pierwszej iteracji został obliczony następująco:

$$\Delta_1 = c_1 - z_1 = 250 - \begin{bmatrix} 0 & 0 & M \end{bmatrix} \begin{bmatrix} 250 \\ \frac{1}{2} \\ 15 \end{bmatrix} = 250 - 15M$$

Wartość Δ_j dla zmiennej niebazowej można traktować jako różnicę między jednostkowym efektem a jednostkowym nakładem potrzebnym, aby dana zmienna miała zostać zmienną bazową. Na tej podstawie zbudowano kryterium optymalności bieżącego rozwiązania. Dla maksymalizacji funkcji celu bieżące rozwiązanie

³ W starszych polskich publikacjach można spotkać odwrotną formułę obliczania wskaźników optymalności, tj. $z_j - c_j$. Wpływa to na interpretację warunku optymalności rozwiązania, ale nie na liczbę iteracji czy samo rozwiązanie.

uznajemy za optymalne, kiedy wszystkie wskaźniki optymalności są niedodatnie. Dla przypadku minimalizacji funkcji celu wszystkie wskaźniki optymalności powinny być niedodatnie. Trzy pierwsze wskaźniki dla pierwszej iteracji w tabeli 3 są ujemne, a więc nie mamy do czynienia z rozwiązaniem optymalnym i możemy spróbować je poprawić.

Krok czwarty. Wybór wektora wchodzącego do bazy

Jeżeli aktualne rozwiązanie bazowe nie spełnia warunku optymalności, konieczna staje się wymiana wektorów w bazie. Najpierw wybierany jest wektor, który znajdzie się w przyszłej bazie. Jako kryterium posłużą znów wskaźniki optymalności; ponownie występuje tu rozdział spowodowany kierunkiem optymalizacji. Do bazy wprowadzamy k -ty wektor taki, że:

$$\begin{cases} \Delta_k = \min_{\Delta_j > 0} \Delta_j, & \text{gdy } f(\mathbf{x}) \rightarrow \max \\ \Delta_k = \max_{\Delta_j < 0} \Delta_j, & \text{gdy } f(\mathbf{x}) \rightarrow \min \end{cases} \quad (36)$$

Tak więc, jeśli funkcja celu jest maksymalizowana, wybieramy wektor \mathbf{p}_k , dla którego wskaźnik optymalności okazał się najwyższy spośród tych, co przyjęły wartość dodatnią. Dla minimalizowanej funkcji będzie to wektor o najniższym wskaźniku z ujemnych. Wektor ten nazwiemy wektorem wyróżnionym. Jeśli dwa wskaźniki (lub więcej) mają taką samą wartość, wybieramy ten o najniższym numerze wektora \mathbf{p}_j . Zatem, w rozwiązywanym przykładzie, do bazy wejdzie wektor \mathbf{p}_3 , dla którego wskaźnik optymalności jest najniższy. Wyróżnioną kolumną jest trzecia kolumna tabeli simpleksowej.

Krok piąty. Wybór wektora opuszczającego bazę

W tym kroku należy zdecydować, która z aktualnych zmiennych bazowych zostanie usunięta, żeby na jej miejsce wprowadzić zmienną wybraną w kroku poprzednim. Dla uprzednio ustalonego wyróżnionego wektora k (k -tej kolumny wnętrza tabeli simpleksowej) obliczamy tzw. ilorazy wyjścia. Sposób ich wyliczania jest niezależny od kierunku kryterium. Z bazy usuwamy l -ty wektor taki, że:

$$l: \frac{x_l^B}{p_{lk}} = \min_{p_{ik} > 0} \left\{ \frac{x_i^B}{p_{ik}} \right\} \quad (37)$$

Ilorazów wyjścia nie obliczamy dla składowych ujemnych i równych 0. Może wystąpić sytuacja, kiedy dwa ilorazy (lub więcej) są jednakowe. Zajmiemy się tym później.

Ilorazy wyjścia dla pierwszej iteracji przykładu 2 są następujące:

$$\min \left\{ \frac{8000}{300}, \frac{200}{1}, \frac{350}{20} \right\} = \min \left\{ 26\frac{2}{3}, 200, 17\frac{1}{2} \right\} = 17\frac{1}{2}$$

Bazę opuści zatem wektor \mathbf{p}_7 (odpowiada mu zmienna t_3), jego miejsce zajmie wektor \mathbf{p}_3 (odpowiada mu zmienna x_3). Wyróżnionym wierszem jest trzeci wiersz tablicy simpleksowej.

Krok szósty. Wymiana wektorów w bazie

Wymiany wektorów w bazie dokonuje się przy pomocy tzw. wierszowych przekształceń elementarnych. W metodzie *simplex* stosuje się je w następujący sposób:

- mnożymy wiersz wyróżniony wnętrza tablicy, łącznie z elementem kolumny odpowiadającym wyróżnionemu wierszowi, przez liczbę różną od 0 (dodatnią lub ujemną) – taką, aby w kolumnie wyróżnionej uzyskać element równy 1;
- do pozostałych wierszy wnętrza tablicy, łącznie z elementami kolumny wartości zmiennych bazowych, dodaje się wiersz wyróżniony, przemnożony przez odpowiednią liczbę różną od 0 w taki sposób, aby w danym wierszu i w wyróżnionej kolumnie uzyskać element równy 0.

Operacje elementarne mają doprowadzić do tego, aby wchodzący do bazy wektor \mathbf{p}_k został przekształcony w wektor jednostkowy, w którym jedynka znajduje się w tym samym wierszu, w którym wprowadzamy zmienną do bazy. Przekształcenia dotyczą wszystkich wierszy tablicy simpleksowej – tak właśnie uzyskaliśmy drugie rozwiązanie bazowe w przykładzie o imporcie herbaty.

Wszystkie wskaźniki optymalności w tej iteracji są mniejsze lub równe 0, co sygnalizuje osiągnięcie rozwiązania optymalnego. W innym przypadku należałoby wrócić do kroku czwartego. Wektorowa wersja rozwiązania optymalnego jest następująca:

$$\mathbf{x}^{opt} = \begin{bmatrix} x_1^{opt} \\ x_2^{opt} \\ x_3^{opt} \\ s_1^{opt} \\ s_2^{opt} \\ s_3^{opt} \\ t_3^{opt} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 17,5 \\ 2750 \\ 182,5 \\ 0 \\ 0 \end{bmatrix}$$

Importer powinien sprowadzić jedynie siedemnaście i pół tony herbaty gatunku trzeciego, co zapewni mu koszt minimalny na poziomie pięciu tysięcy dwustu pięćdziesięciu złotych, czyli o dwa tysiące siedemset pięćdziesiąt złotych mniej niż zakładany limit wydatków. Dostępny limit czasu pracy komory, w której herbatę poddaje się obróbce, nie został w pełni wykorzystany. Niewykorzystany czas pracy wynosi sto osiemdziesiąt dwie i pół godziny, zaś garbnika znajdzie się minimalna wymagana ilość.

Zmienne swobodne, zgodnie z zapisem modelu w postaci kanonicznej, przyjmują wartości nieujemne. Kiedy są równe 0, lewa strona ograniczenia równa się prawej. Ekonomicznie oznacza to dokładne zrealizowanie obowiązującego limitu. Interpretacja zmienia się, gdy zmienna swobodna przyjmuje wartość dodatnią. Niezerowa wartość zmiennej swobodnej w rozwiązaniu optymalnym świadczy o niewykorzystaniu występującego górnego limitu (w przypadku, gdy ograniczenie jest typu „ \leq ”) lub o wielkości przekroczenia dolnego limitu (w sytuacji ograniczenia typu „ \geq ”).

Algorytm *simplex* jest metodą poszukiwania decyzji optymalnej w problemie optymalizacji liniowej. Warto więc wiedzieć, jak rozpoznać inne typy rozwiązań, o których była mowa w poprzednim rozdziale. Mamy do czynienia z czterema możliwymi typami rozwiązań optymalnych, którymi są:

1. Jednoznaczne rozwiązanie optymalne – liczba wskaźników optymalności równych 0 jest równa liczbie zmiennych bazowych (będzie tak wyłącznie dla tych zmiennych).
2. Rozwiązanie niejednoznaczne – liczba wskaźników optymalności równych 0 jest większa od liczby zmiennych bazowych (niektóre zmienne niebazowe mają wskaźniki optymalności równe 0).
3. Brak skończonego rozwiązania optymalnego – nie jest możliwe wyznaczenie ilorazów wyjścia, gdyż wszystkie składowe wektora wchodzącego do bazy są niedodatnie.
4. Zadanie sprzeczne – w rozwiązaniu optymalnym występuje zmienna sztuczna z niezerową wartością.

Szybkie sprawdzenie ostatniej tablicy simpleksowej w tabeli 3 pozwala zauważyć, że przy trzech znajdujących się w bazie zmiennych występują cztery wskaźniki optymalności równe 0. Mamy więc do czynienia z rozwiązaniem niejednoznacznym. Istnieje zatem rozwiązanie bazowe alternatywne, o takiej samej wartości funkcji kryterium. W celu jego wyznaczenia ponownie wykonamy wymianę wektorów w bazie. Wektorem wchodzącym będzie \mathbf{p}_2 , a ilorazy wyjścia wskazują, że zastąpi on wektor \mathbf{p}_3 . Tabela 4 zawiera rozwiązanie z tabeli 3 oraz nowe, powstałe po wykonaniu przekształceń elementarnych.

Zauważmy, że obliczenie ilorazu wyjścia w pierwszym wierszu nie było możliwe, ponieważ ta składowa wektora \mathbf{p}_2 była równa 0. W tym wariantcie należałoby sprowadzić dwadzieścia pięć ton herbaty gatunku drugiego. Koszt minimalny nie uległ rzecz jasna zmianie i nadal wynosi pięć tysięcy dwieście pięćdziesiąt złotych.

Tab. 4. Alternatywne rozwiązanie dla przykładu 2

	c^B	Baza	$[c_j]$	250	210	300	0	0	0	$+M$	Wart. zm. baz.	Ilorazy wyj.
			Zm.	p_1	p_2	p_3	p_4	p_5	p_6	p_7		
			baz.	x_1	x_2	x_3	s_1	s_2	s_3	t_3		
2.	0	p_4	s_1	25	0	0	1	0	15	-15	2750	-
	0	p_5	s_2	-1/4	3/10	0	0	1	1/20	-1/20	182,5	608,3
	300	p_3	x_3	3/4	7/10	1	0	0	-1/20	1/20	17,5	25
	$\Delta_j = c_j - z_j$			25	0	0	0	0	15	$M-15$	5250	
3.	0	p_4	s_1	25	0	0	1	0	15	-15	2750	
	0	p_5	s_2	-4/7	0	-3/7	0	1	1/14	-1/14	192,5	
	210	p_2	x_2	15/14	1	10/7	0	0	-1/14	1/14	25	
	$\Delta_j = c_j - z_j$			25	0	0	0	0	15	$M-15$	5250	

Źródło: opracowanie własne.

Klasyczna wersja algorytmu *simplex*, oraz zaprezentowane w dalszej części rozdziału jej modyfikacje, wymieniają w bazie jeden wektor. Przy dużych zadaniach wydłuża to odnalezienie rozwiązania optymalnego. Dlatego powstają próby przyspieszenia całego procesu przez zwiększenie liczby wymienianych wektorów (Vitor i Easton, 2018), gdzie bazę opuszczają dwie zmienne. Trzeba jednak nadmienić, że autorzy używają odmiennego niż *simplex* algorytmu. W innym podejściu (Ansari, 2019) zaproponowano tzw. *AHA Simplex Algorithm*, który zdaniem autora jest prostszy i szybszy od podejścia klasycznego.

3.2. Degeneracja rozwiązania zadania PL

Wspomnieliśmy wcześniej, że czasem zdarza się, iż dwa ilorazy wyjścia (lub więcej) przyjmują takie same, najniższe wartości. Nazywamy to degeneracją rozwiązywania zadania programowania liniowego. W takiej sytuacji można zastosować jedną z następujących ścieżek postępowania (Murthy, 2007, s. 91).

1. Losowo wybrać wektor (zmienną) opuszczający bazę.
2. Wybrać wektor o najniższym numerze (subskrypcie).
3. Podzielić elementy wybranej kolumny macierzy jednostkowej przez odpowiadające im elementy w kolumnie zmiennej wchodzącej do bazy.

Dwa pierwsze sposoby chyba nie budzą wątpliwości. Dlatego skupimy się na trzecim z nich. Spójrzmy na tablicę simpleksową w tabeli 5, zbudowaną dla pewnego problemu maksymalizacji funkcji celu. Rozwiązanie z iteracji pierwszej da się poprawić, gdyż dwa wskaźniki optymalności są dodatnie. Jako wchodzący do

bazy wybrany został wektor \mathbf{p}_2 , dla którego uzyskano dwa ilorazy wyjścia o takiej samej wartości. W tej chwili wiadomo jedynie, że bazy na pewno nie opuści wektor \mathbf{p}_5 – z powodu odpowiadającej mu składowej wektora \mathbf{p}_2 równej 0.

Wybieramy kolumnę \mathbf{p}_3 i dzielimy pierwszą i ostatnią składową przez jej odpowiedniki w \mathbf{p}_2 , czyli: 1/2 oraz 0/3. Na podstawie drugiego z ilorazów wybieramy wektor \mathbf{p}_6 , jako opuszczający bazę, a następnie dokonujemy wymiany w znany już sposób.

Tab. 5. Degeneracja rozwiązania – wariant 1

	\mathbf{c}^B	Baza	$[c_j]$	5	10	0	0	0	Wart. zm. baz.	Ilorazy wyj.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5		
				x_1	x_2	s_1	s_2	s_3	\mathbf{x}^B	
1.	0	\mathbf{p}_4	s_1	1	2	1	0	0	8	8/2=4
	0	\mathbf{p}_5	s_2	4	0	0	1	0	15	–
	0	\mathbf{p}_6	s_3	3	3	0	0	1	12	12/3=4
	$\Delta_j = c_j - z_j$			5	10	0	0	0	0	<div></div>
2.	0	\mathbf{p}_4	s_1	–1	0	1	0	–2/3	0	
	0	\mathbf{p}_5	s_2	4	0	0	1	0	15	
	10	\mathbf{p}_2	x_2	1	1	0	0	1/3	4	
	$\Delta_j = c_j - z_j$			–5	0	0	0	–10/3	40	<div></div>

Źródło: opracowanie własne.

Tab. 6. Degeneracja rozwiązania – wariant 2

	\mathbf{c}^B	Baza	$[c_j]$	5	10	0	0	0	Wart. zm. baz.	Ilorazy wyj.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5		
				x_1	x_2	s_1	s_2	s_3	\mathbf{x}^B	
1.	0	\mathbf{p}_4	s_1	1	-2	1	0	0	18	-
	0	\mathbf{p}_5	s_2	4	1	0	1	0	4	4/1=4
	0	\mathbf{p}_6	s_3	3	3	0	0	1	12	12/3=4
	$\Delta_j = c_j - z_j$			5	10	0	0	0	0	<div></div>
2.	0	\mathbf{p}_4	s_1	-1	-4	1	0	-2/3	10	
	0	\mathbf{p}_5	s_2	1	0	0	1	-1/3	0	
	10	\mathbf{p}_2	x_2	1	1	0	0	1/3	4	
	$\Delta_j = c_j - z_j$			-5	0	0	0	-10/3	40	<div></div>

Źródło: opracowanie własne.

W drugiej iteracji otrzymujemy rozwiązanie optymalne, którego cechą charakterystyczną jest to, że bazowa zmienna s_1 przyjmuje wartość 0. W rozwiązaniu bazowym zdegenerowanym przynajmniej jedna zmienna bazowa przyjmuje wartość 0.

Przyjrzyjmy się teraz sytuacji w tabeli 6. Ponownie spotykamy się z degeneracją zadania programowania liniowego. Tym razem ilorazy składowych kolumn \mathbf{p}_3 i \mathbf{p}_2 – równe odpowiednio: 0/1 i 0/3 – ponownie nie dają rozstrzygnięcia odnośnie wyboru zmiennej opuszczającej bazę. Dlatego użyjemy kolumny \mathbf{p}_4 , otrzymując: 1/1 i 0/3, co wskazuje na zmienną s_3 jako tą, która opuści bazę.

3.3. Dualizm i analiza wrażliwości

Jedną z najważniejszych cech związanych z rozwiązywaniem problemów programowania liniowego okazuje się występowanie tzw. dualizmu. Z każdym zadaniem PL (od teraz nazywanym prymalnym lub pierwotnym) związane jest pewne szczególne zadanie, zwane dualnym. Zadania te tworzą sprzężoną, symetryczną parę zadań dualnych względem siebie. W ten sposób zadanie dualne względem zadania dualnego będzie identyczne z zadaniem pierwotnym.

Zadanie dualne jest więc, mówiąc w pewnym uproszczeniu, transpozycją zadania prymalnego. Interesuje nas ono z kilku powodów. Po pierwsze pozwala na badanie reakcji optymalnej wartości funkcji celu na marginalne zmiany wybranego środka (limitu), prezentowanego przez wyraz wolny w ograniczeniu zadania PL. Po drugie poznajemy dzięki niemu zakres zmian, jakim może podlegać prawa strona ograniczenia, w ramach którego jesteśmy w stanie przewidzieć zachowanie rozwiązania. Wreszcie, gdy korzystamy z modelu dualnego, rozwiązywanie niektórych problemów jest łatwiejsze. Własności pary zadań dualnych sprawiają, że kiedy liczba ograniczeń w zadaniu prymalnym jest wyższa od liczby zmiennych decyzyjnych, korzystnie będzie rozwiązać zadanie dualne, a następnie wyznaczyć na tej podstawie rozwiązanie zadania pierwotnego. Rozmiary centralnej części tablicy simpleksowej będą bowiem mniejsze. Zresztą określenia „pierwotny” i „dualny” są umowne. W praktyce zdarza się, że stosuje się je wymiennie. Najczęściej przyjmuje się, iż model – odpowiadający wyjściowemu problemowi decyzyjnemu – nazywany jest prymalnym.

Kluczem do budowy i interpretacji zadania dualnego są zmienne dualne⁴ odpowiadające ograniczeniom zadania pierwotnego. Jest ich tyle, ile ograniczeń modelu (bez warunków brzegowych), a oznaczać je będziemy symbolem y_i ($i = 1, 2, \dots, m$). Budowa zadania dualnego podlega następującym regułom (Trzaskalik, 2008, s. 65):

1. Maksymalizacji wartości funkcji celu zadania prymalnego odpowiada minimalizacja wartości funkcji celu zadania dualnego. I odwrotnie.

4 Zwane także wycenami lub cenami dualnymi.

2. Jeżeli w zadaniu dualnym występuje minimalizacja funkcji celu, to wszystkie warunki ograniczające są typu „ \geq ”; w przypadku, gdy w zadaniu dualnym funkcja celu jest maksymalizowana, to warunki ograniczające są typu „ \geq ”.
3. Współczynniki funkcji celu zadania prymalnego stają się wyrazami wolnymi zadania dualnego.
4. Wyrazy wolne ograniczeń zadania pierwotnego stają się współczynnikami funkcji celu zadania dualnego.
5. Macierz współczynników lewych stron ograniczeń zadania dualnego jest transponowaną macierzą A zadania pierwotnego. Wynika stąd, że liczba zmiennych decyzyjnych zadania dualnego jest równa liczbie ograniczeń w zadaniu prymalnym, natomiast liczba ograniczeń w zadaniu dualnym jest równa liczbie zmiennych decyzyjnych w zagadnieniu prymalnym.
6. Warunkowi prymalnemu w postaci nierówności z relacją „ \leq ” odpowiada w zagadnieniu dualnym:
 - nieujemna zmienna dualna ($y_i \geq 0$) – jeżeli prymalna funkcja celu jest maksymalizowana,
 - niedodatnia zmienna dualna ($y_i \leq 0$) – jeżeli prymalna funkcja celu jest minimalizowana.
7. Warunkowi prymalnemu w postaci nierówności z relacją „ \geq ” odpowiada w zagadnieniu dualnym:
 - niedodatnia zmienna dualna ($y_i \leq 0$) – jeżeli prymalna funkcja celu jest maksymalizowana,
 - nieujemna zmienna dualna ($y_i \geq 0$) – jeżeli prymalna funkcja celu jest minimalizowana.
8. Warunkowi prymalnemu w postaci równania odpowiada nieokreślona, co do znaku, zmienna dualna ($y_i \in \mathbb{R}$).

Kierując się powyższymi regułami, zapiszmy model zadania dualnego do zadania (11)–(15). Funkcja celu zadania dualnego jest następująca:

$$g(\mathbf{y}) = 8000y_1 + 200y_2 + 350xy_3 \rightarrow \max \quad (38)$$

Przy ograniczeniach:

$$250y_1 + 0,5y_2 + 15y_3 \leq 250 \quad (39)$$

$$210y_1 + y_2 + 14y_3 \leq 210 \quad (40)$$

$$300y_1 + y_2 + 20y_3 \leq 300 \quad (41)$$

$$y_1 \leq 0, y_2 \leq 0, y_3 \geq 0 \quad (42)$$

Wektor zmiennych dualnych $\mathbf{y} = [y_i]$ ($i = 1, \dots, m$) jest – w przyjętym przez nas dalej zapisie – wektorem wierszowym. Po zbudowaniu modelu zadania dualnego chcielibyśmy poznać jego rozwiązanie. Zadanie można sprowadzić do postaci

kanonicznej i rozwiązać za pomocą algorytmu *simplex*. Ze względu jednak na niedodatniość zmiennych y_1 oraz y_2 , w tym celu należałoby zastąpić je przez dwie zmienne pomocnicze, zdefiniowane jako zmienne o wartościach przeciwnych do y_1 oraz y_2 . Konsekwentnie, znaki parametrów w funkcji kryterium i ograniczeniach należałoby zmienić na przeciwne.

Pomocna okaże się znajomość kilku twierdzeń (Trzaskalik, 2008, s. 65–66).

Twierdzenie 3

Jeżeli X oraz Y są niepustymi zbiorami rozwiązań dopuszczalnych, odpowiednio: zadania pierwotnego i dualnego, z funkcjami celu: $f(\mathbf{x})$ i $g(\mathbf{y})$, to dla dowolnych dwóch rozwiązań $\mathbf{x} \in X$ oraz $\mathbf{y} \in Y$ zachodzi:

$$\max_{\mathbf{x}} f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \leq \min_{\mathbf{y}} g(\mathbf{y}) = \mathbf{y}\mathbf{b}$$

lub:

$$\min_{\mathbf{x}} f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \geq \max_{\mathbf{y}} g(\mathbf{y}) = \mathbf{y}\mathbf{b}$$

w zależności od sformułowanego kryterium optymalizacji.

Twierdzenie 4

Na to, aby rozwiązania dopuszczalne pary zadań dualnych $\mathbf{x}^* \in X$ oraz $\mathbf{y}^* \in Y$ były rozwiązaniami optymalnymi, potrzeba (i wystarczy), aby zachodziły następujące równości:

$$\mathbf{y}^*(\mathbf{A}\mathbf{x}^* - \mathbf{b}) = 0$$

oraz:

$$(\mathbf{x}^*)^T(\mathbf{c} - \mathbf{A}^T(\mathbf{y}^*)^T) = 0$$

Twierdzenie 5 (o dualizmie)

Dla danej pary zadań prymalne-dualne prawdziwe jest tylko jedno z poniższych twierdzeń:

1. Jeżeli jedno z pary zadań posiada rozwiązanie optymalne, to posiada je również drugie z nich, przy czym – optymalne wartości funkcji celu są sobie równe.
2. Jeżeli jedno z pary zadań nie posiada skończonego rozwiązania optymalnego, to drugie z nich jest sprzeczne. Twierdzenie odwrotne nie jest prawdziwe, ponieważ gdy jedno zadanie jest sprzeczne, drugie może być także sprzeczne lub nie posiada skończonego rozwiązania optymalnego.

Twierdzenie 6 (o rozwiązaniu zadania dualnego)

Jeżeli istnieje skończone rozwiązanie optymalne zadania pierwotnego względem bazy \mathbf{B} , to rozwiązanie optymalne zadania dualnego dane jest wzorem:

$$\mathbf{y}^{opt} = (\mathbf{c}^B)^T \mathbf{B}^{-1} \quad (43)$$

Twierdzenie 7

Dla i -tego ograniczenia w postaci nierówności spełniona jest zawsze następująca równość:

$$s_i^{opt} \cdot y_i^{opt} = 0$$

Najważniejszym, płynącym z powyższych twierdzeń wnioskiem jest ten, że nie musimy rozwiązywać zadania dualnego, o ile dysponujemy rozwiązaniem zadania prymalnego. Przypomnijmy sobie pierwszy krok algorytmu *simplex*, czyli utworzenie bazy początkowej. Wiemy, że w pierwszej iteracji, przedstawionej w tabeli 3, tworzą ją kolumny odpowiadające zmiennym swobodnym s_1, s_2 oraz zmiennej sztucznej t_3 . W toku kolejnych przekształceń elementarnych dojdzie do tego, że w ostatniej tablicy simpleksowej te same kolumny utworzą odwróconą macierz bazy \mathbf{B}^{-1} , która ma kluczowe znaczenie przy wyznaczaniu optymalnych wartości zmiennych dualnych w twierdzeniu o rozwiązaniu zadania dualnego. Wyznamy więc wspomniane wartości na potrzeby przykładu o imporcie herbaty.

$$\begin{bmatrix} y_1^{opt} & y_2^{opt} & y_3^{opt} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 300 \end{bmatrix} \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 15 \end{bmatrix}$$

Na mocy twierdzenia 1 znamy maksymalną wartość funkcji celu zadania dualnego, gdyż $g_{\max} = f_{\min} = 5250$.

W tabeli 7 zestawione zostały: prawa strona ograniczenia, optymalna wartość zmiennej swobodnej, optymalna wartość zmiennej dualnej i iloczyn dwóch ostatnich. Jeśli weźmiemy pod uwagę moduły liczb, to w tablicy z optymalnym rozwiązaniem zadania pierwotnego:

- wskaźniki optymalności odpowiadające zmiennym swobodnym wyznaczają optymalne wartości zmiennych dualnych;
- wskaźniki optymalności odpowiadające zmiennym decyzyjnym wyznaczają wartości zmiennych swobodnych w optymalnym rozwiązaniu zadania dualnego.

Tab. 7. Relacje między zmiennymi swobodnymi a dualnymi w przykładzie 2

Ograniczenie	b_i	s_i^{opt}	y_i^{opt}	$s_i^{opt} \cdot y_i^{opt}$
Koszt	8000	2750	0	0
Czas	200	182,5	0	0
Garbnik	350	0	15	0

Źródło: opracowanie własne.

Zależność, o której mowa w twierdzeniu 7, wystąpiła dla wszystkich trzech ograniczeń. Wiąże się ona z interpretacją ekonomiczną zmiennych dualnych. Spójrzmy na wyniki dla ograniczenia opisującego wymaganą ilość garbnika. Rozwiązanie optymalne znalazło się dokładnie na krawędzi wyznaczonej przez to ograniczenie – ponieważ $s_3^{opt} = 0$. Zgodnie z tym, co wiemy z analizy rozwiązania modelu z przykładu 1, jakkolwiek zmiana limitu garbnika potrafi spowodować zmianę optymalnych wartości zmiennych decyzyjnych oraz optymalnej wartości funkcji celu. Z kolei limit kosztu nie został wykorzystany w całości. Jego zwiększenie, np. o dziesięć złotych, nie wpłynie na koszt zakupionej herbaty. Po prostu wzrośnie stopień niewykorzystania limitu.

Warto w tym miejscu zadać sobie dwa pytania: jaki dokładnie efekt wywoła zmiana limitu garbnika oraz czy każda zmiana prawej strony ograniczenia kosztu zakupu herbaty pozostaje bez wpływu na rozwiązanie optymalne? Na pierwsze z pytań odpowiada ekonomiczna interpretacja zmiennej dualnej; na drugie – analiza wrażliwości.

Niech $f(\mathbf{b})$ oznacza optymalną wartość funkcji celu zadania pierwotnego, osiągniętą dzięki znanym ilościom środków (limitów) \mathbf{b} . Przy takim założeniu zachodzi równość z optymalną wartością pierwotnej funkcji celu, tj. $f(\mathbf{b}) = f(\mathbf{x}^{opt})$. Optymalna wartość zmiennej dualnej y_i jest pochodną cząstkową funkcji $f(\mathbf{b})$ względem limitu b_i :

$$y_i^{opt} = \frac{\partial f(\mathbf{b})}{\partial b_i} \quad (44)$$

Zmienna dualna wyraża reakcję optymalnej wartości funkcji celu zadania pierwotnego na niewielkie (krajcowe) zmiany limitu i -tego środka. W praktyce interpretację krajową zastępuje się interpretacją przyrostową (Miszczyński i Miszczyńska, 2002, s. 63). Wtedy optymalna wartość wyceny dualnej y_i^{opt} określa zmianę optymalnej wartości funkcji celu zadania pierwotnego, spowodowaną zmianą wartości wyrazu wolnego b_i i -tego ograniczenia zadania pierwotnego o jednostkę:

$$y_i^{opt} = \frac{\Delta f(\mathbf{b})}{\Delta b_i} \quad (45)$$

Z tego powodu możemy wnioskować, że zwiększenie limitu zawartości garbnika o 1 mg/100 g spowoduje wzrost optymalnego kosztu zakupu o piętnaście złotych; zaś zwiększenie limitu kosztu zakupu o złotówkę nie spowoduje zmiany optymalnej wartości funkcji celu.

Każda z wycen dualnych ma swoją interpretację, ściśle związaną z rozważanym zagadnieniem decyzyjnym. Jest wielkością mianowaną – z mianem będącym stosunkiem jednostki miary funkcji celu do jednostki miary odpowiadającego ograniczenia. Wartości zmiennych dualnych są różne od 0 jedynie w przypadku ograniczeń wiążących – spełnionych w rozwiązaniu równania.

Zmienną dualną odpowiadającą ograniczeniu typu „ \leq ” zadania prymalnego można interpretować jako miarę efektywności wykorzystania ograniczonego zasobu i -tego środka produkcji. Stąd w literaturze pojawia się określenie „cena dualna”. Optymalna wartość tej zmiennej jest krańcową produktywnością danego środka produkcji. Takiej interpretacji nie da się wszakże utrzymać, jeśli mamy do czynienia z nierównością typu „ \geq ”, ponieważ możemy wówczas mówić o wycenie dolnego poziomu danego warunku.

Wiemy już, w jaki sposób wyceny dualne pozwalają ocenić wpływ zmian prawej strony ograniczenia na optymalną wartość funkcji celu. Poznaliśmy aspekt matematyczny i ekonomiczny zmian przyjętych limitów. Nietrudno domyślić się, że dopóki zmiany wartości b_i pozostają niewielkie, interpretacja nadal się utrzymuje. Model decyzyjny z założenia opisuje sytuację statyczną. Nie zakładamy w nim zbyt dużych zmian zachodzących w otoczeniu w najbliższym czasie. Co jednak, jeśli takowe się pojawiają? A nawet jeśli nie, nasza wiedza na temat modelowanego zjawiska zyskałaby, gdybyśmy znali granice, w jakich wartość ceny dualnej pozostaje niezmienna.

Odpowiedzi na pytanie, czy optymalne rozwiązanie bazowe pozostanie optymalne i dopuszczalne – mimo zmiany wartości określonych parametrów – udziela analiza wrażliwości. Zajmuje się ona nie tylko wpływem zmian limitów ograniczeń na wyniki optymalizacji, ale także modyfikacji współczynników funkcji celu czy też parametrów lewych stron ograniczeń. Na początek spróbujemy określić granice przedziałów dopuszczalnych zmian wyrazów wolnych (prawych stron) ograniczeń. Przedział taki jest rozwiązaniem następującego układu nierówności:

$$\mathbf{x}^B = \mathbf{B}^{-1}\mathbf{b} \geq 0 \quad (46)$$

Nierówność we wzorze (46), zwana warunkiem dopuszczalności rozwiązania bazowego, oznacza, że wszystkie zmienne bazowe są nieujemne.

Przez dopuszczalną rozumiemy taką zmianę, przy której skład bazy optymalnej pozostaje wciąż taki sam. Układ nierówności (46) opiera się odwróconej macierzy bazy optymalnej, którą można odczytać z tablicy simpleksowej z ostatniej iteracji. Poniżej przedstawiamy układ nierówności dla pierwszego

ograniczenia z przykładu 2. Pierwszy element wektora \mathbf{b} został zastąpiony symbolem b_1 .

$$\begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} b_1 \\ 200 \\ 350 \end{bmatrix} \geq \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} \Rightarrow b_1 \in \langle 5250; +\infty \rangle$$

Dopóki limit kosztów nie spadnie poniżej pięciu tysięcy dwustu pięćdziesięciu złotych, optymalna wartość zmiennej dualnej nie ulegnie zmianie. Modyfikacje limitu kosztów w wyznaczonym przedziale nie będą wpływały na minimalną wartość kosztu zakupu herbaty (odpowiadająca temu ograniczeniu zmienna dualna pozostanie równa 0). Przedziały dla pozostałych ograniczeń są następujące: $b_2 \in \langle 17,5; +\infty \rangle$, $b_3 \in \left\langle 0; 533\frac{1}{3} \right\rangle$. Przedział dla garbnika jest obustronnie domknięty.

W jego przypadku wycena dualna była różna od 0, więc nawet zmiany w ramach dopuszczalnego przedziału wpływają na wartość funkcji celu.

Widzimy więc, że przeobrażenia prawej strony ograniczenia – w ramach przedziału dopuszczalnych modyfikacji – mogą (choć nie muszą) wpływać na rozwiązanie optymalne modelu. Zmiana wyrazu wolnego w i -tym ograniczeniu, w ramach przedziału dopuszczalnych zmian (przedziału wrażliwości), nie spowoduje zmiany optymalnej wartości ceny dualnej y_i^{opt} . Nie zmieni się też skład wektora zmiennych bazowych – choć ich wartości mogą (nie muszą) ulec zmianie. Zmienne niebazowe pozostają więc niebazowymi z wartością optymalną równą 0. Optymalna wartość funkcji celu może się zmienić, ale wykorzystanie limitów (w sensie: całkowicie/niecałkowicie) pozostaje takie samo.

Przedziały dopuszczalnych zmian da się wyznaczyć także dla współczynników funkcji celu. Sposób, w jaki się to robi, różni się w zależności od tego, czy dana zmienna decyzyjna jest bazowa, czy nie. Zaczniemy od parametru c_3 , stojącego przy zmiennej x_3 , będącej zmienną bazową rozwiązania optymalnego problemu importu herbaty.

Tym razem nie wykorzystujemy macierzy \mathbf{B}^{-1} , lecz wektor \mathbf{c}^B . Nadal jednak opieramy się na ostatniej tablicy sympleksowej. Zmiana wartości parametru c_3 spowoduje zmianę wartości wszystkich wskaźników optymalności dla zmiennych niebazowych (pomijamy zmienne sztuczne), czyli x_1 , x_2 i s_3 . Budujemy i rozwiązujemy odpowiedni układ nierówności dla wskaźników optymalności, przy czym – znak nierówności zależy od kierunku optymalizacji. Jeżeli funkcja celu jest maksymalizowana (minimalizowana), wtedy tworzymy układ z nierówności typu „ \leq ” („ \geq ”). Chodzi o to, żeby rozwiązanie – po zmianie parametru funkcji celu – nadal pozostało optymalne.

$$\left\{ \begin{array}{l} c_1 - z_1 = 250 - \begin{bmatrix} 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} 25 \\ -\frac{1}{4} \\ \frac{3}{4} \end{bmatrix} \geq 0 \Rightarrow c_3 \leq 333\frac{1}{3} \\ c_2 - z_2 = 210 - \begin{bmatrix} 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{3}{10} \\ \frac{7}{10} \end{bmatrix} \geq 0 \Rightarrow c_3 \leq 300 \\ c_6 - z_6 = 0 - \begin{bmatrix} 0 & 0 & c_3 \end{bmatrix} \begin{bmatrix} 15 \\ \frac{1}{20} \\ -\frac{1}{20} \end{bmatrix} \geq 0 \Rightarrow c_3 \geq 0 \end{array} \right. \quad (47)$$

Zmienna x_3 znajduje się na trzecim miejscu w bazie, więc trzeci element wektora \mathbf{c}^B został zastąpiony symbolem c_3 . Rozwiązując układ nierówności (47) względem c_3 , otrzymujemy w wyniku, że $c_3 \in \langle 0; 300 \rangle$. Dopóki cena zakupu herbaty trzeciego gatunku znajduje się w przedziale od zera do trzystu złotych, dopóty (przy założeniu, że pozostałe ceny i inne parametry zadania nie zmieniają się) rozwiązanie optymalne nie ulegnie zmianie. Nie znaczy to jednak, że nie pojawią się żadne inne efekty. Załóżmy, że cena tej herbaty spadnie o dziesięć złotych. Ponowne rozwiązanie modelu wykaże, że nadal należy kupić jej siedemnaście i pół tony, tyle że po niższej cenie. A zatem koszt zakupu spadnie do poziomu pięciu tysięcy siedemdziesięciu pięciu złotych, czyli o $10 \times 17,5 = 175$ [zł].

Zmiany parametru stojącego przy zmiennej niebazowej będą miały wpływ tylko na wskaźnik optymalności dla tej (niebazowej) zmiennej. Dlatego poszukujemy rozwiązania zaledwie jednej nierówności. Zajmijmy się, przykładowo, zmienną x_1 . Znak nierówności ustalamy, podobnie jak w przypadku zmiennej bazowej, w zależności od kierunku optymalizacji. Stąd dla parametru c_1 mamy:

$$c_1 - z_1 = c_1 - \begin{bmatrix} 0 & 0 & 300 \end{bmatrix} \begin{bmatrix} 25 \\ -\frac{1}{4} \\ \frac{3}{4} \end{bmatrix} \geq 0 \Rightarrow c_1 \leq 180$$

Dla parametru c_2 odpowiednia nierówność będzie miała postać:

$$c_2 - z_2 = c_2 - \begin{bmatrix} 0 & 0 & 300 \end{bmatrix} \begin{bmatrix} 0 \\ \frac{3}{10} \\ \frac{7}{10} \end{bmatrix} \geq 0 \Rightarrow c_2 \leq 210$$

Zatem przedziały dopuszczalnych zmian parametrów funkcji celu są następujące: $c_1 \in \langle 180, +\infty \rangle$ oraz $c_1 \in \langle 210, +\infty \rangle$.

Zmiana parametru przy zmiennej decyzyjnej w funkcji celu w ramach przedziału wrażliwości, zwanego przedziałem optymalności, nie powoduje zmiany rozwiązania optymalnego (nie zmienia się wartości zmiennych). Modyfikacjom mogą ulec natomiast: wartość funkcji celu oraz wyceny dualne.

Zauważmy, że krańce przedziałów dopuszczalnych zmian osiągają wartości, które nie zawsze mają ekonomiczne uzasadnienie, jak np. plus lub minus nieskończoność. Wynika to z czysto matematycznego podejścia do ich wyznaczania. Trzeba o tym pamiętać podczas interpretacji.

Wyjście wartości prawej strony ograniczenia lub współczynnika funkcji celu poza przedział dopuszczalnych zmian powoduje, że wzorce interpretacji podane powyżej przestają obowiązywać. Nie jest już możliwa analiza „co, jeśli”. Powstaje zupełnie nowy problem, który trzeba rozwiązać od początku.

Zastanówmy się teraz, jak na rozwiązanie optymalne wpłynie zmiana j -tej kolumny w macierzy $\mathbf{A} = [\mathbf{A}_1, \dots, \mathbf{A}_{n_1}, \dots, \mathbf{A}_n]$, gdzie $j = 1, \dots, n_1$ (kolumny odpowiadające zmiennym decyzyjnym). Zakładamy, że zmieni się przynajmniej jedna ze składowych takiego wektora. Oznaczmy go symbolem \mathbf{A}_j^* . Jeżeli wektor \mathbf{A}_j nie jest wektorem bazowym rozwiązania optymalnego, wtedy należy wyznaczyć nowe składowe odpowiedniej kolumny tablicy simpleksowej:

$$\mathbf{p}_j^* = \mathbf{B}^{-1} \mathbf{A}_j^* \quad (48)$$

dla której obliczamy nowy wskaźnik optymalności:

$$\Delta_j = c_j - (\mathbf{c}^B)^T \mathbf{p}_j^* \quad (49)$$

Jeżeli warunek optymalności zostanie spełniony, rozwiązanie pozostaje optymalne. Jeżeli nie, należy wprowadzić zmieniony wektor do kolejnej bazy i kontynuować postępowanie.

Obliczenia komplikują się, gdy A_j jest wektorem bazowym. Po pierwsze musimy sprawdzić, czy zmieniona macierz bazy (powstała przez zastąpienie wektora p_h wektorem A_j^*) nadal jest nieosobliwa (zachowana zostaje liniowa niezależność wektorów bazy). Jeżeli tak, to staje się ona nową macierzą bazy. Macierz odwrotną do nowej bazy otrzymamy następująco:

$$(B^*)^{-1} = E^{-1}B^{-1} \quad (50)$$

gdzie E jest macierzą przekształcenia elementarnego postaci:

$$E = [e_1 \quad e_2 \quad \dots \quad p_h \quad \dots \quad e_m] \quad (51)$$

a $e_i, i = 1, \dots, m$ – to wektory jednostkowe.

Następnie wyznaczamy nowe rozwiązanie bazowe i sprawdzamy, czy jest ono dopuszczalne i optymalne. Jeśli zmieniona macierz bazy okazuje się osobliwa, nie może być bazą. W takiej sytuacji rozwiązujemy zadanie od początku, podstawiając nowy wektor A_j^* w miejsce A_j , lub wykonujemy inną procedurę (Rogalska i inni, 1991, s. 114) – wychodzimy w niej od dotychczasowego rozwiązania optymalnego (przed zmianą kolumny macierzy A). Postępowanie składa się z następujących etapów:

1. Do układu ograniczeń dołączamy nową zmienną decyzyjną x_{n+1} ze współczynnikiem funkcji celu $c_{n+1} = c_j$, której odpowiada wektor A_j^* .
2. Obliczamy składowe wektora A_j^* względem bazy B , tj. p_j^* , według wzoru (48).
3. Kontynuujemy rozwiązywanie problemu, traktując dotychczasową zmienną x_j jako zmienną sztuczną.

Załóżmy, że czas potrzebny na obróbkę jednej tony herbaty gatunku pierwszego wzrósł do jednej godziny. Zmieniona kolumna macierzy A ma więc postać:

$$A_1^* = \begin{bmatrix} 250 \\ 1 \\ 15 \end{bmatrix}$$

Tak się składa, że zmienna x_1 nie jest w rozwiązaniu optymalnym przykładu 2 zmienną bazową. Wyznaczamy składowe zmienionego wektora p_1^* :

$$p_1^* = B^{-1}A_1^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 250 \\ 1 \\ 15 \end{bmatrix} = \begin{bmatrix} 25 \\ \frac{1}{4} \\ \frac{3}{4} \end{bmatrix}$$

dla którego obliczamy wartość nowego wskaźnika optymalności:

$$\Delta_1 = 250 - \begin{bmatrix} 0 & 0 & 300 \end{bmatrix} \begin{bmatrix} 25 \\ \frac{1}{4} \\ \frac{3}{4} \end{bmatrix} = 25 > 0$$

Wskaźnik optymalności okazał się większy od 0, czyli aktualne rozwiązanie pozostaje nadal optymalne. Kolejny przypadek dotyczy zmiany czasu przebywania w komorze herbaty gatunku trzeciego, który wzrasta do dwóch godzin. Nowa kolumna macierzy \mathbf{A} przybierze wówczas postać:

$$\mathbf{A}_3^* = \begin{bmatrix} 300 \\ 2 \\ 20 \end{bmatrix}$$

Zmienna x_3 jest tym razem zmienną bazową. Utwórzmy nową macierz bazy:

$$\mathbf{B}^* = \begin{bmatrix} \mathbf{p}_4 & \mathbf{p}_5 & \mathbf{A}_3^* \end{bmatrix} = \begin{bmatrix} 1 & 0 & 300 \\ 0 & 1 & 2 \\ 0 & 0 & 20 \end{bmatrix}$$

Jej wyznacznik równa się dwudziestu, czyli mamy do czynienia z macierzą nieosobliwą. Podobnie jak poprzednio, wyznaczmy nowe składowe wektora \mathbf{p}_3^* :

$$\mathbf{p}_3^* = \mathbf{B}^{-1} \mathbf{A}_3^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 200 \\ 2 \\ 20 \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$$

Tworzymy macierz \mathbf{E} , którą następnie odwracamy.

$$\mathbf{E} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{E}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Jesteśmy gotowi do wyznaczenia odwróconej zmienionej macierzy bazy:

$$(\mathbf{B}^*)^{-1} = \mathbf{E}^{-1}\mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{10} \\ 0 & 0 & \frac{1}{20} \end{bmatrix}$$

Sprawdzamy, czy nowe rozwiązanie bazowe jest dopuszczalne:

$$\mathbf{x}^{\mathbf{B}^*} = (\mathbf{B}^*)^{-1}\mathbf{b} = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{10} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix} = \begin{bmatrix} 2750 \\ 165 \\ 17,5 \end{bmatrix} \quad (52)$$

Wszystkie składowe wektora $\mathbf{x}^{\mathbf{B}^*}$ są dopuszczalne, ale nie wiemy, czy rozwiązanie (52) jest optymalne. Obliczmy więc wskaźniki optymalności dla zmiennych niebazowych (dla zmiennych bazowych z definicji są one równe 0). Żeby tego dokonać, potrzebujemy nowych wartości składowych kolumn tablicy simpleksowej.

$$\mathbf{p}_1^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{10} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 250 \\ 0,5 \\ 15 \end{bmatrix} = \begin{bmatrix} 25 \\ -1 \\ \frac{3}{4} \end{bmatrix}$$

$$\mathbf{p}_2^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{10} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 210 \\ 1 \\ 14 \end{bmatrix} = \begin{bmatrix} 0 \\ -\frac{2}{5} \\ \frac{7}{10} \end{bmatrix}$$

$$\mathbf{p}_6^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{10} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = \begin{bmatrix} 15 \\ 0,1 \\ -0,05 \end{bmatrix}$$

$$\mathbf{p}_7^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{10} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} -15 \\ -0,1 \\ 0,05 \end{bmatrix}$$

Odpowiadające im wskaźniki optymalności przyjmują wartości:

$$\Delta_1 = 25, \Delta_2 = 0, \Delta_6 = 15, \Delta_7 = M - 15$$

Widzimy więc, że nowe rozwiązanie jest optymalne i nie ma potrzeby kontynuowania obliczeń, chyba że chcielibyśmy poznać rozwiązanie alternatywne – bo ta cecha jest w naszym przykładzie wciąż obecna.

Gdyby jednak rozwiązanie bazowe \mathbf{x}^B miało przynajmniej jedną niedopuszczalną składową (w tym wypadku ujemną), wtedy należałoby posłużyć się dualną metodą *simplex* (opisaną w podrozdziale 3.6), aby znaleźć rozwiązanie dopuszczalne, po czym – od tego punktu zbioru rozwiązań prowadzić dalsze obliczenia, zgodnie z regułami algorytmu.

Założmy teraz, że zawartość garbnika w herbacie gatunku trzeciego spadnie do 0. Odpowiednia kolumna macierzy \mathbf{A} będzie wówczas miała następującą postać:

$$\mathbf{A}_3^* = \begin{bmatrix} 300 \\ 1 \\ 0 \end{bmatrix} \quad (53)$$

Nowa macierz bazy postaci:

$$\mathbf{B}^* = [\mathbf{p}_4 \quad \mathbf{p}_5 \quad \mathbf{A}_3^*] = \begin{bmatrix} 1 & 0 & 300 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad (54)$$

ma wyznacznik równy 0, a więc jest osobliwa.

Jeżeli nie chcemy rozpoczynać algorytmu *simplex* od samego początku, dołączamy do aktualnego układu ograniczeń nową zmienną decyzyjną x_4 ze współczynnikiem funkcji celu równym trzysta; zaś zmienną x_3 potraktujemy jak zmienną sztuczną, czyli przypiszemy jej w funkcji celu współczynnik równy M . Wyznamy składowe wektora \mathbf{p}_8 dla zmiennej x_4 i wstawimy go do tablicy simpleksowej.

$$\mathbf{p}_8^* = \mathbf{B}^{-1} \mathbf{A}_3^* = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 300 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 300 \\ 1 \\ 0 \end{bmatrix}$$

Tab. 8. Tablica simpleksowa po zmianie zawartości garbnika w przykładzie 2

\mathbf{c}^B	Baza	$[c_j]$	250	210	$+M$	0	0	0	$+M$	300	Wart. zm. baz.	Ilorazy wyj.
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	\mathbf{p}_7	\mathbf{p}_8		
			x_1	x_2	x_3	s_1	s_2	s_3	t_3	x_4	\mathbf{x}^B	
0	\mathbf{p}_4	s_1	25	0	0	1	0	15	15	300	2750	
0	\mathbf{p}_5	s_2	-1/4	3/10	0	0	1	1/20	-1/20	1	182,5	
$+M$	\mathbf{p}_3	x_3	3/4	7/10	1	0	0	-1/20	1/20	0	17,5	
$\Delta_j = c_j - z_j$			250 -3/4M	210 -7/10M	0	0	0	$\frac{1}{20}M$	$\frac{19}{20}M$	300	17,5M	

Źródło: opracowanie własne.

Wskaźniki w tabeli 8 informują, że nie mamy jeszcze do czynienia z rozwiązaniem optymalnym. Trzeba wykonać kolejne iteracje – aż do jego znalezienia lub stwierdzenia, że takowe rozwiązanie nie istnieje.

3.4. Metoda dwóch faz

Algorytm *simplex* bardzo ułatwia poszukiwanie rozwiązania optymalnego. Zwykle nie ma potrzeby odwiedzania wszystkich wierzchołków zbioru rozwiązań dopuszczalnych. Rzadko kiedy trzeba ich sprawdzić więcej niż $3/2m$ (gdzie m to liczba ograniczeń). Więcej rozważań na temat liczby iteracji znajdziemy w literaturze przedmiotu (zob. Sakarovitch, 1983, s. 109–110). Nie znaczy to jednak, że nie podejmowano prób usprawnienia tej metody. Miały one na celu m.in. zmniejszenie wysiłku obliczeniowego, zwiększenie dokładności obliczeń, poprawę stabilności numerycznej, zmniejszenie liczby iteracji, sprawne dochodzenie do pierwszego rozwiązania dopuszczalnego itd. Metoda dwóch faz służy ostatniemu z wymienionych celów. Jest ona szczególnie popularna w implementacjach komputerowych, ponieważ pozwala zmniejszyć zapotrzebowanie na moc obliczeniową i w określonych sytuacjach skrócić czas obliczeń.

Wprowadzenie do modelu zmiennych sztucznych powoduje, że w funkcji celu pojawia się symboliczny parametr M . Operowanie nim bywa niewygodne, szczególnie w programach komputerowych. Trzeba go wówczas zastępować konkretnymi wartościami liczbowymi – co może mieć znaczenie dla procesu obliczeniowego. Zauważmy też, że informację o tym, iż zadanie jest sprzeczne – zgodnie z tym, co wiemy na temat typów rozwiązań problemów optymalizacyjnych – algorytm *simplex* poda dopiero w ostatniej iteracji. W takiej sytuacji cały wysiłek włożony w wykonanie obliczeń można uznać za zmarnowany.

Dążenie do wyeliminowania z rozwiązania zmiennych sztucznych o niezerowych wartościach pozwala, z jednej strony, pozbyć się kary nałożonej na funkcję celu, z drugiej – zyskujemy szansę na wcześniejsze (biorąc pod uwagę liczbę iteracji) uzyskanie potwierdzenia, że mamy do czynienia z nierozwiązywalnym przy danych ograniczeniach problemem. Tu właśnie metoda dwóch faz ma pole do popisu. Jak sama nazwa wskazuje, postępowanie dzieli się na dwie fazy (Murthy, 2007, s. 84):

1. Rozwiązujemy model ze względu na zmienne sztuczne, starając się usunąć je z bazy. Przyjmujemy, że parametry pozostałych zmiennych decyzyjnych i swobodnych w funkcji celu są równe 0.
2. Rozwiązujemy model z pominięciem zmiennych sztucznych, przyjmując rozwiązanie z fazy 1 jako punkt startowy.

Wykorzystanie metody dwóch faz daje konkretne korzyści. Unikamy operowania parametrem M . Jak wspomnieliśmy, przydaje się to, gdy korzystamy z oprogramowania, które nie pozwala na zastosowanie danych symbolicznych. Szybciej niż w klasycznym postępowaniu ustalamy, czy mamy do czynienia z zadaniem sprzecznym. Jeżeli po zakończeniu fazy pierwszej w bazie pozostaje zmienna sztuczna z niezerową wartością, to mamy do czynienia z problemem nieposiadającym rozwiązania. Nie ma wówczas potrzeby przechodzenia do fazy drugiej – i właśnie tutaj pojawia się oszczędność czasu i wysiłku obliczeniowego.

Przykład 3

Rozważmy następujący model PL:



$$\begin{aligned} f(\mathbf{x}) &= 4x_1 + 5x_2 \rightarrow \max \\ 12x_1 + 30x_2 &\leq 60 \\ 10x_1 + 30x_2 &\geq 90 \\ x_1 &\geq 0, x_2 \geq 0 \end{aligned}$$

Postać kanoniczną dla pierwszej fazy zapiszemy:

$$\begin{aligned} f(\mathbf{x}) &= 0x_1 + 0x_2 + 0s_1 + 0s_2 - t_2 \rightarrow \max \\ 12x_1 + 30x_2 + s_1 &= 60 \\ 10x_1 + 30x_2 - s_2 + t_2 &= 90 \\ x_1 \geq 0, x_2 \geq 0, s_1 \geq 0, s_2 \geq 0, t_2 \geq 0 \end{aligned}$$

Jak łatwo zauważyć, funkcja celu postaci kanonicznej w fazie pierwszej wygląda inaczej niż w klasycznej wersji. Zmienne decyzyjne występują w niej z parametrem równym 0, a przy zmiennej sztucznej parametr $-M$ został zastąpiony przez -1 . W tabeli 9 znajdują się iteracje wykonane algorytmem *simplex* dla fazy pierwszej.

Tab. 9. Faza 1 dla przykładu 3

	\mathbf{c}^B	Baza	$[c_j]$	0	0	0	0	-1	Wart. zm. baz.	Ilorazy wyj.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5		
				x_1	x_2	s_1	s_2	t_2	\mathbf{x}^B	
1.	0	\mathbf{p}_4	s_1	12	30	1	0	0	60	2
	-1	\mathbf{p}_5	t_2	10	30	0	-1	1	90	3
	$\Delta_j = c_j - z_j$			10	30	0	-1	0	-90	
2.	0	\mathbf{p}_2	x_2	2/5	1	1/30	0	0	2	
	-1	\mathbf{p}_5	t_2	-2	0	-1	-1	1	30	
	$\Delta_j = c_j - z_j$			-2	0	-1	-1	0	-30	

Źródło: opracowanie własne.

W drugiej iteracji osiągnięte zostało rozwiązanie optymalne – o czym świadczą niedodatnie wskaźniki optymalności. Jednak w bazie pozostała zmienna sztuczna. Zadanie jest sprzeczne, więc nie przechodzimy do fazy drugiej. Kolejny przykład ilustruje nieco odmienną sytuację.

Przykład 4

Rozważmy kolejny model PL:

$$\begin{aligned} f(\mathbf{x}) &= 3x_1 + 5x_2 \rightarrow \max \\ 2x_1 + 3x_2 &\leq 20 \\ 2x_1 + x_2 &\geq 8 \\ x_1 \geq 0, x_2 &\geq 0 \end{aligned}$$

W tabeli 10 znajdziemy dwie iteracje fazy pierwszej – została ona zakończona pomyślnie. Usunęliśmy zmienną sztuczną t_3 z bazy, i możemy przejść do fazy drugiej. Punktem startowym będzie rozwiązanie optymalne z tabeli 10.

Tab. 10. Faza 1 dla przykładu 4

	\mathbf{c}^B	Baza	$[c_j]$	0	0	0	0	-1	Wart. zm. baz.	Ilorazy wyj.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5		
				x_1	x_2	s_1	s_2	t_2	\mathbf{x}^B	
1.	0	\mathbf{p}_4	s_1	2	3	1	0	0	20	10
	-1	\mathbf{p}_5	t_2	2	1	0	-1	1	8	4
		$\Delta_j = c_j - z_j$		2	1	0	-1	0	-8	
2.	0	\mathbf{p}_4	s_1	0	2	1	1	-1	12	
	0	\mathbf{p}_1	x_1	1	1/2	0	-1/2	1/2	4	
		$\Delta_j = c_j - z_j$		0	0	0	0	-1	0	

Źródło: opracowanie własne.

Tab. 11. Faza 2 dla przykładu 4

	\mathbf{c}^B	Baza	$[c_j]$	3	5	0	0	0	Wart. zm. baz.	Ilorazy wyj.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5		
				x_1	x_2	s_1	s_2	t_2	\mathbf{x}^B	
1.	0	\mathbf{p}_4	s_1	0	2	1	1		12	6
	3	\mathbf{p}_1	x_1	1	1/2	0	-1/2		4	8
		$\Delta_j = c_j - z_j$		0	7/2	0	3/2		12	
2.	5	\mathbf{p}_2	x_2	0	1	1/2	1/2		6	
	3	\mathbf{p}_1	x_1	1	0	-1/4	-3/4		1	
		$\Delta_j = c_j - z_j$		0	0	-7/4	-1/4		33	

Źródło: opracowanie własne.

Do wiersza c_j tabeli 11 wracają współczynniki funkcji celu przy zmiennych decyzyjnych, a kolumna p_5 pozostaje pusta. Pierwsza iteracja fazy drugiej nie była jeszcze rozwiązaniem optymalnym. Dlatego wykonana została wymiana wektorów. W jej wyniku zmienna x_2 zastąpiła w bazie zmienną s_2 . Druga iteracja kończy całe postępowanie w metodzie dwóch faz.

3.5. Zrewidowana metoda *simplex*

Klasyczna metoda *simplex* wymaga, aby na koniec każdej iteracji odwrócić zaktualizowaną macierz bazy. Im większa macierz do odwrócenia, tym bardziej pracochłonny jest to proces. Szczególnie uciążliwe będzie odwracanie źle uwarunkowanych macierzy. Jeśli ponadto zechcemy sięgnąć po któryś z języków programowania, pojawią się problemy numeryczne. Więcej na ten temat znajdziemy w literaturze przedmiotu. (zob. He, Bai, Jiang, Ouyang i Jiang, 2018). Zresztą oprogramowanie klasycznego algorytmu *simplex* jest generalnie niewdzięcznym zadaniem (por. Bogdan, 2018).

Podstawową różnicą między klasyczną a zrewidowaną metodą *simplex* okazuje się możliwość sprawdzania optymalności aktualnego rozwiązania bazowego bez konieczności wyznaczania centralnej części tablicy simpleksowej – tej, której kolumny zmieniają się pod wpływem przekształceń elementarnych. Wystarczy do tego znajomość rozwiązania sprzężonego zadania dualnego oraz wyjściowych parametrów problemu decyzyjnego. Odwoływanie się do oryginalnych parametrów zadania jest cechą charakterystyczną metody zrewidowanej. Choć samo postępowanie różni się od wersji klasycznej, to niezmiennie pozostają: sposób oceny optymalności, kryterium wejścia i kryterium wyjścia.

Niech aktualną dopuszczalną bazą zadania PL będzie nieosobliwa macierz \mathbf{B} . Zrewidowana metoda *simplex* wychodzi od tego samego punktu startowego, co klasyczna wersja algorytmu. Macierzą bazy jest więc na początku macierz jednostkowa, złożona z wektorów odpowiadających w ograniczeniach zmiennym swobodnym i sztucznym. Z założenia więc istnieje też macierz \mathbf{B}^{-1} , wymagana w pierwszym kroku algorytmu. Poszczególne etapy zrewidowanej metody *simplex* zostaną ujęte na poniższej liście (Hillier i Lieberman, 2015, s. 187).

1. Wyznacz wektor wartości zmiennych bazowych:

$$\mathbf{x}^B = \mathbf{B}^{-1}\mathbf{b} \quad (55)$$

2. Wyznacz wektor rozwiązania dualnego sprzężonego z bazą \mathbf{B} oraz wartość funkcji celu dla bieżącej bazy:

$$\mathbf{y} = (\mathbf{c}^B)\mathbf{B}^{-1} \quad (56)$$

$$f(\mathbf{x}^B) = \mathbf{y}\mathbf{b} \quad (57)$$

3. Wyznacz wskaźniki optymalności dla bieżącego rozwiązania bazowego, wykorzystując wektor wartości zmiennych bazowych (55):

$$\Delta_j = c_j - z_j = c_j - \mathbf{y}\mathbf{p}_j \quad (58)$$

4. Jeżeli aktualna baza \mathbf{B} jest bazą optymalną, zakończ postępowanie. W przeciwnym razie przejdź do kroku piątego.
 5. Zastosuj kryterium wejścia i ustal zmienną wchodzącą do bazy. Niech będzie to zmienna x_k , której odpowiada wektor \mathbf{p}_k .
 6. Wyznacz postać k -tej kolumny w tablicy simpleksowej dla bazy \mathbf{B} :

$$\mathbf{r}_k = \mathbf{B}^{-1}\mathbf{p}_k \quad (59)$$

7. Wyznacz pozycję wymiany wektorów w bazie (ilorazy wyjścia). Niech będzie to pozycja o numerze l -tym.
 8. Wyznacz macierz odwrotną do nowej bazy \mathbf{B} . Wróć do kroku pierwszego.

$$\mathbf{B}_{\text{nowa}}^{-1} = \mathbf{E}_{\text{aktualna}}^{-1} \mathbf{B}_{\text{aktualna}}^{-1} \quad (60)$$

Macierze \mathbf{E} i \mathbf{E}^{-1} mają postać:

$$\mathbf{E}_{\text{aktualna}} = \begin{bmatrix} 1 & \cdots & r_{1k} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & r_{lk} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & r_{mk} & \cdots & 1 \end{bmatrix} \quad (61)$$

$$\mathbf{E}_{\text{aktualna}}^{-1} = \begin{bmatrix} 1 & \cdots & -\frac{r_{1k}}{r_{lk}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{r_{lk}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\frac{r_{mk}}{r_{lk}} & \cdots & 1 \end{bmatrix} \quad (62)$$

Przykład 5

Wykorzystamy model z przykładu 2, aby przybliżyć postępowanie w zrewidowanej metodzie *simplex*. Na podstawie postaci kanonicznej tworzymy odpowiednie macierze i wektory:

$$\mathbf{P} = [\mathbf{p}_1 \quad \mathbf{p}_2 \quad \mathbf{p}_3 \quad \mathbf{p}_4 \quad \mathbf{p}_5 \quad \mathbf{p}_6 \quad \mathbf{p}_7] = \begin{bmatrix} 250 & 210 & 300 & 1 & 0 & 0 & 0 \\ \frac{1}{2} & 1 & 1 & 0 & 1 & 0 & 0 \\ 15 & 14 & 20 & 0 & 0 & -1 & 1 \end{bmatrix}$$

$$\mathbf{b} = \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix}, \quad \mathbf{c}^T = [250 \quad 210 \quad 300 \quad 0 \quad 0 \quad 0 \quad +M]$$

$$\mathbf{x}^T = [x_1 \quad x_2 \quad x_3 \quad s_1 \quad s_2 \quad s_3 \quad t_3]$$

Tworzymy pierwszą macierz bazy:

$$\mathbf{B}_1 = [\mathbf{p}_4 \quad \mathbf{p}_5 \quad \mathbf{p}_7] = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{B}_1^{-1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Iteracja pierwsza, krok pierwszy – wyznaczenie wektora wartości zmiennych bazowych:

$$\mathbf{x}^{\mathbf{B}_1} = \begin{bmatrix} s_1 \\ s_2 \\ t_3 \end{bmatrix} = \mathbf{B}_1^{-1} \mathbf{b} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix} = \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix}$$

Iteracja pierwsza, krok drugi – wyznaczenie wartości zmiennych dualnych i funkcji celu:

$$\mathbf{y} = (\mathbf{c}^{\mathbf{B}})^T \mathbf{B}_1^{-1} = [0 \quad 0 \quad M] \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [0 \quad 0 \quad M]$$

$$f(\mathbf{x}^{B_1}) = \mathbf{y}\mathbf{b} = \begin{bmatrix} 0 & 0 & M \end{bmatrix} \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix} = 350M$$

Iteracja pierwsza, krok trzeci – wyznaczenie wskaźników optymalności dla zmiennych niebazowych:

$$\Delta_1 = c_1 - \mathbf{y}\mathbf{p}_1 = 250 - \begin{bmatrix} 0 & 0 & M \end{bmatrix} \begin{bmatrix} 250 \\ \frac{1}{2} \\ 15 \end{bmatrix} = 250 - 15M$$

$$\Delta_2 = c_2 - \mathbf{y}\mathbf{p}_2 = 210 - \begin{bmatrix} 0 & 0 & M \end{bmatrix} \begin{bmatrix} 210 \\ 1 \\ 14 \end{bmatrix} = 210 - 14M$$

$$\Delta_3 = c_3 - \mathbf{y}\mathbf{p}_3 = 300 - \begin{bmatrix} 0 & 0 & M \end{bmatrix} \begin{bmatrix} 300 \\ 1 \\ 20 \end{bmatrix} = 300 - 20M$$

$$\Delta_6 = c_6 - \mathbf{y}\mathbf{p}_6 = 0 - \begin{bmatrix} 0 & 0 & M \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = M$$

Iteracja pierwsza, krok czwarty – wskaźniki optymalności informują, że bieżąca baza nie jest optymalna. Nie wszystkie wskaźniki są mniejsze lub równe 0.

Iteracja pierwsza, krok piąty – jako zmienną wchodzącą do bazy wybieramy x_3 .

Iteracja pierwsza, krok szósty – wyznaczamy kolumnę w tablicy simpleksowej dla zmiennej x_3 :

$$\mathbf{r}_3 = \mathbf{B}_1^{-1} \mathbf{p}_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 300 \\ 1 \\ 20 \end{bmatrix} = \begin{bmatrix} 300 \\ 1 \\ 10 \end{bmatrix}$$

Iteracja pierwsza, krok siódmy – ilorazy wyjścia:

$$\min \left\{ \frac{8000}{300}, \frac{200}{1}, \frac{350}{20} \right\} = \min \left\{ 26\frac{2}{3}, 200, 17\frac{1}{2} \right\} = 17\frac{1}{2}$$

Najniższy iloraz wyjścia wypadł dla zmiennej t_3 , a zatem – to ona opuści bazę. Zmienna x_3 wejdzie na jej miejsce.

Iteracja pierwsza, krok ósmy – wyznaczenie macierzy odwrotnej do macierzy bazy:

$$\mathbf{E}_1 = \begin{bmatrix} 1 & 0 & 300 \\ 0 & 1 & 1 \\ 0 & 0 & 20 \end{bmatrix} \Rightarrow \mathbf{E}_1^{-1} = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix}$$

$$\mathbf{B}_2^{-1} = \mathbf{E}_1^{-1} \mathbf{B}_1^{-1} = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix}$$

Iteracja druga, krok pierwszy – wyznaczenie wektora wartości zmiennych bazowych:

$$\mathbf{x}^{\mathbf{B}_2} = \begin{bmatrix} s_1 \\ s_2 \\ x_3 \end{bmatrix} = \mathbf{B}_2^{-1} \mathbf{b} = \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix} = \begin{bmatrix} 2750 \\ 182,5 \\ 17,5 \end{bmatrix}$$

Iteracja pierwsza, krok drugi – wyznaczenie wartości zmiennych dualnych i funkcji celu:

$$\mathbf{y} = (\mathbf{c}^{\mathbf{B}})^T \mathbf{B}_2^{-1} = \begin{bmatrix} 0 & 0 & 300 \end{bmatrix} \begin{bmatrix} 1 & 0 & -15 \\ 0 & 1 & -\frac{1}{20} \\ 0 & 0 & \frac{1}{20} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 15 \end{bmatrix}$$

$$f(\mathbf{x}^{B_2}) = \mathbf{y}\mathbf{b} = \begin{bmatrix} 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} 8000 \\ 200 \\ 350 \end{bmatrix} = 5250$$

Iteracja druga, krok trzeci – wyznaczanie wskaźników optymalności dla zmiennych niebazowych:

$$\Delta_1 = c_1 - \mathbf{y}\mathbf{p}_1 = 250 - \begin{bmatrix} 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} 250 \\ \frac{1}{2} \\ 15 \end{bmatrix} = 25$$

$$\Delta_2 = c_2 - \mathbf{y}\mathbf{p}_2 = 210 - \begin{bmatrix} 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} 210 \\ 1 \\ 14 \end{bmatrix} = 0$$

$$\Delta_6 = c_6 - \mathbf{y}\mathbf{p}_7 = 0 - \begin{bmatrix} 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ -1 \end{bmatrix} = 15$$

$$\Delta_7 = c_7 - \mathbf{y}\mathbf{p}_7 = M - \begin{bmatrix} 0 & 0 & 15 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} = M - 15$$

Wszystkie wskaźniki optymalności są niedodatnie. Otrzymane rozwiązanie jest więc optymalne. Bez trudu zauważymy dwie rzeczy. Po pierwsze liczba iteracji w klasycznej i zrewidowanej metodzie *simplex* jest taka sama. Po drugie rozwiązanie i wnioski z niego płynące również pokrywają się ze sobą. Nic nie stoi też na przeszkodzie, aby połączyć metodę zrewidowaną z metodą dwóch faz.

3.6. Dualna metoda *simplex*

W prymalnej metodzie *simplex* zaczynamy postępowanie od dowolnej bazy dopuszczalnej i, przeglądając kolejne bazy dopuszczalne, poszukujemy takiej, która okaże się bazą optymalną. Metoda prymalna zajmuje się bezpośrednio rozwiązaniami bazowymi problemu pierwotnego, które są prymalnie dopuszczalne, ale nie są dopuszczalne dualnie. Następnie dąży do znalezienia rozwiązania optymalnego, próbując osiągnąć dopuszczalność dualną. W metodzie dualnej rozpoczynamy od dowolnej bazy dualnie optymalnej, a procedurę kończymy w momencie,

gdy pewna baza optymalna jest jednocześnie dopuszczalna. Dualna metoda *simplex* rozpoczyna się od rozwiązania bazowego problemu pierwotnego, które jest dopuszczalne dualnie, ale już nie prymalnie.

Metoda dualna przydaje się w niektórych szczególnych sytuacjach. Zwykle łatwiej znaleźć początkowe rozwiązanie bazowe dopuszczalne prymalnie, a nie dualnie. Jednak czasami konieczne bywa wprowadzenie do modelu wielu zmiennych sztucznych. W takich przypadkach łatwiej zacząć od dualnego dopuszczalnego rozwiązania bazowego i zastosować właśnie dualną metodę *simplex*. Jeżeli przy tej okazji unikniemy konieczności sprowadzenia do 0 wielu zmiennych sztucznych, to bardzo spadnie liczba iteracji algorytmu.

Metoda przydaje się też podczas przeprowadzania analizy wrażliwości. Załóżmy, że analizujemy wpływ zmian w modelu na rozwiązanie optymalne. Jeśli to, które uzyskaliśmy, przestaje być prymalnie dopuszczalne z powodu wprowadzonych zmian (ale nadal spełnia kryteria optymalności), możemy wykorzystać metodę dualną jako punkt startowy – przyjmując rozwiązanie dualnie dopuszczalne. Zastosowanie dualnej metody *simplex* zwykle prowadzi do nowego optymalnego rozwiązania znacznie szybciej niż rozwiązanie nowego problemu od samego początku klasyczną metodą *simplex*.

Metoda dualna może być również przydatna w rozwiązywaniu od podstaw pewnych ogromnych problemów programowania liniowego – z racji swojej wydajności. Jest ona znacznie bardziej efektywna numerycznie, co przy dużych problemach ma znaczenie, które trudno przecenić. Schemat działania dualnej metody *simplex* jest następujący (Rogalska i inni, 1991, s. 99):

1. Wyznacz bazę dualnie dopuszczalną **B**.
2. Sprawdź, czy baza **B** jest prymalnie dopuszczalna. Jeżeli tak, zakończ postępowanie. W przeciwnym razie przejdź do kroku trzeciego.
3. Wyznacz wektor opuszczający bazę.
4. Sprawdź, czy można wyznaczyć wektor wchodzący do bazy. Jeżeli to niemożliwe, zadanie jest sprzeczne. Jeżeli można wyznaczyć wektor wchodzący, przejdź do kroku piątego.
5. Wyznacz nową bazę **B** i wróć do kroku drugiego.

Aby baza była na pewno dualnie dopuszczalna, wymagany jest model programowania liniowego z minimalizowaną funkcją celu oraz wszystkimi ograniczeniami typu „ \geq ”. W pozostałych przypadkach uzyskanie bazy dualnie dopuszczalnej wymaga wprowadzenia dodatkowego, sztucznego ograniczenia zdefiniowanego następująco:

$$\sum_{j=1}^n x_j \leq M \quad (63)$$

gdzie $M \gg 0$.

Jeżeli sztuczne ograniczenie nie zostało dodane, wtedy z bazy usuwamy wektor l taki, że:

$$l : x_l^B = \min_i \{x_i^B\} \quad (64)$$

Wybieramy pozycję najmniejszej ujemnej składowej w rozwiązaniu bazowym.

Po wprowadzeniu sztucznego ograniczenia w pierwszej iteracji przyjmujemy $l = m + 1$, czyli bazę opuszcza zmienna swobodna sztucznego ograniczenia – mimo jej dodatniej wartości. Jako zmienną wchodzącą do bazy wybieramy zmienną o najmniej korzystnej wartości wskaźnika optymalności.

Może dojść do sytuacji, w której – po zakończeniu procedowania dualnej metody *simplex* – zmienna swobodna sztucznego ograniczenia będzie niebazowa. Oznacza to przypadek braku skończonego rozwiązania optymalnego danego modelu PL.

Wektor wchodzący do bazy wybieramy, opierając się na ilorazach wejścia, obliczanych jako stosunek wskaźnika optymalności do odpowiadającego mu elementu w l -tym równaniu tablicy simpleksowej. Do bazy wchodzi wektor k , dla którego moduł ilorazu wejścia jest najmniejszy:

$$k : \left| \frac{c_k - z_k}{y_{lk}} \right| = \min_{y_{lj} < 0} \left\{ \frac{c_j - z_j}{y_{lj}} \right\} \quad (65)$$

Jeżeli zdarzy się, że powyższe kryterium nie daje jednoznacznego wyboru, to do bazy wchodzi wektor, dla którego dzielnik (y_{lj}) jest bliższy 0. Jeśli zaś w wierszu l nie występuje $y_{lj} < 0$, to nie można wymienić wektorów w bazie – mamy do czynienia z zadaniem sprzecznym.

Przykład 6

Znajdźmy rozwiązanie poniższego modelu programowania liniowego przy pomocy dualnej metody *simplex*:

$$f(\mathbf{x}) = x_1 + 2x_2 + 3x_3 \rightarrow \min \quad (66)$$

$$2x_1 - x_2 + 2x_3 \geq 4 \quad (67)$$

$$x_1 + x_2 + 2x_3 \geq 8 \quad (68)$$

$$x_2 - x_3 \geq 2 \quad (69)$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0 \quad (70)$$

Model (66)–(70) nie wymaga wprowadzenia sztucznego ograniczenia, dlatego postać kanoniczna jest następująca:

$$\begin{aligned}
f(\mathbf{x}) &= x_1 + 2x_2 + 3x_3 + 0s_1 + 0s_2 + 0s_3 \rightarrow \min \\
2x_1 - x_2 + 2x_3 - s_1 &= 4 \\
x_1 + x_2 + 2x_3 - s_2 &= 8 \\
x_2 - x_3 - s_3 &= 2 \\
x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, s_1 \geq 0, s_2 \geq 0, s_3 \geq 0
\end{aligned}$$

Baza początkowa wykorzystuje wektory współczynników przy zmiennych swobodnych. Jest ona dualnie dopuszczalna i prymalnie niedopuszczalna. Macierze utworzone na podstawie postaci kanonicznej są następujące:

$$\mathbf{P} = \begin{bmatrix} 2 & -1 & 2 & -1 & 0 & 0 \\ 1 & 1 & 2 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 & 0 & -1 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} 4 \\ 8 \\ 2 \end{bmatrix}$$

$$\mathbf{B} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \Rightarrow \mathbf{B}^{-1} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Wektory tworzące centralną część tablicy simpleksowej i wektor wartości zmiennych bazowych obliczamy ze wzorów:

$$\mathbf{Y} = \mathbf{B}^{-1}\mathbf{p} \quad (71)$$

$$\mathbf{x}^B = \mathbf{B}^{-1}\mathbf{b} \quad (72)$$

Stąd:

$$\mathbf{Y} = \begin{bmatrix} -2 & 1 & -2 & 1 & 0 & 0 \\ -1 & -1 & -2 & 0 & 1 & 0 \\ 0 & -1 & 1 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{x}^B = \begin{bmatrix} -4 \\ -8 \\ -2 \end{bmatrix}$$

W tabeli 12 znalazły się iteracje dualnej metody *simplex* dla przykładu 6. Wskaźniki optymalności w pierwszej iteracji są większe lub równe 0. Baza jest bazą dopuszczalną dla zadania dualnego, ale niedopuszczalną dla prymalnego. Zostanie z niej usunięta zmienna s_2 , ponieważ zachodzi:

$$\min\{-4, -8, -2\} = -8$$

Możemy więc obliczyć ilorazy wejścia (wiersz pod wskaźnikami optymalności). Przypominamy, że wskaźniki optymalności dzielimy przez ujemne elementy wiersza s_2 . Najniższy iloraz wejścia otrzymaliśmy dla zmiennej x_1 . Dokonujemy wymiany wektorów w bazie przy pomocy przekształceń elementarnych.

Druga baza nadal jest dopuszczalna dualnie, lecz niedopuszczalna prymalnie (ujemna wartość zmiennej bazowej s_3). Z bazy usunięta zostanie zmienna s_3 , a na jej miejsce wejdzie zmienna x_2 , która jako jedyna ma w tym wierszu ujemną składową.

W trzeciej iteracji wszystkie wskaźniki optymalności są nieujemne, zaś rozwiązanie bazowe jest prymalnie dopuszczalne. Znaleźliśmy więc rozwiązanie optymalne problemu (66)–(70), które przedstawia się następująco:

$$x_1^{opt} = 6, x_2^{opt} = 2, x_3^{opt} = 0, s_1^{opt} = 6, s_2^{opt} = 0, s_3^{opt} = 0, f_{\min} = 10$$

Znamy również optymalne wartości zmiennych dualnych:

$$y_1^{opt} = 0, y_2^{opt} = 1, y_3^{opt} = 1, g_{\max} = 10$$

Tab. 12. Dualna metoda *simplex* dla przykładu 6

	c^B	Baza	$[c_j]$	1	2	3	0	0	0	Wart. zm. baz.
			Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	
				x_1	x_2	x_3	s_1	s_2	s_3	x^B
1.	0	p_4	s_1	-2	1	-2	1	0	0	-4
	0	p_5	s_2	-1	-1	-2	0	1	0	-8
	0	p_6	s_3	0	-1	1	0	0	1	-2
	$\Delta_j = c_j - z_j$			1	2	3	0	0	0	0
	$ (c_j - z_j)/y_{ij} $			1	2	1,5				
2.	0	p_4	s_1	0	3	2	1	-2	0	12
	1	p_1	x_1	1	1	2	0	-1	0	8
	0	p_6	s_3	0	-1	1	0	0	1	-2
	$\Delta_j = c_j - z_j$			0	1	1	0	1	0	8
	$ (c_j - z_j)/y_{ij} $				1					
3.	0	p_4	s_1	0	0	5	1	-2	3	6
	1	p_1	x_1	1	0	3	0	-1	1	6
	2	p_2	x_2	0	1	-1	0	0	-1	2
	$\Delta_j = c_j - z_j$			0	0	2	0	1	1	10

Źródło: opracowanie własne.

Przykład 7

W tym przykładzie użyjemy dualnej metody *simplex* do znalezienia rozwiązania problemu decyzyjnego z przykładu 1. Model wymaga wprowadzenia do postaci kanonicznej dodatkowego, sztucznego ograniczenia. Przyjeliśmy, że zachodzi dla niego $M = 100\ 000$. W związku z tym mamy:

$$\begin{aligned} f(\mathbf{x}) &= 70x_1 + 40x_2 + 0s_1 + 0s_2 + 0s_3 + 0s_4 \rightarrow \max \\ x_1 + 2x_2 + s_1 &= 20\ 000 \\ 2x_1 + x_2 + s_2 &= 25\ 000 \\ 3x_1 + 3x_2 - s_3 &= 15\ 000 \\ x_1 + x_2 + s_4 &= 100\ 000 \\ x_1 \geq 0, x_2 \geq 0, s_1 \geq 0, s_2 \geq 0, s_3 \geq 0, s_4 \geq 0 \end{aligned}$$

Odpowiednie wyjściowe macierze są następujące:

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \Rightarrow \mathbf{B}^{-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} 1 & 2 & 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 1 & 0 & 0 \\ -3 & -3 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{x}^B = \begin{bmatrix} 20000 \\ 25000 \\ -15000 \\ 100000 \end{bmatrix}$$

Tab. 13. Dualna metoda *simplex* dla przykładu 7

	\mathbf{c}^B	Baza	$[c_j]$	70	40	0	0	0	0	Wart. zm. baz.
			Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	
				x_1	x_2	s_1	s_2	s_3	s_4	\mathbf{x}^B
1.	0	\mathbf{p}_3	s_1	1	2	1	0	0	0	20 000
	0	\mathbf{p}_4	s_2	2	1	0	1	0	0	25 000
	0	\mathbf{p}_5	s_3	-3	-3	0	0	1	0	-15 000
	0	\mathbf{p}_6	s_4	1	1	0	0	0	1	100 000
$\Delta_j = c_j - z_j$				70	40	0	0	0	0	0

Tab. 13 (cd.)

	c^B	Baza	$[c_j]$	70	40	0	0	0	0	Wart. zm. baz.
			Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	x^B
				x_1	x_2	s_1	s_2	s_3	s_4	
2.	0	p_3	s_1	0	1	1	0	0	-1	-80 000
	0	p_4	s_2	0	-1	0	1	0	-2	-175 000
	0	p_5	s_3	0	0	0	0	1	3	285 000
	70	p_1	x_1	1	1	0	0	0	1	100 000
$\Delta_j = c_j - z_j$				0	-30	0	0	0	-70	7 000 000
$ (c_j - z_j)/y_{ij} $					30				35	

Źródło: opracowanie własne.

Wymiana wektorów w bazie wygląda teraz inaczej, ponieważ obligatoryjnie usuwamy z bazy zmienną swobodną s_4 (dla sztucznego ograniczenia), a do bazy wchodzi zmienna x_1 , jako że odpowiada jej najmniej korzystny (tu najbardziej dodatni) wskaźnik optymalności. Poczynając od iteracji drugiej, postępowanie jest podobne do tego z przykładu 6.

W rozwiązaniu optymalnym (czwarta iteracja) zmienna swobodna s_4 ze sztucznego ograniczenia występuje jako zmienna bazowa. Oznacza to, że nasze zadanie posiada skończone rozwiązanie optymalne:

$$x_1^{opt} = 10\,000, x_2^{opt} = 5\,000, s_1^{opt} = 0, s_2^{opt} = 0, s_3^{opt} = 30\,000 \\ s_4^{opt} = 85\,000, f_{\max} = 900\,000$$

Tab. 14. Dualna metoda *simplex* dla przykładu 7 cd.

	c^B	Baza	$[c_j]$	70	40	0	0	0	0	Wart. zm. baz.
			Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	x^B
				x_1	x_2	s_1	s_2	s_3	s_4	
3.	0	p_3	s_1	0	0	1	1	0	-3	-255 000
	40	p_2	x_2	0	1	0	-1	0	2	175 000
	0	p_5	s_3	0	0	0	0	1	3	285 000
	70	p_1	x_1	1	0	0	1	0	-1	-75 000
$\Delta_j = c_j - z_j$				0	0	0	-30	0	-10	1 750 000
$ (c_j - z_j)/y_{ij} $									10/3	
4.	0	p_6	s_4	0	0	-1/3	-1/3	0	1	85 000
	40	p_2	x_2	0	1	2/3	-1/3	0	0	5 000
	0	p_5	s_3	0	0	1	1	1	0	30 000
	70	p_1	x_1	1	0	-1/3	2/3	0	0	10 000
$\Delta_j = c_j - z_j$				0	0	-10/3	-100/3	0	0	900 000

Źródło: opracowanie własne.

Dualna metoda *simplex* jest szybka i wydajna, co nie znaczy, że nie można próbować jej poprawić – jak zresztą uczyniono (zob. Vaidya, 2017), oponując jednocześnie wymianę dwóch wektorów w bazie.

4. Parametryczne programowanie liniowe

4.1. Parametryzacja współczynników funkcji celu

Do tej pory zakładaliśmy, że współczynniki funkcji celu czy prawe strony ograniczeń modeli decyzyjnych są ustalone i znane oraz przyjmują konkretne wartości liczbowe. Rozważmy jednak sytuację, w której stanowią one funkcje pewnych zmiennych, zwanych dalej parametrami. Przykładowo, zysk jednostkowy może zależeć od liczby pracowników przydzielonych do danej aktywności. Zagadnienia PL, w których znamy postać funkcji opisującej zależność między elementem modelu a pewnym parametrem, nazywamy parametrycznym programowaniem liniowym (PPL). W naszych rozważaniach zakładamy, że zależność ta ma charakter liniowy oraz będzie dotyczyć współczynników funkcji celu lub składowych wektora ograniczeń \mathbf{b} .

Przyjmijmy, że symbolem λ oznaczmy parametr, od którego zależy wartość współczynnika funkcji celu. Liniowy model decyzyjny z parametryczną funkcją celu zapiszemy wtedy następująco:

$$f(\mathbf{x}, \lambda) = [\mathbf{c} + \alpha\lambda]^T \mathbf{x} \rightarrow \max/\min \quad (73)$$

$$\mathbf{A}_1 \mathbf{x} \leq \mathbf{b}_1 \quad (74)$$

$$\mathbf{A}_2 \mathbf{x} = \mathbf{b}_2 \quad (75)$$

$$\mathbf{A}_3 \mathbf{x} \geq \mathbf{b}_3 \quad (76)$$

$$\mathbf{x} \geq 0 \quad (77)$$

Założmy, że znane jest rozwiązanie bazowe \mathbf{x}^B problemu (73)–(77) dla ustalonej wartości $\lambda = \lambda_0$. Składowe tego rozwiązania są niezależne od λ , co oznacza, że \mathbf{x}^B będzie dopuszczalne dla każdej jego wartości. W tej sytuacji interesuje nas, dla jakich wartości λ rozwiązanie jest optymalne. Przypomnijmy – rozwiązanie uznajemy za optymalne, kiedy wskaźniki optymalności przyjmują odpowiednie wartości. Te zaś obliczane są na podstawie współczynników funkcji celu. W konsekwencji wpływa na nie wartość parametru λ . Rozwiązanie optymalne dla $\lambda = \lambda_0$ nie musi już takie być dla $\lambda \neq \lambda_0$. Formalnie należałoby wyznaczyć rozwiązania optymalne dla wszystkich wartości przyjmowanych przez λ .

Postać funkcji celu, danej wzorem (73), wskazuje, że jeśli $\lambda = 0$, to otrzymujemy zwykle zadanie programowania liniowego. Dlatego właśnie taką wartość λ przyjmuje się jako wyjściową. Schemat postępowania w przypadku PPL da się zamknąć w następujących krokach (Rogalska i inni, 1991, s. 150):

1. Rozwiąż zadanie PPL, przyjmując $\lambda = 0$. Jeśli istnieje skończone rozwiązanie optymalne \mathbf{x}^B , przejdź do kroku drugiego, jeśli nie istnieje – przejdź do kroku trzeciego.
2. Wyznacz wskaźniki optymalności rozwiązania optymalnego jako funkcje parametru λ . Następnie na ich podstawie wyznacz kres dolny $\underline{\lambda}^B$ i kres górny $\bar{\lambda}^B$ przedziału zmienności λ . Możliwe są następujące przypadki:
 - jeśli $\underline{\lambda}^B < \lambda < \bar{\lambda}^B$, to rozwiązanie \mathbf{x}^B pozostaje optymalne i jednoznaczne;
 - jeśli $\lambda = \underline{\lambda}^B$ lub $\lambda = \bar{\lambda}^B$, to rozwiązanie \mathbf{x}^B pozostaje optymalne, ale jest niejednoznaczne;
 - jeśli $\lambda < \underline{\lambda}^B$ lub $\lambda > \bar{\lambda}^B$, to można otrzymać lepsze rozwiązanie bazowe, wprowadzając do bazy wektor \mathbf{p}_k , dla którego $\lambda = \underline{\lambda}^B$ lub $\lambda = \bar{\lambda}^B$.
3. Wyznacz zależność wskaźnika optymalności k -tego wektora, który należałoby wprowadzić do bazy, od parametru λ : $\Delta_k = \alpha_k + \beta_k \lambda$, a następnie oblicz $\lambda^B = -\frac{\alpha_k}{\beta_k}$. Możliwe są następujące przypadki:
 - jeśli $\beta_k = 0$ dla wszystkich wartości λ , to nie istnieje skończone rozwiązanie optymalne;
 - jeśli $\beta_k > 0$, to nie istnieje skończone rozwiązanie optymalne, gdy $\lambda > \lambda^B$ dla maksymalizowanej lub gdy $\lambda < \lambda^B$ dla minimalizowanej funkcji celu;
 - jeśli $\beta_k < 0$, to nie istnieje skończone rozwiązanie optymalne, gdy $\lambda < \lambda^B$ dla maksymalizowanej lub gdy $\lambda > \lambda^B$ dla minimalizowanej funkcji celu.
4. Jeśli $\lambda = \lambda^B$, to należy oddzielnie sprawdzić, czy istnieje skończone rozwiązanie optymalne.

Przykład 8

Wróćmy do przykładu 1. Decydent doszedł do wniosku, że cena jednostkowa filtra produkowanego przez firmę Paralel S.A. zależy od parametru λ , wyrażającego wpływ zakładanej marży. Wykonano odpowiednie analizy i stwierdzono, że funkcje cen jednostkowych zależą od λ w sposób liniowy. Oszacowanie współczynników dało w rezultacie następujące funkcje: $70 - \lambda$ dla pierwszego modelu i $40 + 3\lambda$ dla drugiego modelu filtra. W związku z tym model (1)–(5) przybierze postać:

$$f(\mathbf{x}) = (70 - \lambda)x_1 + (40 + 3\lambda)x_2 \rightarrow \max \quad (78)$$

$$x_1 + 2x_2 \leq 20\,000 \quad (79)$$

$$2x_1 + x_2 \leq 25\,000 \quad (80)$$

$$3x_1 + 3x_2 \geq 15\,000 \quad (81)$$

$$x_1 \geq 0, x_2 \geq 0 \quad (82)$$

Interesuje nas, dla jakich wartości parametru λ rozwiązanie jest optymalne.

Rozpocniemy od rozwiązania modelu przy założeniu, że $\lambda = 0$. W tabeli 15 znajduje się finalna tablica simpleksowa tej wersji modelu. Samo rozwiązanie optymalne jest takie samo jak to, które uzyskaliśmy wcześniej dwiema metodami. Wyznamy je teraz (zgodnie z drugim krokiem opisanej wyżej procedury) w zależności od parametru λ . Różnica polega na wprowadzeniu do tablicy zmodyfikowanych współczynników funkcji celu, jak miało to miejsce w tabeli 16. Centralna część tablicy oraz rozwiązanie bazowe \mathbf{x}^B nie uległy, jak widać, modyfikacji. Zmieniły się natomiast wskaźniki optymalności dla zmiennych swobodnych s_1 i s_2 . Teraz ich wartości zależą od wartości λ , ale kryterium optymalności rozwiązania pozostaje nadal takie samo – powinny one przyjmować wartości niedodatnie (funkcja celu jest maksymalizowana). Również wartość funkcji celu zależy od wprowadzonego parametru.

Tab. 15. Rozwiązanie optymalne modelu (78)–(82) dla $\lambda = 0$

\mathbf{c}^B	Baza	$[c_j]$	70	40	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	
			x_1	x_2	s_1	s_2	s_3	t_3	\mathbf{x}^B
40	\mathbf{p}_2	x_2	0	1	2/3	-1/3	0	0	5 000
0	\mathbf{p}_5	s_3	0	0	1	1	1	-1	30 000
70	\mathbf{p}_1	x_1	1	0	-1/3	2/3	0	0	10 000
$\Delta_j = c_j - z_j$			0	0	-10/3	-100/3	0	$-M$	900 000

Źródło: opracowanie własne.

Tab. 16. Rozwiązanie optymalne z tabeli 15 zależne od λ

\mathbf{c}^B	Baza	$[c_j]$	$70-\lambda$	$40+3\lambda$	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	
			x_1	x_2	s_1	s_2	s_3	t_3	\mathbf{x}^B
$40+3\lambda$	\mathbf{p}_2	x_2	0	1	2/3	-1/3	0	0	5 000
0	\mathbf{p}_5	s_3	0	0	1	1	1	-1	30 000
$70-\lambda$	\mathbf{p}_1	x_1	1	0	-1/3	2/3	0	0	10 000
$\Delta_j = c_j - z_j$			0	0	$-10/3+$ $-(7/3)\lambda$	$-100/3+$ $+(5/3)\lambda$	0	$-M$	720 000+ +5000 λ

Źródło: opracowanie własne.

Musimy zatem rozwiązać następujący układ nierówności:

$$\begin{cases} -\frac{10}{3}-\frac{7}{3}\lambda \leq 0 \\ -\frac{100}{3}+\frac{5}{3}\lambda \leq 0 \end{cases}$$

Po tym dowiadujemy się, że rozwiązanie z tabeli 15 będzie optymalne, skończone i jednoznaczne dla wszystkich wartości $\lambda \in \left(-\frac{10}{7}; 20\right)$.

Podstawmy teraz w tabeli 16 $\lambda = 20$. Powstanie w ten sposób tabela 17, w której wskaźnik optymalności dla niebazowej zmiennej s_2 jest równy 0. Możemy więc dokonać wymiany wektorów w bazie. Po wyznaczeniu ilorazów wyjścia okaże się, że s_2 wejdzie do bazy na miejsce zmiennej x_1 . Przy pomocy przekształceń elementarnych tworzymy kolejną tablicę simpleksową, dla której ponownie określimy przedział wartości parametru λ na podstawie formuł na wskaźniki optymalności.

Tab. 17. Rozwiązanie optymalne modelu (74)–(78) dla $\lambda = 20$

c^B	Baza	$[c_j]$	50	100	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	
			x_1	x_2	s_1	s_2	s_3	t_3	x^B
100	p_2	x_2	0	1	2/3	-1/3	0	0	5 000
0	p_5	s_3	0	0	1	1	1	-1	30 000
50	p_1	x_1	1	0	-1/3	2/3	0	0	10 000
$\Delta_j = c_j - z_j$			0	0	-50	0	0	$-M$	1 000 000

Źródło: opracowanie własne.

Tab. 18. Rozwiązanie optymalne z tabeli 17 zależne od λ

c^B	Baza	$[c_j]$	70- λ	40+3 λ	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	
			x_1	x_2	s_1	s_2	s_3	t_3	x^B
40+3 λ	p_2	x_2	1/2	1	1/2	0	0	0	10 000
0	p_5	s_3	-3/2	0	3/2	0	1	-1	15 000
0	p_4	s_2	3/2	0	-1/2	1	0	0	15 000
$\Delta_j = c_j - z_j$			50+ -(5/2) λ	0	-20+ -(3/2) λ	0	0	$-M$	400 000+ +30 000 λ

Źródło: opracowanie własne.

Poszukajmy rozwiązania poniższego układu nierówności:

$$\begin{cases} 50 - \frac{5}{2}\lambda \leq 0 \\ -20 - \frac{3}{2}\lambda \leq 0 \end{cases}$$

Rozwiązanie w tabeli 18 pozostanie zatem optymalne dopóty, dopóki $\lambda \in (20; +\infty)$. Zauważmy, że dla $\lambda = 20$ rozwiązanie stanie się niejednoznaczne. Funkcja celu osiągnie tę samą wartość maksymalną w dwóch wierzchołkach zbioru: tym z tabeli 17 oraz 18. Ponadto kraniec przedziału wyrażony jako nieskończoność nie daje się zinterpretować merytorycznie. Sam przedział wyznacza się czysto analitycznie – więc jego rozpiętość nie zawsze uda się pogodzić z otoczeniem problemu. Dlatego podczas interpretacji należy wziąć to pod uwagę i w razie konieczności przyjąć kraniec mający sens merytoryczny.

Wróćmy teraz do rozwiązania przedstawionego w tabeli 16 i przyjmijmy, że $\lambda = -\frac{10}{7}$. Po podstawieniu otrzymamy tabelę 19. Tym razem to przy zmiennej s_1 wartość wskaźnika optymalności informuje o istnieniu rozwiązania alternatywnego. Żeby je wyznaczyć, trzeba tę zmienną wprowadzić do bazy na miejsce zmiennej x_2 . Po wykonaniu odpowiednich przekształceń, powstanie nowa tablica simpleksowa, na podstawie której będziemy w stanie określić krańce trzeciego już przedziału wartości parametru λ . Dla $\lambda = -\frac{10}{7}$ istnieje więcej niż jedno rozwiązanie optymalne.

Tab. 19. Rozwiązanie optymalne modelu (78)–(82) dla $\lambda = -10/7$

c^B	Baza	$[c_j]$	500/7	250/7	0	0	0	-M	Wart. zm. baz.
		Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	
			x_1	x_2	s_1	s_2	s_3	t_3	x^B
250/7	p_2	x_2	0	1	2/3	-1/3	0	0	5 000
0	p_5	s_3	0	0	1	1	1	-1	30 000
500/7	p_1	x_3	1	0	-1/3	2/3	0	0	10 000
$\Delta_j = c_j - z_j$			0	0	0	-250/7	0	-M	6 250 000/7

Źródło: opracowanie własne.

Tab. 20. Rozwiązanie optymalne z tabeli 19 zależne od λ

$\mathbf{c^B}$	Baza	$[c_j]$	$70-\lambda$	$40+3\lambda$	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_3}$	$\mathbf{p_4}$	$\mathbf{p_5}$	$\mathbf{p_6}$	
			x_1	x_2	s_1	s_2	s_3	t_3	$\mathbf{x^B}$
0	$\mathbf{p_3}$	s_1	0	3/2	1	-1/2	0	0	12 500
0	$\mathbf{p_5}$	s_3	0	-3/2	0	3/2	1	-1	17 500
$70-\lambda$	$\mathbf{p_1}$	x_1	1	1/2	0	1/2	0	0	425 000/3
$\Delta_j = c_j - z_j$			0	$5 + (7/2)\lambda$	0	$-35 + (1/2)\lambda$	0	$-M$	$9916666\ 2/3 +$ $-(141666\ 2/3)\lambda$

Źródło: opracowanie własne.

Ze wskaźników optymalności dla zmiennych x_2 i s_2 utworzony zostanie układ nierówności:

$$\begin{cases} 5 + \frac{7}{2}\lambda \leq 0 \\ -35 + \frac{1}{2}\lambda \leq 0 \end{cases}$$

Wynika z niego, że bieżące rozwiązanie bazowe pozostanie optymalne, jeśli $\lambda \in \left(-\infty; -\frac{10}{7}\right)$. Lewy kraniec przedziału wypada w minus nieskończoności.

Zatem podczas interpretacji należy wziąć pod uwagę dolne ograniczenie, które będzie miało sens merytoryczny w kontekście otoczenia problemu.

Podsumujmy otrzymane wyniki. W zależności od wartości parametru λ , otrzymaliśmy przedziały i odpowiadające im rozwiązania zebrane w tabeli 21.

Tab. 21. Podsumowanie analizy dla współczynników funkcji celu

Lp.	Wartość λ	Rozwiązanie
1.	$\lambda < -\frac{10}{7}$	$x_1^{opt} = \frac{425000}{3}$, $x_2^{opt} = 0$, $s_1^{opt} = 12500$, $s_2^{opt} = 0$, $s_3^{opt} = 17500$
2.	$\lambda = -\frac{10}{7}$	Rozwiązanie niejednoznaczne. Tworzą je kombinacje liniowe rozwiązań nr 1 i nr 3
3.	$\lambda \in \left(-\frac{10}{7}; 20\right)$	$x_1^{opt} = 10\ 000$, $x_2^{opt} = 5000$, $s_1^{opt} = 0$, $s_2^{opt} = 0$, $s_3^{opt} = 30\ 000$
4.	$\lambda = 20$	Rozwiązanie niejednoznaczne. Tworzą je kombinacje liniowe rozwiązań nr 3 i nr 5
5.	$\lambda > 20$	$x_1^{opt} = 0$, $x_2^{opt} = 15\ 000$, $s_1^{opt} = 0$, $s_2^{opt} = 15\ 000$, $s_3^{opt} = 15\ 000$

Źródło: opracowanie własne.

Uważny czytelnik na pewno zauważył, że jeśli $\lambda > \bar{\lambda}^B$, to dolna granica przedziału dla parametru w nowym rozwiązaniu optymalnym będzie identyczna z górną granicą przedziału z rozwiązania sąsiedniego. Podobnie, jeśli $\lambda < \underline{\lambda}^B$, to górna dolna granica przedziału dla parametru w nowym rozwiązaniu optymalnym okaże się tożsama z dolną granicą przedziału z rozwiązania sąsiedniego. Właśnie dlatego osiągnięcie przez λ wartości któregoś z krańców kończy się rozwiązaniem niejednoznacznym.

4.2. Parametryzacja prawych stron ograniczeń

Drugim elementem modelu optymalizacyjnego, który zostanie uzależniony od pewnego parametru λ , będzie wektor prawych stron ograniczeń. W jego wypadku uzmiennienie wpływa na poziom przyjętych limitów ograniczeń – wpływ mogą mieć na nie wahania popytu, charakterystyka zastosowanej technologii, wywiązywanie się dostawców z umów itd. Model decyzyjny z wyrazem wolnym ograniczeń, uzależnionym od parametru λ , zapiszemy w pewnym uproszczeniu, pomijając mieszaną postać samych ograniczeń.

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \rightarrow \max \quad (83)$$

$$\mathbf{A}\mathbf{x} \leq \mathbf{b} + \boldsymbol{\beta}\lambda \quad (84)$$

$$\mathbf{x} \geq 0 \quad (85)$$

albo:

$$f(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \rightarrow \min \quad (83')$$

$$\mathbf{A}\mathbf{x} \geq \mathbf{b} + \boldsymbol{\beta}\lambda \quad (84')$$

$$\mathbf{x} \geq 0 \quad (85')$$

Model (83)–(85) albo (83')–(85') można zamienić na postać dualną. Wtedy parametr λ znajdzie się w funkcji celu, a problem stanie się analogiczny do tego, który opisaliśmy w poprzednim podrozdziale. Da się go więc rozwiązać zgodnie z procedurą stosowaną w przypadku parametrycznej funkcji celu. Istnieje też drugi sposób wyznaczania rozwiązania optymalnego zadania z parametryczną prawą stroną ograniczeń, opierający się na dualnej metodzie *simplex*. Można go przeprowadzić w następujących krokach (Rogalska i inni, 1991, s. 162):

1. Rozwiąż zadanie PPL, przyjmując $\lambda = 0$. Jeśli istnieje skończone rozwiązanie optymalne \mathbf{x}^B , przejdź do kroku drugiego; jeśli zadanie jest sprzeczne, przejdź do kroku trzeciego.
2. Dla zmiennych bazowych o nieujemnych wartościach wyznacz krańce przedziału $\underline{\lambda}^B$ i $\bar{\lambda}^B$ dopuszczalnego przedziału zmienności parametru λ . Jeśli $\lambda \in (\underline{\lambda}^B; \bar{\lambda}^B)$ lub gdy $\underline{\lambda}^B = -\infty$ i $\bar{\lambda}^B = +\infty$, rozwiązanie \mathbf{x}^B pozostaje dopuszczalne.

Dla pozostałych wartości λ rozwiązanie staje się niedopuszczalne. Przejdź do kroku czwartego.

3. Dla k zmiennych bazowych o ujemnych wartościach wyznacz: $x_k^B = \alpha_k + \beta_k \lambda$. Możliwe są następujące przypadki:

- jeśli $\beta_k = 0$ dla wszystkich k , to zadanie PPL jest sprzeczne dla wszystkich wartości parametru λ ;
- jeśli istnieją $\beta_k > 0$, to zadanie PPL jest sprzeczne dla wszystkich wartości

$$\lambda < \min_{\beta_k > 0} \left\{ -\frac{\alpha_k}{\beta_k} \right\};$$

- jeśli istnieją $\beta_k < 0$, to zadanie PPL jest sprzeczne dla wszystkich wartości

$$\lambda > \max_{\beta_k < 0} \left\{ -\frac{\alpha_k}{\beta_k} \right\};$$

- jeśli dla $\beta_k \neq 0$ parametr λ przybiera inne wartości niż w dwóch powyższych podpunktach, przejdź do kroku czwartego w celu znalezienia innego, dopuszczalnego rozwiązania bazowego.

4. Dokonaj wymiany wektorów bazowych zgodnie z regułami dualnej metody *simplex* i wróć do kroku drugiego.

Przykład 9

Kierownictwo firmy Paralel S.A. doszło do wniosku, że model decyzyjny dotyczący produkcji filtrów do odkurzaczy należy zmodyfikować o parametr wyrażający wpływ sytuacji rynkowej na przyjęte limity ograniczeń. Po konsultacjach z analitykami powstała zmodyfikowana wersja modelu opisana wzorami (82)–(86). Decydent chciałby dowiedzieć się, jak wprowadzone zmiany przełożą się na rozwiązanie optymalne.

$$f(\mathbf{x}) = 70x_1 + 40x_2 \rightarrow \max \quad (86)$$

$$x_1 + 2x_2 \leq 20\,000 + 3\lambda \quad (87)$$

$$2x_1 + x_2 \leq 25\,000 + \lambda \quad (88)$$

$$3x_1 + 3x_2 \geq 15\,000 - \lambda \quad (89)$$

$$x_1 \geq 0, x_2 \geq 0 \quad (90)$$

Zgodnie z pierwszym krokiem postępowania, rozwiązujemy zadanie (86)–(90), przyjmując, że $\lambda = 0$. Uzyskamy wówczas rozwiązanie optymalne przedstawione już wcześniej w tabeli 1. Pozwoli nam ono utworzyć odwróconą macierz bazy \mathbf{B}^{-1} , którą przemnożymy przez wektor prawych stron ograniczeń uzależnionych od λ .

Stąd wynika, że:

$$\mathbf{B}^{-1}[\mathbf{b} + \beta\lambda] = \begin{bmatrix} \frac{2}{3} & -\frac{1}{3} & 0 \\ 1 & 1 & -1 \\ -\frac{1}{3} & \frac{2}{3} & 0 \end{bmatrix} \begin{bmatrix} 20000 + 3\lambda \\ 25000 + \lambda \\ 15000 - \lambda \end{bmatrix} = \begin{bmatrix} 5000 + \frac{5}{3}\lambda \\ 30000 + 5\lambda \\ 10000 - \frac{1}{3}\lambda \end{bmatrix}$$

Otrzymaliśmy wektor rozwiązań bazowych dla tablicy simpleksowej, uzależniony od parametru λ . Rozwiązanie optymalne dla $\lambda = 0$ pozostanie dopuszczalne dopóki wartości wszystkich zmiennych bazowych będą większe lub równe 0. Zgodnie z poleceniem, w kroku drugim wyznaczmy granice jego przedziału zmienności, rozwiązując układ nierówności.

$$\begin{cases} 5000 + \frac{5}{3}\lambda \geq 0 \\ 30000 + 5\lambda \geq 0 \\ 10000 - \frac{1}{3}\lambda \geq 0 \end{cases}$$

Zatem dopóki $\lambda \in \langle -3\,000; 30\,000 \rangle$, nasze aktualne rozwiązanie pozostanie dopuszczalne. Sprawdzimy teraz, co się stanie jeśli $\lambda = -3000$. Zapiszmy tabelę 15 w nieco zmieniony sposób. Rozszerzamy ją o dodatkowy wiersz i kolumnę.

Ostatnia kolumna zawiera wartości zmiennych bazowych \mathbf{x}^B , gdy λ równa się przyjętej liczbie. Dzięki niej dowiemy się, która zmienna opuści bazę. Dodatkowy wiersz jest uzupełniany w taki sam sposób i pełni tę samą funkcję, jak w tablicach dualnej metody *simplex* – wskazuje wektor opuszczający bazę.

Tab. 22. Rozwiązanie optymalne dla pierwszej iteracji

\mathbf{c}^B	Baza	$[c_j]$	70	40	0	0	0	$-M$	Wart. zm. baz.	Rozw. baz. dla:
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6		
		x_1	x_2	s_1	s_2	s_3	t_3	\mathbf{x}^B		
40	\mathbf{p}_2	x_2	0	1	2/3	-1/3	0	0	$5\,000 + (5/3)\lambda$	0
0	\mathbf{p}_5	s_3	0	0	1	1	1	-1	$30\,000 + 5\lambda$	15 000
70	\mathbf{p}_1	x_1	1	0	-1/3	2/3	0	0	$10\,000 - \lambda$	11 000
$\Delta_j = c_j - z_j$			0	0	-10/3	-100/3	0	$-M$	$750\,000 + (310/3)\lambda$	440 000
$ \Delta_j/y_{ij} $						100				

Źródło: opracowanie własne.

Zgodnie z regułami dualnej metody *simplex*, zmienną opuszczającą bazę będzie x_2 . W tym wierszu tylko jeden element jest ujemny – odpowiadający zmiennej s_2 , a więc to ona wejdzie do bazy na pierwsze miejsce. Dokonujemy przekształceń elementarnych, aby otrzymać kolejną tablicę simpleksową. Uwaga! Przekształcenia te wykonujemy na kolumnie \mathbf{x}^B , ale już nie na ostatniej kolumnie. Tę uzupełnimy dopiero po wyznaczeniu przedziału dla λ .

Tab. 23. Rozwiązanie optymalne dla drugiej iteracji

\mathbf{c}^B	Baza	$[c_j]$	70	40	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6	
			x_1	x_2	s_1	s_2	s_3	t_3	\mathbf{x}^B
0	\mathbf{p}_4	s_2	0	-3	-2	1	0	0	-15 000 - 5 λ
0	\mathbf{p}_5	s_3	0	3	3	0	1	-1	45 000 + 10 λ
70	\mathbf{p}_1	x_1	1	2	1	0	0	0	20 000 + 3 λ
$\Delta_j = c_j - z_j$			0	-100	-70	0	0	$-M$	1 400 000 + 3 λ

Źródło: opracowanie własne.

Sprawdzamy, dla jakich wartości λ rozwiązanie z tabeli 23 pozostanie dopuszczalne, rozwiązując układ nierówności.

$$\begin{cases} -15000 - 5\lambda \geq 0 \\ 45000 + 10\lambda \geq 0 \\ 20000 + 3\lambda \geq 0 \end{cases}$$

Dopóki $\lambda \in \langle -4\,500; -3\,000 \rangle$, dopóty rozwiązanie zaprezentowane w tabeli 23 pozostanie dopuszczalne. Zauważmy, że jeśli $\lambda < -4\,500$, to zmienna s_3 będzie ujemna. Czyni to taką bazę niedopuszczalną – zadanie dla λ z tego przedziału jest sprzeczne.

Tab. 24. Rozwiązanie optymalne dla trzeciej iteracji

\mathbf{c}^B	Baza	$[c_j]$	70	40	0	0	0	$-M$	Wart. zm. baz.	Rozw. baz. dla:
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	\mathbf{p}_6		
			x_1	x_2	s_1	s_2	s_3	t_3	\mathbf{x}^B	$\lambda = 30\,000$
40	\mathbf{p}_2	x_2	0	1	2/3	-1/3	0	0	5000 + (5/3) λ	55 000
0	\mathbf{p}_5	s_3	0	0	1	1	1	-1	30 000 + 5 λ	180 000
70	\mathbf{p}_1	x_1	1	0	-1/3	2/3	0	0	10 000 - λ	0
$\Delta_j = c_j - z_j$			0	0	-10/3	-100/3	0	$-M$	750 000 + (310/3) λ	2 200 000
$ \Delta_j/y_{ij} $					10					

Źródło: opracowanie własne.

Zajmijmy się teraz drugim krańcem pierwszego wyznaczonego przedziału. Rozważymy przypadek, w którym $\lambda = 30\,000$. Po podstawieniu do finalnej tabeli simpleksowej dla $\lambda = 0$, otrzymujemy tabelę 24 (por. tabela 22).

Wymiana objęmię zmienne x_1 (wychodząca z bazy) i s_1 (wchodząca do bazy). Po wykonaniu stosownych obliczeń powstaje nowa tablica simpleksowa, przedstawiona w tabeli 25.

Tab. 25. Rozwiązanie optymalne dla czwartej iteracji

c^B	Baza	$[c_j]$	70	40	0	0	0	$-M$	Wart. zm. baz.
		Zm. baz.	p_1	p_2	p_3	p_4	p_5	p_6	
			x_1	x_2	s_1	s_2	s_3	t_3	x^B
40	p_2	x_2	2	1	0	1	0	0	$25\,000 + \lambda$
0	p_5	s_3	3	0	0	3	1	-1	$60\,000 + 4\lambda$
0	p_3	s_1	-3	0	1	-2	0	0	$-30\,000 + \lambda$
$\Delta_j = c_j - z_j$			-10	0	0	-40	0	$-M$	$1\,000\,000 + 40\lambda$

Źródło: opracowanie własne.

Ponownie wyznaczamy przedział zmienności λ dla sparymetryzowanych wartości wektora x^B .

$$\begin{cases} 25000 + \lambda \geq 0 \\ 60000 + 4\lambda \geq 0 \\ -30000 + \lambda \geq 0 \end{cases}$$

Okazuje się, że $\lambda \in (30\,000; +\infty)$. Zatem jeśli λ będzie wyższe niż 30 000, każde rozwiązanie bazowe okaże się dopuszczalne oraz optymalne. Podsumujmy zachowanie rozwiązania w zależności od wartości parametru λ .

Tab. 26. Podsumowanie analizy dla prawych stron ograniczeń

Lp.	Wartość λ	Rozwiązanie
1.	$\lambda < -4500$	Zadanie sprzeczne
2.	$\lambda \in (-4500; -3000)$	$x_1^{opt} = 20000 + 3\lambda$, $x_2^{opt} = 0$, $s_1^{opt} = 0$, $s_2^{opt} = -15000 - 5\lambda$ $s_3^{opt} = 45000 + 10\lambda$
3.	$\lambda \in (-3000; -30\,000)$	$x_1^{opt} = 10000 - \frac{1}{3}\lambda$, $x_2^{opt} = 5000 + \frac{5}{3}\lambda$, $s_1^{opt} = 0$ $s_2^{opt} = 0$, $s_3^{opt} = 30000 + 5\lambda$
4.	$\lambda > 30\,000$	Wszystkie rozwiązania są dopuszczalne i optymalne

Źródło: opracowanie własne.

Po raz kolejny zwracamy uwagę na fakt, że krańce przedziałów wyznaczono w sposób analityczny. λ równa, powiedzmy, minus dwóm tysiącom minut nie ma interpretacji merytorycznej i, analizując wpływ zmian parametru na wyniki modelu, trzeba wziąć to pod uwagę.

5. Programowanie liniowe w liczbach całkowitych

5.1. Konsekwencje wprowadzenia zmiennych o wartościach całkowitych

Optymalna decyzja dotycząca wielkości produkcji filtrów przez firmę Paralel S.A. (przykład 1) nie budzi naszych wątpliwości, co do poprawności merytorycznej. Zgodnie z rozwiązaniem – oba wyroby produkowane mają być w ilościach będących liczbami całkowitymi. Można odnieść wrażenie, że model zadbał o to, aby optymalna decyzja nie wskazywała na konieczność wytworzenia np. połowy filtra modelu pierwszego. Byłoby to jednak błędne przekonanie, gdyż w modelu nie uwzględniono wymagania, by produkcję wyrazić w liczbach całkowitych. Po prostu warunki, w jakich przyszło podejmować decyzję, ułożyły się w korzystny sposób. Dodanie tego niepozornego i intuicyjnie zrozumiałego warunku do modelu programowania liniowego powoduje pojawienie się poważnych komplikacji. Przyjrzyjmy się następującemu problemowi decyzyjnemu.

Przykład 10

Firma Paralel S.A. podjęła się realizacji zamówienia na wyprodukowanie czterech części montowanych w ekspresach do kawy. Jednostkowy zysk z wytworzenia poszczególnych części wynosi odpowiednio: osiemdziesiąt pięć, siedemdziesiąt, siedemdziesiąt i osiemdziesiąt złotych. Do produkcji wykorzystywany jest półfabrykat, którego zapas w magazynie znajduje się na poziomie tysiąca kilogramów, i na chwilę obecną nie ma możliwości zwiększenia tej ilości. Jednostkowe zużycie półfabrykatu to odpowiednio: pięć, dwa, trzy i dwa kilogramy. Łącznie części pierwszej i trzeciej należy wytworzyć co najmniej sto trzydzieści sztuk. Części drugiej i czwartej trzeba z kolei wyprodukować w sumie minimum trzysta sztuk. Produkcja części drugiej powinna wynieść co najmniej 30% łącznej wielkości produkcji. Należy opracować plan realizacji zamówienia gwarantujący jak najwyższy zysk przy świadomości, że każda wyprodukowana liczba części zostanie sprzedana.

Na polecenie decydenta analityk przygotował i przedstawił model programowania liniowego. Zawierał on cztery zmienne decyzyjne:

- x_1 – wielkość produkcji części I [szt.],
- x_2 – wielkość produkcji części II [szt.],

- x_3 – wielkość produkcji części III [szt.],
- x_4 – wielkość produkcji części IV [szt.].

Funkcja celu wyrażająca łączny koszt produkcji dana jest wzorem (91). nierówności (92)–(95) opisują otoczenie wpływające na podejmowaną decyzję.

$$f(\mathbf{x}) = 85x_1 + 70x_2 + 70x_3 + 80x_4 \rightarrow \max \quad (91)$$

$$5x_1 + 2x_2 + 3x_3 + 2x_4 \leq 1\,000 \quad (92)$$

$$x_1 + x_3 \geq 130 \quad (93)$$

$$x_2 + x_4 \geq 300 \quad (94)$$

$$x_2 \geq 0,3(x_1 + x_2 + x_3 + x_4) \quad (95)$$

$$x_1 \geq 0, x_2 \geq 0, x_3 \geq 0, x_4 \geq 0 \quad (96)$$

Rozwiązanie optymalne modelu (91)–(96) jest następujące:

$$\begin{aligned} x_1^{opt} = 0, \quad x_2^{opt} = 130,5, \quad x_3^{opt} = 130, \quad x_4^{opt} = 174,5 \\ s_1^{opt} = 0, \quad s_2^{opt} = 0, \quad s_3^{opt} = 5, \quad s_4^{opt} = 0, \quad f_{\max} = 32\,195 \end{aligned}$$

Model PL posiada jednoznaczne rozwiązanie optymalne, ale decydent nie może go zaakceptować – ponieważ zmienne decyzyjne x_2 i x_4 przyjmują niedopuszczalne merytorycznie wartości ułamkowe. Rozważono więc najprostsze wyjście z tej kłopotliwej sytuacji i zaokrąglono wyniki tam, gdzie było to konieczne. Ponownie obliczono wartości funkcji celu i zmiennych swobodnych, uzyskując wyniki:

$$\begin{aligned} x_1 = 0, \quad x_2 = 131, \quad x_3 = 130, \quad x_4 = 175 \\ s_1 = -2, \quad s_2 = 0, \quad s_3 = 6, \quad s_4 = 0,2, \quad f(\mathbf{x}) = 32\,270 \end{aligned}$$

Takie postępowanie doprowadziło do znalezienia rozwiązania niedopuszczalnego. Naruszono ograniczenie zapasu półfabrykatu. W tej sytuacji zdecydowano się na odrzucenie części ułamkowej liczb z pierwszego rozwiązania optymalnego. Po wykonaniu stosownych obliczeń otrzymano:

$$\begin{aligned} x_1 = 0, \quad x_2 = 130, \quad x_3 = 130, \quad x_4 = 174 \\ s_1 = 2, \quad s_2 = 0, \quad s_3 = 4, \quad s_4 = -0,2, \quad f(\mathbf{x}) = 32\,120 \end{aligned}$$

Tym razem naruszono ograniczenie udziału części drugiej. Rozwiązanie wciąż okazuje się niedopuszczalne.

Dlatego analityk dokonał ponownie optymalizacji, tym razem jednak dodał do modelu (91)–(96) warunek narzucający zmiennym decyzyjnym przyjmowanie wyłącznie wartości całkowitoliczbowych. W rezultacie powstała nowa decyzja optymalna:

$$x_1^{opt} = 0, x_2^{opt} = 131, x_3^{opt} = 130, x_4^{opt} = 174$$

$$s_1^{opt} = 0, s_2^{opt} = 0, s_3^{opt} = 5, s_4^{opt} = 0,5, f_{\max} = 32\,190$$

Rozwiązanie optymalne jest nie tylko dopuszczalne, ale nie budzi też wątpliwości merytorycznych – choć wartość funkcji celu okazała się być nieco niższa w porównaniu z wyjściową wersją.

Przytoczony przykład pokazał, że trywialny z pozoru problem, jakim jest spełnienie warunku całkowitoliczbowości przez zmienne decyzyjne, w rzeczywistości wymaga zmiany podejścia do poszukiwania rozwiązania optymalnego. Można pominąć ten kłopotliwy warunek i użyć którejś z metod PL (nazywamy to relaksacją zadania). O ile jednak takie rozwiązanie nie okaże się całkowitoliczbowe, narażamy się na problemy ze wskazaniem decyzji spełniającej wszystkie nasze oczekiwania.

Przyczyna różnic między zwykłym problemem programowania liniowego a problemem programowania liniowego w liczbach całkowitych (PLC) pojawia się już u samych fundamentów formułowania modelu. Pozornie są one bardzo do siebie podobne – w obu występuje liniowa funkcja celu i także ograniczenia. Zupełnie inaczej wyglądają jednak zbiory rozwiązań dopuszczalnych. Zmienne decyzyjne w PLC mogą przyjmować wyłącznie wartości całkowite. Przyjmijmy tu założenie twierdzące, że warunek całkowitoliczbowości nałożony został na wszystkie zmienne decyzyjne. To w gruncie rzeczy niewielkie uproszczenie. Graficznie zbiór tworzą więc punkty o całkowitych współrzędnych. Konsekwencją powyższego jest to, że zbiór rozwiązań dopuszczalnych zadania PLC (X_C) okazuje się niewypukły – choć zawiera się w zbiorze X zadania PL. Z faktu, że zbiór X_C nie jest wypukły, wynika niemożność wykorzystania twierdzenia Weierstrassa do znajdowania rozwiązania optymalnego zadania PLC. Odpadają więc wszystkie omówione do tej pory metody. Konieczne było opracowanie nowych, dających się pogrupować w następujący sposób:

1. Metody proste:
 - przegląd zupełny zbioru X_C ,
 - regularyzacja (zaokrąglenie) rozwiązania optymalnego zadania PL.
2. Metody złożone:
 - metody płaszczyzn odcinających (metody cięć),
 - przegląd zbioru X_C o charakterze kombinatorycznym,
 - poszukiwania przypadkowe i metody przybliżone.

Przegląd zupełny sens ma tylko wtedy, gdy zbiór X_C jest mały i skończony. Sprowadza się to do ustalenia całkowitoliczbowych współrzędnych punktów, będących jednocześnie decyzjami dopuszczalnymi. Kiedy już znamy elementy tego zbioru, szukamy punktu, w którym funkcja celu osiąga swoją wartość największą (najmniejszą).

Regularyzację można zaliczyć do klasy metod przybliżonych. Przykład 10 pokazał, że postępowanie takie kryje w sobie niebezpieczeństwo generowania roz-

wiazań PLC nienależących do zbioru X_C . Skala odstępstwa zależy od wielkości liczb opisujących rozwiązanie optymalne. Inny wpływ na rozwiązanie spowoduje zaokrąglenie liczby rzędu tysięcy, a jeszcze odmienny – liczby z przedziału od jednego do dziesięciu. Częściej, zamiast klasycznego matematycznego zaokrąglenia, stosuje się odcięcie części ułamkowej. Takie rozwiązanie zazwyczaj przesuwają nas w głąb zbioru rozwiązań i zwiększa szansę uniknięcia rozwiązań niedopuszczalnych. Wiemy już jednak, że nie jest to szansa stuprocentowa. Dlatego niektórzy autorzy (Trzaskalik, 2008, s. 106) odradzają takie postępowanie.

W metodach płaszczyzn odcinających do zadania dołącza się nowe ograniczenia, które muszą spełnić następujące warunki:

- odetną (nie obejmą) uzyskane wcześniej rozwiązanie zadania PL, gdyż nie było to rozwiązanie w liczbach całkowitych;
- będzie wiadomo, że nowe ograniczenia nie odetną wszystkich rozwiązań w liczbach całkowitych, tj. obejmą przynajmniej jedno rozwiązanie dopuszczalne w liczbach całkowitych.

Algorytm Gomory'ego, omówiony w dalszej części rozdziału, zalicza się do tej grupy. Płaszczyznami odcinającymi będą odpowiednio przekształcone równania z tablicy sympleksowej, zawierającej rozwiązanie optymalne rozszerzonego zadania PL.

W przeglądzie o charakterze kombinatorycznym rozwiązuje się zadanie PL z pominięciem warunku całkowitoliczbowości. Jeśli rozwiązanie nie jest całkowitoliczbowe, to buduje się kolejne zadania. Są one tak skonstruowane, że suma ich zbiorów rozwiązań dopuszczalnych odpowiada zbiorowi rozwiązań dopuszczalnych zadania wyjściowego. Reprezentantką tej grupy metod jest np. metoda podziału i ograniczeń – bardzo popularna w pakietach komputerowych. Zadanie PL dzielimy na kolejne zadania. Zbiór rozwiązań dopuszczalnych X jest w wyniku każdego podziału zawężany. Po skończonej liczbie podziałów zadania PL uzyskuje się rozwiązanie zadania PLC, o ile takie istnieje. Do rozwiązania zadań podzielnym wykorzystuje się takie metody, jak klasyczny algorytm *simplex*, zrewidowany algorytm *simplex* itd.

Do ostatniej grupy metod zalicza się zastosowania wykorzystujące algorytmy heurystyczne, jak np. algorytmy genetyczne. Często zawierają one uproszczone fragmenty sygnalizowanych wcześniej postępowań.

5.2. Metoda podziału i ograniczeń

Metoda podziału i ograniczeń jest jedną z bardziej popularnych metod poszukiwania rozwiązania optymalnego zadań PLC, mimo że wymaga niekiedy wykonania dużej ilości obliczeń. Ogólna jej idea polega na ciągłym podziale i rozwiązywaniu zadań programowania liniowego. Tu chyba należy szukać przyczyny jej popularności, szczególnie w zastosowaniach informatycznych. Wykorzystujemy bowiem znany i sprawdzony algorytm metody *simplex*.

Optymalna wartość funkcji celu zagadnienia PLC, z maksymalizowaną funkcją kryterium, jest nie większa niż optymalna wartość funkcji celu tego samego zadania z usuniętym warunkiem całkowitoliczbowości. Innymi słowy, wartość funkcji celu, odpowiadająca rozwiązaniu optymalnemu zadania PL, okazuje się górnym ograniczeniem maksymalnej wartości funkcji celu zadania PLC. Na tej własności opiera się metoda podziału i ograniczeń.

Dla uproszczenia dalszych rozważań zakładamy, że zadanie polega na poszukiwaniu wartości największej funkcji celu (maksymalizacja). Jeżeli jest odwrotnie (minimalizacja), to mnożymy funkcję celu przez -1 , a po zakończeniu postępowania jej wartość optymalną należy ponownie pomnożyć przez -1 .

Do zadania PLC dołączamy dodatkowe warunki, tzw. ograniczenia widełkowe, ustanowione dla wszystkich zmiennych decyzyjnych:

$$d_j \leq x_j \leq g_j \quad (97)$$

Granice warunków widełkowych powinny być liczbami całkowitymi. Najczęściej przyjmuje się, że dolne ograniczenia dla zmiennych są równe 0 ($d_j = 0$). Z kolei dla górnych ograniczeń przyjmuje się dostatecznie dużą całkowitą liczbę M . W sensie geometrycznym dobór parametrów określających dolną i górną wartość zmiennej x_j jest taki, że hiperprostopadłościan generowany przez te parametry pokrywa na początek zbiór rozwiązań dopuszczalnych X zadania programowania liniowego.

Zanim użyjemy samego algorytmu, rozwiązujemy zadanie programowania liniowego z dodatkowymi warunkami widełkowymi. Jeżeli zadanie to jest sprzeczne lub rozwiązanie okazuje się całkowitoliczbowe, kończymy postępowanie. Gdy z kolei optymalna wartość jakiegokolwiek zmiennej decyzyjnej jest równa oryginalnej (początkowej) wartości jej górnego ograniczenia (g_j), to zadanie PLC nie posiada skończonego rozwiązania optymalnego. Jeśli zaś istnieje skończone rozwiązanie, lecz przynajmniej jedna zmienna decyzyjna z nałożonym warunkiem całkowitoliczbowości nie przyjęła wartości całkowitej, przechodzimy do kroku czwartego opisanego niżej schematu działania. Sam algorytm metody podziału i ograniczeń da się rozpisać w następujących krokach:

1. Uporządkuj listę zadań – z listy usuń: zadania już podzielone, zadania sprzeczne oraz te, które mają wartość funkcji celu mniejszą lub równą wartości funkcji celu zadań spełniających warunki całkowitoliczbowości. Pozostałe na liście zadania nazywamy zadaniami aktywnymi.
2. Sprawdź, czy istnieje takie zadanie aktywne, którego rozwiązanie optymalne spełnia warunki całkowitoliczbowości, a jednocześnie na liście nie ma żadnego innego zadania aktywnego lub wszystkie pozostałe zadania aktywne mają wartość funkcji celu nie większą niż w takim zadaniu. Jeżeli istnieje takie zadanie aktywne, zakończ postępowanie. Zadanie to jest rozwiązaniem optymalnym zadania PLC. Jeżeli nie, to przejdź do kolejnego kroku.

3. Wybierz zadania do podziału – jako zadanie do podziału wybieramy to, które ma największą wartość funkcji celu i nie spełnia warunków całkowitoliczbowości.
4. Wybierz zmienną, według której wykonany zostanie podział zadania – podział zadania dokonuj zawsze ze względu na dowolnie wybraną zmienną, która w rozwiązaniu optymalnym nie miała wartości całkowitej.
5. Podział zadania – w wyniku podziału zadania z kroku trzeciego (zadanie „matka”) powstaną zawsze dwa nowe zadania (zadanie „córka” oraz zadanie „syn”). Są one kopiami zadania dzielonego i różnią się wyłącznie ograniczeniem widełkowym dla zmiennej k , które modyfikujemy następująco:
 - dla pierwszego z zadań („córka”) przyjmujemy: $d_k \leq x_k \leq [x_k^{opt}]$,
 - dla drugiego z zadań („syn”) przyjmujemy: $[x_k^{opt}] + 1 \leq x_k \leq g_k$.
6. Rozwiąż zadania z aktualnego podziału – po rozwiązaniu obu nowych zadań przejdź do kroku pierwszego.

W sensie geometrycznym w zbiorze rozwiązań dopuszczalnych X zadania „matka” wycinane jest pasmo $[x_k^{opt}] < x_k < [x_k^{opt}] + 1$, co prowadzi do podziału tego zbioru na dwa podzbiory związane odpowiednio z zadaniami „córka” i „syn”.

Przykład 11

Rozważmy model (zob. Szapiro, 2000), w którym maksymalizacji podlega zysk z produkcji dwóch wyrobów. Ograniczenia tworzą dwa surowce $U1$ i $U2$, których posiadane zapasy są znane. Model decyzyjny ma postać:

$$f(\mathbf{x}) = 3x_1 + 4,5x_2 \rightarrow \max \quad (98)$$

$$U1: \quad 6x_1 + 10x_2 \leq 36 \quad (99)$$

$$U2: \quad 5x_1 + 4x_2 \leq 20 \quad (100)$$

$$x_1 \geq 0, x_2 \geq 0, x_1, x_2 - \text{całkowite} \quad (101)$$

Kolejne etapy rozwiązywania będziemy ilustrować wykresami znanymi z metody graficznej, które świetnie pokażą czemu w nazwie metody pojawiają się terminy „podział” i „ograniczenia”. Utworzone oznaczać będziemy $Z_{nr \text{ bieżącego zad./nr zad. „matki”}}$

Na początek rozszerzamy model o ograniczenia widełkowe. Uznaliśmy za wystarczające, żeby $M = 100$. Dodatkowe ograniczenia mają więc postać:

$$0 \leq x_1 \leq 100 \quad (102)$$

$$0 \leq x_2 \leq 100 \quad (103)$$

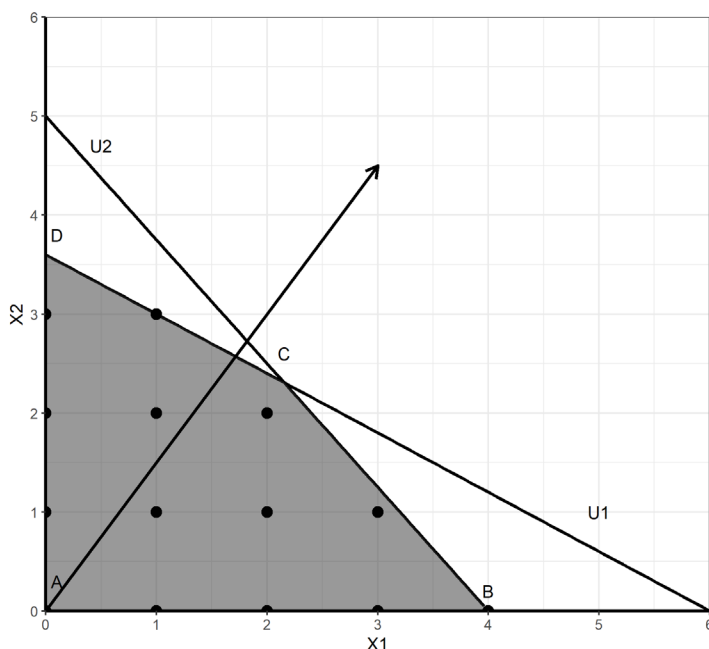
Rysunek 10 przedstawia zbiór rozwiązań zadania (98)–(101) z pominięciem warunku całkowitoliczbowości (wielobok $ABCD$) oraz punkty będące rozwiązaniami dopuszczalnymi o całkowitych współrzędnych. Wyraźnie widać różnicę

między owymi zbiorami rozwiązań. Dołączenie warunku całkowitoliczbowości sprawia, że tylko niektóre punkty należy sprawdzić pod kątem rozwiązania optymalnego. Istnieją za to obszary nieciągłości, które podczas przeszukiwania trzeba pominąć. Możemy więc zapomnieć o obliczaniu pochodnych i wykorzystywaniu gradientu. Oczywiście przy tak małym przykładzie dałoby się wykonać przegląd zupełny, ale naszym zamiarem pozostaje zaprezentowanie metody podziału i ograniczeń.

Na początek rozwiązujemy zadanie PL z dołączonymi ograniczeniami widełkowymi. Wyniki zapiszemy w tabeli 27. Jest to rozwiązanie pierwszego zadania „matka” i dlatego druga część symbolu została pominięta.

Funkcja celu osiągnęła wartość maksymalną w wierzchołku C (por. rysunek 10). Rozwiązanie optymalne jest jednoznaczne i skończone, ale nie całkowitoliczbowe. Przechodzimy do kroku czwartego algorytmu. Podzielimy zadanie $Z_{1/-}$. Podziału dokonamy dla zmiennej x_1 . Dzielimy ograniczenie widełkowe dla zmiennej x_1 (krok piąty). W zadaniu „córka” ($Z_{2/1}$) będzie ono wyglądało następująco:

$0 \leq x_1 \leq \left\lfloor 2 \frac{2}{13} \right\rfloor$, czyli $0 \leq x_1 \leq 2$; zaś w zadaniu „syn”: ($Z_{3/1}$): $\left\lfloor 2 \frac{2}{13} \right\rfloor + 1 \leq x_1 \leq 100$, czyli $3 \leq x_1 \leq 100$.



Rys. 10. Zbiór rozwiązań modelu (98)–(101)

Źródło: opracowanie własne.

Tab. 27. Wyniki przykładu 11 dla zadania $Z_{1/-}$

$Z_{1/-}$	wartość	d_j	g_j
x_1	$2\frac{2}{13}$	0	100
x_2	$2\frac{4}{13}$	0	100
f_{\max}	16,85		

Źródło: opracowanie własne.

W efekcie powstaną dwa modele optymalizacyjne, które należy rozwiązać (krok szósty).

$Z_{2/1}$:

$$f(\mathbf{x}) = 3x_1 + 4,5x_2 \rightarrow \max$$

$$U1: \quad 6x_1 + 10x_2 \leq 36$$

$$U2: \quad 5x_1 + 4x_3 \leq 20$$

$$3a: \quad x_1 \leq 2$$

$$x_1 \geq 0, x_2 \geq 0$$

$Z_{3/1}$:

$$f(\mathbf{x}) = 3x_1 + 4,5x_2 \rightarrow \max$$

$$U1: \quad 6x_1 + 10x_2 \leq 36$$

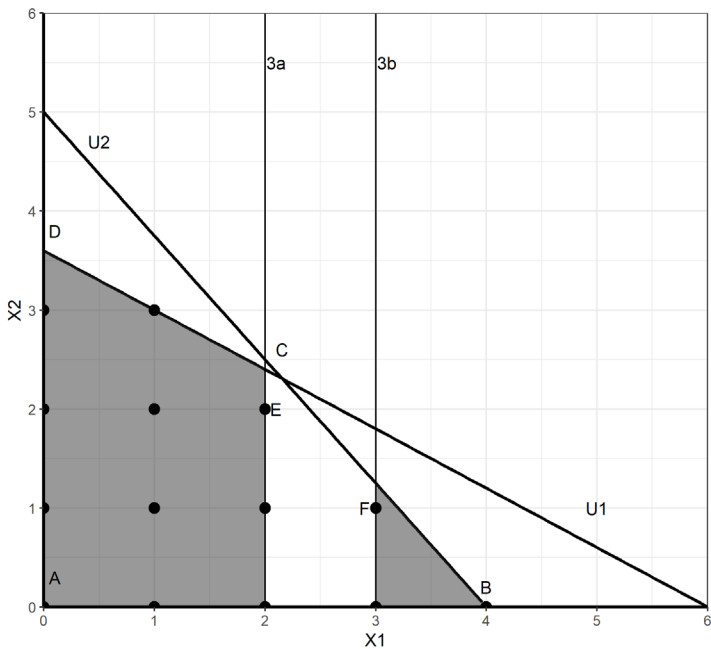
$$U2: \quad 5x_1 + 4x_3 \leq 20$$

$$3b: \quad x_1 \geq 3$$

$$x_1 \geq 0, x_2 \geq 0$$

Wykres na rysunku 11 przedstawia zbiory rozwiązań zadań $Z_{2/1}$ i $Z_{3/1}$. Ograniczenia 3a i 3b podzieliły wyjściowy zbiór na dwie części. Oba zadania posiadają skończone rozwiązania optymalne, znajdujące się w tabeli 28. Na rysunku są to punkty E i F. Wracamy do kroku pierwszego algorytmu i przystępujemy do porządkowania listy zadań. Aktualna lista jest następująca:

- $Z_{1/-}, f_{\max} = 16,85$: rozwiązanie niecałkowitoliczbowe;
- $Z_{2/1}, f_{\max} = 16,8$: rozwiązanie niecałkowitoliczbowe;
- $Z_{3/1}, f_{\max} = 14,625$: rozwiązanie niecałkowitoliczbowe.



Rys. 11. Zbiory rozwiązań modeli $Z_{2/1}$ i $Z_{3/1}$

Źródło: opracowanie własne.

Tab. 28. Wyniki przykładu 11 dla zadań $Z_{2/1}$ i $Z_{3/1}$

$Z_{2/1}$	wartość	d_j	g_j	$Z_{3/1}$	wartość	d_j	g_j
x_1	2	0	2	x_1	3	3	100
x_2	2,4	0	100	x_2	1,25	0	100
f_{\max}	16,8			f_{\max}	14,625		

Źródło: opracowanie własne.

Usuwamy z listy zadanie $Z_{1/-}$ (jest już podzielone). Uporządkowaną listę zadań tworzą $Z_{2/1}$ i $Z_{3/1}$. Tu kończy się krok pierwszy procedury. W kroku drugim sprawdzamy, czy można zakończyć postępowanie. Brak zadań aktywnych o rozwiązaniu w liczbach całkowitych oznaczają, że przechodzimy do kroku trzeciego – wyboru zadania do podziału. Decydujemy się na to o największej wartości funkcji celu spośród zadań nadających się do podziału – w tym wypadku $Z_{2/1}$. W kroku czwartym wybieramy zmienną, według której dokonamy podziału. Nasz wybór to oczywiście x_2 .

Dzielimy ograniczenie widelkowe dla zmiennej x_2 (krok piąty). Dla zadania „córka” ($Z_{4/2}$) będzie ono wyglądało następująco: $0 \leq x_2 \leq [2,4]$, czyli $0 \leq x_2 \leq 2$; zaś

w zadaniu „syn”: $(Z_{5/2}): [2,4] + 1 \leq x_1 \leq 100$, czyli $3 \leq x_2 \leq 100$. W efekcie powstaną następujące dwa modele optymalizacyjne, które należy rozwiązać (krok szósty).

$Z_{4/2}$:

$$f(\mathbf{x}) = 3x_1 + 4,5x_2 \rightarrow \max$$

$$U1: 6x_1 + 10x_2 \leq 36$$

$$U2: 5x_1 + 4x_2 \leq 20$$

$$3a: x_1 \leq 2$$

$$4a: x_2 \leq 2$$

$$x_1 \geq 0, x_2 \geq 0$$

$Z_{5/2}$:

$$f(\mathbf{x}) = 3x_1 + 4,5x_2 \rightarrow \max$$

$$U1: 6x_1 + 10x_2 \leq 36$$

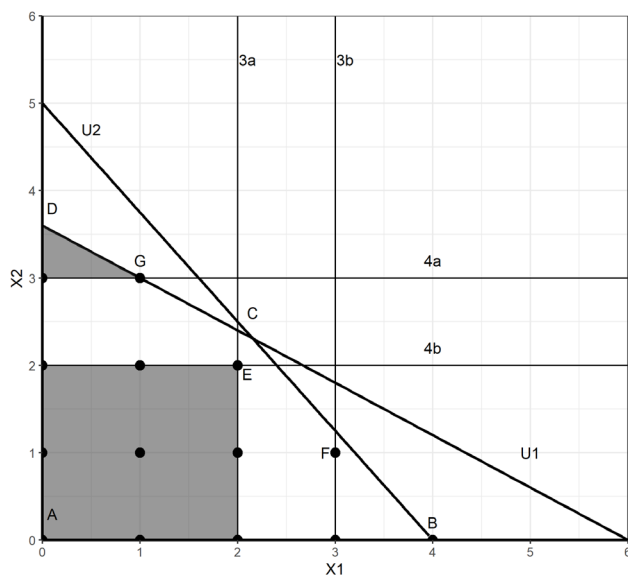
$$U2: 5x_1 + 4x_2 \leq 20$$

$$3b: x_1 \geq 3$$

$$4b: x_2 \geq 3$$

$$x_1 \geq 0, x_2 \geq 0$$

Wykres na rysunku 12 przedstawia efekty drugiego podziału. Zaznaczone zostały tylko te obszary, które odpowiadają zbiorom rozwiązań dopuszczalnych zadań $Z_{4/2}$ i $Z_{5/2}$.



Rys. 12. Zbiory rozwiązań modeli $Z_{4/2}$ i $Z_{5/2}$

Źródło: opracowanie własne.

Tab. 29. Wyniki przykładu 11 dla zadań $Z_{4/2}$ i $Z_{5/2}$

$Z_{4/2}$	wartość	d_j	g_j	$Z_{5/2}$	wartość	d_j	g_j
x_1	2	0	2	x_1	1	0	2
x_2	2	0	2	x_2	3	3	100
f_{\max}	15			f_{\max}	16,5		

Źródło: opracowanie własne.

Aktualna, uporządkowana lista zadań wygląda następująco:

- $Z_{2/1}, f_{\max} = 16,8$: rozwiązanie niecałkowitoliczbowe;
- $Z_{4/2}, f_{\max} = 15$: rozwiązanie całkowitoliczbowe;
- $Z_{5/2}, f_{\max} = 16,5$: rozwiązanie całkowitoliczbowe.

Usuujemy z listy zadanie $Z_{2/1}$ (jako już podzielone). Uporządkowaną listę zadań tworzą $Z_{4/2}$ i $Z_{5/2}$. Podczas sprawdzenia warunku zakończenia postępowania stwierdzamy, że wszystkie zadania aktywne mają rozwiązania w liczbach całkowitych. Nie ma też zadań aktywnych nadających się do dalszego podziału. Rozwiązaniem optymalnym zadania PLC jest rozwiązanie zadania $Z_{5/2}$, ponieważ w jego przypadku funkcja celu jest wyższa niż dla $Z_{4/2}$, czyli:

$$x_1^{opt} = 1, \quad x_2^{opt} = 3, \quad s_1^{opt} = 0, \quad s_2^{opt} = 3, \quad f_{\max} = 16,5$$

5.3. Algorytm Gomory'ego

Algorytm Gomory'ego należy do grupy metod cięć. W każdej jego iteracji dodawane jest kolejne ograniczenie, pozwalające nadać wartość całkowitą tylko jednej zmiennej. Ograniczenie to nazywamy płaszczyzną tnącą. Określenie wzięło się stąd, że – wprowadzając kolejne ograniczenia – dokonujemy odcinania fragmentów zbioru rozwiązań dopuszczalnych. Płaszczyzna odcina ze zbioru dotychczasowe rozwiązanie optymalne. W odróżnieniu od metody podziału i ograniczeń nie tworzymy rozłącznych zbiorów, lecz cały czas operujemy na jednym. Zmieniają się za to współrzędne wierzchołków wieloboku wypukłego.

Algorytm Gomory'ego wymaga, aby wszystkie parametry lewych i prawych stron ograniczeń były liczbami całkowitymi. Jeżeli tak nie jest, należy pomnożyć każde z ograniczeń z osobna przez odpowiednio dobraną dodatnią stałą, a po uzyskaniu rozwiązania optymalnego wartość zmiennej swobodnej dzielimy przez tą stałą.

Podobnie jak to miało miejsce przy metodzie podziału i ograniczeń, na wstępnym etapie rozwiązujemy zadanie programowania liniowego z pominięciem warunku całkowitoliczbowości. Jeżeli zadanie okaże się sprzeczne albo nie posiada skończonego rozwiązania optymalnego, to kończymy postępowanie. Gdy rozwiązanie optymalne dane jest w liczbach całkowitych, to także kończymy postępowanie.

nie. Jeśli zaś rozwiązanie nie spełnia warunku całkowitoliczbowości, to przechodzimy do właściwej procedury.

W każdej iteracji wykonujemy następujące kroki (Rogalska i inni, 1991, s. 177):

1. Wybierz zmienną, której w danej iteracji nadajemy wartość całkowitoliczbową.
2. Zbuduj płaszczyznę tnącą.
3. Dodaj płaszczyznę tnącą do modelu i rozwiąż go, korzystając z dualnej metody *simplex*.
4. Sprawdź, czy w otrzymanym rozwiązaniu optymalnym zmienne decyzyjne przyjęły wartości całkowite. Jeśli tak, zakończ postępowanie. Jeśli nie, wróć do kroku pierwszego.

W kroku pierwszym wśród zmiennych bazowych \mathbf{x}^B , wyrażonych jako ułamki, znajdujemy tę o największej części ułamkowej. Niech taką zmienną będzie zmienna bazowa o numerze l .

$$l: (x_l^B - [x_l^B]) = \max_i \{ (x_i^B - [x_i^B]) \} \quad (104)$$

W przypadku niejednoznacznego wyboru kierujemy się zasadą niższego numeru (niższej pozycji na liście zmiennych bazowych). Nawias $[]$ oznacza część całkowitą liczby.

Płaszczyznę tnącą z kroku drugiego opisuje wzór:

$$\sum_{j=1}^n ([y_{lj}] - y_{lj}) x_j \leq ([x_l^B] - x_l^B) \quad (105)$$

Elementy y_{lj} oraz x_l^B pochodzą z l -tego wiersza tablicy simpleksowej zawierającej rozwiązanie optymalne zadania PL, rozwiązywanego w iteracji poprzedniej. Do tablicy tej dodajemy równanie:

$$\sum_{j=1}^n ([y_{lj}] - y_{lj}) x_j + x_{n+1} = ([x_l^B] - x_l^B) \quad (106)$$

Jako kolejną (ostatnią) zmienną bazową w nowej tablicy przyjmujemy zmienną x_{n+1} . Tablica taka zawsze zawiera rozwiązanie bazowe, które jest dualnie dopuszczalne, ale niedopuszczalne prymalnie – gdyż dodana zmienna bazowa ma ujemną wartość. Wykonujemy iterację dualnej metody *simplex*, która spowoduje, że aktualne rozwiązanie stanie się prymalnie dopuszczalne. W tym momencie istnieje jedna z poniższych sytuacji:

1. Jesteśmy w stanie wykonać iterację dualnej metody *simplex* i otrzymane rozwiązanie jest całkowitoliczbowe. Oznacza to koniec postępowania.

2. Jesteśmy w stanie wykonać iterację dualnej metody *simplex*, ale otrzymane rozwiązanie nie jest całkowitoliczbowe. Wracamy do kroku pierwszego i rozpoczynamy kolejną iterację.
3. Nie jesteśmy w stanie wykonać iteracji dualnej metody *simplex*, ponieważ w wierszu l -tym brak elementów ujemnych. Koniec postępowania – rozwiązane zadanie nie posiada rozwiązania optymalnego w liczbach całkowitych.

Jeżeli w problemie optymalizacyjnym występowało m ograniczeń, to w pierwszej iteracji algorytmu Gomory'ego rozwiązujemy zadania z $m+1$ ograniczeniami i $n+1$ zmiennymi, w drugiej – z $m+2$ i $n+2$ itd. Tym samym – szybko rosną rozmiary samego modelu i tablicy simpleksowej. Jest to niewątpliwą słabością tej metody.

Przykład 12

Rozważmy następujący model optymalizacyjny:

$$f(\mathbf{x}) = 3x_1 + 4x_2 \rightarrow \max \quad (107)$$

$$2x_1 + 2x_2 \leq 11 \quad (108)$$

$$x_2 \leq 4,5 \quad (109)$$

$$x_1 \geq 0, x_2 \geq 0, x_1, x_2 - \text{całkowite} \quad (110)$$

Zauważmy, że prawa strona ograniczenia (109) nie jest liczbą całkowitą. Dlatego mnożymy całe ograniczenie przez stałą $\lambda = 2$. Otrzymujemy zmodyfikowany model (111)–(114), który rozwiązujemy klasycznym algorytmem *simplex*, pomijając warunek całkowitoliczbowy. W tabeli 30 znajduje się finalna tablica simpleksowa tego etapu postępowania.

$$f(\mathbf{x}) = 3x_1 + 4x_2 \rightarrow \max \quad (111)$$

$$2x_1 + 2x_2 \leq 11 \quad (112)$$

$$x_2 \leq 4,5 \quad (113)$$

$$x_1 \geq 0, x_2 \geq 0, x_1, x_2 - \text{całkowite} \quad (114)$$

Rozwiązanie nie jest całkowitoliczbowe – przechodzimy do kroku pierwszego algorytmu. Ułamekowa wartość wystąpiła dla zmiennej x_2 . Dlatego to właśnie ona wskazuje na pozycję w bazie o największej części ułamkowej, czyli wiersz $l = 2$. Równanie cięcia będzie miało postać:

$$([0]-0)x_1 + ([1]-1)x_2 + ([0]-0)s_1 + \left(\left[\frac{1}{2}\right] - \frac{1}{2}\right)s_2 + s_3 = \left(\left[4\frac{1}{2}\right] - 4\frac{1}{2}\right)$$

$$0x_1 + 0x_2 + 0s_1 - \frac{1}{2}s_2 + s_3 = -\frac{1}{2}$$

Tab. 30. Rozwiązanie optymalne modelu (111)–(114)

\mathbf{c}^B	Baza	$[c_j]$	3	4	0	0	Wart. zm. baz.
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	
			x_1	x_2	s_1	s_2	\mathbf{x}^B
3	\mathbf{p}_1	x_1	1	0	1/2	-1/2	1
4	\mathbf{p}_2	x_2	0	1	0	1/2	9/2
$\Delta_j = c_j - z_j$			0	0	-3/2	-1/2	21

Źródło: opracowanie własne.

Rozszerzamy tablicę simpleksową o równanie cięcia i zmienną s_3 , która zajmie w bazie ostatnie miejsce. Tablicę konstruujemy zgodnie z regułami dualnej metody *simplex*.

Tab. 31. Dualna metoda *simplex* dla pierwszej iteracji przykładu 12

\mathbf{c}^B	Baza	$[c_j]$	3	4	0	0	0	Wart. zm. baz.
		Zm. baz.	\mathbf{p}_1	\mathbf{p}_2	\mathbf{p}_3	\mathbf{p}_4	\mathbf{p}_5	
			x_1	x_2	s_1	s_2	s_3	\mathbf{x}^B
3	\mathbf{p}_1	x_1	1	0	1/2	-1/2	0	1
4	\mathbf{p}_2	x_2	0	1	0	1/2	0	9/2
0	\mathbf{p}_5	s_3	0	0	0	-1/2	1	-1/2
$\Delta_j = c_j - z_j$			0	0	-3/2	-1/2	0	21
$ (c_j - z_j)/y_{ij} $						1		
3	\mathbf{p}_1	x_1	1	0	1/2	0	-1	1/2
4	\mathbf{p}_2	x_2	0	1	0	0	1	4
0	\mathbf{p}_4	s_2	0	0	0	1	-2	1
$\Delta_j = c_j - z_j$			0	0	-3/2	0	-1	17,5

Źródło: opracowanie własne.

Zmienna s_3 w pierwszej części tabeli przyjmuje wartość ujemną, zaś wskaźniki optymalności informują o otrzymaniu rozwiązania optymalnego. Jest ono niedopuszczalne prymalnie, ale dopuszczalne dualnie. Zgodnie z zasadami dualnej metody *simplex* wybieramy zmienną bazową o najbardziej ujemnej wartości. To ona opuści bazę, a na jej miejsce wejdzie zmienna wskazana przez kryterium wejścia – będzie nią zmienna s_2 . Dokonujemy wymiany wektorów w bazie. Nowe rozwiązanie (zob. druga część tabeli 31) jest prymalnie dopuszczalne i optymalne, ale nie jest całkowitoliczbowe. Przechodzimy do iteracji drugiej.

Tab. 32. Dualna metoda *simplex* dla drugiej iteracji przykładu 12

c ^B	Baza	[c _j]	3	4	0	0	0	0	Wart. zm. baz.
		Zm. baz.	p ₁	p ₂	p ₃	p ₄	p ₅	p ₆	
			x ₁	x ₂	s ₁	s ₂	s ₃	s ₄	x ^B
3	p ₁	x ₁	1	0	1/2	0	-1	0	1/2
4	p ₂	x ₂	0	1	0	0	1	0	4
0	p ₄	s ₂	0	0	0	1	-2	0	1
0	p ₆	s ₄	0	0	-1/2	0	0	1	-1/2
Δ _j = c _j - z _j			0	0	-3/2	0	-1	0	17,5
(c _j -z _j)/y _{ij}					3				
3	p ₁	x ₁	1	0	0	0	-1	1	0
4	p ₂	x ₂	0	1	0	0	1	0	4
0	p ₄	s ₂	0	0	0	1	-2	0	1
0	p ₃	s ₁	0	0	1	0	0	-2	1
Δ _j = c _j - z _j			0	0	0	0	-1	-3	16

Źródło: opracowanie własne.

Pozycją w bazie o największej części ułamkowej, przy wartości zmiennej bazowej, jest teraz $l = 1$. Tworzymy na podstawie tego wiersza równanie cięcia:

$$([1]-1)x_1 + ([0]-0)x_2 + \left(\left[\frac{1}{2}\right] - \frac{1}{2}\right)s_1 + ([0]-0)s_2 + ([-1]+1)s_3 + s_4 = \left(\left[\frac{1}{2}\right] - \frac{1}{2}\right)$$
$$0x_1 + 0x_2 - \frac{1}{2}s_1 + 0s_2 + 0s_3 + s_4 = -\frac{1}{2}$$

Tablica simpleksowa z tabeli 32 zostanie rozszerzona o jeden wiersz i jedną kolumnę.

Rozwiązanie z tabeli 32 jest całkowitoliczbowe, a zatem zakończyliśmy postępowanie. Otrzymane rozwiązanie optymalne jest następujące:

$$x_1^{opt} = 0, x_2^{opt} = 4, s_1^{opt} = 1, x_1^{opt} = \frac{1}{\lambda} = \frac{1}{2}, f_{\max} = 16$$

W jednym z artykułów (Mei Lee, Adi, Mohd Rizal i Effendi, 2018) porównano efektywność różnych algorytmów PL, PLC oraz metody heurystycznej podczas rozwiązywania problemów optymalizacyjnych różniących się rozmiarami. Konkluzja płynąca z badań (zresztą spodziewana) była taka, że klasyczne algorytmy

przestają się sprawdzać przy dużych zadaniach. Tu szybsze są metody heurystyczne, które jednak nie dają dokładnych rozwiązań, lecz przybliżone. Z tego powodu prowadzone są analizy mające na celu poprawę wydajności metod dokładnych. Niektórzy autorzy (zob. He i Lee, 2017) podjęli się próby zmodyfikowania algorytmu Gomory'ego poprzez wykorzystanie prymalnej metody *simplex*. W innej zaś pracy (Tiwari i Kumar, 2018) dokonano przeglądu literatury na temat programowania liniowego oraz celowego – pod kątem porównania przydatności tych metod przy wspomaganiu procesu podejmowania decyzji.

Rozdział II

Problemy przydziału

1. Natura problemów przydziału

Problem przydziału (*assignment*) dotyczy dużej ilości sytuacji decyzyjnych, które pojawiają się w przedsiębiorstwach i innych organizacjach. Dotyczyć mogą przyporządkowania zadań do pracowników czy też maszyn, typów pojazdów do wykonania konkretnych zadań przewozowych itd. (Krawczyk, 2001)¹. Cechą charakterystyczną problemu przydziału jest to, że każde zadanie musi zostać przydzielone jednemu i tylko jednemu wykonawcy (Jędrzejczyk, Kukuła, Skrzypek i Walkosz, 2016, s. 156 i nn.; por. Kalichman, 1971)². Jako kryterium optymalności rozwiązania można wybrać, np.:

1. Minimalizację kosztów związanych z przydziałem czy łącznego czasu wykonania zadań.
2. Maksymalizację zysku związanego z wykonaniem zadań, sprzedaży, udziału w rynku czy efektywności.

Problem przydziału jest zadaniem kombinatorycznym, ale może być przedstawiony jako specyficzne zadanie programowania liniowego – zadanie liniowego przydziału LAP. Rozważmy sytuację polegającą na przydziale m pracowników do n projektów. Koszt związany z wykonaniem projektu j -tego przez i -tego pracownika wynosi c_{ij} . Zanim przystąpi się do rozwiązywania problemu, należy go zbilansować. Sposób bilansowania zależy od relacji między liczbą pracowników a liczbą projektów.

1. Więcej projektów do wykonania niż pracowników, $m < n$. Oznacza to, że nie wszystkie projekty będą wykonane. W takiej sytuacji przekształcamy macierz kosztów, dodając dodatkowe wiersze (oznaczające fikcyjnych pracowników)

1 Stanisław Krawczyk omawia problem przydziału (nazywając go zadaniem przyporządkowania) na przykładzie rozdziału zadań przewozowych.

2 Przez problem przydziału w przywołanych pracach rozumiany jest ogólny problem alokacji zasobów przy kryterium minimalizacji nakładów bądź maksymalizacji efektów. Ograniczenia dotyczą zapewnienia popytu odbiorców i nieprzekroczenia możliwości dostawców. Tak rozumiany problem przydziału można modelować i rozwiązać jak klasyczne zadanie programowania liniowego.

w ilości $(n - m)$. Koszty odpowiadające dodatkowym wierszom są równe 0, chyba że zależy nam, aby zagwarantować wykonanie j -tego projektu – wówczas blokujemy przydział fikcyjnych pracowników do tego projektu, wstawiając relatywnie wysoki koszt c_{ij} , gdzie indeksy i odpowiadają fikcyjnym pracownikom.

2. Więcej pracowników niż projektów, $m > n$. Oznacza to, że nie wszyscy pracownicy będą mieli przydzielone zadania. W takiej sytuacji przekształcamy macierz danych, dodając dodatkowe kolumny (oznaczające fikcyjne zadania do wykonania) w ilości $(m - n)$. Koszty odpowiadające dodatkowym kolumnom są równe 0, chyba że zależy nam, aby zagwarantować przydział projektu wybranemu i -temu pracownikowi – wówczas blokujemy przydział tego pracownika do fikcyjnych projektów, wstawiając odpowiednio relatywnie wysoki koszt c_{ij} , gdzie indeksy j odpowiadają fikcyjnym projektom.

Przyjmijmy, że po zbilansowaniu otrzymujemy problem przydziału n^* pracowników do wykonania n^* projektów, gdzie $n^* = \max(m, n)$. Zmienne decyzyjne problemu, oznaczające przydział i -tego pracownika do wykonania j -tego projektu, oznaczmy przez x_{ij} . Zmienne decyzyjne zadania mają charakter zero-jedynkowy:

- $x_{ij} = 0$ – oznacza, że pracownik i -ty nie jest przydzielony do wykonania projektu j -tego,
- $x_{ij} = 1$ – oznacza, że pracownikowi i -temu przydzielono wykonanie projektu j -tego.

Niech $\mathbf{X} = [x_{ij}]$ oznacza macierz $n^* \times n^*$ zmiennych decyzyjnych zagadnienia przydziału. Model matematyczny liniowego problemu optymalnego przydziału LAP, można przedstawić następująco:

$$\min z(\mathbf{X}) = \sum_{i=1}^{n^*} \sum_{j=1}^{n^*} c_{ij} x_{ij} \quad (115)$$

$$\sum_{j=1}^{n^*} x_{ij} = 1, \quad i = 1, \dots, n^* \quad \text{równania typu podaźowego} \quad (116)$$

$$\sum_{i=1}^{n^*} x_{ij} = 1, \quad j = 1, \dots, n^* \quad \text{równania typu popytowego} \quad (117)$$

$$x_{ij} \in \{0, 1\} \quad i, j = 1, \dots, n^* \quad (118)$$

Liczba rozwiązań dopuszczalnych zadania (115)–(118) jest znana, skończona i wynosi dokładnie $n^!$

Macierz parametrów układu ograniczeń (116)–(117) jest macierzą unimodularną³. Macierz odwrotna do macierzy unimodularnej istnieje i posiada własność, że jeżeli elementy wektora prawych stron układu równań stanowiącego ograni-

³ Macierz unimodularna to macierz kwadratowa o elementach całkowitych, posiadająca wyznacznik równy +1 albo -1.

czenia zadania są całkowite – a w tym przypadku zostało to oczywiście spełnione, bo prawe strony równań (116)–(117) są równe 1 – to rozwiązanie układu jest również całkowitoliczbowe (czyli zmienne decyzyjne przyjmą wartości 0 albo 1). Warunek (118) binarności zmiennych decyzyjnych możemy zatem zastąpić warunkiem nieujemności – otrzymujemy wtedy postać modelu, która jest zadaniem programowania liniowego z rzeczywistymi wartościami zmiennych:

$$\min z(\mathbf{X}) = \sum_{i=1}^{n^*} \sum_{j=1}^{n^*} c_{ij} x_{ij} \quad (115')$$

$$\sum_{j=1}^{n^*} x_{ij} = 1, \quad i = 1, \dots, n^* \quad \text{równania typu podażowego} \quad (116')$$

$$\sum_{i=1}^{n^*} x_{ij} = 1, \quad j = 1, \dots, n^* \quad \text{równania typu popytowego} \quad (117')$$

$$x_{ij} \geq 0, \quad i, j = 1, \dots, n^* \quad (118')$$

Parametry c_{ij} dla $i, j = 1, \dots, n^*$ funkcji kryterium (115') tworzą macierz \mathbf{C} o wymiarach $n^* \times n^*$ i wyrażają koszty/straty związane z – odpowiadającym indeksom – przydziałem (pracownika i do projektu j). Zadanie (115')–(118') może również przedstawiać problem minimalizacji łącznego czasu wykonania projektów – wówczas współczynnik c_{ij} oznaczać będzie czas wykonania j -tego projektu przez i -tego pracownika.

2. Algorytm węgierski

2.1. Schemat postępowania w algorytmie węgierskim

Autorami algorytmu rozwiązującego problem optymalnego przydziału są węgierscy matematycy: Dénes König (König, 1931) i Jenő Egerváry (Egerváry, 1931). Oparty jest on na twierdzeniu Königa–Egerváry'ego, znanemu w teorii grafów (zob. Ford i Fulkerson, 1969, s. 76–82). Metoda została przybliżona szerszemu gronu odbiorców po ukazaniu się publikacji Harolda W. Kuhna w języku angielskim (Kuhn, 1955). Algorytm ten, znany jako algorytm węgierski, opracowany został dla problemu minimalizacji funkcji kryterium (koszt, czas). W przypadku zadania z kryterium maksymalizacji poleca się przekształcić macierz danych wejściowych (zysków), odejmując wszystkie jej elementy od elementu o wartości maksymalnej. Wówczas zadanie maksymalizacji zysku przekształca się do zadania minimalizacji relatywnych strat.

Rozwiązaniem zadania przydziału (115)–(118) jest macierz przydziału $X = [x_{ij}]$, $i, j = 1, \dots, n^*$, w której dokładnie n^* elementów jest równych jedności, a pozostałe są równe 0.

Algorytm węgierski można interpretować, wykorzystując koncepcję kosztów utraconych możliwości. Koszt utraconych możliwości (*opportunity cost*, *penalty cost*) związany jest z relatywną stratą związaną z rzeczywiście podjętą decyzją względem najlepszej możliwej do podjęcia decyzji – celem metody jest zminimalizowanie strat związanych z wyborem nie najtańszego sposobu przydziału. Kroki metody są następujące.

Krok pierwszy. Konstrukcja macierzy kosztów utraconych możliwości

Należy przekształcić macierz kosztów $C = [c_{ij}]$ $i, j = 1, \dots, n^*$ do postaci macierzy kosztów utraconych możliwości w taki sposób, aby w każdym jej wierszu i w każdej kolumnie występowało przynajmniej jedno 0. W tym celu od każdego wiersza macierzy C odejmuje się jego najmniejszy element i następnie, jeżeli trzeba, od każdej kolumny odejmuje się jej najmniejszy element.

Krok drugi. Analiza dopuszczalności

Należy skreślić w przekształconej macierzy kosztów wiersze i kolumny zawierające 0 możliwie najmniejszą liczbą linii (poziomych i pionowych). Jeżeli najmniejsza liczba linii niezbędnych do pokrycia wszystkich 0 jest równa wymiarowi macierzy, czyli n^* , to można przejść do kroku czwartego i wyznaczyć rozwiązanie optymalne.

W przeciwnym przypadku należy przejść do kroku trzeciego.

Krok trzeci. Konstrukcja zrewidowanej macierzy kosztów utraconych możliwości

Jeżeli liczba skreśleń jest mniejsza od rozmiaru macierzy n^* , w bieżącej macierzy kosztów należy znaleźć najmniejszy nieskreślony element i ten element:

- odjąć od elementów nieskreślonych,
- dodać do elementów podwójnie skreślonych.

Elementy skreślone jedną linią zostawiamy bez zmian. Należy wrócić do kroku drugiego.

Krok czwarty. Ustalenie rozwiązania optymalnego

Ustalenie optymalnego przydziału polega na takiej konstrukcji macierzy $X = [x_{ij}]$, aby jedynki znalazły się tylko na tych miejscach, na których w przekształconej macierzy kosztów występują zera (pamiętając, że w każdym wierszu i każdej kolumnie macierzy przydziału może występować tylko jedna 1). Przydział rozpoczynamy od wierszy lub kolumn, w których – w przekształconej macierzy kosztów – występuje tylko jedno 0.

Zaprezentujemy trzy przykłady zastosowania algorytmu węgierskiego:

- w przykładzie 1 przedstawiony jest problem, w którym rozwiązanie jest niejednoznaczne;

- w przykładzie 2 jako kryterium przyjęto maksymalizację oczekiwanej sprzedaży;
- w przykładzie 3 mamy do czynienia z problemem niebilansowanym – $m < n$ (liczba pracowników mniejsza od liczby projektów) oraz zastosowaniem preferencji.

Przykład 1

Firma taksówkarska posiada taksówki oczekujące na klientów na czterech postojach. Czterech klientów zgłosiło chęć przejazdu. W tabeli przedstawione są odległości w kilometrach od oczekujących na postojach taksówek do klientów. Wyznaczyć optymalny przydział taksówek do klientów, aby zminimalizować dystans, jaki taksówki muszą pokonać, aby dotrzeć do klientów.

Rozwiązanie:

Tab. 33. Zestawienie odległości pomiędzy klientami i postojami taksówek w kilometrach w przykładzie 1

Miejsce postoju	Klient				min.
	1	2	3	4	
A	16	8	6	14	6
B	10	9	8	13	8
C	12	18	14	9	9
D	8	7	12	10	7

Źródło: opracowanie własne.

W tabeli 33 w każdym wierszu znajdujemy wartość minimalną, odpowiadającą przydziałowi taksówki do klienta będącego najbliżej. Konstruujemy skorygowaną macierz odległości, odejmując od elementów każdego wiersza odpowiednie minimum. Otrzymujemy macierz przedstawioną w tabeli 34.

Tab. 34. Skorygowana macierz odległości w przykładzie 1

Miejsce postoju	Klient			
	1	2	3	4
A	10	2	0	8
B	2	1	0	5
C	3	9	5	0
D	1	0	5	3
min.	1	0	0	0

Źródło: opracowanie własne.

W macierzy w tabeli 34 w każdym wierszu występuje co najmniej jedno 0. Sprawdzamy występowanie 0 w kolumnach tej macierzy – z wyjątkiem kolumny pierwszej – w każdej kolumnie występuje 0. Oznacza to, że korektę musimy kontynuować – od elementów kolumny pierwszej odejmujemy wartość minimalną, wynoszącą 1. Efekt obu korekt zawiera tabela 35.

Przystępujemy do kroku drugiego – analizy dopuszczalności rozwiązania sugerowanego przez pozycje zerowe w zrewidowanej macierzy kosztów – skreślamy 0 liniami pionowymi i poziomymi, w sposób wykorzystujący najmniejszą liczbę skreśleń. W tabeli 35 zaznaczono trzy takie skreślenia – ponieważ $n^* = 4$, oznacza to, że postępowanie musimy kontynuować według procedury opisanej w kroku trzecim.

Tab. 35. Macierz kosztów utraconych korzyści i linie skreślające 0 w przykładzie 1

Miejsce postoj	Klient			
	1	2	3	4
A	9	2	0	8
B	1	1	0	5
C	2	9	5	0
D	0	0	5	3

Źródło: opracowanie własne.

Minimalny nieskreślony element w tabeli 35 ma wartość 1. Odejmujemy tę wartość od elementów nieskreślonych i dodajemy do elementów podwójnie skreślonych. Efekt postępowania znajduje się w tabeli 36.

Tab. 36. Zrewidowana macierz kosztów i linie skreślające 0 w przykładzie 1

Miejsce postoj	Klient			
	1	2	3	4
A	8	1	0	7
B	0	0	0	4
C	2	9	6	0
D	0	0	6	3

Źródło: opracowanie własne.

W tabeli 36 nie można skreślić 0 za pomocą mniejszej liczby linii niż cztery. Oznacza to, że możemy przystąpić do konstrukcji macierzy przydziału. Rozpoczynamy od przydziału w wierszu pierwszym i w kolumnie czwartej – tam wystę-

pują w zrewidowanej macierzy kosztów pojedyncze 0. Przydział taksówki z postoju A do trzeciego klienta oraz taksówki z postoju C do czwartego klienta jest oczywisty i jednoznaczny. Taksówkę z postoju B można przydzielić do pierwszego klienta i – w konsekwencji – taksówkę z postoju D do klienta drugiego. Takie rozwiązanie wskazuje macierz przydziału w tabeli 37. Optymalny przydział jest niejednoznaczny – warunki ograniczające zadania są również spełnione, gdy taksówkę z postoju B wyślemy do klienta drugiego, a taksówkę z postoju D do klienta pierwszego. Takie rozwiązanie zaproponowane zostało w tabeli 38.

Tab. 37. Macierz przydziału X_1 – wariant pierwszy

Miejsce postoju	Klient			
	1	2	3	4
A	0	0	1	0
B	1	0	0	0
C	0	0	0	1
D	0	1	0	0

Źródło: opracowanie własne.

Tab. 38. Macierz przydziału X_2 – wariant drugi

Miejsce postoju	Klient			
	1	2	3	4
A	0	0	1	0
B	0	1	0	0
C	0	0	0	1
D	1	0	0	0

Źródło: opracowanie własne.

Długość trasy, jaką muszą przejechać taksówki w wariantcie pierwszym, wynosi $6 + 10 + 9 + 7 = 32$ [km]. W wariantcie drugim długość odpowiedniej trasy to $6 + 9 + 9 + 8 = 32$ [km]. Oba rozwiązania są równoważne.

Przykład 2

Firma ubezpieczeniowa zatrudnia pięciu agentów ubezpieczeniowych i chce dokonać ich przydziału do pięciu regionów sprzedaży. Biorąc pod uwagę poprzednie kontakty agentów w rejonach sprzedaży, oszacowano ich efektywność mierzoną za pomocą oczekiwanej wielkości sprzedaży w tysiącach złotych. Wyniki szacunków przedstawione są w tabeli 39.

Tab. 39. Oczekiwana sprzedaż agentów w regionach sprzedaży

Agent	Region sprzedaży				
	A	B	C	D	E
1	30	24	28	32	36
2	28	20	30	20	28
3	32	26	20	38	32
4	36	34	24	34	28
5	24	32	20	28	26

Źródło: opracowanie własne.

Należy dokonać optymalnego przydziału agentów do rejonów sprzedaży, kierując się kryterium maksymalizacji wielkości sprzedaży.

Rozwiązanie:

Ze względu na przyjęte kryterium optymalizacji – maksymalizacja wielkości sprzedaży – należy wstępnie dokonać transformacji macierzy wartości kryterium, aby przekształcić problem do poszukiwania rozwiązania zadania minimalizacji. W tabeli danych o oczekiwanej sprzedaży znajdujemy wartość maksymalną – wynosi ona trzydzieści osiem tysięcy złotych. Wartości poszczególnych elementów macierzy odejmujemy od wartości maksymalnej. Uzyskujemy macierz kosztów przedstawioną w tabeli 40, której elementy możemy interpretować jako relatywne „straty” w porównaniu z wartością maksymalnej, potencjalnie możliwej do osiągnięcia sprzedaży w pojedynczym regionie.

Tab. 40. Macierz kosztów z przykładu 2

Agent	Region sprzedaży					min.
	A	B	C	D	E	
1	8	14	10	6	2	2
2	10	18	8	18	10	8
3	6	12	18	0	6	0
4	2	4	14	4	10	2
5	14	6	18	10	12	6

Źródło: opracowanie własne.

Od elementów każdego wiersza macierzy kosztów odejmujemy odpowiednio wartość minimalną w wierszu. Wynik przedstawia tabela 41.

Tab. 41. Skorygowana macierz kosztów z przykładu 2

Agent	Region sprzedaży				
	A	B	C	D	E
1	6	12	8	4	0
2	2	10	0	10	2
3	6	12	18	0	6
4	0	2	12	2	8
5	8	0	12	4	6

Źródło: opracowanie własne.

W każdym wierszu i w każdej kolumnie macierzy z tabeli 41 występuje wartość 0. Możemy zweryfikować dopuszczalność projektowanego rozwiązania za pomocą skreśleń – okazuje się, że do wykreślenia wszystkich 0 niezbędnych jest dokonanie pięciu skreśleń, co odpowiada wymiarowi zadania przydziału. W macierzy jest dokładnie pięć 0 – rozwiązanie jest jednoznaczne i konstrukcja macierzy optymalnego przydziału okazuje się oczywista. Wynik przedstawia tabela 42.

Tab. 42. Macierz optymalnego przydziału z przykładu 2

Agent	Region sprzedaży				
	A	B	C	D	E
1	0	0	0	0	1
2	0	0	1	0	0
3	0	0	0	1	0
4	1	0	0	0	0
5	0	1	0	0	0

Źródło: opracowanie własne.

Szacowana wartość łącznej sprzedaży produktów ubezpieczeniowych związana z optymalnym przydziałem wynosi sto siedemdziesiąt dwa tysiące złotych miesięcznie.

Przykład 3

Zarząd firmy inwestycyjnej musi przydzielić cztery zespoły konsultantów do nadzorowania pięciu projektów inwestycyjnych o podobnym horyzoncie czasowym. Biorąc pod uwagę doświadczenie i efektywność pracy konsultantów oraz złożoność projektów, oszacowano możliwe zyski dla firmy wynikające z dokonania przydziału zadań. Oczekiwane zyski roczne w tysiącach złotych przedsta-

wione zostały w tabeli 43. Strategia firmy wymaga, aby nadzór nad projektem A był prowadzony przez konsultantów przedsiębiorstwa; rozważa się przekazanie jednego z pozostałych projektów firmie zewnętrznej. Kryterium optymalizacji to maksymalizacja oczekiwanego zysku. Należy dodatkowo ocenić, ile utraconych korzyści związanych jest ze strategią polegającą na utrzymaniu nadzoru nad projektem A.

Tab. 43. Oczekiwane zyski roczne z nadzoru projektów przez konsultantów w tys. PLN – przykład 3

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	100	150	120	200	190
2	150	200	170	190	220
3	120	180	190	200	180
4	170	220	200	230	210

Źródło: opracowanie własne.

Rozwiązanie:

Problem rozwiązujemy dwukrotnie – w wersji a) bez zastosowania preferencji dla projektu A oraz w wersji b) z zablokowaniem możliwości przekazania nadzoru nad projektem A firmie zewnętrznej. Na wstępie przekształcamy macierz oczekiwanych zysków, odejmując jej elementy od wartości maksymalnej równej dwustu trzydziestu tysiącom złotych – aby zastąpić kryterium maksymalizacji oczekiwanego zysku przez kryterium minimalizacji utraconych korzyści. Dokonujemy również bilansowania problemu – poprzez dodanie fikcyjnego wykonawcy; w macierzy kosztów w ostatnim piątym wierszu wstawiamy w wersji a) elementy równe 0. Wynik przedstawiony jest w tabeli 44.

Tab. 44. Utracone korzyści w tys. PLN – przykład 3a

Zespół konsultantów	Projekt					min.
	A	B	C	D	E	
1	130	80	110	30	40	30
2	80	30	60	40	10	10
3	110	50	40	30	50	30
4	60	10	30	0	20	0
5 – fikcyjny	0	0	0	0	0	0

Źródło: opracowanie własne.

Postępując zgodnie z krokami algorytmu węgierskiego, od elementów wierszy 1–3 odejmujemy odpowiednią wartość minimalną. W skorygowanej macierzy skreślamy wszystkie 0 – minimalna liczba linii wynosi trzy. Należy zatem dokonać korekty zgodnie z krokiem trzeciego algorytmu. Minimalny nieskreślony element jest równy dziesięciu – odejmujemy go od elementów nieskreślonych i dodajemy do elementów podwójnie skreślonych. W skorygowanej ponownie macierzy dokonujemy skreśleń 0 – tym razem liczba skreśleń wynosi pięć. Oznacza to, że możemy skonstruować macierz optymalnego przydziału. Wyniki obliczeń przedstawiają tabele 45–47.

Tab. 45. Macierz kosztów z przykładu 3a

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	100	50	80	0	10
2	70	20	50	30	0
3	80	20	10	0	20
4	60	10	30	0	20
5 – fikcyjny	0	0	0	0	0

Źródło: opracowanie własne.

Tab. 46. Skorygowana macierz kosztów z przykładu 3a

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	90	40	70	0	10
2	60	10	40	30	0
3	70	10	0	0	20
4	50	0	20	0	20
5 – fikcyjny	0	0	0	10	10

Źródło: opracowanie własne.

Tab. 47. Macierz optymalnego przydziału z przykładu 3a

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	0	0	0	1	0
2	0	0	0	0	1
3	0	0	1	0	0
4	0	1	0	0	0
5 – fikcyjny	1	0	0	0	0

Źródło: opracowanie własne.

Optymalny przydział zespołów do projektów wymaga przekazania projektu A firmie zewnętrznej. Oczekiwane zyski, wynikające z optymalnego przydziału zespołów do pozostałych projektów, szacowane są wówczas na osiemset trzydzieści tysięcy złotych.

W wersji b) rozwiązania blokujemy przydział fikcyjnego zespołu do projektu A poprzez wprowadzenie do ostatniego wiersza w macierzy z tabeli 44 relatywnie wysokiej wartości kosztu odpowiadającego projektowi A – otrzymujemy macierz kosztów przedstawioną w tabeli 48. Wyniki obliczeń przeprowadzonych za pomocą algorytmu węgierskiego prezentowane są w tabelach 49–51. Minimalna liczba linii wymagana dla skreślenia 0 w macierzy z tabeli 49 wynosi cztery. Minimalny nieskreślony element ma wartość dziesięć. Należy dokonać dalszej korekty według kroku trzeciego algorytmu. Skorygowana macierz kosztów ma postać jak w tabeli 50 – aby skreślić wszystkie 0, wymaganych jest minimum pięć linii – można przystąpić do konstrukcji macierzy optymalnego przydziału. Przedstawiona jest ona w tabeli 51.

Tab. 48. Utracone korzyści w tys. PLN – przykład 3b

Zespół konsultantów	Projekt					min.
	A	B	C	D	E	
1	130	80	110	30	40	30
2	80	30	60	40	10	10
3	110	50	40	30	50	30
4	60	10	30	0	20	0
5 – fikcyjny	M	0	0	0	0	0

Źródło: opracowanie własne.

Tab. 49. Macierz kosztów z przykładu 3b

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	40	50	80	0	10
2	10	20	50	30	0
3	20	20	10	0	20
4	0	10	30	0	20
5 – fikcyjny	M=60	0	0	0	0

Źródło: opracowanie własne.

Tab. 50. Skorygowana macierz kosztów z przykładu 3b

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	30	40	70	0	10
2	0	10	40	30	0
3	10	10	0	0	20
4	0	10	30	10	30
5 – fikcyjny	10	0	0	10	10

Źródło: opracowanie własne.

Tab. 51. Macierz optymalnego przydziału z przykładu 3b

Zespół konsultantów	Projekt				
	A	B	C	D	E
1	0	0	0	1	0
2	0	0	0	0	1
3	0	0	1	0	0
4	1	0	0	0	0
5 – fikcyjny	0	1	0	0	0

Źródło: opracowanie własne.

W wariancie b) firmie zewnętrznej przekazany będzie nadzór nad projektem B. Ocena oczekiwanego zysku rocznego wynosi w tym wariancie siedemset osiemdziesiąt tysięcy złotych. Porównując szacunkowy zysk, w przypadku przydziału zadań z blokadą przekazania nadzoru nad projektem A firmie zewnętrznej oraz bez zastosowania wspomnianej strategii firmy, możemy ocenić, że przyjęcie strategii oznacza relatywną stratę w wysokości pięćdziesięciu tysięcy złotych rocznie.

2.2. Algorytm węgierski jako wariant metody *simplex*

Algorytm węgierski może być wyjaśniony w kategoriach metody *simplex* rozwiązywania zadań programowania liniowego. Model dualny do modelu optymalnego przydziału (115')–(118') da się zapisać następująco:

$$\max w = \sum_{i=1}^{n^*} u_i + \sum_{j=1}^{n^*} v_j \quad (115'')$$

$$u_i + v_j \leq c_{ij}, \quad i, j = 1, \dots, n^* \quad (116'')$$

$$u_i, v_j, i, j = 1, \dots, n^* \quad \text{liczby rzeczywiste, bez ograniczeń co do znaku} \quad (117'')$$

Komplementarne warunki swobody, wynikające z twierdzenia o komplementarności Dantziga–Ordena (Dantzig i Orden, 1953), w przypadku dualnych wzajemnie modeli (115')–(118') oraz (115'')–(117'') można sformułować następująco: warunki konieczne i wystarczające na to, aby wartości zmiennych zadań dualnych (115')–(118') oraz (115'')–(117''), tj. $u_i, v_j, x_{ij}, i, j = 1, \dots, n^*$, były optymalne i określone jako:

$$(u_i + v_j - c_{ij})x_{ij} = 0, \quad i, j = 1, \dots, n^* \quad (118'')$$

Rozwiązanie optymalne problemu przydziału pozostaje niezmiennione w przypadku dodania lub odjęcia stałej od dowolnego wiersza lub kolumny macierzy kosztów; transformacja taka wywiera wpływ jedynie na wartość funkcji kryterium, odpowiednio ją zwiększając lub zmniejszając. Można zatem poszukiwać rozwiązania optymalnego, bazując na skorygowanej macierzy C' (Taha, 1992, s. 216). Łatwo to wykazać, wykorzystując ograniczenia zadania przydziału (116) i (117). Korekta macierzy C , poprzez odjęcie od każdego wiersza stałej p_i i od każdej kolumny stałej q_j , prowadzi jedynie do modyfikacji wartości funkcji kryterium o stałą:

$$\begin{aligned} z' &= \sum_i \sum_j (c_{ij} - p_i - q_j)x_{ij} = \sum_i \sum_j c_{ij}x_{ij} - \sum_i p_i \sum_j x_{ij} - \sum_j q_j \sum_i x_{ij} = \\ &= z - \sum_i p_i - \sum_j q_j \end{aligned}$$

Dopuszczalne rozwiązanie zadania dualnego można uzyskać, poszukując wartości:

$$p_i = \min_j \{c_{ij}\} \quad i, j = 1, \dots, n^*$$

oraz:

$$q_j = \min_i \{c_{ij} - p_i\} \quad i, j = 1, \dots, n^*$$

Jeżeli uda się dokonać korekty macierzy kosztów w taki sposób, aby elementy równe 0 lub ich podzbiór odpowiadały rozwiązaniu dopuszczalnemu, to uzyskane rozwiązanie dopuszczalne będzie jednocześnie optymalne. Gdy okaże się, że zachodzi $p_i + q_j - c_{ij} = 0$, to p_i oraz q_j stanowią wartości zmiennych dualnych, a odpowiadająca zmienna x_{ij} może przyjąć wartość 1. Jeśli proces wyznaczania wartości p_i oraz q_j nie doprowadzi do rozwiązania dopuszczalnego problemu przydziału, należy poszukać nowego układu wartości zmiennych dualnych, spełniających warunki komplementarności twierdzenia Dantziga–Ordena. Procedura proponowana przez algorytm węgierski prowadzi właśnie do takiego rozwiązania.

3. Rozszerzenia problemu przydziału

3.1. Uogólniony problem przydziału GAP

GAP jest zadaniem całkowitoliczbowego, zero-jedynkowego problemu programowania liniowego (0-1 ILP). Niech n oznacza liczbę zadań, które należy przydzielić m wykonawcom, $n \geq m$. Parametry modelu możemy zdefiniować jako:

- c_{ij} – koszt wykonania zadania j przez wykonawcę i ,
- r_{ij} – wielkość zasobu wymagana dla wykonania zadania j przez wykonawcę i ,
- b_i – ilość jednostek zasobu dostępna wykonawcy i .

Zmienne decyzyjne zadania GAP mają charakter zero-jedynkowy, podobnie jak w klasycznym zagadnieniu przydziału (115)–(118):

$$x_{ij} = \begin{cases} 1, & \text{jeżeli zadanie } j \text{ jest przydzielone wykonawcy } i \\ 0, & \text{w przeciwnym przypadku} \end{cases}$$

Postać matematyczna modelu GAP jest następująca:

$$\min z(\mathbf{X}) = \sum_{i=1}^m \sum_{j=1}^n c_{ij} x_{ij} \quad (119)$$

$$\sum_{i=1}^m x_{ij} = 1 \quad j = 1, \dots, n \quad (120)$$

$$\sum_{j=1}^n r_{ij} x_{ij} \leq b_i \quad i = 1, \dots, m \quad (121)$$

$$x_{ij} \in \{0, 1\} \quad i = 1, \dots, m; j = 1, \dots, n \quad (122)$$

Ograniczenia (120) oznaczają, że każde zadanie musi zostać przydzielone dokładnie jednemu wykonawcy. Każdemu wykonawcy można przydzielić natomiast więcej zadań niż jedno, w zależności od wielkości zasobu, którym dysponuje (np. czas pracy).

Problem GAP sformułowali i prace nad nim rozpoczęli badacze w latach siedemdziesiątych ubiegłego wieku (Srinivasan i Thompson, 1973; Ross i Soland, 1975). Model GAP znajduje wiele zastosowań w dziedzinie planowania i organizacji produkcji, harmonogramowania, alokacji zasobów, organizacji transportu. Problem doczekał się także wielu rozszerzeń. W literaturze przedmiotu (Kundakcioglu i Alizamir, 2008) znajdziemy dość obszerną bibliografię stosowanych w praktyce wersji GAP i ich możliwości. Do najważniejszych modyfikacji należą:

- problem GAP z wieloma rodzajami zasobów MRGAP;
- wielopoziomowy problem GAP (MGAP) – dopuszcza się zróżnicowanie wydajności wykonawców przy wykonywaniu zadań, zróżnicowanie kosztów

oraz wymagań odnośnie zasobów; każde zadanie musi być przydzielone dokładnie jednemu wykonawcy i każdy z wykonawców dysponuje zasobami o ograniczonej wielkości;

- dynamiczny uogólniony problem przydziału DGAP – w modelu uwzględnia się dodatkowo warunki czasowe wykonywania zadań, wymagania popytowe, dodatkowe koszty związane z niedotrzymaniem wymaganych terminów lub wykonaniem zadań zbyt wcześnie;
- GAP z kryterium minimaxowym (Bottleneck GAP, BGAP) – kryterium optymalizacji to minimalizacja maksymalnej straty związanej z przydziałem wykonawcy każdego z zadań; o ile kryterium (119) minimalizacji sumy kosztów jest zwykle stosowane dla przedsiębiorstw sektora prywatnego, to kryterium minimaxowe wykorzystywane jest często w sektorze publicznym;
- problem GAP ze specjalnym uporządkowaniem zbiorów zadań (GAPS2) – jest to problem alokacji zadań do przedziałów czasowych w taki sposób, aby każde zadanie było przydzielone do określonego przedziału czasowego lub ewentualnie rozdzielone na dwa sąsiadujące przedziały;
- stochastyczny problem GAP – problem, w którym dokładna informacja dotycząca dostępnych zasobów, niezbędnych do wykonania zadań, nie jest znana z wyprzedzeniem lub nie wszystkie zadania są znane z wyprzedzeniem;
- wielokryteriowy problem GAP – w którym rozważa się co najmniej dwa kryteria oceny każdego z przydziałów, np. czas i koszt;
- uogólniony problem wielokrotnego przydziału (GMAP) – uwzględnia się możliwość, że każde z zadań może być przydzielone wielu wykonawcom, a nie tylko jednemu;
- dla konkretnych modyfikacji uogólnionych problemów przydziału opracowano wiele algorytmów poszukiwania rozwiązań optymalnych – dokładnych i przybliżonych; ich złożoność obliczeniowa – ze względu na binarny charakter zmiennych decyzyjnych – jest bardzo duża; wymienić należy algorytmy podziału i ograniczeń (*branch-and-bound*) w różnych wersjach, wiele wariantów metod heurystycznych, algorytmy genetyczne (por. Kundakcioglu i Alizamir, 2008).

3.2. Kwadratowy problem przydziału

Kwadratowy problem przydziału QAP (ang. *Quadratic Assignment Problem*) został sformułowany w literaturze przedmiotu (Koopmans i Beckmann, 1957) i polega na przydziale zbioru n obiektów (fabryk, magazynów) do zbioru n lokalizacji. Znane są odległości pomiędzy poszczególnymi lokalizacjami oraz zależność (przepływ towarów) pomiędzy obiektami. Celem jest tu dokonanie przydziału w taki sposób, aby zminimalizować sumę iloczynów przepływu i odległości między obiektami.

Macierze zmiennych decyzyjnych $\mathbf{X} = [x_{ij}]$ w modelach przydziału LAP oraz QAP, ze względu na naturę problemu przydziału, są macierzami permutacji. Każda permutacja ϕ zbioru $N = \{1, 2, \dots, n\}$ może być przedstawiona w postaci macierzy o wymiarach $n \times n$, spełniających warunek:

$$x_{ij} = \begin{cases} 1, & \text{jeżeli } \phi(i) = j \\ 0, & \text{w przeciwnym przypadku} \end{cases}$$

gdzie zapis $\phi(i) = j$ należy rozumieć w ten sposób, że w i -tym wierszu macierzy \mathbf{X} , będącym wektorem jednostkowym, element równy 1 znajduje się w j -tej kolumnie. Przykładowo, następującej permutacji ϕ zbioru trzeylementowego:

$$\phi = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 3 & 1 \end{bmatrix}, \text{ tj. } \phi(1) = 2, \phi(2) = 3, \phi(3) = 1$$

odpowiada macierz permutacji:

$$\mathbf{X} = \mathbf{X}(\phi) = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

Dane są trzy macierze o wymiarach $n \times n$ o elementach rzeczywistych:

- $\mathbf{F} = [f_{ij}]$, gdzie f_{ij} oznacza przepływ pomiędzy obiektem i a obiektem j ;
- $\mathbf{D} = [d_{kl}]$, gdzie d_{kl} oznacza odległość między lokalizacją k oraz lokalizacją l ;
- $\mathbf{B} = [b_{ik}]$, gdzie b_{ik} oznacza koszt umieszczenia obiektu i w lokalizacji k .

Model QAP w wersji Koopmans–Beckmann można sformułować, za badaczami (Burkard, Cela, Pardalos i Pitsoulis, 1998, s. 243 i nn.), następująco. Oznaczmy przez $N = \{1, 2, \dots, n\}$ zbiór numerów obiektów i lokalizacji.

$$\min_{\phi \in \delta_n} \sum_{i=1}^n \sum_{j=1}^n f_{ij} d_{\phi(i)\phi(j)} + \sum_{i=1}^n b_{i\phi(i)} \quad (123)$$

gdzie δ_n to zbiór wszystkich permutacji $\phi: N \rightarrow N$. Każdy z iloczynów $f_{ij} d_{\phi(i)\phi(j)}$ jest kosztem przydzielenia obiektu i do lokalizacji $\phi(i)$ oraz obiektu j do lokalizacji $\phi(j)$. Macierze \mathbf{F} oraz \mathbf{D} są macierzami symetrycznymi z 0 na głównej przekątnej i wszystkie macierze mają elementy nieujemne. Model przydziału z funkcją kryterium (123) można zapisać symbolicznie jako QAP(\mathbf{F} , \mathbf{D} , \mathbf{B}).

Bardziej ogólną wersję QAP zaproponował inny autor (Lawler, 1963). W miejsce macierzy \mathbf{F} oraz \mathbf{D} wprowadził czterowymiarową macierz parametrów

$\mathbf{C} = [c_{ijkl}]$, gdzie $c_{ijkl} = f_{ij}d_{kl}$ dla wszystkich indeksów i, j, k, l , dla których $i \neq j, k \neq l$ oraz $c_{iikk} = f_{ii}d_{kk} + b_{ik}$ w przeciwnym przypadku. Funkcja kryterium przyjmuje w tej wersji modelu postać:

$$\min_{\phi \in \delta_n} \sum_{i=1}^n \sum_{j=1}^n c_{ij\phi(i)\phi(j)} + \sum_{i=1}^n b_{i\phi(i)} \quad (124)$$

Zapis modelu QAP (Lawler, 1963):

$$\min z(\mathbf{X}) = \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n \sum_{l=1}^n c_{ijkl} x_{ik} x_{jl} + \sum_{i=1}^n \sum_{j=1}^n b_{ij} x_{ij} \quad (125)$$

$$\sum_{i=1}^n x_{ij} = 1, \quad j = 1, \dots, n \quad (126)$$

$$\sum_{j=1}^n x_{ij} = 1, \quad i = 1, \dots, n \quad (127)$$

$$x_{ij} \in \{0, 1\} \quad i, j = 1, \dots, n \quad (128)$$

Uzyskanie rozwiązania optymalnego dla problemu QAP ciągle stanowi wyzwanie dla naukowców. Jak donoszą badacze algorytmów poszukiwania rozwiązania problemu QAP (Burkard, Dell'Amico i Martello, 2009; Truetsch, 2014), jedynym sposobem uzyskania rozwiązania optymalnego w rozsądnym czasie obliczeniowym dla $n \geq 15$ są zmodyfikowane algorytmy podziału i ograniczeń (*branch-and-bound*). Ich idea opiera się na poszukiwaniu krok po kroku optymalnej permutacji, bez potrzeby przeszukiwania wszystkich permutacji.

Model QAP doczekał się wielu aplikacji: lokalizacja pomieszczeń szpitalnych, budynków w kampusie uniwersyteckim, połączenia w sieciach komputerowych, optymalizacja problemów transportowych, optymalizacja układu wskaźników na tablicach sterowniczych, optymalne projektowanie układów scalonych itd. W literaturze przedmiotu (Povh, 2008) można znaleźć informacje na temat znanych zastosowań modeli przydziału i efektywnych algorytmów dla rozwiązywania problemów przydziału.

Rozdział III

Zarządzanie projektem w logistyce¹

1. Istota i charakterystyka projektów

W literaturze przedmiotu twierdzi się, że pierwszym projektem była budowa kolei transkontynentalnej w Stanach Zjednoczonych. Teorię tę należy traktować jednak dosyć elastycznie, ponieważ określenie „projekt” znane jest od bardzo dawna. Szybki rozwój badań nad projektami przypada na przełom XIX i XX wieku (Koźmiński, Piotrowski, 2004). Ich prekursorami byli tacy uczeni, jak Frederick Winslow Taylor (Taylor, 1911)² czy Henry Laurence Gantt (Gantt, 1913)³. Nowoczesne metody badań nad projektami zaczęto stosować w okresie II wojny światowej (Hoffman, 2016, s. 12). Do najbardziej znanych projektów z tego czasu można zaliczyć projekt Manhattan oraz budowę radaru.

Nie można wskazać jednego znaczenia terminu „projekt”. Występuje wiele definicji, które mają, oczywiście, pewne wspólne cechy, do których zaliczamy:

- cel – projekt to „działanie podejmowane dla spowodowania rezultatów oczekiwanych przez stronę zamawiającą” (Oberlander, 2000, s. 4–5); powinien być sformułowany zgodnie z zasadą SMART, czyli: proste (ang. *simple*), mierzalne (ang. *measurable*), dające się ocenić jakościowo (ang. *asseable*), realne (ang. *realistic*), określone w czasie (ang. *time-bound*);
- niepowtarzalność – projekt „ma charakter niepowtarzalny zarówno w zakresie koncepcji, jak i realizacji” i „jest odpowiedzią na jakąś jednostkową potrzebę” (Strategor, 2001, s. 365); natomiast według PMI projekt to działanie „podejmowane w celu stworzenia niepowtarzalnego wyrobu lub usługi” (Duncan, 2013, s. 4);

1 Rozdział w dużej części został oparty na publikacjach, których autorem lub współautorem jest Michał Trocki, mający bardzo duże osiągnięcia w omawianej dziedzinie.

2 Frederick W. Taylor stwierdził, że każdą pracę można podzielić na działania, które składają się z pojedynczych czynności.

3 Henry L. Gantt opracował sposób graficznej prezentacji harmonogramu wielu różnych działań, wykonywanych w różnym czasie i składających się na konkretny projekt. Należy zauważyć, że wykresy Gantta są stosowane do dziś.

- złożoność – projekt to złożone przedsięwzięcie, które według definicji prakseologicznej oznacza „działanie złożone, wielopodmiotowe, przeprowadzane zgodnie z planem, który ze względu na skomplikowanie bywa sporządzany przy pomocy specjalnych metod” (Kotarbiński, 1970, s. 193);
- określoność – projekt to działanie „zawarte w skończonym przedziale czasu, z wyróżnionym początkiem i końcem” (Jędrzejczyk, Kukula, Skrzypek i Walkosz, 2016, s. 223);
- autonomia (niezależność) – według grupy francuskich uczonych publikujących jako Strategor (Strategor, 1995) realizacja projektów „przebiega w sposób niezależny od reszty działalności przedsiębiorstwa”.

Opierając się na polskiej literaturze przedmiotu, precyzyjną definicję projektu można by stworzyć, łącząc podejście Karola Kukuli i Michała Trockiego: projekt to zorganizowane działanie ludzkie zmierzające do osiągnięcia określonego celu, niepowtarzalne (realizowane jednorazowo), złożone, zawarte w skończonym przedziale czasu – z wyróżnionym początkiem i końcem, realizowane przez skończoną liczbę osób, określoną ilość środków technicznych, energii, materiałów, środków finansowych i informacji, realizowane w sposób względnie niezależny od powtarzalnej działalności przedsiębiorstwa, za pomocą specjalnych metod i technik, a także składające się z ciągu czynności wzajemnie ze sobą powiązanych (Jędrzejczyk, Kukula, Skrzypek i Walkosz, 2016, s. 223; Trocki, Grucza i Ogonek, 2009).

Z definicji tej wynikają następujące szczegółowe cechy działań projektowych (Łada, Kozarkiewicz, 2010, s. 14):

- realizowane zadanie powinno mieć charakter unikatowy, zarówno na etapie koncepcji, jak i samej realizacji;
- cele projektu powinny być specyficzne, a także określone w kategoriach czasu, kosztów i parametrów jakościowych, co oznacza, że ich sformułowanie powinno być zgodne z zasadą SMART;
- wykonanie projektu powinno dawać określony efekt;
- faza realizacji projektu nie powinna zależeć od pozostałej działalności przedsiębiorstwa;
- projekt odznacza się najczęściej sporą złożonością, dużym zakresem i skalą działania, a także znaczną wartością zaangażowanych zasobów;
- organizacja zasobów oraz działań powinna dotyczyć specyfiki zadań, jakie należy wykonywać;
- projekt powinien być ograniczony w czasie, mieć określony początek i koniec, a jego wykonanie związane powinno być z presją czasową;
- ilość czasu powinna zostać ustalona z góry i kontrolowana;
- realizacja każdego projektu jest związana z ponoszeniem określonych kosztów, a jednocześnie występuje ograniczony zakres finansowania.

Definicja projektu powinna zawierać następujące elementy:

- tło projektu (ang. *background*),
- cele (ang. *objectives*),

- zakres (ang. *scope*),
- ograniczenia (ang. *constraints*),
- założenia (ang. *assumptions*),
- ryzyka (ang. *risks*),
- rezultaty (ang. *deliverables*).

Podójście takie można określić akronimem BOSCARD.

„Ekspozowanie w definicjach projektu unikatowości i traktowanie go jako zadania indywidualnego (jednorazowego), jest zasadniczym ograniczeniem zakresu i rodzaju przedsięwzięć, jakie w praktyce można utożsamiać z pojęciem »projekt«. W rzeczywistości większość projektów ma charakter typowy i powtarzalny, aczkolwiek z cechami, które mogą nadawać im również specyfikę i znamiona oryginalności. W większości projektów występują zarówno rozwiązania wyjątkowe oraz spektakularne, jak i stereotypowe. Jednak wszystkie projekty (np. budowli, urządzeń, strategii przedsiębiorstwa, planu biznesowego, projektu prawnego itd.), które na ogół nie mają jakichś cech nadzwyczajności, powinny się cechować racjonalnością i innowacyjnością” (Stabryła, 2006).

Istotą projektów jako złożonych przedsięwzięć jest ich określoność, dotycząca podstawowych parametrów projektu, takich jak (Trocki, 2012, s. 21):

- zakres (ang. *scope*) – zestaw czynności niezbędnych do osiągnięcia zamierzonych rezultatów;
- spełnienie wymagań (ang. *performance*) – odnosi się do wymogów jakościowych dotyczących wyniku projektu;
- koszty realizacji (ang. *cost*) – występuje limit nakładów i kosztów (dla całego projektu i jego poszczególnych części), który nie powinien być przekroczony;
- czas realizacji (ang. *time*) – okres, w którym projekt powinien być wykonany lub terminy realizacji projektu (w obu przypadkach dla całego projektu i jego poszczególnych części).

W wyniku realizacji projektu należy osiągnąć wszystkie określone wyżej parametry na założonym poziomie, czyli:

- zakres – w całości,
- spełnienie wymagań – dobrze,
- koszty – tanio,
- czas – szybko.

Spełnienie wymagań dotyczy jakościowej strony wyniku projektu, co w przypadku obiektów złożonych, jakimi są projekty właśnie, powoduje konieczność sporządzenia odpowiednich katalogów wymagań. Parametr kosztowy oznacza określenie limitu nakładów i kosztów dla całego projektu oraz jego głównych części składowych, który nie powinien być przekroczony. Parametr czasowy to określenie dla projektu i jego głównych składowych:

- okresu, w którym powinna zakończyć się jego realizacja;
- terminów kalendarzowych realizacji.

Parametrem niepodlegającym optymalizacji, ale mającym zasadnicze znaczenie określające rozmiar projektu, jest jego zakres, związany z wymaganiami jakościowymi, dotyczącymi zamierzonego rezultatu oraz ograniczeniami kosztów i czasu (termin wykonania). Podstawowe parametry projektu są ze sobą ściśle powiązane – i dlatego nie mogą być określane autonomicznie, lecz z uwzględnieniem wzajemnych ograniczeń i powiązań.

Cykl życia projektu to model realizacji projektu w czasie, określający różnicowanie sytuacji (faz życia projektu) występujących w trakcie jego realizacji. Nie można wskazać jednego ogólnie akceptowalnego modelu – wielu specjalistów zarządzania projektem tworzyło modele cyklu życia projektu, składające się z mniejszej lub większej liczby faz.

2. Rola projektów w zarządzaniu przedsiębiorstwem

Zarządzanie projektami można określać w dwóch perspektywach:

- węższej – zespół działań kierowniczych związanych z realizacją projektów oraz zbiór wykorzystywanych w tych działaniach zasad, metod i środków;
- szerszej – dziedzina wiedzy teoretycznej i praktycznej znajdującej zastosowanie w zarządzaniu projektami.

Zgodnie z definicją Project Management Institute zarządzanie projektami to „dziedzina zarządzania zajmująca się zastosowaniem dostępnej wiedzy, umiejętności, narzędzi oraz technik w celu spełnienia potrzeb i oczekiwań zleceniodawców projektów” (Duncan, 2013, s. 6). Często cytowane definicje zarządzania projektami przedstawione zostały w tabeli 52.

Tab. 52. Porównanie wybranych definicji pojęcia „zarządzanie projektami”

Źródła	Definicja	Najważniejsze aspekty
Szyjewski (2001)	Proces z pogranicza nauki o zarządzaniu i sztuki; polega na definiowaniu celów i ich osiąganiu przez równoczesną realizację prac prowadzących do osiągnięcia tych celów, organizowanie zespołów i kierowanie niezbędnym personelem, kontrolę postępów i zakończenie wszystkich prac.	<i>nauka i sztuka zarządzania, równoczesna realizacja prac</i>
Gareis (2004)	Proces zachodzący w organizacji zorientowanej projektowo, który obejmuje następujące subprocesy: inicjowanie projektu, ciągłą koordynację, monitorowanie, rozwiązywanie problemów związanych z ciągłością prac (kryzysy, zmiany organizacyjne, zmiany celów itd.).	<i>proces i subprocesy</i>

Źródła	Definicja	Najważniejsze aspekty
Kerzner (2005)	Planowanie, harmonogramowanie i kontrola ciągu powiązanych ze sobą działań, pozwalające na realizację celów projektu skutecznie i w sposób możliwie najbardziej zgodny z oczekiwaniami jego interesariuszy.	<i>planowanie, harmonogramowanie i kontrola</i>
Shenhar, Dvir (2008)	Zbiór czynności menedżerskich koniecznych do doprowadzenia projektu do pomyślnego zakończenia.	<i>czynności menedżerskie</i>

Źródło: Kozarkiewicz, 2012, s. 25.

Pod koniec lat osiemdziesiątych XX wieku w licznych przedsiębiorstwach można było zauważyć znaczne zainteresowanie projektami. Realizacja poszczególnych zadań, a także wymagania, jakie stawia rzeczywistość, sprawiają, że organizacje coraz więcej działań wcielają w formę projektów. Spowodowało to ewolucję sposobu zarządzania organizacjami.

Organizacje nieprojektowe	Organizacje mieszane	Organizacje projektowe	Organizacje w formie projektu
<ol style="list-style-type: none"> 1. Produkcja powtarzalna 2. Nacisk na produkcję i sprzedaż 3. Długie cykle życia produktów 4. Struktury sztywne, najczęściej funkcjonalne 	<ol style="list-style-type: none"> 1. Produkcja powtarzalna, ale realizująca wiele projektów 2. Nacisk na tworzenie nowych produktów 3. Krótki cykl życia produktów 4. Struktury funkcjonalne i zespoły projektowe 	<ol style="list-style-type: none"> 1. Działalność w postaci kolejnych projektów 2. Okresowo, nierytmicznie realizowane projekty 3. Orientacja na różnorodne potrzeby klientów 4. Nacisk na efektywne zarządzanie projektami 5. Struktury projektowe 	<ol style="list-style-type: none"> 1. Organizacja powołana do określonego projektu 2. Zarządzanie przedsiębiorstwem sprostowane do zarządzania projektami 3. Struktura dostosowana do potrzeb projektu
Zarządzanie produktami	Zarządzanie projektami	Zarządzanie przez projekty	Zarządzanie przedsiębiorstwem – projektowym

Rys. 13. Rodzaje organizacji z punktu widzenia roli projektów

Źródło: Łada i Kozarkiewicz, 2010, s. 46.

Jeżeli organizacja realizuje kilka projektów, mamy do czynienia z portfelem projektów. Najczęściej cytowane definicje tego pojęcia przedstawia tabela 53.

Tab. 53. Porównanie wybranych definicji pojęcia „portfel projektów”

Źródło	Definicja	Najważniejsze aspekty
Platje, Seidel, Wadman (1994)	Zbiór projektów zarządzanych i koordynowanych w taki sposób, że poziom łącznych dostarczanych korzyści nie byłby możliwy do osiągnięcia w przypadku indywidualnego zarządzania każdym projektem.	<i>zbiór projektów, koordynowanie, łączne korzyści</i>
Archer, Ghasem-zadch (1999)	Grupa projektów realizowanych pod kierunkiem jednego wspólnego sponsora lub jednej organizacji.	<i>grupa projektów, wspólny sponsor</i>
Gareis (2004)	Portfel projektów tworzą wszystkie projekty realizowane w przedsiębiorstwie w określonym czasie oraz istniejące między nimi powiązania.	<i>wszystkie projekty, określony czas, powiązania</i>
Carroll (2006)	Zrównoważony zbiór projektów w sensie dopasowania do celów strategicznych oraz kompromisu między ryzykiem a korzyściami w całej organizacji.	<i>zbiór projektów, równowaga, cele strategiczne, kompromis</i>
PMI (2008)	Zbiór projektów lub programów oraz innych działań, które zostały zgrupowane razem, aby ułatwić efektywne zarządzanie i osiągnięcie celów strategicznych.	<i>zbiór projektów, programów i innych działań, efektywne zarządzanie, cele strategiczne</i>

Źródło: Kozarkiewicz, 2012, s. 28.

Uzupełnieniem definicji „portfela projektów” jest analiza kategorii projektów tworzących portfel. Kategorie te mogą dotyczyć wielkości projektu, związanego z nim ryzyka, wymaganych nakładów, rodzaju sponsora, poziomu innowacyjności itd. W trakcie realizacji projektów znajdujących się w portfelu szczególnie istotna jest analiza występujących między nimi zależności. Pod tym względem można wyróżnić następujące rodzaje projektów (Kozarkiewicz, 2012, s. 29):

- projekty niezależne, czyli takie, które mogą być realizowane w tym samym czasie;
- projekty komplementarne, czyli takie, które są ze sobą powiązane względami technicznymi lub strategicznymi;
- projekty wykluczające się – tworzy je para lub grupa projektów, z których tylko jeden może zostać zrealizowany;

Zarządzanie portfelem projektów musi uwzględniać wszystkie funkcje kierowania projektami w nim zawartymi. Wybrane definicje zarządzania portfelem projektów przedstawia tabela 54.

Tab. 54. Porównanie wybranych definicji pojęcia „zarządzanie portfelem projektów”

Źródło	Definicja	Najważniejsze aspekty
Artto, Dietrich, Ikonen (2002)	Zarządzanie organizacją multiprojektową, w której podstawowym wyzwaniem jest połączenie celów projektów z celami organizacji.	<i>łącznie celów projektów i organizacji</i>
Kerzmer (2005)	Ma na celu określenie, wybór, finansowanie, monitorowanie i realizację odpowiedniego zestawu projektów i przedsięwzięć potrzebnych do osiągnięcia celów organizacji.	<i>określenie, wybór, finansowanie i realizacja</i>
Thiry (2006)	Proces analizy i alokacji zasobów organizacji między projekty i programy tak, aby osiągać cele organizacji i maksymalizować wartość dla interesariuszy.	<i>alokacja zasobów, maksymalizacja wartości dla interesariuszy</i>
Shenhar, Dvir (2008)	Obejmuje działania i decyzje podejmowane przez firmę w celu wybrania lub odrzucenia projektów, które mają być włączone do zestawu projektów realizowanych, oraz sposoby najbardziej efektywnego podziału zasobów między projektami; naczelnym celem zarządzania portfelem projektów powinna być zawsze realizacja celów biznesowych.	<i>selekcja projektów, alokacja zasobów, cele biznesowe</i>
PMI (2008)	Scentralizowane zarządzanie, które obejmuje zebranie, identyfikację, priorytetyzację, selekcję, autoryzację, równoważenie, zarządzanie, kontrolę (składników portfela), aby osiągnąć założone cele biznesowe.	<i>ciąg określonych działań, cele biznesowe</i>

Źródło: Kozarkiewicz, 2012, s. 28.

Zarządzanie portfelami projektów powinno być ciągłym, powtarzalnym procesem decyzyjnym, umożliwiającym kształtowanie składu portfela obejmującego wprowadzanie nowych projektów oraz monitorowanie i zmianę sposobu realizacji projektów będących w trakcie realizacji (przyspieszenie, opóźnienie, rezygnacja). Zarządzanie portfelami projektów winno również w szerokiej perspektywie prowadzić do wzrostu wartości firmy, a tym samym umocnienia jej pozycji na rynku. Można to osiągnąć poprzez taki strategiczny dobór projektów, który zapewni równowagę między różnymi ich rodzajami, zminimalizuje ryzyko związane z ich realizacją, zmaksymalizuje zwrot z poniesionych nakładów oraz zapewni najlepsze wykorzystanie dostępnych zasobów przedsiębiorstwa. Zarządzanie projektami rozumiane jest jako „robienie projektów w sposób właściwy” (ang. *doing project right*), podczas gdy zarządzanie portfelami projektów to „robienie właściwych projektów” (ang. *doing the right project*) (Kozarkiewicz, 2012, s. 30).

3. Projekty logistyczne jako szczególny rodzaj projektów

W działalności współczesnych przedsiębiorstw i łańcuchów dostaw coraz częściej wyodrębnia się projekty realizowane w obszarze logistyki, których celem jest wdrożenie nowych rozwiązań pozwalających na uniknięcie lub złagodzenie problemów związanych z przepływem materiałów, informacji, środków finansowych, wiedzy i osób (Pisz i Łapuńka, 2015, s. 85).

Projekty logistyczne są realizowane lub koordynowane przez działy logistyki w ramach jednego przedsiębiorstwa pomiędzy bezpośrednimi ogniwami łańcucha dostaw – dostawcą i odbiorcą, a także w całym łańcuchu dostaw: na skalę krajową lub międzynarodową (Kasperek 2006; Kisperska-Moroń i Krzyżaniak 2009; Nowosielski 2008; Pisz i Łapuńka 2015, s. 87).

Podejście projektowe w logistyce jest wynikiem postępującej globalizacji, przenikania się kultur, rozwoju biznesu globalnego, a tym samym – integracji łańcuchów dostaw. Podstawową przesłanką realizacji projektów logistycznych okazuje się podniesienie poziomu obsługi klienta, zwiększenie efektywności gospodarowania w obszarze zaopatrzenia, transportu, magazynowania, produkcji, dystrybucji w pojedynczych przedsiębiorstwach i/lub łańcuchach dostaw. Zainteresowanie planowaniem i realizacją projektów logistycznych oraz towarzyszącym im problemom wyraża się wzrostem liczby publikacji naukowych. Wynikiem tego jest powstanie różnych definicji projektu logistycznego. Najbardziej trafne wydają się dwie:

- według Jarosława Witkowskiego i Bartłomieja Rodawskiego (Witkowski i Rodawski, 2007) projekty logistyczne stanowią jednorazowe, ograniczone czasowo i budżetowo przedsięwzięcia, których realizacja służy poprawie sprawności i efektywności przepływów produktów oraz towarzyszących im informacji w przedsiębiorstwach, łańcuchach dostaw lub układach przestrzennych;
- według Marka Kasperka (Kasperek, 2012) projekt logistyczny to nierutynowe przedsięwzięcie, wyodrębnione czasowo i kosztowo z innych przedsięwzięć, mające na celu realizację jednorazowego i unikatowego działania, w wyniku którego dochodzi do zmiany systemu logistycznego przedsiębiorstwa lub łańcucha dostaw, w którym to przedsiębiorstwo do tej pory funkcjonowało.

Uogólniając, projekt logistyczny można zdefiniować jako tymczasową działalność podmiotu gospodarczego lub podmiotów gospodarczych będących ogniwami łańcuchów dostaw, dotyczącą zmian w procesach logistycznych, podejmowaną w celu wytworzenia unikatowego wyrobu, dostarczenia unikatowej usługi lub osiągnięcia unikatowego rezultatu. Niemalże każdą znaczącą modyfikację procesu logistycznego w pewnych warunkach należy traktować jako przedsięwzięcie, którego efektywność uzależniona jest od planu wykonania (Pisz i Łapuńka, 2015, s. 87).

Przykładami projektów logistycznych realizowanych w obrębie logistyki i zarządzania łańcuchem dostaw są m.in. (Kisperska-Moroń i Krzyżaniak, 2009; Nowosielski, 2008; Pisz i Łapuńka, 2015, s. 91) – zob. tabela 35:

- lokalizacja zakładów produkcyjnych,
- lokalizacja magazynów,
- wdrożenie systemu WMS,
- wybór systemu magazynowania,
- wybór środków transportu,
- optymalizacja tras przewozu,
- zmiana operatora logistycznego,
- zmiana procesów logistycznych,
- wdrożenie i rozbudowa systemu IT wspierającego logistykę,
- wdrożenie lub rozbudowa systemu klasy ERP,
- wybór nowego dostawcy.

Tab. 55. Typologia projektów logistycznych

Kryterium	Rodzaj projektu logistycznego	Przykłady projektów logistycznych
Czas i znaczenie realizacji projektu logistycznego	strategiczne	<i>centralizacja sieci dystrybucji, wdrożenie kompleksowego pakietu usług logistycznych w przedsiębiorstwach sektora TSL, zmiana lokalizacji zakładu produkcyjnego, wdrożenie koncepcji ECR</i>
	taktyczne	<i>opracowanie normatywów sterowania zapasami, optymalizacja systemu sortowania odpadów suchych i mokrych, wdrożenie techniki kanban</i>
	operacyjne	<i>optymalizacja trasy przewozu, szkolenia w zakresie zarządzania zapasami</i>
Obszar funkcjonalny projektu logistycznego	lokalizacyjne	<i>lokalizacja spalarni śmieci</i>
	transportowe	<i>wybór modelu obsługi transportowej</i>
	magazynowe	<i>budowa magazynu</i>
	zarządzanie zapasami	<i>wybór modelu sterowania zapasami</i>
	poprawy jakości obsługi klientów	<i>sformułowanie strategii obsługi klientów</i>
Rodzaj działań i ich rezultatów w projekcie logistycznym	logistyki zwrotnej	<i>opracowanie systemu gospodarowania opakowaniami zwrótnymi</i>
	szkoleniowe	<i>szkolenie z zakresu technologii IT w logistyce</i>
	doradczo-wdrożeniowe	<i>audyt i racjonalizacja systemu logistycznego</i>
	inwestycyjne (w tym infrastrukturalne)	<i>modernizacja systemu transportu wewnętrznego</i>

Tab. 55 (cd.)

Kryterium	Rodzaj projektu logistycznego	Przykłady projektów logistycznych
Liczba uczestników w projekcie logistycznym	projekty realizowane przez pojedyncze przedsiębiorstwo	<i>wdrożenie techniki kanban</i>
	projekty międzyorganizacyjne – realizowane przez kilka współpracujących organizacji w ramach łańcucha dostaw	<i>wdrożenie koncepcji VMI między dostawcą a odbiorcą</i>
Zasięg przestrzenny projektu logistycznego	lokalne	<i>optymalizacja systemu komunikacji miejskiej</i>
	regionalne	<i>budowa regionalnego centrum dystrybucji</i>
	krajowe	<i>założenie elektronicznej giełdy spedycyjnej</i>
	międzynarodowe	<i>budowa tunelu pod kanałem La Manche</i>
Zasięg podmiotowy projektu	projekty oddziałujące na organizację bezpośrednio w nich uczestniczące	<i>wdrożenie koncepcji JIT</i>
	projekty oddziałujące na otoczenie: sektorowe, regionalne lub makroekonomiczne	<i>budowa regionalnego/międzynarodowego centrum logistycznego</i>
Źródła finansowania projektu logistycznego	finansowanie ze środków własnych	<i>zakup systemu informatycznego klasy ERP</i>
	z wykorzystaniem montażu finansowego (kapitał prywatny i środki publiczne)	<i>budowa centrum logistycznego według modelu PPP</i>
	finansowane ze środków publicznych	<i>budowa obwodnicy miejskiej</i>
Wielkość budżetu projektu logistycznego	mikroprojekty	<i>szkolenie z zakresu logistyki biznesu</i>
	projekty wysokonakładowe	<i>budowa centrum logistycznego</i>

Źródło: Pisz i Łapuńska, 2015; Witkowski i Rodawski, 2007, s. 4.

4. Planowanie sieciowe

Planowanie jest najważniejszą funkcją zarządzania. Według Tadeusza Kotarbińskiego (Kotarbiński, 1982, s. 160–162) warunki konieczne do stworzenia dobrego planu można przedstawić następująco. Plan jest dobry wtedy, kiedy prowadzi do działania sprawnego, a w szczególności, gdy okazuje się:

- celowy, tzn. doprowadzający do celu;
- wykonalny, tzn. możliwy do zrealizowania;
- teoretycznie i praktycznie konsekwentny, tzn. zgodny wewnętrznie i niezawierający sprzeczności;
- operatywny, czyli przejrzysty i czytelny, tzn. niesprawiający trudu w rozumieniu (komunikatywny);
- racjonalny, czyli ugruntowany poznawczo, tzn. oparty na rzetelnej wiedzy;
- giętki, tzn. dopuszczający zmiany w trakcie realizacji;
- optymalnie szczegółowy, tzn. nie za szczegółowy i niezbyt ogólny;
- odpowiednio długodystansowy, tzn. obejmujący możliwie najdłuższy przedział czasu;
- czasowo określony, tzn. zawierający termin prekluzyjny (nieprzekraczalny) wykonania;
- kompletny, tzn. obejmujący całość zadania i istotne jego elementy.

W planowaniu projektów możemy wyróżnić następujące czynności (Trocki, Grucza i Ogonek, 2009, s. 53):

- analiza zadań – podział celów projektu na cele cząstkowe i zadania,
- analiza czynności – przyporządkowanie zadaniom właściwych działań,
- analiza obiektów działania – określenie obiektów działania,
- analiza zasobów – przydział środków do poszczególnych zadań.

Pierwszym projektem, do realizacji którego wykorzystano metody sieciowe, był Manhattan (rozpoczęta w 1941 roku budowa bomby atomowej). Ze względu na tajemnicę wojskową metody te nie mogły być szerzej rozpowszechniane. Pierwsze cywilne zastosowanie miało miejsce na przełomie lat czterdziestych i pięćdziesiątych, przy okazji uruchomienia tzw. planu Marshalla. W latach następnych coraz popularniejsze stało się używanie wspomnianych metod do planowania skomplikowanych i złożonych projektów (m.in. program Polaris, program Apollo, produkcja komputera osobistego IBM, produkcja samolotu Airbus A320). Podstawowe metody planowania sieciowego zostały opracowane i udoskonalone na przełomie lat pięćdziesiątych i sześćdziesiątych XX wieku. Zaliczały się do nich (Trocki, Grucza i Ogonek, 2009, s. 29):

- CPM (ang. *Critical Path Method*) z 1957 roku,
- LESS (ang. *Least Cost Estimating and Scheduling*) z lat pięćdziesiątych,
- PERT (ang. *Program Evaluation and Review Technique*) z 1958 roku,
- MPM (ang. *Metra Potential Method*) z 1958 roku,
- GAN (ang. *Generalized Activity Network*) z 1962 roku,
- PDM (ang. *Precedence Diagramming Method*) z 1964 roku,
- GERT (ang. *Graphical Evaluation and Review Technique*) z 1966 roku,
- RAMPS (ang. *Resource Allocation in Multi-Project and Scheduling*) z lat sześćdziesiątych.

Metody planowania sieciowego wykorzystywane są w zarządzaniu projektami do planowania przebiegu projektu oraz zasobów koniecznych do jego realizacji.

Definicje przytoczone na początku tego rozdziału określają zarządzanie projektami jako dziedzinę teorii organizacji i zarządzania. Z punktu widzenia badań operacyjnych należałoby zwrócić większą uwagę na fakt, iż projekt to tak naprawdę złożone przedsięwzięcie wieloczynnościowe, które można opisać za pomocą skończonej liczby pojedynczych czynności oraz określić rodzaj i stopień ich wzajemnych powiązań (Miszczyńska i Miszczyński, 2002, s. 112).

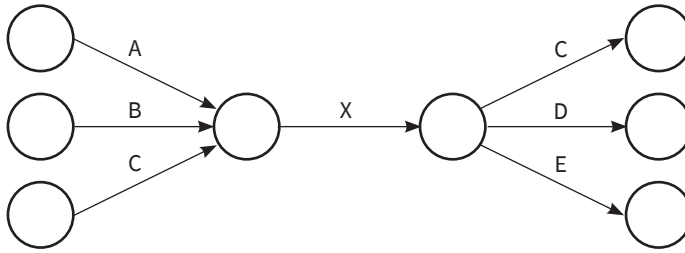
Wyodrębnienie czynności pozwala przede wszystkim na przedstawienie projektu w postaci modelu matematycznego. Do opisywania tego rodzaju modeli stosuje się sieci budowane na podstawie teorii grafów. „Graf jest obiektem składającym się z niepustego zbioru węzłów V i zbioru strzałek lub krawędzi E , oraz zdefiniowanego na zbiorze E odwzorowania, które każdemu elementowi $e \in E$ przyporządkowuje dokładnie dwa elementy $i, j \in V$. Gdy każda para elementów i, j jest uporządkowana, a więc jest wyróżniony węzeł wcześniejszy i następujący po nim, mówimy że graf jest skierowany. Węzeł poprzedzający jest nazywany poprzednikiem, a węzeł następujący następnikiem. Gdy węzły grafu są połączone krawędziami bez wyróżniania zwrotu, graf nazywamy nieskierowanym” (Krawczyk, 2001, s. 94). Grafy, w których strzałkom lub krawędziom przyporządkowane są liczby, tworzą klasę grafów wartościowanych. Graf jest spójny, kiedy dla każdej pary węzłów istnieje ciąg następujących po sobie strzałek lub krawędzi (Krawczyk, 2001, s. 95). Można wyróżnić dwa podejścia ze względu na interpretację węzłów i strzałek (krawędzi), będących składowymi grafu:

1. Notację *Activity on the Arrow* (AoA), występującą także pod nazwą *Activity Diagramming Method* (ADM) lub „model strzałkowy”. Została ona po raz pierwszy zastosowana w pierwszej połowie lat pięćdziesiątych ubiegłego wieku w programie raketowym Polaris (*mfiles.pl*). W tym modelu sieciowym każda strzałka oznacza czynność odwzorowującą wykonanie dowolnego zadania częściowego. Jest procesem trwającym w czasie, którego realizacja wymaga określonych środków, co z kolei powoduje generowanie kosztów. Węzły reprezentują natomiast zdarzenia występujące w projekcie, które określają rozpoczęcie lub zakończenie jednej lub wielu czynności. Powiązania między czynnościami mogą przybierać dwie formy: szeregową i równoległą (zob. rysunek 14 i 15).



Rys. 14. Czynność A jest bezpośrednim poprzednikiem czynności B (połączenie szeregowe)

Źródło: Krawczyk, 2001, s. 116.

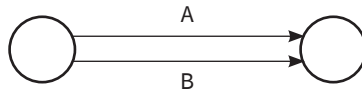


Rys. 15. Czynności A, B i C mogą być wykonywane równolegle i jednocześnie poprzedzają czynność X; a czynności C, D i E także mogą być wykonywane równolegle i poprzedzane są bezpośrednio przez czynność X

Źródło: opracowanie własne na podstawie Krawczyk, 2001, s. 116–118.

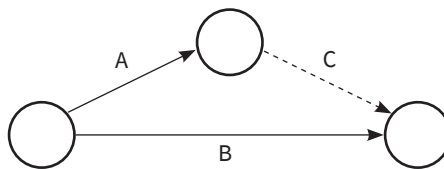
Sieci sporządzone zgodnie z metodą AoA muszą spełniać dwie dodatkowe reguły:

- a) Każde dwa zdarzenia nie mogą być bezpośrednio połączone więcej niż jedną czynnością rzeczywistą⁴. Rysunek 16 przedstawia błędny sposób zapisu, z kolei na rysunku 17 zaprezentowano poprawny sposób zapisu połączeń, realizujący relację bezpośredniego poprzedzania czynności. Zwiualizowana na rysunku 17 czynność C to czynność pozorna, do realizacji której nie potrzeba ani czasu, ani kosztów.



Rys. 16. Błędny zapis

Źródło: Krawczyk, 2001, s. 117.

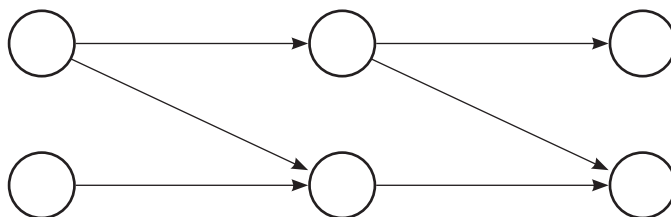


Rys. 17. Prawidłowy zapis

Źródło: Krawczyk, 2001, s. 117.

⁴ Czynność rzeczywista to taka, do realizacji której wymagany jest określony czas i potrzebne są odpowiednie środki.

- b) Każda sieć musi posiadać jedno zdarzenie początkowe i jedno zdarzenie końcowe.

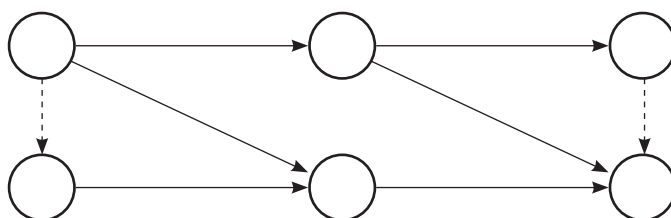


Rys. 18. Błędny zapis

Źródło: opracowanie własne.

W sieci przedstawionej na rysunku 18 występują dwa zdarzenia początkowe i dwa zdarzenia końcowe – jest to niedopuszczalne w sieci AoA. Prawidłowy zapis ilustruje rysunek 19.

Zakończenia czynności A (relacja koniec–początek).



Rys. 19. Prawidłowy zapis

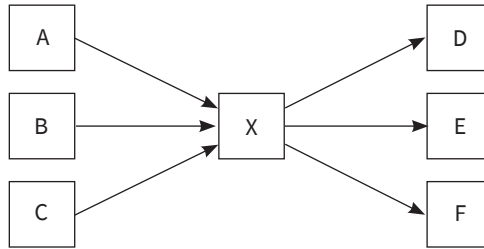
Źródło: opracowanie własne.

- Notację *Activity on the Node* (AoN), występującą także pod nazwą *Precedence Diagramming Method* (PDM) i „metoda diagramowania pierwszeństwa”. Została ona spopularyzowana wraz z upowszechnieniem komputerów, zastępując notację AoA. W tym modelu sieciowym każdy węzeł oznacza czynność, a strzałki przedstawiają relacje zachodzące między analizowaną czynnością a czynnościami bezpośrednio ją poprzedzającymi i bezpośrednio po niej następującymi (*mfiles.pl*).



Rys. 20. Czynność A jest bezpośrednim poprzednikiem czynności B (połączenie szeregowe)

Źródło: Trocki, 2012, s. 161.

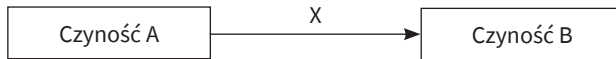


Rys. 21. Czynności A, B i C mogą być wykonywane równolegle i jednocześnie poprzedzają czynność X; czynności D, E i F także mogą być wykonywane równolegle i poprzedzane są bezpośrednio przez czynność X

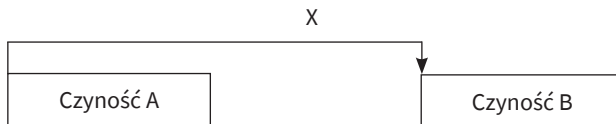
Źródło: opracowanie własne na podstawie Trocki, 2012, s. 161.

Dla każdych dwóch czynności A i B, występujących w sieci w bezpośrednim układzie szeregowym (zob. rysunek 20), można rozpatrywać następujące przypadki (Trocki, 2012, s. 161):

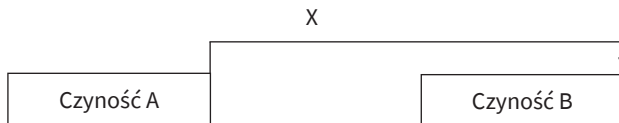
- a) Czynność B może się rozpocząć nie wcześniej niż po upływie X jednostek czasu od zakończenia czynności A (relacja koniec–początek).



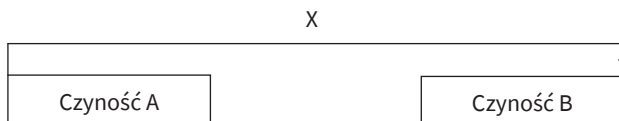
- b) Czynność B może się rozpocząć nie wcześniej niż po upływie X jednostek czasu od rozpoczęcia czynności A (relacja początek–początek).



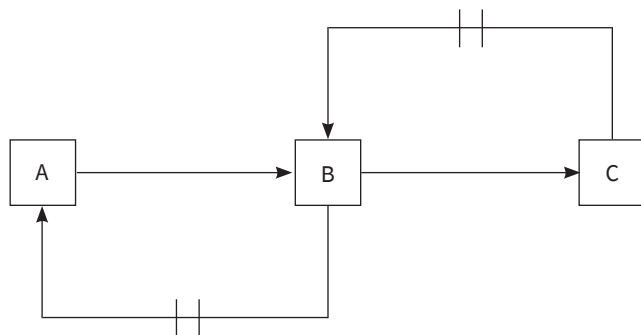
- c) Czynność B może być zakończona nie wcześniej niż po upływie X jednostek czasu od zakończenia czynności A (relacja koniec–koniec).



- d) Czynność B może być zakończona nie wcześniej niż po upływie X jednostek czasu od rozpoczęcia czynności A (relacja początek–koniec)



W sieci typu AoN sprzężenia zwrotne są niedopuszczalne (zob. rysunek 22).



Rys. 22. Sprzężenia zwrotne w sieci typu AoN – niedopuszczalne

Źródło: Trocki, 2012, s. 162.

W przypadku, gdy występuje więcej niż jedna czynność początkowa i/lub końcowa, wykorzystujemy czynności pozorne – mogą być one także stosowane do połączenia dwóch lub większej liczby grafów.

Dzięki planowaniu sieciowemu mamy możliwość (Bładowski, 1970, s. 9):

- przedstawienia projektu w postaci sieci czynności wzajemnie ze sobą powiązanych,
- sporządzenia harmonogramu realizacji projektu (określenie terminów rozpoczęcia i kończenia poszczególnych czynności),
- analizy tolerancji czasu rozpoczynania i kończenia poszczególnych czynności (oszacowanie zapasów czasu i ich analiza),
- określenia czynności krytycznych mających decydujące znaczenie w realizacji projektu,
- optymalizacji rozdziału środków niezbędnych do realizacji projektu,
- oszacowania prawdopodobieństwa dotrzymania założonego terminu realizacji projektu,
- oceny alternatywnych planów realizacji projektu,
- bieżącej kontroli realizacji projektu.

Podczas realizacji większości projektów mamy do czynienia z ograniczonymi zasobami (środki finansowe, środki produkcji, materiały, siła robocza), czego skutkiem jest konieczność odpowiedniego uszeregowania czynności wchodzących w skład projektu. Równie ważnym ograniczeniem okazuje się możliwy do zaakceptowania czas trwania projektu

5. Planowanie przebiegu projektu

Planowanie przebiegu projektu jest jednym z elementów drugiej fazy cyklu życia projektu (faza planowania i organizowania wykonawstwa projektu). Chodzi o powiązanie elementów struktury projektu z czasem i miejscem ich realizacji. Planowanie to obejmuje następujące czynności (Trocki, 2012, s. 171):

- identyfikacja czynności (zadań),
- określenie czasów trwania czynności projektu,
- obliczenie terminów wykonania poszczególnych czynności,
- określenie czasu realizacji całego projektu oraz terminów jego rozpoczęcia i zakończenia,
- obliczenie rezerw czasu w projekcie,
- wyznaczenie ścieżki krytycznej,
- sporządzenie harmonogramu projektu,
- określenie kamieni milowych.

Punktem wyjścia okazuje się tu identyfikacja czynności (zadań) wchodzących w skład projektu. W zależności od stopnia innowacyjności projektu występują trzy sposoby tworzenia listy czynności (Krawczyk, 2001, s. 96–97):

- przypadkowy,
- systematyczny,
- wyróżniający czynności w trakcie tworzenia listy zależności.

Wszystkie wyróżnione czynności tworzą zbiorowość określającą projekt – dlatego też musi zostać zachowana skala pomiarów umożliwiającą ich agregację; powinny one posiadać również wspólną jednostkę czasu. Drugim parametrem, który należy uwzględnić, jest potencjał wykonawczy. Ponieważ czynności mogą być realizowane przez ludzi, maszyny, a także z wykorzystaniem sił naturalnych, najczęściej parametrami pokazującymi stronę ich wykonania są koszty. W sytuacji, kiedy okazują się one wielkościami decyzyjnymi, wprowadza się je jako funkcje czynników wykonawczych i tworzy zadania optymalizacyjne, których funkcją kryterium jest z kolei ich minimalizacja. Po wyspecyfikowaniu wszystkich wchodzących w skład projektu czynności należy je uporządkować. Postępowanie to można podzielić na dwa etapy (Krawczyk, 2001, s. 98):

1. Etap pierwszy – dynamiczne tworzenie listy, na której dowolna czynność X ma numer większy od wszystkich czynności ją poprzedzających i niższy od tych, które po niej następują.
2. Etap drugi – wskazanie wśród czynności poprzedzających czynność X jej bezpośrednich poprzedników.

Następnie szacujemy przewidywany czas trwania czynności tworzących strukturę projektu. W zależności od tego, czy projekt cechuje się niskim stopniem innowacyjności, czy też jest przedsięwzięciem unikatowym, stosuje się różne techniki. W pierwszym przypadku można wykorzystać doświadczenia wynikające

z realizacji podobnych projektów, co pozwala na prawie dokładne (deterministyczne) określenie czasu trwania czynności. W drugiej sytuacji, z powodu braku doświadczeń wynikających z realizacji podobnych projektów, stosuje się metody szacunkowe, wykorzystujące zasady rachunku prawdopodobieństwa. Z reguły nakład czasu potrzebny do realizacji czynności ustala się najpierw w jednostkach naturalnych, a następnie w kalendarzowych (Trocki, 2009, s. 54–55). Stosowane mogą być różnorodne techniki – przedstawia je tabela 56.

Tab. 56. Techniki określania czasu realizacji czynności projektu

Techniki	Charakterystyka
Analityczna	nakłady czasu są określone na podstawie obliczeń analitycznych czasów trwania elementów czynności
Analogii	nakłady czasu są określone na podstawie analogii do czynności zrealizowanych w podobnych projektach
Multiplikacyjna	nakłady czasu są określone przez pomnożenie produktów/wyników cząstkowych czynności przez nakłady dla tych produktów
Oceny ważonej	nakłady czasu są określone przez ustalenie wag dla zestawu czynników wpływających na czas trwania czynności (obliczenie wyniku następuje na podstawie odpowiednich wzorów)
Parametryczna	nakłady czasu są określone na podstawie analizy korelacji
Procentowa	nakłady czasu są określone za pomocą struktury procentowej nakładów podobnych projektów
Punktów funkcjonalnych	specjalna technika szacowania nakładów wykorzystująca specjalistyczne oprogramowanie

Źródło: Trocki, 2009, s. 55.

W zależności od rodzaju struktury projektu stosuje się różne techniki planowania jego przebiegu (zob. tabela 57).

Tab. 57. Przegląd technik planowania przebiegu projektu

Struktura projektu Czynności projektu	Hierarchiczna	Koooperacyjna	
		Zdeterminowana	Stochastyczna
Zdeterminowane	Technika LOB Technika Pattern Technika CPE Techniki drzewa zależności	Harmonogramy – Wykresy Gantta Technika CPM Technika MPM	
Stochastyczne	Techniki drzewa decyzyjnego	Technika PERT Technika łańcucha krytycznego	Technika Gert

Źródło: Trocki, Grucza i Ogonek, 2009, s. 56.

5.1. Technika CPM

Metoda ta została opracowana w latach 1956–1957 w Stanach Zjednoczonych przez firmę Du Pont de Nemours & Co. Była ona wykorzystywana do zarządzania konserwacją urządzeń stosowanych w procesach ciągłych koncernu. Metoda ścieżki krytycznej miała doprowadzić do uporządkowania organizacyjnego, co miało spowodować wzrost efektywności firmy. Jest to metoda deterministyczna, czyli taka, w której zakłada się, że określone są czasy realizacji poszczególnych czynności i ich uporządkowana lista. Mając dane czasy trwania każdej czynności oraz sieć powiązań między nimi, wyznacza się – na podstawie ścieżki krytycznej – najkrótszy czas realizacji całego projektu, nazywany czasem krytycznym.

Ścieżka krytyczna jest to droga prowadząca od zdarzenia początkowego do końcowego i charakteryzująca się najdłuższym czasem trwania; tworzą ją te czynności, których zapas czasu całkowity jest równy luzowi czasowemu zdarzenia końcowego.

Jeżeli stosujemy notację AoA, proces wyznaczania ścieżki krytycznej jest następujący. Etap pierwszy – przedstawienie struktury projektu w postaci wykresu sieciowego typu AoA. Sieć składa się z dwóch elementów: zdarzeń (węzłów) i łuków (czynności) – zob. rysunek 23.



Rys. 23. Elementarna składowa sieci – czynność (i,j)

Źródło: opracowanie własne.

gdzie:

- (i,j) – czynność rozpoczynająca się zdarzeniem i oraz kończąca się zdarzeniem j ,
- t_{ij} – czas trwania czynności (i,j) ,
- t_i^0 – najwcześniejszy moment rozpoczęcia czynności (i,j) ,
- t_i^1 – najpóźniejszy moment rozpoczęcia czynności (i,j) ,
- t_j^0 – najwcześniejszy moment zakończenia czynności (i,j) ,
- t_j^1 – najpóźniejszy moment zakończenia czynności (i,j) ,
- L_i – luz czasowy zdarzenia i .

Etap drugi – wyznaczenie terminów najwcześniejszych dla zdarzeń (t_i^0). Dla zdarzenia początkowego całego przedsięwzięcia, tj. dla zdarzenia o numerze 1, najczęściej przyjmuje się, że termin ten jest równy 0⁵.

5 Jeżeli terminy związane są ściśle z kalendarzem, jako moment rozpoczęcia projektu przyjąć można np. numer tygodnia, w którym planowane jest jego rozpoczęcie. W tym przypadku

$$t_i^0 = 0 \quad (129)$$

Dla każdego zdarzenia j , ($j = 2, 3, \dots, n$) najwcześniejszy termin zajścia oblicza się, analizując czasy trwania czynności, dla których dane zdarzenie jest zdarzeniem końcowym. Terminem najwcześniejszego zajścia zdarzenia będzie termin najbarziej oddalony w czasie. Wynika to z definicji zdarzenia⁶. Procedura wyznaczania najwcześniejszych terminów dla zdarzeń jest zgodna z następującym wzorem:

$$t_j^0 = \max_{i < j} \{t_i^0 + t_{ij}\} \quad j = 2, 3, \dots, n \quad (130)$$

Proces wyznaczania najwcześniejszych terminów dla zdarzeń przebiega w sposób narastający, czyli w kolejności od zdarzenia pierwszego (początkowego dla całego procesu) do zdarzenia ostatniego (kończącego proces). Inne podejście do wyznaczania terminów w drugim etapie metody CPM przedstawia Tadeusz Trzaskalik (Trzaskalik, 2008). Odmienność polega na tym, że zamiast wyznaczania najwcześniejszych terminów zajścia poszczególnych zdarzeń, wyznacza się najwcześniejsze momenty rozpoczęcia i zakończenia poszczególnych czynności. Najwcześniejszy moment rozpoczęcia czynności jest tożsamy z najwcześniejszym momentem zajścia zdarzenia początkowego. Najwcześniejszy moment zakończenia czynności okazuje się natomiast równy najwcześniejszemu momentowi zajścia zdarzenia końcowego dla danej czynności.

Etap trzeci – wyznaczenie terminów najpóźniejszych dla zdarzeń (t_i^1). Proces wyznaczania terminów najpóźniejszych zajść zdarzeń przebiega w kolejności od ostatniego (zdarzenie o numerze n) do pierwszego (zdarzenie o numerze 1). Pierwszym krokiem tego etapu jest wyznaczenie najpóźniejszego terminu zakończenia przedsięwzięcia, tzw. terminu dyrektywnego ($TD = t_n^1$). Ze względu na deterministyczny charakter metody, termin dyrektywny nie może być krótszy od najwcześniejszego momentu zajścia zdarzenia końcowego n .

$$t_n^0 \leq TD \quad (131)$$

Najczęściej za termin dyrektywny przyjmuje się najwcześniejszy moment zajścia zdarzenia końcowego n ($TD = t_n^0$). Na tym poziomie przyjmuje się również najpóźniejszy termin dla zdarzenia końcowego ($TD = t_n^1$). Z założeń tych wynika, że $t_n^1 = t_n^0$.

Dla zdarzeń o kolejnych, niższych numerach najpóźniejszy termin określa się jako najmniejszą z różnic pomiędzy terminami najpóźniejszymi dla zdarzeń kończących wszystkie czynności rozpoczynające się w analizowanym węźle, a czasami trwania tych czynności. Proces ten można ogólnie zapisać następująco:

wszystkie czasy rozpoczęcia i zakończenia poszczególnych czynności wyrażone będą numerami tygodni.

6 Mówimy, że zdarzenie zaszło, jeżeli zakończyły się wszystkie czynności, dla których jest ono zdarzeniem końcowym.

$$t_i^1 = \min_{i < j} \{t_j^1 - t_{ij}\} \quad j = n-1, n-2, \dots, 1 \quad (132)$$

Skuteczną kontrolą obliczeń w tym etapie jest uzyskanie takich samych różnic pomiędzy terminem najwcześniejszym a najpóźniejszym dla zdarzeń pierwszego (o numerze 1) i ostatniego (o numerze n).

$$t_1^1 - t_1^0 = t_n^1 - t_n^0 \quad (133)$$

Etap czwarty – wyznaczenie luzów czasowych dla zdarzeń (L_i). Dla każdego z węzłów określone zostały dwa terminy: najwcześniejszych i najpóźniejszych momentów zajścia zdarzenia. Różnicę między tymi wielkościami nazywamy luzem czasowym dla zdarzenia i . Wyznacza się go następująco:

$$L_i = t_i^1 - t_i^0 \quad (134)$$

Luz czasowy pokazuje, w jakim przedziale czasu należy spodziewać się zakończenia wszystkich czynności, dla których zdarzenie i jest zdarzeniem końcowym. Na tym etapie wskazać można zdarzenia krytyczne. Są to te zdarzenia, dla których luz czasowy $L_i = 0$. Leżą one na ścieżce krytycznej.

Etap piąty – wyznaczanie zapasów czasu dla czynności. Po wyznaczeniu najwcześniejszych i najpóźniejszych momentów dla zdarzeń obliczamy pewne charakterystyki nazywane zapasami czasu wykonywania czynności. Ze względu na to, że każdą czynność można teoretycznie rozpocząć zarówno w najwcześniejszym, jak i w najpóźniejszym momencie zajścia zdarzenia początkowego dla czynności, jak również dwojako ją zakończyć, istnieją w sumie cztery rodzaje zapasów czasowych – zob. tabela 58.

Tab. 58. Rodzaje zapasów czasu dla czynności analizowanych w technice CPM

Nazwa	Symbol	Definicja
Zapas czasu całkowity	ZC_{ij}	Odcinek czasu pomiędzy najpóźniejszym a najwcześniejszym terminem wystąpienia czynności.
Zapas czasu warunkowy	ZW_{ij}	Odcinek czasu, o który czynność może być przesunięta w stosunku do swojego najwcześniejszego terminu wystąpienia – bez wpływu na terminy innych czynności.
Zapas czasu względny	ZWZ_{ij}	Odcinek czasu, o który czynność może być przesunięta w stosunku do swojego najpóźniejszego terminu wystąpienia – bez wpływu na terminy innych czynności.
Zapas czasu niezależny	ZN_{ij}	Odcinek czasu, o który czynność może być przesunięta, gdy czynność ją poprzedzająca znajduje się w najpóźniejszym terminie, a czynność następująca w najwcześniejszym terminie.

Źródło: opracowanie własne.

Z zapasem całkowitym mamy do czynienia w sytuacji, gdy czynność (i,j) rozpoczynana jest w najwcześniejszym momencie zajścia zdarzenia początkowego (t_i^0) , a kończona w najpóźniejszym momencie zajścia zdarzenia końcowego (t_j^1) . Oblicza się go w następujący sposób:

$$ZC_{ij} = t_j^1 - t_i^0 - t_{ij} \quad (135)$$

Zapas niezależny jest przeciwieństwem zapasu całkowitego. Powstaje w chwili, gdy czynność (i,j) rozpoczynana jest w najpóźniejszym momencie zajścia zdarzenia początkowego (t_i^1) , a kończona w najwcześniejszym momencie rozpoczęcia zdarzenia końcowego (t_j^0) . Oblicza się go w następujący sposób:

$$ZN_{ij} = t_j^0 - t_i^1 - t_{ij} \quad (136)$$

Kolejnym rodzajem zapasu czasowego jest zapas swobodny. Wynika on z polityki rozpoczynania czynności (i,j) najwcześniej jak to możliwe (t_i^0) i zakończenia jej również najwcześniej jak się da (t_j^0) . Zapasem swobodnym nazywamy luz czasowy powstający w następujący sposób:

$$ZS_{ij} = t_j^0 - t_i^0 - t_{ij} \quad (137)$$

Ostatnią możliwą kombinacją jest rozpoczęcie czynności (i,j) jak najpóźniej (t_i^1) i zakończenia jej również jak najpóźniej (t_j^1) . Rezerwę czasu powstającą w ten sposób nazywamy zapasem warunkowym. Oblicza się go następująco:

$$ZW_{ij} = t_j^1 - t_i^1 - t_{ij} \quad (138)$$

Etap szósty – wyznaczenie harmonogramu przedsięwzięcia. Etap ten polega na określeniu dla każdej czynności najwcześniejszych oraz najpóźniejszych terminów jej rozpoczęcia i zakończenia.

Przyjmujemy następujące oznaczenia:

NWP_{ij} – najwcześniejszy termin rozpoczęcia czynności (i,j) :

$$NWP_{ij} = t_i^0 \quad (139)$$

NPP_{ij} – najpóźniejszy termin rozpoczęcia czynności (i,j) :

$$NPP_{ij} = t_i^0 + ZC_{ij} \quad (140)$$

NWK_{ij} – najwcześniejszy termin zakończenia czynności (i,j) :

$$NWK_{ij} = t_j^1 + ZC_{ij} \quad (141)$$

NPK_{ij} – najpóźniejszy termin zakończenia czynności (i,j) :

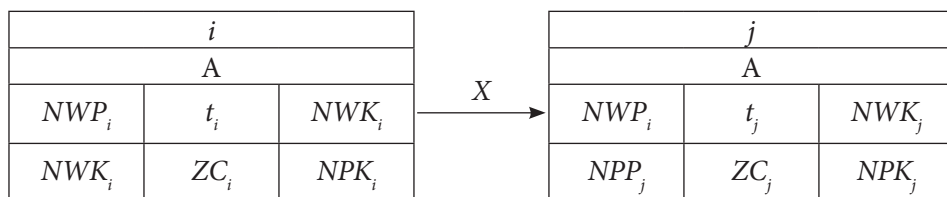
$$NPK_{ij} = t_j^1 \quad (142)$$

Etap siódmy – wyznaczenie ścieżki krytycznej przedsięwzięcia. W etapie drugim przyjęto założenie, że termin dyrektywny dla całego przedsięwzięcia nie może być krótszy niż najwcześniejszy moment zajścia zdarzenia końcowego n . Wynika stąd, że luz czasowy dla zdarzenia n jest nieujemny.

$$L_n = t_n^1 - t_n^0 \geq 0 \quad (143)$$

Ścieżka krytyczna to zbiór tych czynności, dla których zapas całkowity równy jest luzowi czasowemu ostatniego zdarzenia. Opóźnienie którejkolwiek z czynności wchodzących w skład ścieżki krytycznej (czynności krytycznych) jest równoznaczne z takim samym opóźnieniem realizacji całego projektu. Z kolei skrócenie czasu trwania czynności krytycznej może prowadzić do przyspieszenia realizacji całego przedsięwzięcia.

Jeżeli stosujemy notację AoN, to proces wyznaczania ścieżki krytycznej jest z kolei następujący: Etap pierwszy – przedstawienie struktury projektu w postaci wykresu sieciowego typu AoN.



gdzie:

- i – numer czynności $i = 1, 2, \dots, n$;
- A – słowny opis czynności;
- NWP_i – najwcześniejszy początek i -tej czynności;
- NPP_i – najpóźniejszy początek i -tej czynności;
- NWK_i – najwcześniejszy koniec i -tej czynności;
- NPK_i – najpóźniejszy koniec i -tej czynności;
- t_i – czas trwania i -tej czynności;
- ZC_i – zapas czasu całkowity i -tej czynności;
- X – przyspieszenie/opóźnienie rozpoczęcia/zakończenia czynności.

Etap drugi – wyznaczenie terminów najwcześniejszych dla czynności. Dla pierwszej czynności, w zależności od wyboru postępowania, przyjmujemy że:

$$NWP_i = 0 \text{ albo } NWP_i = 1 \quad (144)$$

Dla kolejnych:

$$NWP_j = NWP_i + t_i \text{ lub } NWK_j = NWK_i + t_i \quad (145)$$

Etap trzeci – wyznaczenie terminów najpóźniejszych dla czynności. Dla ostatniej czynności, w zależności od wyboru postępowania, przyjmujemy że:

$$NPK_n = NWK_n \text{ lub } NPK_n = TD \quad (146)$$

gdzie: TD – termin dyrektywny.

Jednocześnie musi być spełniony warunek mówiący, że:

$$TD > NWK_n \quad (147)$$

Dla kolejnych:

$$NPP_i = NPP_j - t_i \text{ lub } NPK_i = NPK_j - t_i \quad (148)$$

Etap czwarty – wyznaczanie zapasów czasu dla czynności.

Zapas czasu całkowity:

$$ZC_i = NPK_i - NWK_i \text{ lub } ZC_i = NPP_i - NWP_i \quad (149)$$

Zapas czasu niezależny:

$$ZN_i = NWP_i - NWK_i \text{ lub } ZN_i = NPP_i - NPK_i \quad (150)$$

Zapas czasu swobodny:

$$ZS_i = NWK_i - NWP_i - t_i \text{ lub } ZS_i = NPK_i - NPP_i - t_i \quad (151)$$

Zapas czasu warunkowy:

$$ZW_i = NPK_i - NWK_i - t_i \text{ lub } ZW_i = NPP_i - NWP_i - t_i \quad (152)$$

Etap piąty – wyznaczanie ścieżki krytycznej. Ścieżkę krytyczną tworzą te czynności, których zapas czasu całkowity jest równy:

$$ZC_i = NPK_n - NWK_n \quad (153)$$

5.2. Technika MPM

MPM jest kolejną deterministyczną techniką planowania sieciowego. Projekt przedstawiany jest tu w postaci wykresu sieciowego, ilustrującego wszystkie czynności wchodzące w jego skład oraz zależności między nimi. Czasy ich realizacji oraz terminy rozpoczęcia czynności w odniesieniu do czynności bezpośrednio

poprzedzających określa się na podstawie przyjętych założeń oraz dotychczasowych doświadczeń.

Technika MPM (franc. *Méthode des potentiels Métra*) została opracowana we Francji w 1958 roku przez Bernarda Roya, a następnie spopularyzowana przez *Société d'Economie et de Mathématiques Appliquées* (SEMA) – przy okazji prac nad samolotem Concorde oraz elektrowniami jądrowymi (Bładowski, 1970).

Do podstawowych zadań techniki MPM należą:

- sprawdzenie, jaki będzie czas realizacji projektu – w jakim najkrótszym czasie może zostać on zrealizowany;
- wyodrębnienie czynności, których zakończenie można opóźnić, nie naruszając przy tym harmonogramu projektu;
- możliwość ustalenia czynności krytycznych – wymagających szczególnej uwagi z punktu widzenia realizacji przedsięwzięcia.

Podobnie jak w przypadku techniki CPM, w technice MPM mamy do czynienia z siecią o deterministycznej strukturze oraz analizą czynności krytycznych. Metoda MPM wykorzystuje jednak jednopunktowe modele sieciowe, które umożliwiają przedstawienie dodatkowych zależności między czynnościami. W odróżnieniu od dwupunktowych modeli stosowanych m.in. w metodach CPM i PERT, w technice MPM czynności przedstawia się za pomocą prostokątów z zaznaczonymi wewnątrz numerami i opisami. Łuki oznaczane są strzałkami pionowymi, poziomymi oraz ukośnymi. Postępowanie w metodzie MPM podzielić można na siedem etapów.

Etap pierwszy – określenie projektu i przygotowanie analizy jego struktury. Należy wyznaczyć czynności ze ściśle określonym początkiem i końcem, których realizacja wymaga czasu i odpowiednich środków.

Etap drugi – określenie zależności pomiędzy poszczególnymi czynnościami wchodzącymi w skład projektu. Podobnie jak w metodzie CPM, dla każdej czynności należy wskazać czynności ją poprzedzające i występujące po jej zakończeniu. Szczególnie ważne jest powiązanie analizowanej czynności z czynnościami bezpośrednio ją poprzedzającymi.

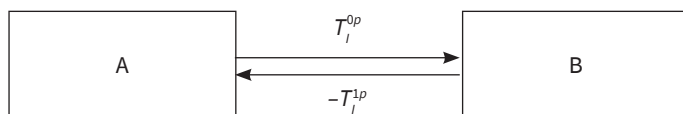
Etap trzeci – sporządzenie wykresu sieciowego łączącego wszystkie czynności. W tym przypadku wykorzystuje się jednopunktowe modele sieciowe, składające się z węzłów (prostokąty) i łuków (strzałki). Węzły oznaczają czynności. Wewnątrz wpisuje się ich oznaczenia oraz początek czynności, koniec lub czas trwania. Łuki przedstawiają strukturalne zależności między określonymi punktami czasu różnych czynności. Wykres sieciowy w tej metodzie ma węzeł początkowy oznaczany jako „start” (ST), od którego rozpoczyna się obliczanie terminów rozpoczęcia poszczególnych czynności. Czas trwania tego węzła równa się 0. Sieć nie powinna zawierać pętli łączących dwukrotnie te same węzły oraz czynności pozorne. Czynności powiązane między sobą, które łączy zależność przyczynowo-skutkowa, tworzą ścieżkę krytyczną. Poszukujemy ścieżki prowadzącej od czynności początkowej do końcowej, którą nazywamy pełną (Trocki, Grucza i Ogonek, 2009).

Etap czwarty – przypisanie czynnościom zakładanego czasu realizacji. Pierwsza czynność grafu reprezentuje „start” (ST) sieci i w praktyce nie oznacza czynności, która fizycznie musi zostać wykonana. Jest ona natomiast o tyle istotna, że od niej rozpoczyna się obliczanie terminów dla kolejnych czynności. Czas trwania czynności ST jest równy 0.

W opisywanej metodzie terminy rozpoczęcia danej czynności wyznacza się w stosunku do momentu rozpoczęcia czynności bezpośrednio ją poprzedzającej⁷. Przyjmuje się następujące zasady oznaczeń łuków:

- wartość dodatnia (umieszczana nad łukiem) oznacza najwcześniejszy termin rozpoczęcia danej czynności, licząc od momentu rozpoczęcia czynności poprzedzającej (moment ten oznacza się następująco jako T_i^{0p});
- wartość ujemna (umieszczana pod łukiem) oznacza najpóźniejszy termin rozpoczęcia danej czynności w stosunku do czynności poprzedzającej (oznacza się go jako T_i^{1p}).

Graficzną prezentację przedstawionych oznaczeń przedstawia rysunek 24:



Rys. 24. Graficzna prezentacja oznaczeń łuków

Źródło: opracowanie własne.

Wielkości T_i^{0p} i T_i^{1p} nie są tożsame z używanymi w metodzie CPM terminami NWP_{ij} , NPP_{ij} , NWK_{ij} , NPK_{ij} , gdyż odnoszą się jedynie do czynności poprzedzających, nie zaś do całości projektu.

Dla powyższych zależności możliwe są następujące sytuacje:

1. Określony jest jedynie najwcześniejszy termin rozpoczęcia czynności następującej w stosunku do poprzedzającej, tj. T_i^{0p} . Czynność następująca może się rozpocząć najwcześniej w T_i^{0p} jednostce czasu trwania czynności poprzedzającej.
2. Określony jest dokładny termin rozpoczęcia czynności następującej w stosunku do poprzedzającej, najwcześniejszy oraz najpóźniejszy termin rozpoczęcia czynności są sobie równe, tj. $T_i^{0p} = T_i^{1p}$.
3. Określony jest zarówno najwcześniejszy, jak i najpóźniejszy moment rozpoczęcia czynności następującej w stosunku do czynności poprzedzającej,

⁷ Wszystkie czasy trwania oraz momenty dotyczące czynności opracowywane są przez ekspertów oraz osoby związane z projektem – na podstawie uwarunkowań technicznych, organizacyjnych oraz wcześniejszych doświadczeń.

- tj. $T_i^{0p} < T_i^{1p}$. Czynność następująca może zacząć się najwcześniej w T_i^{0p} , a najpóźniej w T_i^{1p} jednostce trwania czynności poprzedzającej.
4. Czynność następująca musi zacząć się w tym samym momencie, co czynność poprzedzająca, tj. $T_i^0 = T_i^1 = 0$. Omawianie pozostałych kombinacji terminów T_i^{0p} i T_i^{1p} nie jest celowe, gdyż w praktyce oznacza zamianę kolejności rozpatrywanych czynności.

Etap piąty – obliczanie najwcześniejszych terminów rozpoczęcia i zakończenia czynności. Aby obliczyć najwcześniejsze momenty rozpoczęcia czynności, tj. NWP_i , należy zbudować macierz powiązań dla sieci zależności opisującej dany projekt, ilustrującą relacje czasowe między czynnościami. Do poszczególnych pól macierzy należy wpisać oznaczenia z łuków między czynnościami. Wiersze oznaczają początek łuku, czyli czynność początkową, natomiast kolumny – koniec łuku (czynność końcową). Przed obliczeniem najwcześniejszych terminów rozpoczęcia i zakończenia czynności należy wyznaczyć wartości pomocnicze T_i^{0p+} oraz T_i^{0p-} . Parametr T_i^{0p+} uwzględnia dodatnie relacje między czynnościami (bierze pod uwagę jedynie wartości umieszczane nad łukami; w macierzy – wartości powyżej przekątnej).

Parametr ten oblicza się, korzystając z przygotowanej wcześniej macierzy powiązań między czynnościami, zaczynając od lewej strony macierzy (czynność ST), dla każdej czynności. Postępowanie polega na sumowaniu wartości leżących na skrzyżowaniu danej kolumny z wartością T_i^{0p+} , wyznaczoną wcześniej dla czynności odpowiadającej numerowi wiersza, w którym znajduje się wartość liczbowa. W przypadku, gdy w danej kolumnie znajduje się więcej niż jedna wartość, na T_i^{0p+} wybiera się tę sumę, która jest większa. Dla czynności ST przyjmuje się $T_i^{0p+} = 0$. Następnie oblicza się wartości T_i^{0p-} , uwzględniając jedynie wartości ujemne (leżące pod łukami, a w macierzy – poniżej przekątnej). Procedura jest analogiczna i również przebiega w kolejności od lewej do prawej. Sumuje się wartości leżące na skrzyżowaniu danej kolumny z wartościami T_i^{0+} odpowiadającymi tej czynności, w której wierszu (w danej kolumnie) znajduje się wartość liczbowa. W przypadku kilku wartości w jednej kolumnie, za T_i^0 przyjmuje się większą sumę. Wartość NWP_i dla każdej czynności wyznacza się, wybierając większą z obliczonych wartości T_i^{0p+} lub T_i^{0p-} .

$$NWP_i = \max\{T_i^{0+}; T_i^{0-}\} \quad (154)$$

Etap szósty – obliczanie terminu zakończenia całego projektu oraz najpóźniejszych terminów rozpoczęcia i zakończenia czynności. W tej fazie również zostaje wykorzystana skonstruowana w poprzednim etapie macierz powiększona o dodatkowe trzy kolumny, w których umieszczone będą dla każdej czynności NWP_i , t_p , czyli czas trwania czynności oraz kolumna z sumą $NWP_i + t_p$. Dodatkowo w macierzy uwzględnia się przygotowane do obliczeń trzy wiersze T_i^{1p+} , T_i^{1p-} oraz NPP .

Obliczenia rozpoczyna się od umieszczenia w polach macierzy, leżących na przecięciu wierszy T_i^{1p+} oraz T_i^{1p-} z kolumną t_p , terminu zakończenia projektu, który jest największą z wyliczonych w poprzednim etapie sum $NWP_i + t_i$. Dalsze obliczenia T_i^{1p+} przeprowadza się wierszami, począwszy od wiersza dolnego, w następujący sposób:

- od terminu realizacji przedsięwzięcia odejmuje się czas trwania czynności, dla której obliczana jest bieżąca wartość T_i^{1p+} , t_p ;
- od obliczonego wcześniej terminu T_i^{1p+} , odpowiadającego czynności zapisanej w tej samej kolumnie co dodatnia wartość liczbową znajdująca się w analizowanym wierszu, odejmuje się tę właśnie wartość; w przypadku, gdy w wierszu odpowiadającym danej czynności znajduje się kilka wartości dodatnich w różnych kolumnach, obliczenia przeprowadza się dla wszystkich, a jako wartość T_i^{1p+} wybiera się tę najmniejszą.

W dalszej kolejności wyznacza się parametr T_i^{1p-} , który uwzględnia jedynie wartości ujemne macierzy. Procedura jest analogiczna. Począwszy od wiersza dolnego, od obliczonego wcześniej terminu T_i^{1p+} , odpowiadającemu czynności zapisanej w tej kolumnie co ujemna wartość liczbową znajdująca się w analizowanym wierszu, odejmuje się tę właśnie wartość. W przypadku kilku wartości w jednym wierszu dla każdej kolumny przeprowadza się odejmowanie i wybiera wartość mniejszą. Ostatnim krokiem tego etapu jest wyznaczenie wartości NPP_i .

$$NPP_i = \max\{T_i^{1+}; T_i^{1-}\} \quad (155)$$

Etap siódmy – obliczanie zapasów czasu i wskazanie czynności krytycznych. Po dokonaniu wcześniejszych obliczeń możliwe jest wyznaczenie najwcześniejszych oraz najpóźniejszych terminów zakończenia czynności. Korzysta się z następujących zależności:

$$NWK_i = NWP_i + t_i \quad (156)$$

$$NPK_i = NPP_i + t_i \quad (157)$$

Zapasy czasu dla poszczególnych czynności, dzięki którym wskazać można czynności krytyczne, oblicza się, korzystając z jednego ze wzorów:

$$Z_i = NPK_i - NWK_i \quad (158)$$

$$Z_i = NPP_i - NWP_i \quad (159)$$

5.3. Technika PERT

Technikę PERT (ang. *Program Evaluation and Review Technique*) opracowano na potrzeby marynarki wojennej USA podczas realizacji projektu budowy nuklearnej łodzi podwodnej Polaris (Trocki, Grucza i Ogonek, 2009). Należy ona do

stochastycznych metod sieciowych zarządzania projektami. Oznacza to, że czasy trwania czynności wchodzących w skład przedsięwzięcia są zmiennymi losowymi. Technika PERT daje odpowiedzi m.in. na następujące pytania:

1. Kiedy nastąpi oczekiwane zakończenie projektu?
2. Które z czynności są krytyczne dla projektu i opóźnienie których z nich wydłuży jego oczekiwane zakończenie?
3. Jakie jest prawdopodobieństwo zakończenia projektu w określonym czasie?

Postępowanie w metodzie PERT okazuje się w dużej mierze analogiczne do postępowania w CPM. PERT również należy do dwupunktowych metod zarządzania projektami. Podstawową różnicę stanowi określenie czasów trwania czynności. W technice CPM czasy te określone są z góry, na podstawie danych historycznych lub doświadczeń ekspertów. W metodzie PERT zakłada się natomiast, że czasy trwania czynności są losowe⁸. Do oszacowania parametrów rozkładów prawdopodobieństwa czasów realizacji czynności wykorzystuje się trzy charakterystyki:

1. Optymistyczny czas trwania czynności, oznaczany jako a_{ij} . Odnosi się on do sytuacji, gdy czynność wykonywana jest w najbardziej sprzyjających warunkach i określa najkrótszy możliwy czas jej trwania, tj.:

$$P(t_{ij} < a_{ij}) = 0 \quad (160)$$

2. Najbardziej prawdopodobny czas trwania czynności, oznaczany jako m_{ij} . Odnosi się do sytuacji, kiedy czynność wykonywana jest w normalnych warunkach i oznacza najbardziej prawdopodobny czas realizacji czynności.
3. Pesymistyczny czas trwania czynności, oznaczany jako b_{ij} , zakładający realizację czynności w warunkach najmniej sprzyjających. Oznacza najpóźniejszy możliwy czas realizacji czynności, tj.:

$$P(t_{ij} > b_{ij}) = 0 \quad (161)$$

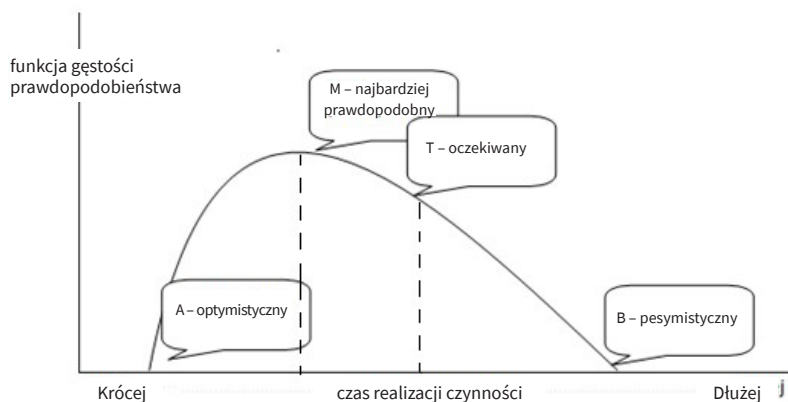
W niektórych opracowaniach prawdopodobieństwo, że czas trwania czynności t_{ij} będzie wartością spoza przedziału (a_{ij}, b_{ij}) przyjmowane jest jako mniejsze niż 1%. Zakładamy, że zmienna losowa, którą jest czas trwania czynności, ma rozkład *beta*. Założenie to wynika z dotychczasowych doświadczeń, na podstawie których wywnioskowano, że czasy wykonania czynności często przekraczają wartość podawaną jako najbardziej prawdopodobną. Jest to konsekwencja zbyt optymistycznego podejścia do szacowania wartości czasów trwania czynności. Oczekiwany czas realizacji czynności znacznie częściej jest dłuższy (niż krótszy) od najbardziej prawdopodobnego. Ta własność prowadzi do wyboru rozkładu

8 Podstawowe informacje na temat prawdopodobieństwa i zmiennych losowych można znaleźć w dodatku 4 niniejszego tomu, pt. *Wybrane elementy teorii prawdopodobieństwa*.

prawdopodobieństwa o prawostronnej (dodatniej) asymetrii, jakim jest rozkład *beta*. Na podstawie przyjętego założenia o rozkładzie prawdopodobieństwa czasu trwania czynności przyjmuje się, że czas oczekiwany $t_{ij}^e = E(t_{ij})$ można oszacować następująco:

$$t_{ij}^e = \frac{a_{ij} + 4m_{ij} + b_{ij}}{6} \quad (162)$$

Graficzną prezentację rozkładu *beta* oraz przedstawionych parametrów przedstawiamy na rysunku 25:



Rys. 25. Rozkład *beta* szacowanego czasu wykonania czynności

Źródło: opracowanie własne na podstawie Trocki, Grucza i Ogonek, 2009, s. 202.

Wykorzystując oczekiwane czasy realizacji poszczególnych czynności, wyznaczamy najwcześniejsze momenty ich rozpoczęcia (oczekiwane). Następnie przyjmujemy arbitralnie wyznaczony czas dyrektywny lub zakładamy, że:

$$TD = t_n^0 = t_n^1 \quad (163)$$

Po wyznaczeniu ścieżki krytycznej (jednej lub więcej) powinniśmy być w stanie wyznaczyć parametry rozkładu czasu trwania projektu, który ma rozkład asymptotycznie normalny (dla projektów z dużą liczbą czynności) – z wartością oczekiwaną $E(t_n)$ równą wartości oczekiwanej terminu najwcześniejszego ukończenia projektu (t_n^0) i z wariancją $\sigma^2(t_n)$ równą sumie wariancji czasów trwania czynności należących do zbioru czynności krytycznych C (przy dodatkowym założeniu, że czasy trwania czynności krytycznych są losowo niezależne).

$$t_n: N\left(E(t_n) = t_n^0; \sigma^2(t_n) = \sum_{(i,j) \in C} \sigma_{ij}^2\right) \quad (164)$$

W tym celu konieczne jest obliczenie wariancji czasów trwania czynności krytycznych zgodnie z następującym wzorem:

$$\sigma_{ij}^2 = \left(\frac{b_{ij} - a_{ij}}{6}\right)^2 \quad (165)$$

Znając oczekiwany czas realizacji projektu $E(t_n)$ oraz jego odchylenie standardowe (pierwiastek z wariancji) $\sigma(t_n)$, można w prosty sposób obliczyć prawdopodobieństwo zakończenia projektu w dowolnym czasie dyrektywnym TD . Korzysta się tu z dystrybuanty Φ rozkładu normalnego $N(0, 1)$, odczytując z tablic statystycznych odpowiednią wartość⁹:

$$\begin{aligned} P(t_n < TD) &= P(u < u_{TD}) = \Phi(u_{TD}) \\ u_{TD} &= \frac{TD - E(t_n)}{\sigma(t_n)} \\ u &\sim N(0, 1) \end{aligned} \quad (166)$$

Przeważnie przyjmuje się harmonogram projektu, dla którego prawdopodobieństwo dotrzymania terminu znajduje się w przedziale od 30 do 60%. Harmonogram, dla którego prawdopodobieństwo to wynosi poniżej 30%, nazywany jest harmonogramem ryzykanta, a powyżej 60% – harmonogramem asekuranta.

5.4. Zastosowanie optymalizacji liniowej do wyznaczania ścieżki krytycznej

Analizę czasową sieci deterministycznej przedstawionej w notacji AoA można przeprowadzić, wykorzystując model optymalizacji liniowej. W wyniku rozwiązania liniowego modelu optymalizacyjnego, dotyczącego analizy czasowej projektu, poznamy terminy zajścia poszczególnych zdarzeń, które są konieczne do zbudowania harmonogramu projektu oraz najkrótszy czas jego trwania. Wynika z tego, że zmiennymi decyzyjnymi modelu są terminy zajścia poszczególnych zdarzeń, a funkcja celu składa się z jednego elementu, jakim jest termin zajścia zdarzenia końcowego.

9 W arkuszu kalkulacyjnym Excel 2016 wartość dystrybuanty zmiennej $u \sim N(0, 1)$ można znaleźć za pomocą standardowej funkcji `=ROZKŁ.NORMALNY.S(u_{TD} ;1)`, gdzie u_{TD} oznacza standardyzowaną wartość terminu TD .

Zatem model przyjmuje następującą postać:

$$t_n \rightarrow \min \quad (167)$$

$$t_j - t_i \geq t_{ij} \quad (i, j) \in G \quad (168)$$

$$t_i \geq 0 \quad (i = 1, 2, \dots, n) \quad (169)$$

gdzie:

- t_i – termin zajścia zdarzenia i ,
- t_j – termin zajścia zdarzenia j ,
- G – zbiór wszystkich czynności projektu.

Wartość funkcji celu wyznacza najkrótszy termin realizacji projektu. Czynności krytyczne odpowiadają tym ograniczeniom, dla których zmienne swobodne są równe 0, a sprzężone z nimi wyceny dualne przyjmują wartości większe od 0. Zerowa wartość zmiennej swobodnej dla danego ograniczenia oznacza, że czynność mu odpowiadająca nie posiada zapasu czasu. Natomiast niezerowa wartość zmiennej dualnej związanej z danym ograniczeniem mówi, o ile wydłuży się czas trwania projektu na skutek zwiększenia czasu trwania czynności t_{ij} o jednostkę (Miszczyńska i Miszczyński, 2002, s. 144).

6. Planowanie zasobów projektu

Jest to kolejny etap zarządzania projektami. Według Ralpha Keelinga (Keeling, 2000, s. 67) zasoby wykorzystywane do realizacji projektu można określić symbolem 7M:

- środki finansowe (ang. *money*),
- materiały (ang. *materials*),
- towary (ang. *merchandise*),
- maszyny i urządzenia (ang. *machinery*),
- zasoby ludzkie (ang. *manpower*),
- kierownictwo i specjaliści (ang. *management, professionals and specialists*),
- ruch (ang. *movement*).

Etap ten jest z reguły określany jako: planowanie kosztów. Budżetowanie zasobów koniecznych do realizacji projektu w jednostkach naturalnych jest niemożliwe. Dlatego postępowanie staje się dwuetapowe:

- określenie potrzebnych zasobów w jednostkach naturalnych,
- określenie wartości zasobów przedstawionych w jednostkach naturalnych.

Z ekonomicznego punktu widzenia powinno się dążyć do tego, aby projekt był zrealizowany przy minimalnym zużyciu zasobów, których wykorzystanie jest ściśle powiązane z czasem jego realizacji. Powoduje to konieczność wykorzystania technik planowania przebiegu projektu. Z uwagi na ograniczone zasoby może wystąpić konieczność zmiany planu przebiegu projektu – zob. tabela 59.

Tab. 59. Przegląd technik planowania zasobów

Metody planowania \ Zasoby	Ludzkie i materialne	Nominalne (nakłady, koszty)
Harmonogramy (wykresy Gantta)	Harmonogramy wykorzystania poszczególnych rodzajów zasobów	Planowanie wykorzystania zasobów
Techniki sieciowe	CPM, PERT	LESS, PERT-COST
Techniki drzewa decyzyjnego	PATTERN, CPE	PATTERN, CPE

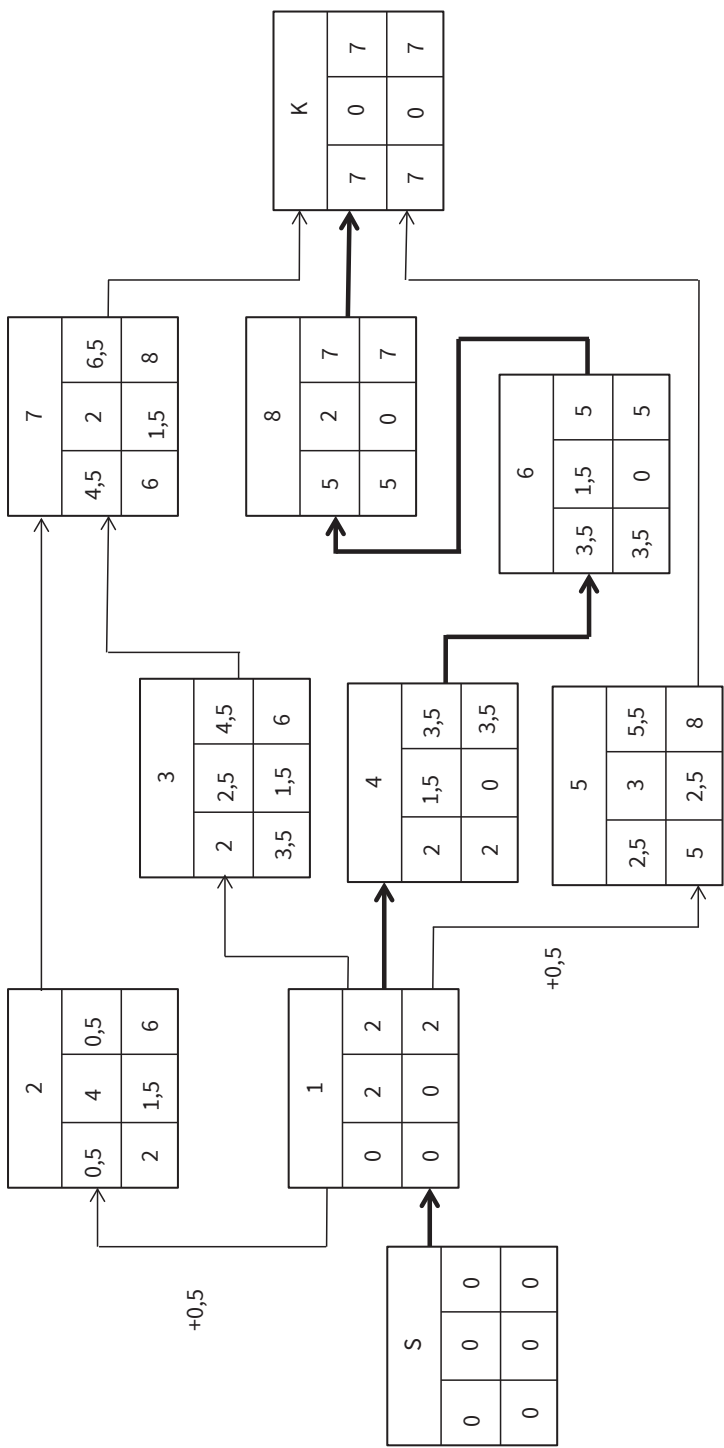
Źródło: Trocki, Grucza i Ogonek, 2009.

Przykład wyrównywania zapotrzebowania na zasoby projektu.

Tab. 60. Dane projektu niezbędne do wyznaczenia ścieżki krytycznej

Zadanie	Poprzedniki	Czas trwania (miesiące)	Zasoby (tys. PLN)
1	–	2	10
2	1 (RR+0,5)	4	12
3	1	2,5	2,5
4	1	1,5	6
5	1 (ZR+0,5)	3	6
6	4	1,5	9
7	2,3	2	2
8	6	2	3

Źródło: opracowanie własne.



Rys. 26. Ścieżka krytyczna
Źródło opracowanie własne.

Tab. 61. Wykres Gantta

1	K	K	K	K										
2										Z				
3										Z				
4					K	K	K							
5													Z	Z
6								K	K	K				
7														Z
8											K	K	K	K
	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7

Źródło opracowanie własne.

Tab. 62. Profil zasobów w tys. PLN przy maksymalnych zasobach w każdych dwóch tygodniach wynoszących 5,5 tys. PLN

tys. PLN	7														
	6,5														
	6														
	5,5														
	5														
	4,5														
	4														
	3,5														
	3														
	2,5														
	2														
	1,5														
	1														
	0,75														
	0,5														
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	
Jednostka: dwa tygodnie															

Jednostka: dwa tygodnie

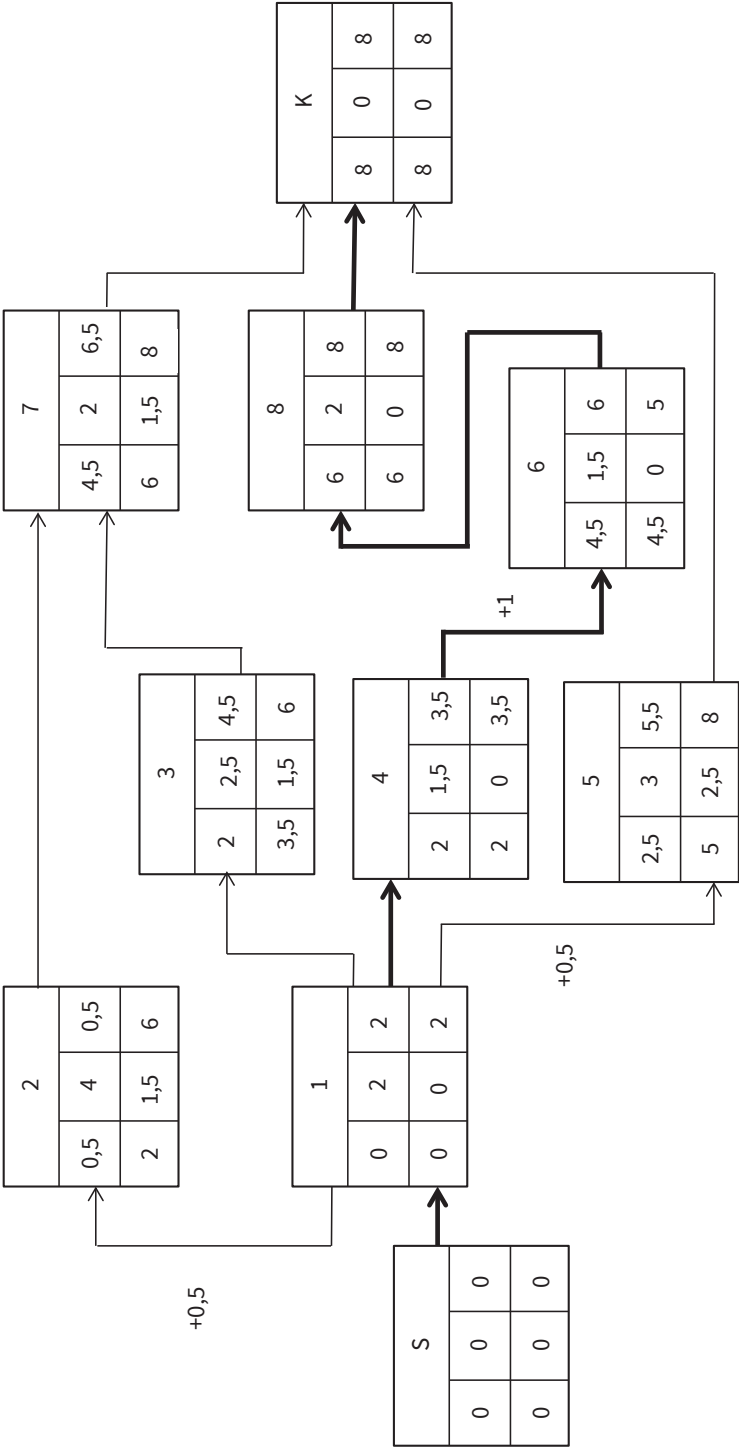
Źródło: opracowanie własne.

Ponieważ przekraczamy limit dostępnych środków, przesuwamy czynność szóstą o cztery tygodnie do przodu.

Tab. 63. Dane do ścieżki krytycznej po zmianach

Zadanie	Poprzedniki	Czas trwania (miesiące)	Zasoby (tys. PLN)
1	–	2	10
2	1 (RR+0,5)	4	12
3	1	2,5	2,5
4	1	1,5	6
5	1 (ZR+0,5)	3	6
6	4 (ZR+1)	1,5	9
7	2,3	2	2
8	6	2	3

Źródło: opracowanie własne.



Rys. 27. Ścieżka krytyczna po zmianach
Źródło: opracowanie własne.

Tab. 64. Wykres Gantta po zmianach

1	K	K	K	K												
2										Z						
3										Z						
4					K	K	K									
5													Z	Z	Z	Z
6										K	K	K				
7														Z	Z	Z
8													K	K	K	K
	0,5	1	1,5	2	2,5	3	3,5	4	4,5	5	5,5	6	6,5	7	7,5	8

Źródło opracowanie własne.

Tab. 65. Profil zasobów w tys. PLN przy maksymalnych zasobach w każdych dwóch tygodniach wynoszących 5,5 tys. PLN po zmianach

tys. PLN	7,5															
	7															
	6,5															
	6															
	5,5															
	5															
	4,5															
	4															
	3,5															
	3															
	2,5															
	2															
	1,5															
	1,25															
	1															
0,75																
0,5																
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Jednostka: dwa tygodnie																

Źródło: opracowanie własne.

6.1. Technika LESS

Czas realizacji większości projektów można przyspieszyć. Wynika to z tego, że pewna liczba czynności może być wykonywana krócej niż to pierwotnie zakładano. Nie jest to związane z błędnym oszacowaniem czasu trwania, a ze świadomym podejmowaniem decyzji o wykonaniu ich w czasie krótszym od założonego. Staje się to możliwe pod warunkiem poniesienia pewnych dodatkowych kosztów. Jedną z uwzględniających taką możliwość metod jest LESS (ang. *Least Cost Estimating*

and Scheduling). W celu rozpoczęcia analizy czasowo-kosztowej należy rozróżnić podstawowe rodzaje kosztów związanych z realizacją projektu. Na koszt całkowity projektu składają przede wszystkim:

- 1) Koszty bezpośrednie, które powiązane są z poszczególnymi czynnościami; w miarę skracania czasu trwania czynności koszty te wzrastają. Wiąże się to z koniecznością wykorzystania większej ilości materiałów, zatrudnienia większej liczby ludzi lub zastosowania nowych technologii w ramach realizacji konkretnej czynności wchodzącej w skład projektu;
- 2) Koszty pośrednie, które generowane są w związku z obsługą realizacji projektu. Do tego rodzaju kosztów zaliczyć można: koszty dzierżawy budynków czy środków transportu, koszty administracyjne i inne koszty ogólne. Z charakteru tych kosztów wynika, że ich wysokość jest wprost proporcjonalna do czasu trwania przedsięwzięcia (Miszczyńska i Miszczyński, 2002, s. 133–134).

Koszt całkowity projektu jest zatem wypadkową kosztów bezpośrednich i pośrednich. Głównym celem stosowania metody czasowo-kosztowej staje się wyznaczenie takiego terminu zakończenia projektu, przy którym koszt całkowity jego wykonania będzie najmniejszy. Dokonuje się tego poprzez wyznaczanie czynności, których czas należy skrócić, określenie o ile należy je przyspieszyć oraz jakie dodatkowe koszty się przy tym ponosi. W związku z głównym, postawionym przed metodą LESS zadaniem konieczne okazuje się zdefiniowanie jeszcze jednej wielkości, a mianowicie: jednostkowego kosztu akceleracji czynności. Wielkość ta określa koszt przyspieszenia danej czynności o jednostkę czasową. Aby w prosty sposób wyznaczyć koszty akceleracji poszczególnych czynności, należy przyjąć następujące założenia:

- 1) Zmiana czasu trwania jednej konkretnej czynności nie ma wpływu na trwanie czynności pozostałych; czasy ich trwania są od siebie niezależne;
- 2) Funkcje kosztów pośrednich i bezpośrednich są funkcjami liniowymi;
- 3) Koszt akceleracji wyznaczyć można jedynie dla tych czynności, dla których $t_{ij}^g < t_{ij}^n$, czyli występuje możliwość skrócenia czasu trwania czynności.

Na mocy przyjętych powyżej założeń jednostkowy koszt akceleracji można wyznaczyć jako:

$$s_{ij} = \frac{\Delta K_{ij}}{\Delta t_{ij}} = \frac{K_{ij}^g - K_{ij}^n}{t_{ij}^n - t_{ij}^g} \quad (170)$$

gdzie:

- t_{ij}^n – normalny czas trwania czynności (i, j);
- t_{ij}^g – graniczny czas trwania czynności;
- t_{ij} – czas trwania czynności, przy czym $t_{ij}^g \leq t_{ij} \leq t_{ij}^n$;
- K_{ij}^n – koszt wykonania czynności w czasie normalnym t_{ij}^n ;
- K_{ij}^g – koszt wykonania czynności w czasie granicznym;
- K_{ij} – koszt bezpośredni wykonania czynności (i, j) w czasie t_{ij} .

Najpopularniejszą metodą wyznaczania harmonogramu o najkrótszym możliwym terminie zakończenia jest metoda iteracyjna. W pierwszym kroku zakłada się, że czynności wykonywane są w normalnym czasie trwania. W każdym kolejnym skraca się czas trwania wybranej czynności o jednostkę oraz oblicza koszt całkowity projektu. Postępowanie takie prowadzi się do momentu, w którym koszt całkowity zaczyna wzrastać, przy czym należy kierować się następującymi zasadami:

- 1) Przyspieszać należy jedynie czynności krytyczne. Przyspieszenie czynności nienależących do ścieżki krytycznej spowoduje podwyższenie kosztów bezpośrednich, a co za tym idzie – kosztu całkowitego projektu (ale nie skróci terminu jego zakończenia);
- 2) Wśród czynności krytycznych przyspieszyć należy tę, której jednostkowy koszt akceleracji jest najniższy;
- 3) W przypadku kilku równoległych ścieżek krytycznych należy przyspieszyć wszystkie, przy czym – w każdej wybiera się do przyspieszenia tę czynność, której koszt przyspieszenia jest najniższy (Miszczyńska i Miszczyński, 2002, s. 136).

6.2. Zastosowanie optymalizacji liniowej do analizy czasowo-kosztowej projektu

Funkcja celu liniowego modelu optymalizacyjnego analizy czasowo-kosztowej realizacji projektu przedstawia koszty całkowite jego wykonania, składające się z kosztów bezpośrednich (K_{ij}), które dla dowolnej czynności (i,j) realizowanej w czasie t_{ij} , spełniają warunek $t_{ij}^g \leq t_{ij} \leq t_{ij}^n$. Można ją zapisać następująco:

$$K_{ij} = K_{ij}^n + s_{ij}\Delta t_{ij} = K_{ij}^n + s_{ij}(t_{ij}^n - t_{ij}) = K_{ij}^n + s_{ij}t_{ij}^n - s_{ij}t_{ij} \quad (171)$$

Koszty bezpośrednie projektu opisuje następująca funkcja:

$$KB = \sum_{(i,j) \in G} K_{ij} = \sum_{(i,j) \in G} (c_{ij} - s_{ij}t_{ij}) = \sum_{(i,j) \in G} c_{ij} - \sum_{(i,j) \in G} s_{ij}t_{ij} \quad (172)$$

gdzie:

$$c_{ij} = K_{ij}^n + s_{ij}t_{ij}^n$$

Drugim elementem kosztów całkowitych są koszty pośrednie (KP). Przyjmujemy założenie, że są one liniowo zależne od czasu trwania projektu, który określa termin zajścia zdarzenia końcowego (t_n). Możemy zapisać je następująco:

$$KP = at_n + b \quad (173)$$

Optymalizacyjny model liniowy, pozwalający odpowiedzieć na pytanie, których czynności czas trwania powinniśmy skrócić i o ile, aby koszt całkowity realizacji projektu był możliwie najkrótszy, przyjmuje następującą postać:

$$KC = \sum_{(i,j) \in G} c_{ij} - \sum_{(i,j) \in G} s_{ij} t_{ij} + at_n + b \quad (174)$$

$$t_j - t_i - t_{ij} \geq 0 \quad (i, j) \in G \quad (175)$$

$$t_{ij}^g \leq t_{ij} \leq t_{ij}^n \quad (i, j) \in G \quad (176)$$

$$t_n \leq TD \quad (177)$$

gdzie: TD – termin dyrektywny (graniczny) zakończenia projektu.

6.3. Technika PERT-COST

Służy ona do wyznaczania optymalnego skracania czasu realizacji projektu w sieciach PERT przy minimalnych kosztach. Założeniem tej metody jest takie postępowanie, dzięki któremu największe przyspieszenie realizuje się w przypadku czynności o najniższym jednostkowym koszcie przyspieszenia. W odróżnieniu od przedstawionej wcześniej metody LESS, technika PERT-COST bierze pod uwagę jedynie bezpośrednie koszty przyspieszenia czynności. Krzywa kosztów jest zatem funkcją parametrów danych czynności, takich jak: a_{ij} , m_{ij} , b_{ij} , t_{ij}^e , σ_{ij}^2 .

W porównaniu z metodą LESS – tu wprowadzamy dodatkowe oznaczenia:

- t_{ij}^{en} – oczekiwany normalny czas trwania czynności, któremu odpowiadają najniższe koszty wykonania czynności K_{ij}^n ;
- t_{ij}^{eg} – oczekiwany graniczny czas trwania czynności, czyli możliwie najkrótszy, wynikający z technicznych oraz organizacyjnych ograniczeń projektu, któremu odpowiadają koszty graniczne K_{ij}^g .

Założeniem techniki PERT-COST, podobnie jak techniki LESS, jest liniowa zależność kosztów akceleracji czynności od czasu jej trwania. Przyrost kosztów wykonania czynności, spowodowany skróceniem czasu jej trwania o jednostkę, można opisać wzorem:

$$s_{ij} = \frac{K_{ij}^g - K_{ij}^n}{t_{ij}^{en} - t_{ij}^{eg}} \quad (178)$$

Dodatkowym założeniem metody są stałe relacje czasów trwania optymistycznego a_{ij} oraz pesymistycznego b_{ij} poszczególnych czynności:

$$\frac{a_{ij}^n}{t_{ij}^{en}} = \frac{a_{ij}^g}{t_{ij}^{eg}} = r_1 \quad (179)$$

$$\frac{b_{ij}^n}{t_{ij}^{en}} = \frac{b_{ij}^g}{t_{ij}^{eg}} = r_2 \quad (180)$$

Z relacji tych wyliczyć można odchylenie standardowe skróconego czasu trwania czynności:

$$\sigma = \sqrt{\frac{(r_2 - r_1)}{36}} (t_{ij}^{es}) \quad (181)$$

gdzie: t_{ij}^{es} – skrócony czas trwania czynności, dla którego koszty wynoszą:

$$K_{ij} = (t_{ij}^{en} - t_{ij}^{es}) \cdot S_{ij} \quad (182)$$

Etapy postępowania są podobne, jak w technice LESS. Rozpoczyna się od obliczenia normalnych czasów dla wszystkich czynności oraz określenia czynności krytycznych. Po obliczeniu kosztów przyspieszenia dla poszczególnych czynności krytycznych należy wyeliminować te, których nie da się przyspieszyć (tzn. te, dla których $t_{ij}^{en} = t_{ij}^{eg}$). W kolejnych krokach skraca się poszczególne czynności, wybierając te z najmniejszymi kosztami przyspieszenia. Jeżeli w sieci występuje więcej niż jedna ścieżka krytyczna, trzeba skracać je jednocześnie, o taką samą wartość przyspieszenia. Postępowanie kończy się uzyskaniem harmonogramu, w którym na każdej ścieżce krytycznej wszystkie czynności osiągnęły czasy graniczne t_{ij}^{eg} . Dalsze skracanie jest w tym przypadku niemożliwe.

Łączne koszty przyspieszenia realizacji całości projektu oblicza się poprzez zsumowanie kosztów – według wzoru (182) – ponoszonych w kolejnych etapach przyspieszania poszczególnych czynności.

7. Studium przypadku – wdrożenie systemu komputerowego w firmie handlowej

Prezentowana w przykładzie firma to spółka handlowa zatrudniająca ok. dwudzieśtu osób, posiadająca oddziały regionalne. Jest wyłącznym przedstawicielem na rynku polski kilku wiodących producentów w branży, w której działa. W związku z dynamicznym rozwojem podjęto decyzję o wdrożeniu systemu komputerowego umożliwiającego zintegrowane działanie oraz kontrolę nad większością aspektów działalności przedsiębiorstwa. Obecnie firma zarządzana jest przy pomocy

trzech programów: finansowo-księgowego, obsługującego sprzedaż i magazyny oraz programu umożliwiającego administrowanie bazą danych klientów. Wyprecyzowano następujące czynności, które muszą być wykonane w trakcie realizacji projektu:

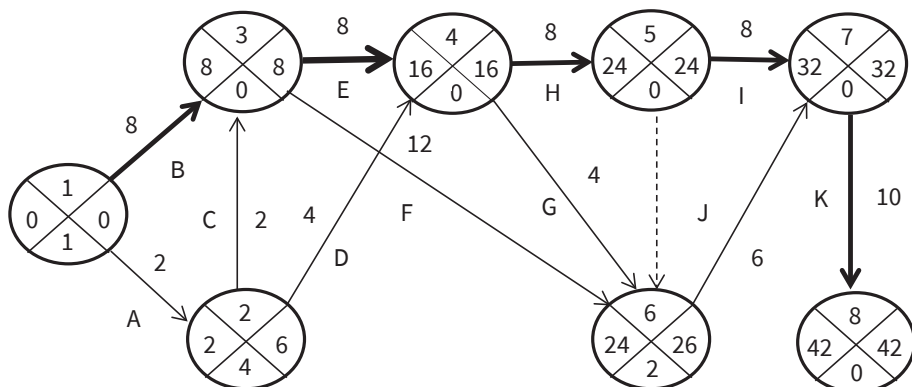
- A. Instalacja i wstępna inicjalizacja systemu.
- B. Gromadzenie danych.
- C. Parametryzacja modułów standardowych.
- D. Opracowanie i stworzenie rachunku zysków i strat, bilansu i rachunku kosztów.
- E. Import danych.
- F. Stworzenie możliwości połączenia z oddziałami regionalnymi oraz test połączeń.
- G. Szkolenia użytkowników systemu.
- H. Tworzenie podstawowych wydruków.
- I. Stworzenie pól i raportów dodatkowych.
- J. Start produkcyjny.
- K. Stworzenie protokołu rozbieżności.

Zastosowanie techniki CPM do określenia najkrótszego czasu trwania projektu – zob. tabela 66 oraz rysunek 28.

Tab. 66. Dane konieczne do sporządzenia wykresu sieciowego typu AoA oraz zapasy czasu całkowite uzyskane w wyniku zastosowania techniki CPM

Symbol	Nazwa czynności	Czynności bezpośrednio poprzedzające	Czas trwania	Zapasy czasu całkowite
A	Instalacja i wstępna inicjalizacja systemu	–	2	4
B	Gromadzenie danych	–	8	0
C	Parametryzacja modułów standardowych	A	2	4
D	Opracowanie i stworzenie rachunków zysków i strat, bilansu i rachunku kosztów	A	4	10
E	Import danych	B, C	8	0
F	Stworzenie możliwości połączenia z oddziałami regionalnymi oraz test połączeń	B, C	12	6
G	Szkolenia użytkowników systemu	D, E	4	6
H	Tworzenie podstawowych wydruków	D, E	8	0
I	Stworzenie pól i raportów dodatkowych	H	8	0
J	Start produkcyjny	F, G, H	6	2
K	Stworzenie protokołu rozbieżności	I, J	10	0

Źródło: opracowanie własne.



Rys. 28. Wykres sieciowy typu AoA z przeprowadzoną analizą CPM

Źródło: opracowanie własne.

Ścieżkę krytyczną tworzą czynności B–E–H–I–K. Najkrótszy czas trwania projektu to czterdzieści dwie jednostki czasowe.

Liniowy model optymalizacyjny do analizy czasowej – zob. równania (183)–(195).

$$t_8 \rightarrow \min \quad (183)$$

$$t_2 - t_1 \geq 2 \quad (184)$$

$$t_3 - t_1 \geq 8 \quad (185)$$

$$t_3 - t_2 \geq 2 \quad (186)$$

$$t_4 - t_2 \geq 4 \quad (187)$$

$$t_4 - t_3 \geq 8 \quad (188)$$

$$t_6 - t_3 \geq 12 \quad (189)$$

$$t_6 - t_4 \geq 4 \quad (190)$$

$$t_5 - t_4 \geq 8 \quad (191)$$

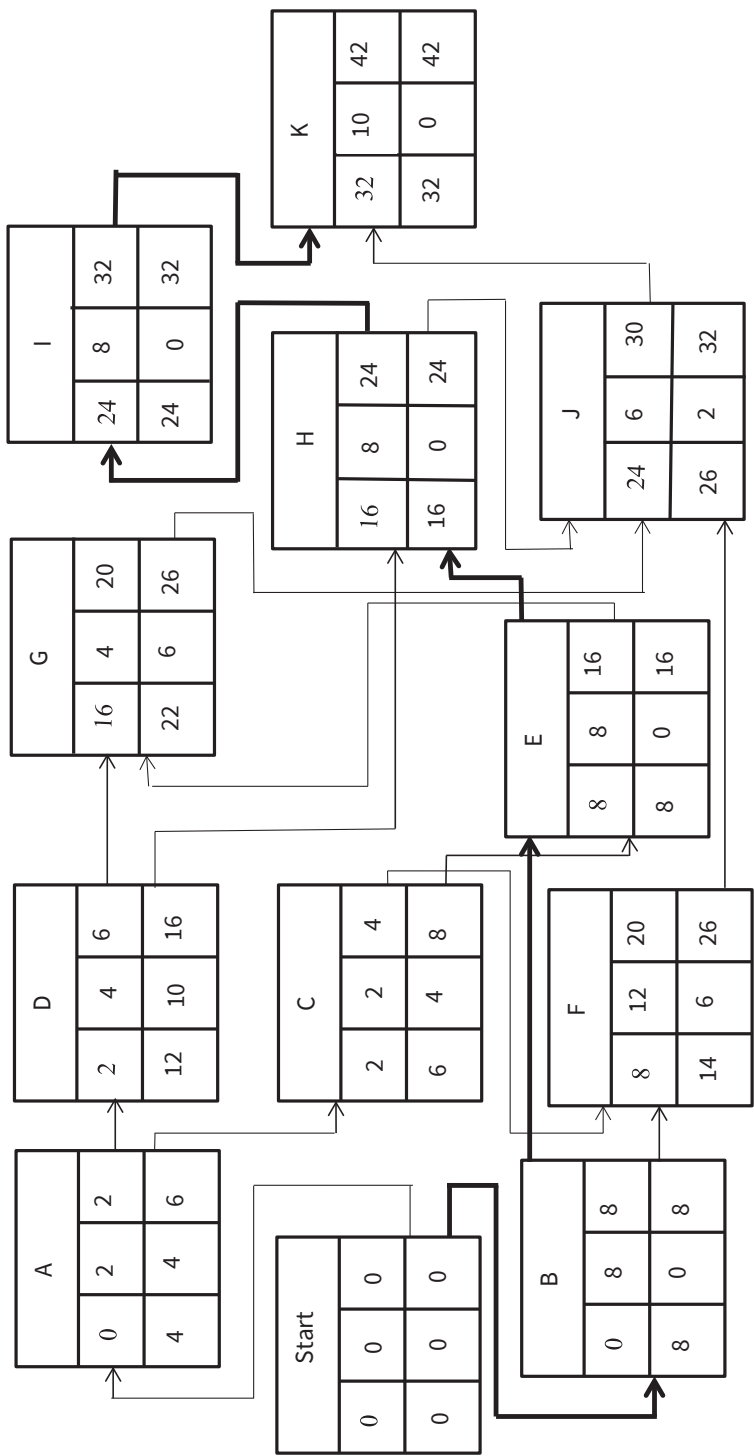
$$t_7 - t_5 \geq 8 \quad (192)$$

$$t_7 - t_6 \geq 6 \quad (193)$$

$$t_8 - t_7 \geq 10 \quad (194)$$

$$t_i \geq 0 \quad (i = 1, 2, \dots, 8) \quad (195)$$

Model rozwiązujemy przy pomocy programu Excel Solver. Wartość funkcji celu $F(t_8) = 42$, co oznacza, że najkrótszy czas realizacji projektu wynosi czterdzieści dwie jednostki czasowe (tygodnie). Pozostałe wyniki przedstawia tabela 67.



Rys. 29. Wykres sieciowy typu AoN przedstawiający analizę CPM
Źródło: opracowanie własne.

Tab. 67. Wyniki rozwiązania modelu analizy czasowej

Czynność (ograniczenie)	Zmienna swobodna	Zmienna dualna
A	0	0
B	0	1
C	4	0
D	0	0
E	0	1
F	0	0
G	0	0
H	0	1
I	0	1
J	6	0
K	0	1

Źródło: opracowanie własne.

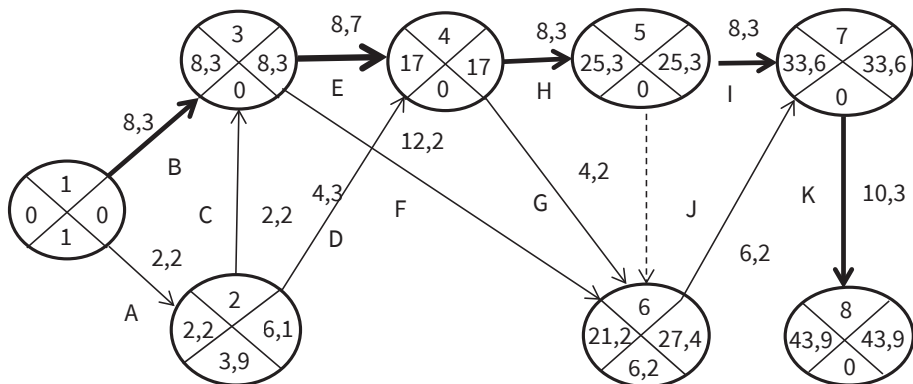
Ścieżkę krytyczną tworzą te czynności, dla których zmienna swobodna przyjmuje wartość równą 0, a zmienna dualna wartość różną od 0. Są to czynności B, E, H, I, J oraz K. Otrzymaliśmy takie samo rozwiązanie, jak w graficznej metodzie poszukiwania ścieżki krytycznej. Zmienne dualne dla czynności krytycznych są równe 1. Oznacza to, że opóźnienie którejkolwiek z czynności krytycznych o jedną jednostkę czasową spowoduje wydłużenie czasu realizacji projektu także o jedną jednostkę czasową. Grafy na rysunku 28 i rysunku 29 przedstawiają rozwiązanie graficzne w konwencji AoA oraz AoN. Długość ścieżki krytycznej wynosi czterdzieści dwie jednostki czasowe.

Zastosowanie techniki PERT do określenia oczekiwanego czasu trwania projektu i jego parametrów – zob. tabela 68 i rysunek 30.

Tab. 68. Dane do analizy PERT

Symbol	<i>i</i>	<i>j</i>	t_{ij}^a	t_{ij}^m	t_{ij}^b	t_{ij}^e	σ_{ij}^2	ZC_{ij}
A	1	2	1	2	4	2,17	0,25	3,9
B	1	3	6	8	12	8,33	1,00	0
C	2	3	1	2	4	2,17	0,25	3,9
D	2	4	3	4	7	4,33	0,44	10,5
E	3	5	4	8	16	8,67	4,00	0
F	3	7	11	12	17	12,17	1,00	6,9
G	5	7	3	4	6	4,17	0,25	6,2
H	5	6	6	8	12	8,33	1,00	0
I	6	8	6	8	12	8,33	1,00	0
J	7	8	5	6	8	6,17	0,25	6,2
K	8	9	8	10	14	10,33	1,00	0

Źródło: opracowanie własne.



Rys. 30. Wykres sieciowy typu AoA przedstawiający analizę PERT

Źródło: opracowanie własne.

Rozkład czasu trwania projektu jest rozkładem asymptotycznie normalnym o wartości oczekiwanej równej momentowi zajścia zdarzenia końcowego i odchyleniu standardowym równym pierwiastkowi kwadratowemu z sumy wariancji czynności krytycznych. Czynności krytyczne to te, których oczekiwane całkowite zapasy czasu są równe luzowi czasowemu zdarzenia końcowego (w tabeli 68 zostały one zacięniowane). Zapasy czasu w opisywanym w tabeli 68 przykładowym projekcie są równe 0 dla czynności krytycznych. Suma wariancji czasów trwania czynności krytycznych wynosi 8 – zatem odchylenie standardowe to w przybliżeniu 2,83. Dla rozpatrywanego projektu czas trwania jest zmienną losową o rozkładzie prawdopodobieństwa:

$$t_8 \sim asN(43,9; 2,83) \quad (196)$$

Wykorzystując parametry tego rozkładu oraz tablice statystyczne, możemy wyznaczyć:

- 1) Prawdopodobieństwo zakończenia projektu w czasie nie dłuższym niż jakiś czas zadany TD , np. prawdopodobieństwo zakończenia projektu w czasie nie dłuższym niż czterdzieści pięć jednostek czasowych. W tym celu standaryzujemy wartość czasu TD :

$$u_{TD} = \frac{TD - t_8^e}{\sigma(t_8)} = \frac{45 - 43,9}{2,83} = 0,39 \quad (197)$$

$$P(t_8 \leq TD) = F(TD) = \Phi(u_{TD}) = 0,6513 \quad (198)$$

Wynik oznacza, że prawdopodobieństwo ukończenia projektu w czasie nie dłuższym niż czterdzieści pięć jednostek czasowych wynosi 65,13%.

- 2) Czas trwania projektu, dla którego prawdopodobieństwo dotrzymania terminu nie przekracza określonej wartości, np. szukamy terminu ukończenia projektu z prawdopodobieństwem przekraczającym 80%:

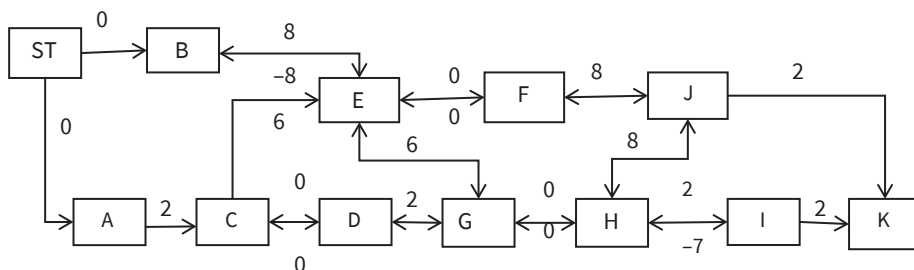
$$P(t_8 < TD) = F(TD) = \Phi(u_{TD}) \geq 0,80 \Rightarrow u_{TD} = 0,84^{10} \quad (199)$$

$$TD = u_{TD} \sigma(t_8) + t_8^e = 0,84 \cdot 2,83 + 43,9 \approx 46,12 \quad (200)$$

Czas trwania projektu, dla którego prawdopodobieństwo dotrzymania terminu jest nie mniejsze niż 80% wynosi, w przybliżeniu, 46,12 jednostek czasowych – rozpatrując wartości całkowite, należy wynik zaokrąglić w górę do wartości czterdziestu siedmiu jednostek czasowych.

Zastosowanie metody MPM do określenia najkrótszego czasu trwania projektu – metoda ta, mimo że należy do metod deterministycznych, różni się znacznie od metody CPM. Główna różnica polega na sposobie konstruowania sieci. Czynności przedstawiane są tu za pomocą węzłów w sieci – dla odróżnienia oznaczane prostokątami. Poza tym, sieć ta uwzględnia także sytuację, w której czynności mogą się zaczynać przed zakończeniem czynności je poprzedzających. Określić trzeba, w którym najwcześniejszym (i najpóźniejszym) momencie czynności poprzedzającej można rozpocząć realizację czynności następującej. Dane te uzyskiwane są na podstawie doświadczeń oraz sugestii ekspertów. W wyniku postępowania uzyskuje się również ścieżkę krytyczną przedsięwzięcia.

Sieć zawierającą powiązania między czynnościami oraz najwcześniejsze i najpóźniejsze terminy rozpoczęcia, zgodnie z założeniami metody, przedstawia rysunek 31.



Rys. 31. Sieć typu AoN zawierająca informacje potrzebne do zastosowania techniki MPM

Źródło opracowanie własne.

10 Do wyznaczenia granicznej wartości u_{TD} odpowiadającej $\Phi(u_{TD}) = 0,80$ można wykorzystać np. funkcję arkusza Excel 2016: =ROZKŁ.NORMALNY.S.ODWR(0,8).

Zgodnie z teoretycznymi zasadami postępowania w metodzie MPM do obliczenia najwcześniejszego i najpóźniejszego momentu rozpoczęcia czynności (w odniesieniu do czynności bezpośrednio poprzedzającej) oraz wyznaczenia ścieżki krytycznej wykorzystuje się macierze pomocnicze między czynnościami – zob. tabele 69–71.

Tab. 69. Macierz obliczeń pomocniczych 1

	ST	A	B	C	D	E	F	G	H	I	J	K	T_i^{0+}	T_i^{0-}	NWP_i
ST	0	0	0										0		0
A				2									0		0
B						8							0	0	0
C					0	6							2		2
D				0				2					2		2
E			-8				0	6					8		8
F						0					8		8		8
G													14		14
H										2	8		14	9	14
I									-7			2	16		16
J												2	22		22
K													24		24

Źródło: opracowanie własne.

Tab. 70. Macierz obliczeń pomocniczych 2

	ST	A	B	C	D	E	F	G	H	I	J	K	t_i	NWP_i	$NWP_i + t_i$
ST	0	0	0										0	0	0
A				2									2	0	2
B						8							8	0	8
C					0	6							2	2	4
D				0				2					4	2	6
E			-8				0	6					8	8	16
F						0					8		12	8	20
G									0				4	14	18
H								0		2	8		8	14	22
I									-7			2	8	16	24
J												2	6	22	28
K													10	24	34
T_i^{0+}	0	0	0	2	12	8	14	14	14	22	22	24	34		
T_i^{0-}						8				21			34		
NPP_i	0	0	0	2	12	8	14	14	14	21	22	24	34		

Źródło: opracowanie własne.

Tab. 71. Obliczenia zapasów czasu w projekcie

Czynność	t_i	NWP_i	NWK_i	NPP_i	NPK_i	$Z_i = NPP_i - NWP_i$	$Z_i = NPK_i - NWK_i$
ST	0	0	0	0	0	0	0
A	2	0	2	0	2	0	0
B	8	0	8	0	8	0	0
C	2	2	4	2	4	0	0
D	4	2	6	12	16	10	10
E	8	8	16	8	16	0	0
F	12	8	20	14	26	6	6
G	4	14	18	14	18	0	0
H	8	14	22	14	22	0	0
I	8	16	24	21	29	5	5
J	6	22	28	22	28	0	0
K	10	24	34	24	34	0	0

Źródło: opracowanie własne.

Czynnościami krytycznymi w projekcie są czynności: A, B, C, E, G, H, J i K. Opóźnienie wykonania którejkolwiek z tych czynności opóźni realizację całości projektu. Czas trwania projektu jest krótszy – wynosi trzydzieści cztery jednostki czasowe – niż ten wyznaczony metodą CPM. Wynika to z faktu, że pojawia się tu możliwość realizacji czynności następujących w trakcie trwania czynności poprzedzających.

Analiza czasowo-kosztowa LESS – w tabeli 72 przedstawione zostały dane konieczne do przeprowadzenia deterministycznej analizy czasowo-kosztowej LESS.

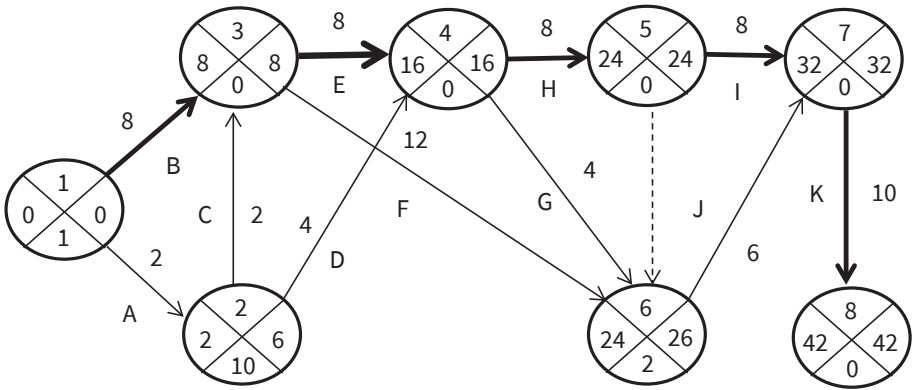
Tab. 72. Dane niezbędne do analizy LESS

Czynność	i	j	t_{ij}^n	t_{ij}^g	K_{ij}^n	K_{ij}^g	s_{ij}	ZC_{ij}
A	1	2	2	1	5	7,5	2,50	4
B	1	3	8	6	12	15	1,50	0
C	2	3	2	1	8	10	2,00	4
D	2	4	4	2	10	13	1,50	10
E	3	4	8	5	3	5	0,67	0
F	3	6	12	8	50	62	3,00	6
G	4	6	4	3	6	9	3,00	6
H	4	5	8	6	40	51	5,50	0
I	5	7	8	5	11	14	1,00	0
J	6	7	6	3	11	15	1,33	2
K	7	8	10	7	10	12	0,67	0

Źródło: opracowanie własne.

Koszty podane w tabeli 72 to koszty bezpośrednie wyrażone w tysiącach złotych, a czas – w tygodniach.

Krok pierwszy – wyznaczamy najkrótszy czas trwania projektu, biorąc pod uwagę normalne czasy trwania poszczególnych czynności.



Rys. 32. Ścieżka krytyczna przy normalnym czasie trwania czynności

Źródło: opracowanie własne.

Dla kosztów pośrednich KP przyjmujemy założenie, że $KP = 0,75t_g + 3,5$, co oznacza, że w przypadku, gdy wszystkie czynności są realizowane w normalnym czasie trwania i $t_g = 42$ (zob. rysunek 32), to wynoszą one trzydzieści pięć tysięcy złotych. Każde skrócenie czasu realizacji projektu o jedną jednostkę czasową powoduje zmniejszenie kosztów pośrednich o siedemset pięćdziesiąt złotych – zob. tabela 73.

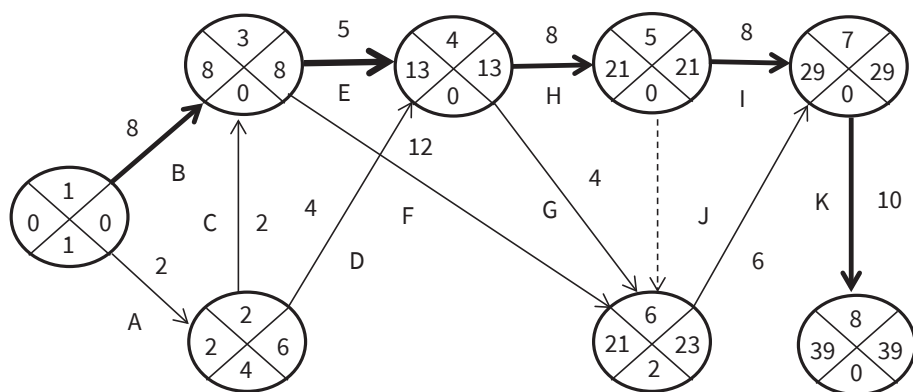
Tab. 73. Koszty realizacji projektu dla czasu jego trwania równego czterdzieści dwa tygodnie

t_g	KB	KP	KC	Czynność skracana	Koszt skrócenia	Ścieżka krytyczna
42	166,00	35,00	201	–	–	B, E, H, I, J

Źródło: opracowanie własne.

Krok drugi – skracamy czas trwania projektu. W tym celu spośród czynności krytycznych wybieramy tę, której odpowiada najmniejszy koszt skrócenia ($min s_{ij}$). Najmniejszy koszt skracania wynosi sześćset siedemdziesiąt złotych i odpowiada czynnościom E i K. Do skracania wybieramy jedną z nich. Niech to będzie czynność E. Skracanie czasu trwania czynności powinno się odbywać w sposób umożliwiający identyfikację nowych ścieżek krytycznych, powstających w czasie

procesu skracania. Jeżeli nie będziemy tak postępowali, to może okazać się, że skrócenie czynności o x jednostek czasowych nie spowoduje skrócenia czasu trwania projektu o taką samą wartość – ponieważ czas trwania projektu będzie blokowany przez nowo powstałą ścieżkę krytyczną, do której nie należała skracana czynność. Dlatego najbezpieczniej jest skracać czynność o jedną jednostkę czasową. Następnie należy sprawdzić, czy nie powstały nowe ścieżki krytyczne, jeżeli tak – to w celu skrócenia czasu trwania projektu należy skrócić czas trwania każdej z nich o taką samą ilość jednostek czasowych. Postępowanie to można przyspieszyć, wiedząc, że w wyniku skracania danej czynności krytycznej zmniejszeniu mogą ulec całkowite zapasy czasu wszystkich tych czynności, które nie należą do ścieżek prowadzących od zdarzenia początkowego do końcowego sieci, a w skład których wchodzi skracana czynność. W analizowanym przykładzie są to czynności: A, C, D i F – ich zapasy czasu całkowite (zob. tabela 71) wynoszą odpowiednio cztery, cztery, dziesięć i sześć tygodni. Wynika z tego, że czynność E można byłoby skrócić maksymalnie o cztery tygodnie. Jednakże, biorąc pod uwagę graniczny i normalny czas jej trwania, należy ją skrócić tylko o trzy tygodnie – zob. rysunek 33 i tabela 74.



Rys. 33. Ścieżka krytyczna po skróceniu czynności E o trzy tygodnie

Źródło: opracowanie własne.

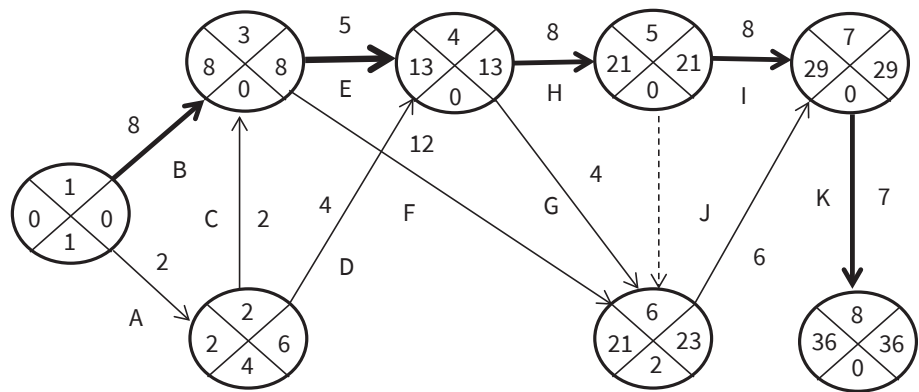
Tab. 74. Koszty realizacji projektu dla czasu jego trwania równego trzydziści dziewięć tygodni

t_s	KB^{11}	KP	KC	Czynność skracana	Koszt skrócenia	Ścieżka krytyczna
42	166,00	35,00	201,00	–	–	B, E, H, I, J
39	168,00	32,75	200,75	E	2,00	B, E, H, I, J

Źródło: opracowanie własne.

11 Koszty bezpośrednie zaokrąglone do liczby całkowitej.

Następnie skracamy czas trwania czynności *K*, która jest ostatnią czynnością w sieci. Oznacza to, że wszystkie czynności niekrytyczne leżą na ścieżkach, do których ona należy. Wynika z tego, że nie ma dodatkowych ograniczeń czasu trwania czynności *K* – oprócz jej czasu granicznego. Skracamy ją o trzy tygodnie – zob. rysunek 34 oraz tabela 75.



Rys. 34. Ścieżka krytyczna po skróceniu czynności *K* o trzy tygodnie

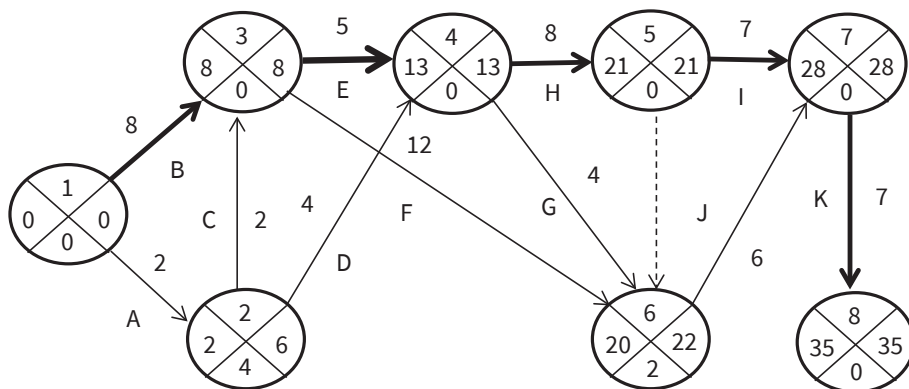
Źródło: opracowanie własne.

Tab. 75. Koszty realizacji projektu dla czasu jego trwania równego trzydzieści sześć tygodni

t_s	<i>KB</i>	<i>KP</i>	<i>KC</i>	Czynność skracana	Koszt skrócenia	Ścieżka krytyczna
42	166,00	35	201,00	–	–	B, E, H, I, J
39	168,00	32,75	200,75	E	2,00	B, E, H, I, J
36	170,00	30,50	200,50	K	2,00	B, E, H, I, J

Źródło: opracowanie własne.

Skracanie czasu trwania projektu, mające na celu określenie terminu jego realizacji charakteryzującego się minimalnymi kosztami całkowitymi, odbywa się do momentu, w którym przyrost kosztów bezpośrednich jest mniejszy od spadku kosztów pośrednich. W analizowanym przykładzie następną czynnością, jaką powinniśmy skrócić, jest czynność *I* – z kosztem skrócenia o jednostkę (tydzień) wynoszącym tysiąc złotych. Zobaczmy, jaki efekt przyniesie skrócenie jej o tydzień – zob. rysunek 35 i tabela 76.

**Rys. 35.** Ścieżka krytyczna po skróceniu czynności *i* o tydzień

Źródło: opracowanie własne.

Tab. 76. Koszty realizacji projektu dla czasu jego trwania równego trzydziści pięć tygodni

t_8	KB	KP	KC	Czynność skracana	Koszt skrócenia	Ścieżka krytyczna
42	166,00	35,00	201,00	–	–	B, E, H, I, J
39	168,00	32,75	200,75	E	2,00	B, E, H, I, J
36	170,00	30,50	200,50	K	2,00	B, E, H, I, J
35	171,00	29,75	200,75	I	1,00	B, E, H, I, J

Źródło: opracowanie własne.

Widzimy, że koszty całkowite przy czasie realizacji projektu równym trzydziści pięć tygodni są większe niż przy trzydziestu sześciu tygodniach. Oznacza to, że dalsze skracanie czasu trwania – zgodne z przyjętą procedurą – powodowało będzie wzrost kosztów całkowitych. Tak więc optymalny czas realizacji analizowanego projektu – z punktu widzenia kosztów całkowitych – wynosi trzydziści sześć tygodni. W porównaniu do czasu realizacji projektu, w sytuacji, gdy wszystkie czynności były realizowane w normalnym czasie trwania, został on skrócony o sześć tygodni, a koszty całkowite zmalały o pięćset złotych.

Liniowy model optymalizacyjny dla przeprowadzenia analizy czasowo-kosztowej – zob. równania (201)–(224) i tabela 77.

$$KC = 313 - 2,5t_{12} - 1,5t_{13} - 2t_{23} - 1,5t_{24} - 0,67t_{34} - 3t_{36} - 3t_{46} - 5,5t_{45} - t_{57} - 1,33t_{67} - 0,67t_{78} + 0,75t_8 + 3,5 \rightarrow \min \quad (201)$$

$$t_2 - t_1 - t_{12} \geq 0 \quad (202)$$

$$1 \leq t_{12} \leq 2 \quad (203)$$

$$t_3 - t_1 - t_{13} \geq 0 \quad (204)$$

$$6 \leq t_{13} \leq 8 \quad (205)$$

$$t_3 - t_2 - t_{23} \geq 0 \quad (206)$$

$$1 \leq t_{23} \leq 2 \quad (207)$$

$$t_4 - t_2 - t_{24} \geq 0 \quad (208)$$

$$2 \leq t_{24} \leq 4 \quad (209)$$

$$t_4 - t_3 - t_{34} \geq 0 \quad (210)$$

$$5 \leq t_{34} \leq 8 \quad (211)$$

$$t_6 - t_3 - t_{36} \geq 0 \quad (212)$$

$$8 \leq t_{36} \leq 12 \quad (213)$$

$$t_6 - t_4 - t_{46} \geq 0 \quad (214)$$

$$3 \leq t_{46} \leq 4 \quad (215)$$

$$t_5 - t_4 - t_{45} \geq 0 \quad (216)$$

$$6 \leq t_{45} \leq 8 \quad (217)$$

$$t_7 - t_5 - t_{57} \geq 0 \quad (218)$$

$$5 \leq t_{57} \leq 8 \quad (219)$$

$$t_7 - t_6 - t_{67} \geq 0 \quad (220)$$

$$3 \leq t_{67} \leq 6 \quad (221)$$

$$t_8 - t_7 - t_{78} \geq 0 \quad (222)$$

$$7 \leq t_{78} \leq 10 \quad (223)$$

$$t_i \geq 0 \quad (i = 1, \dots, 8) \quad (224)$$

Tab. 77. Rozwiązanie optymalne zadania (201)–(224)

Czynność	Czas trwania	Zmienna swobodna	Zmienna dualna
A	2	4	0
B	8	0	0,75
C	2	0	0
D	4	0	0
E	5	0	0,75
F	12	3	0
G	4	6	0
H	8	0	0,75
I	8	0	0,75
J	6	0	0
K	7	0	0,75

Źródło: opracowanie własne.

Model został rozwiązany za pomocą programu Excel Solver. Optymalny czas realizacji projektu, odpowiadający minimalizacji kosztów całkowitych, wyniósł $t_8 = 36$. Wartość stałej w funkcji celu (201) wynosi $(\sum_{(i,j) \in G} K_{ij}^n + s_{ij} t_{ij}^n) + b = 313 + 3,5 = 316,5$. Suma pozostałych składowych (201) dla rozwiązania optymalnego równa się 116. Zatem wartość kosztów całkowitych odpowiadających rozwiązaniu optymalnemu wynosi $316,5 - 116 = 200,5$ [tys. PLN]. Porównując je z kosztami całkowitymi, w przypadku, gdy wszystkie czynności wykonywane są w czasie normalnym, obserwujemy, że:

$$\Delta KC_{ij} = 200,5 - 201 = -0,50 \text{ [tys. PLN]}$$

Zmienne dualne dla czynności krytycznych są równe współczynnikowi kierunkowemu funkcji kosztów pośrednich. Oznacza to, że wydłużenie którejkolwiek z nich o tydzień spowoduje wzrost kosztów całkowitych o siedemset pięćdziesiąt złotych.

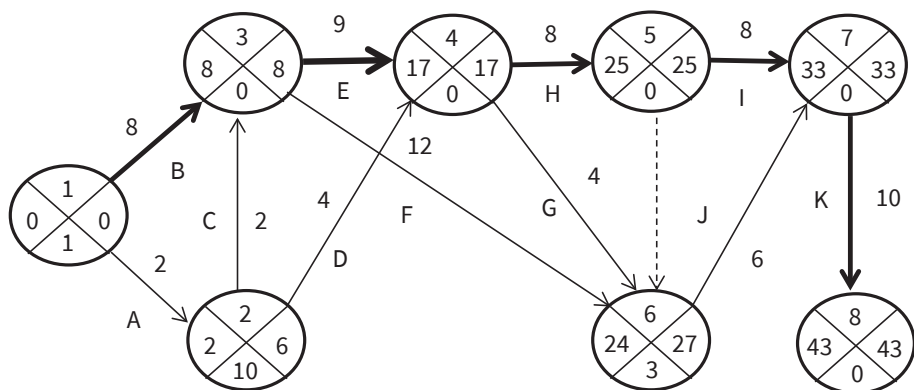
Zastosowanie metody PERT-COST – metodę PERT-COST stosujemy do obliczania optymalnego, z punktu widzenia kosztów całkowitych (minimalizacja), programu przyspieszania czynności i wyznaczania najkrótszego czasu realizacji projektu. W tym przypadku nie bierzemy pod uwagę kosztów pośrednich – zob. tabela 78.

Tab. 78. Dane niezbędne do analizy PERT-COST

Symbol	t_{ij}^a	t_{ij}^b	t_{ij}^{en}	t_{ij}^{eg}	K_{ij}^n	K_{ij}^g	s_{ij}	r_1	r_2	σ_{ij}^2	ZC_{ij}
A	1	4	2	1	5	7,5	2,5	0,50	2,00	0,25	4
B	6	12	8	6	12	15	1,5	0,75	1,50	1,00	0
C	1	4	2	1	8	10	2,0	0,50	2,00	0,25	4
D	3	7	4	3	10	13	3,0	0,75	1,75	0,44	11
E	4	16	9	4	3	5	1,0	0,44	1,78	4,00	0
F	11	17	12	11	50	62	12,0	0,92	1,42	1,00	7
G	3	6	4	3	6	9	3,0	0,75	1,50	0,25	6
H	6	12	8	6	40	51	5,5	0,75	1,50	1,00	0
I	6	12	8	6	11	14	1,5	0,75	1,50	1,00	0
J	5	8	6	5	11	15	4,0	0,83	1,33	0,25	3
K	8	14	10	8	10	12	1,0	0,80	1,40	1,00	0

Źródło: opracowanie własne.

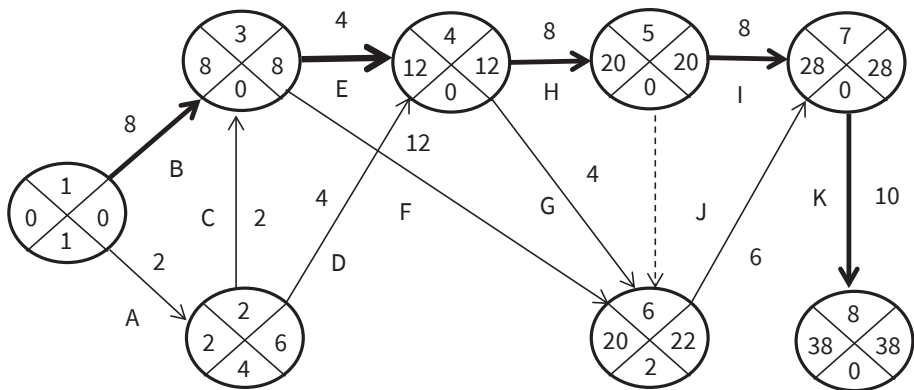
Krok zerowy – zob. rysunek 36.



Rys. 36. Ścieżka krytyczna przy normalnym oczekiwanym czasie trwania czynności

Źródło: opracowanie własne.

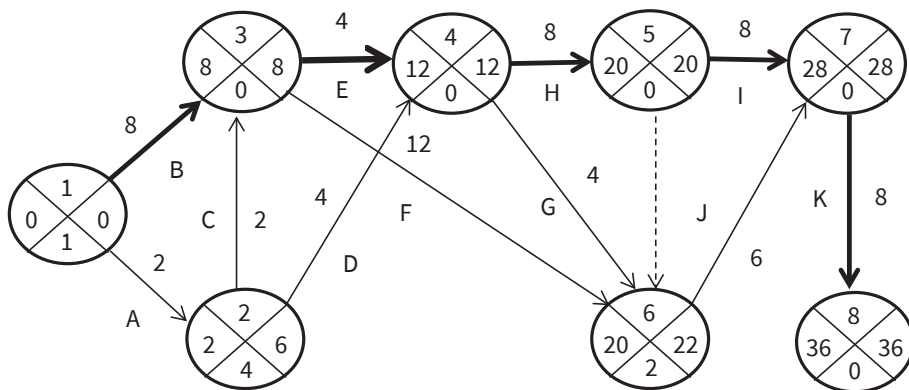
Krok pierwszy – skracamy czynność E o pięć jednostek czasowych (zob. rysunek 37).



Rys. 37. Ścieżka krytyczna po skróceniu czynności E

Źródło: opracowanie własne.

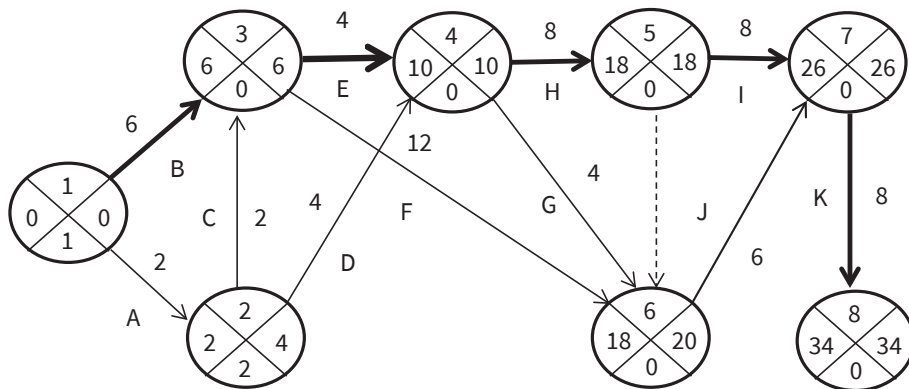
Krok drugi – skracamy czynność *K* o dwie jednostki czasowe (zob. rysunek 38).



Rys. 38. Ścieżka krytyczna po skróceniu czynności *K*

Źródło: opracowanie własne.

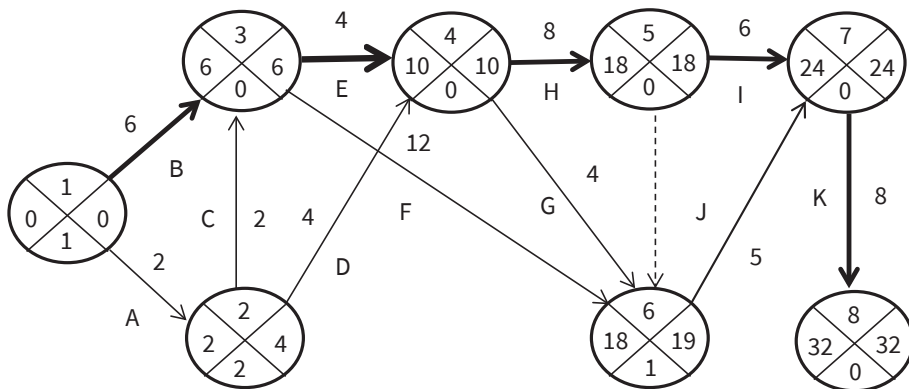
Krok trzeci – skracamy czynność *B* o dwie jednostki czasowe (zob. rysunek 39).



Rys. 39. Ścieżka krytyczna po skróceniu czynności *B*

Źródło: opracowanie własne.

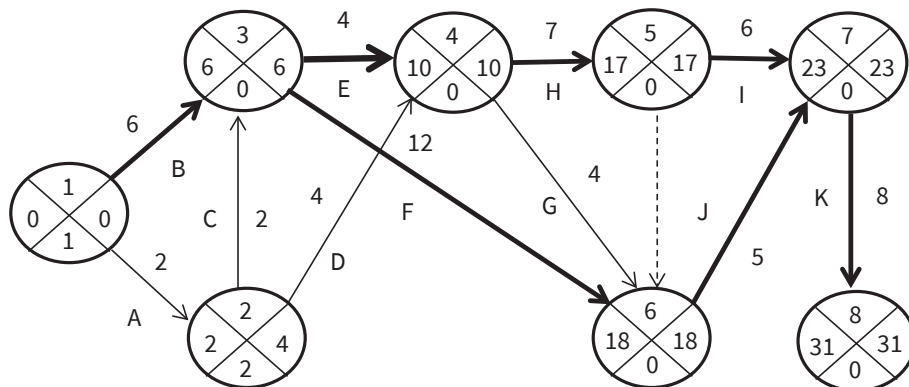
Krok czwarty – skracamy czynność I o dwie jednostki czasowe (zob. rysunek 40).



Rys. 40. Ścieżka krytyczna po skróceniu czynności I

Źródło: opracowanie własne.

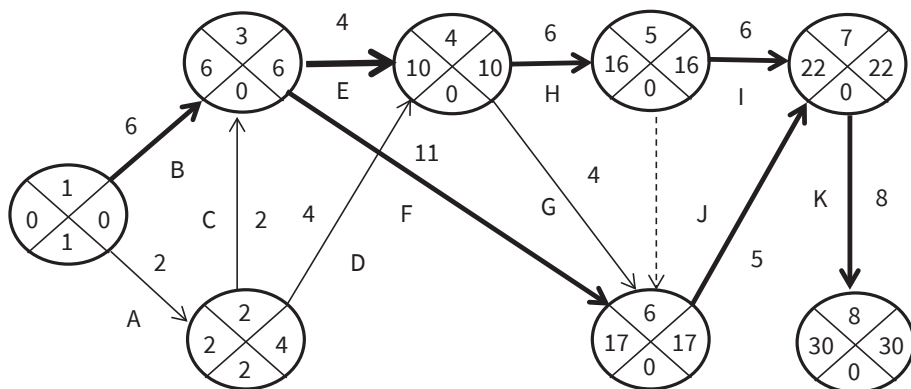
Krok piąty – skracamy czynność H o jedną jednostkę czasową (zob. rysunek 41).



Rys. 41. Ścieżka krytyczna po skróceniu czynności H

Źródło: opracowanie własne.

Krok szósty – skracamy czynność H i F po jednej jednostce czasowej (zob. rysunek 42).



Rys. 42. Ścieżka krytyczna po skróceniu czynności H i F

Źródło: opracowanie własne.

Dalsze skracanie nie ma sensu, ponieważ wszystkie czynności leżące na ścieżkach krytycznych osiągnęły graniczne czasy trwania. Można zauważyć, że w miarę skracania czasu trwania projektu maleje odchylenie standardowe czasu trwania projektu.

Następnie obliczymy wariancje czasu i koszty realizacji poszczególnych realizacji projektu – zob. tabela 79.

Tab. 79. Oczekiwane czasy graniczne, wariancje czasu i koszty wyspecyfikowanych programów realizacji projektu

Program	t_s^{eg}	Przyrost kosztów całkowitych			$\sum_{(i,j) \in C} \sigma_{ij}^2$	σ_{t_s}
		ΔK	$\Sigma \Delta K$	$KC(t_{gr})$		
0	43	0	0	166	8,04	2,84
1	38	5	5	171	4,8	2,19
2	36	2	7	173	4,44	2,11
3	34	3	10	176	4,00	2,00
4	32	3	13	179	3,56	1,89
5	31	5,5	18,5	184,5	3,33	1,82
6	30	17,5	36	220,5	3,12	1,77

Źródło: opracowanie własne na podstawie Bładowski, 1970, s. 202.

Oznaczmy przez t_n^{max} maksymalny dopuszczalny najkrótszy termin zakończenia projektu, a przez r – poziom ryzyka rozumiany jako prawdopodobieństwo, że rzeczywisty czas trwania projektu przekroczy tę wartość, tj. $P(t_n > t_n^{max}) = r$. Wzór (225) jest pomocny do obliczania terminów realizacji projektu obciążonych założonym poziomem ryzyka r :

$$t_n^{max} = t_n^{eg} + u\sigma_{t_n} \quad (225)$$

gdzie u jest wartością zmiennej losowej $U \sim N(0, 1)$, dla której wartość dystrybucyjności wynosi: $\Phi(u) = 1 - r$.

Biorąc pod uwagę programy skracania oczekiwanego czasu trwania projektu do różnych poziomów t_n^{eg} oraz oszacowane wartości odchyłeń standardowych czasu trwania odpowiadające tym programom σ_{t_n} , możemy zadać pytanie: jaka jest dopuszczalna maksymalna wartość najkrótszego czasu trwania projektu, w przypadku różnych programów, związana z założonym poziomem ryzyka? Dla przykładowego projektu wartości te zawarte są w tabeli 80. Poziomy ryzyka przedstawione zostały w skali procentowej. Wraz ze wzrostem akceptowalnego poziomu ryzyka, terminy te są coraz krótsze. Zmniejszanie się wartości wskaźnika r , oznaczające spadek skłonności do ryzyka, wydłuża dopuszczalny termin oczekiwanego zakończenia projektu. Z kolei skracanie oczekiwanego czasu trwania projektu związane jest ze skracaniem się maksymalnych dopuszczalnych wartości najkrótszego czasu trwania projektu oraz ze zmniejszaniem odchylenia standardowego σ_{t_n} .

Tab. 80. Maksymalne dopuszczalne terminy realizacji projektu zgodnie z wyspecyfikowanymi programami, w zależności od poziomu podejmowanego ryzyka

Program	t_8^{eg}	σ_{t_8}	Poziom ryzyka r				
			10%	30%	50%	60%	90%
0	43	2,84	46,46	44,48	43	42,29	39,36
1	38	2,19	40,67	39,14	38	37,47	35,20
2	36	2,11	38,57	37,10	36	35,47	33,30
3	34	2,00	36,44	35,04	34	33,50	31,44
4	32	1,89	34,41	32,98	32	31,53	29,58
5	31	1,82	33,22	31,95	31	30,55	28,67
6	30	1,77	32,16	30,92	30	29,56	27,73

Źródło: opracowanie własne na podstawie Bładowski, 1970, s. 203.

Porównując terminy w tabeli 80 dla różnych poziomów ryzyka, np. z terminem dyrektywnym, który wynosi $t_8^{eg} = 36$ jednostek czasowych, możemy dokonać wyboru programu skracania, odpowiadającego temu terminowi oraz założonemu

poziomowi ryzyka. Ponadto, dokonując wyboru, musimy wziąć pod uwagę, że wraz ze skracaniem projektu w czasie rosną koszty całkowite jego realizacji. Dlatego, poszukując w kolumnach tabeli 80 programu, dla którego termin $t_n^{max} \leq 36$, trzeba zastosować program, któremu odpowiada największa z wartości spełniających tę nierówność. Do wyboru mamy następujące programy:

- Program 4 – z ryzykiem 10% (koszt realizacji 179 j.p.¹²),
- Program 3 – z ryzykiem 30% (koszt realizacji 176 j.p.),
- Program 2 – z ryzykiem 60% (koszt realizacji 173 j.p.),
- Program 1 – z ryzykiem 90% (koszt realizacji 171 j.p.).

Oznacza to, że im większe jest prawdopodobieństwo dotrzymania terminu dyrektywnego, tym większe są przewidywane koszty (Bładowski, 1970, s. 204).

12 j.p. – jednostki pieniężne.

Rozdział IV

Teoria masowej obsługi

1. Teoretyczne aspekty systemów kolejkowych

Teoria kolejek – w polskiej literaturze często używa się terminu „teoria masowej obsługi” (por. Koźniewska i Włodarczyk, 1978; Kozubski, 2000) – opisuje zjawiska spotykane w zaskakująco wielu dziedzinach życia i działalności gospodarczej. Przeciętnemu człowiekowi pojęcie „kolejki” kojarzy się ze sklepami samoobsługowymi, przychodniami, pocztą itd., czyli codziennymi czynnościami, w których grupa osób w tym samym czasie oczekuje przed jakimś stanowiskiem obsługi. Teoria kolejek ma jednak mnóstwo zastosowań także w innych dziedzinach: informatyce (obsługa poleceń przez procesor, zapytań skierowanych do strony internetowej), telekomunikacji (połączenia telefoniczne), logistyce (obsługa klientów, zarządzanie dostępem do zapasów w magazynie), a nawet w strukturach administracyjnych (Filipowicz i Kwiecień, 2008; Nogalski, Czerska i Klimek, 2010).

Aspekt logistyczny, który interesuje nas najbardziej, wiąże się z dynamicznymi zmianami gospodarczymi ostatnich lat. Przenoszenie działalności do sieci powoduje, że bardzo ważną staje się sprawna obsługa klientów. Za pośrednictwem Internetu uzyskują oni łatwy dostęp do oferty, w dogodnym dla siebie czasie, i oczekują, że złożone zamówienia zostaną zrealizowane równie szybko i sprawnie. Obsługa klienta staje się zatem ważnym narzędziem w walce konkurencyjnej, przyczyniając się do podjęcia ostatecznej decyzji, np. przy wyborze dostawcy. A to tylko jeden z aspektów wykorzystania teorii masowej obsługi w logistyce. Są też inne: oczekujące na realizację zlecenia produkcyjne, zamówienia na magazynowane części składowane, pojazdy, które mają zostać załadowane lub rozładowane (Matuszak, Bundz i Homik, 2014; Jacyna i Żak, 2012) i wiele innych. Teoria masowej obsługi pomaga określić potencjał stanowiska pracy, poziom jego efektywności oraz pokazuje, jakie działania należy wdrożyć, aby zwiększyć satysfakcję uczestników systemu obsługi.

Jak wspomnieliśmy w poprzednich rozdziałach, w literaturze przedmiotu (Jędrzejczyk, Kukuła, Skrzypek i Walkosz, 2016, s. 208) jako początek badań nad

systemami kolejkowymi podaje się II wojnę światową i problem rozdysponowania dużej liczby samolotów bombowych, wracających z akcji bojowych, pomiędzy ograniczoną liczbę lotnisk. Nie jest to jednak prawda. Pierwsze analizy naukowe są o kilkadziesiąt lat wcześniejsze i dotyczą prac duńskiego matematyka Agnera K. Erlanga. W 1909 roku podjął się on opisanie procesu powstawania kolejki w trakcie nawiązywania połączeń przechodzących przez centralę telefoniczną w Kopenhadze. Udowodnił, że pojawiające się w sposób losowy połączenia telefoniczne podlegają rozkładowi Poissona, który wyraża prawdopodobieństwo wystąpienia niezależnych, następujących po sobie zdarzeń z daną częstotliwością α – liczbą zdarzeń na jednostkę czasową. W publikacji z 1917 roku przedstawił wzory na prawdopodobieństwo blokady połączenia w centralach telefonicznych, znane potem jako model Erlanga.

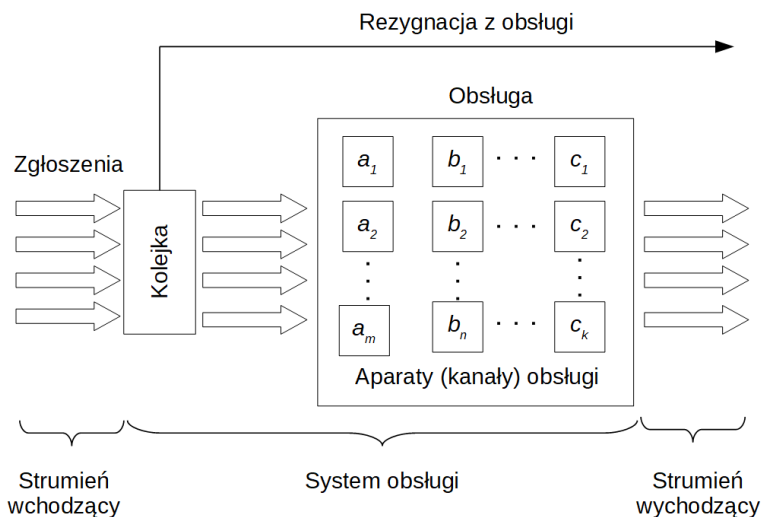
Jak zatem powstaje kolejka? Zgłoszenia wymagające obsługi (np. klienci, którzy skierowali się do kasy) pojawiają się w pewnych odstępach czasu w systemie obsługi (np. przy kasie). Jeśli nie zostają natychmiast obsłużone, dołączają do kolejki. Co pewien czas przebywające w kolejce zgłoszenie jest wybierane do obsługi według określonej reguły (dyscypliny kolejki). Obsługę dokonuje mechanizm zwany serwerem. Przyjmuje się, że każdy serwer może obsługiwać tylko jedno zgłoszenie na raz. Jeśli wszystkie serwery są zajęte, zgłoszenie (klient) czeka w kolejce na obsługę lub rezygnuje z obsługi (systemy bez oczekiwania). Po zakończeniu lub po rezygnacji z obsługi zgłoszenie opuszcza system. Podczas obsługi klient może przejść przez jeden lub więcej etapów, zanim odejdzie z systemu.

Proces masowej obsługi składa się więc z następujących elementów (por. schemat na rysunku 43):

1. Strumień wchodzący (strumień wejściowy, strumień zgłoszeń) – zgłoszenia nadchodzące do systemu.
2. System obsługi (kanały obsługi, aparaty obsługi) – zbiór urządzeń lub stanowisk świadczących obsługę zgłoszenia wraz z kolejką zgłoszeń oczekujących na obsługę.
3. Strumień wychodzący (strumień wyjściowy) – zbiór zgłoszeń po obsłużeniu oraz zbiór zgłoszeń, które zrezygnowały z obsługi.

Kanały obsługi mogą być ułożone w sposób szeregowy, równoległy lub (jak na rysunku 43) mieszany. Nie każde zgłoszenie musi czekać w kolejce aż zostanie obsłużone. Klient w kolejce do kasy ma prawo zniecierpliwić się i opuścić ją. Zachowanie się zgłoszenia (klienta) pozwala więc wyróżnić:

- systemy ze stratami – zgłoszenie opuszcza po upływie pewnego czasu system, rezygnując z obsługi (gdy wszystkie stanowiska obsługi są zajęte obiekt opuszcza system nieobsłużony);
- systemy bez strat – zgłoszenie w systemie przebywa do czasu obsłużenia (obiekt oczekuje w kolejce na obsługę);
- system mieszany – występują w nim warunki pośrednie w stosunku do poprzednich systemów, np. czas przebywania w systemie jest ograniczony lub liczba miejsc w kolejce okazuje się ograniczona.



Rys. 43. Schemat procesu masowej obsługi

Źródło: opracowanie własne na podstawie Miszczyński i Miszczyńska, 2002.

W systemach mieszanych występują pewne szczególne wzorce zachowań. Po pierwsze zdarza się, że klient dołącza do kolejki, lecz po odczekaniu określonego czasu traci cierpliwość i opuszcza ją. Po drugie kilku klientów pozostaje ze sobą w zмовie i tylko jeden z nich czeka w kolejce. Długość kolejki pozornie będzie niewielka, ale czas obsługi pojedynczego klienta się wydłuża. Wtedy inni oczekujący tracą cierpliwość – co prowadzi do następnego typu zachowania. Klient w jednej z kolejek po zobaczeniu innej, krótszej, mając nadzieję na rychlejsze skorzystanie z usługi, może opuścić obecną i dołączyć do drugiej. Problem ten analizowany jest w literaturze przedmiotu (Engel i Hassin, 2017) jako sytuacja, gdy klient wybiera pomiędzy dwoma typami kolejek.

Z punktu widzenia rozmiarów kolejki spotykamy systemy z kolejką ograniczoną lub nieograniczoną oraz systemy z kolejką zabronioną lub niezabronioną. Innym kryterium klasyfikacji jest liczba źródeł. Rozróżniamy dwa typy takich systemów: otwarte i zamknięte. Te pierwsze mogą mieć nieograniczenie wielką liczbę zgłoszeń przychodzących do systemu. W przypadku systemów zamkniętych maksymalna liczba zgłoszeń do obsługi jest z góry ograniczona i stała w czasie.

Wspomnieliśmy, że zgłoszenia znajdujące się w kolejce podlegają pewnej dyscyplinie obsługi. W zależności od sposobu organizacji wyodrębnimy dyscyplinę typu:

- FIFO (ang. *First In First Out*) – zgłoszenie stojące na pierwszym miejscu w kolejce jest obsługiwane jako pierwsze (kolejka naturalna);
- LIFO (ang. *Last In First Out*) – zgłoszenie stojące na ostatnim miejscu w kolejce jest obsługiwane jako pierwsze;

- SIRO (ang. *Selection In Random Order*) – losowy dobór zgłoszenia do obsługi;
- obsługa z priorytetem – pierwszeństwo mają zgłoszenia uprzywilejowane.

W ostatnim z wymienionych przypadków obiekty mają przypisane priorytety, które decydują o kolejności wyboru do obsługi. Dzielią się one na klasy – klasa pierwsza ma najwyższy priorytet obsługi. Pojawienie się zgłoszenia priorytetowego wpływa na aktualnie trwającą obsługę. W systemie bez przerywanej obsługi (ze względnym priorytetem) nowe zgłoszenie o wyższym priorytecie nie powoduje przerywania obsługi zadania z niższym priorytetem, tylko czeka aż jego obsługa się zakończy. Z kolei system z przerywaną obsługą (bezwzględny priorytet) charakteryzuje się tym, że nowe zgłoszenie o wyższym priorytecie powoduje przerywanie obsługi zadania z niższym priorytetem, które wraca do kolejki.

Kolejki opisywane są szeregiem miar statystycznych, dających się podzielić na dwie podstawowe grupy: opisujące zgłoszenia i obsługę. Charakterystyki liczbowe systemów masowej obsługi zależą w dużej mierze od dwóch właściwości statystycznych, a mianowicie: rozkładu prawdopodobieństwa odstępów czasu zgłoszeń i rozkładu prawdopodobieństwa czasu obsługi. W realnych zastosowaniach rozkłady te mogą przybierać prawie dowolną formę. Jedynym ograniczeniem jest to, że nie przyjmują one wartości ujemnych.

Stopę zgłoszeń (przybyć) rozumiemy jako liczbę zgłoszeń napływających do systemu obsługi w jednostce czasu (średnio λ). Najczęściej zakłada się, że stopa ta jest stała dla wszystkich zgłoszeń. Ich liczba zależy w dużej mierze od wielkości populacji, która może być skończona lub nie. Intensywność z kolei wyraża odstęp czasu pomiędzy kolejnymi zgłoszeniami (średnio $1/\lambda$). Większość modeli zakłada, że pewne rozproszenie czasu przybycia dotyczy wszystkich zgłoszeń.

Stopa obsługi to liczba zgłoszeń obsługiwanych w ustalonej jednostce czasu (średnio μ). Najczęściej zakłada się, że stopa obsługi jest stała dla wszystkich zgłoszeń. Intensywność obsługi z kolei to czas obsługi zgłoszenia przez jeden z równoległych kanałów obsługi (średnio $1/\mu$). Stopę zgłoszeń i stopę obsługi łączy zależność zwana stałą Erlanga lub intensywnością ruchu. Jest to stosunek średniej liczby zgłoszeń, jaka napływa do systemu w jednostce czasu, do średniej liczby zgłoszeń, jaka może być obsłużona w tej samej jednostce czasu. Zakładając, że w systemie funkcjonuje s kanałów obsługi, stałą Erlanga opisuje wzór (226):

$$\rho = \frac{\lambda}{s\mu} \quad (226)$$

Możliwe są trzy sytuacje:

1. Jeżeli $\rho < 1$ – układ jest stabilny, prawdopodobieństwo tego, że kolejka ma określoną długość jest stałe w każdej jednostce czasu.
2. Jeżeli $\rho \geq 1$ – układ jest niestabilny, prawdopodobieństwo długiej kolejki rośnie.
3. Jeżeli obie stopy są stałe, można uznać, że układ zmierza do stanu równowagi, tzn. kolejka ma stałą, określoną długość w każdej jednostce czasu.

W tym miejscu warto zastanowić się, co rozumiemy przez stabilność układu. W ramach codziennych aktywności nierzadko obserwujemy sytuacje, w których pojawiają się klienci chcący skorzystać z usług świadczonych przez pewien system obsługi. Byłoby idealnie, gdyby zgłoszenie natychmiast podlegało obsłudze. Praktyka pokazuje jednakże coś innego. Kiedy liczba zgłoszeń przekracza możliwości ich obsługi, pojawia się kolejka, którą traktujemy jako moment utraty stabilności przez system. Dopóki jednak długość kolejki utrzymuje się w rozsądnych granicach, czyli jej długość nie rośnie gwałtownie (stan systemu nie zależy od początkowej długości kolejki), system nadal możemy traktować jako stabilny. Dzieje się tak właśnie wtedy, gdy stopa obsługi przewyższa stopę zgłoszeń. W odwrotnej sytuacji serwery (np. kasjerki w markecie) nie nadążają z obsługą zgłoszeń. Tych szybko przybywa, a kolejka rośnie i rośnie, teoretycznie nawet w nieskończoność. Doprowadzi to do rezygnacji z obsługi i opuszczenia systemu przynajmniej przez część klientów. Nierównowagi między klientem a placówką usługową, czyli zatoru, nie da się całkowicie wyeliminować, ale można rozwinąć wysiłki/techniki w celu zmniejszenia natężenia zatłoczenia lub czasu oczekiwania na kolejne przybycia do systemu. Najprostsze i najbardziej oczywiste rozwiązanie – zwiększanie liczby serwerów – zmniejsza zatory, ale powoduje wydłużenie czasu bezczynności serwera oczekującego na zgłoszenie, co generuje koszty. Zatem zarówno sytuacja, w której pojawia się nieuzasadniona, długa kolejka, jak i otwieranie dodatkowych serwerów bywają nieopłacalne dla osób zarządzających systemem. Dlatego stała Erlanga jest zawsze pierwszą charakterystyką liczbową obliczaną dla systemu kolejkowego.

Kolejną ważną charakterystyką systemów kolejkowych jest tzw. prawo Little'a (Little, 1961), mówiące, że średnia liczba zgłoszeń w systemie jest równa iloczynowi średniego czasu przebywania w systemie oraz średniego tempa ich przybywania. Prawo to pozwala na wyprowadzenie następujących formuł (227)–(229):

$$N = \lambda R \quad (227)$$

$$Q = \lambda W \quad (228)$$

$$R = W + \frac{1}{\mu} \quad (229)$$

gdzie:

- N – oczekiwana liczba zgłoszeń w systemie,
- R – oczekiwany czas pobytu w systemie,
- Q – oczekiwana długość kolejki.

Zgłoszenia w systemie oraz ich obsługa są procesami losowymi i opisują je określone rozkłady zmiennych losowych. Zmienna losowa charakteryzuje albo czas upływający między dwoma kolejnymi zgłoszeniami, albo czas obsługi jedne-

go zgłoszenia. Odstępy czasu pomiędzy zgłoszeniami mogą być (i zazwyczaj są) różne, ale dla uproszczenia rozważań zakłada się ich jednakowość. Podobnie czyni się w przypadku czasu obsługi pojedynczego zgłoszenia. Jeśli wymagana obsługa jest zasadniczo identyczna dla każdego zgłoszenia, a serwer zawsze wykonuje tę samą sekwencję czynności, wówczas rzeczywiste czasy obsługi są zbliżone do oczekiwanego czasu obsługi. Mogą wystąpić niewielkie odchylenia od średniej, ale zwykle odpowiadają za nie drobne różnice w wydajności serwera.

Schemat pojawiania się zgłoszeń i czas ich obsługi zależą od wielu czynników, z których przynajmniej część pozostaje poza kontrolą decydenta. Zjawisko kolejki wynika bezpośrednio z losowości w działaniu aparatu obsługi i losowego schematu przybywania zgłoszeń. Moment pojawienia się zgłoszenia nie jest z góry ustalony, a czas potrzebny na obsłużenie każdego zgłoszenia zależy od wielkości wymaganej usługi.

Na bazie dotychczasowych rozważań okazuje się jasne, że analizy systemów masowej obsługi są silnie związane z wpływem czasu między pojawianiem się kolejnych zgłoszeń i czasem obsługi samego zlecenia. Ponadto są to analizy o charakterze statystycznym, wykorzystujące rozkłady zmiennych losowych. David G. Kendall (Kendall, 1951) zaproponował ujednoliconą notację modeli masowej obsługi, która później podlegała dalszym modyfikacjom. Zgodnie z nią każdy model tego typu opisuje się następującym zestawem parametrów – zob. równanie (230):

$$f(\tau_1)/g(\tau_2)/s/P/L/K \quad (230)$$

gdzie:

- $f(\tau_1)$ – rozkład strumienia zgłoszeń (τ_1 – czas upływający między dwoma kolejnymi zgłoszeniami),
- $g(\tau_2)$ – rozkład czasu obsługi (τ_2 – czas obsługi jednego zgłoszenia),
- s – liczba równoległych kanałów obsługi,
- P – liczebność obsługiwanej populacji,
- L – maksymalna długość kolejki,
- K – kolejność obsługi (dyscyplina kolejki).

Jeżeli w systemie obowiązuje dyscyplina kolejki typu FIFO (tzw. kolejka naturalna), wtedy w zapisie danym wzorem (230) pomija się ostatni parametr. Z kolei zmienne losowe stanowiące dwa pierwsze parametry mogą posiadać jeden z poniższych rozkładów:

- D – proces nielosowy (deterministyczny),
- M – rozkład wykładniczy lub Poissona,
- E_n – rozkład Erlanga n -tego rzędu,
- N – rozkład normalny,
- G – dowolny rozkład czasu obsługi.

Powyższa lista zawiera najczęściej wykorzystywane w praktyce rozkłady, ale oczywiście nie wyczerpuje wszystkich możliwości. Spośród wymienionych za niezwykle użyteczne uważa się rozkład wykładniczy Poissona. Z jednej strony dobrze opisuje on zachowanie systemów masowej obsługi, a z drugiej – jego prostota ułatwia zapis matematyczny. Odstęp czasu (t) pomiędzy dwoma kolejnymi zgłoszeniami, który jest zmienną losową o tzw. ujemnym rozkładzie wykładniczym, opisuje następująca funkcja gęstości:

$$f(t) = \lambda e^{-\lambda t}, \quad \text{dla } t \geq 0 \quad (231)$$

Wartość oczekiwana i wariancja tego rozkładu wynoszą:

$$E(t) = \frac{1}{\lambda}, \quad D^2(t) = \left(\frac{1}{\lambda}\right)^2 \quad (232)$$

Liczba zgłoszeń (n), pojawiająca się w systemie w jednostce czasu o długości (T), jest wtedy zmienną losową o rozkładzie Poissona:

$$P(n=k) = \frac{(\lambda T)^k e^{-\lambda T}}{k!}, \quad \text{dla } k = 0, 1, 2, \dots \quad (233)$$

z wartością oczekiwaną i wariancją równymi odpowiednio:

$$E(n) = D^2(n) = \lambda T \quad (234)$$

Czas (t) obsługi zgłoszenia okazuje się zmienną losową o ujemnym rozkładzie wykładniczym z poniższą funkcją gęstości:

$$f(t) = \mu e^{-\mu t}, \quad \text{dla } t \geq 0 \quad (235)$$

z wartością oczekiwaną i wariancją równymi odpowiednio:

$$E(t) = \frac{1}{\mu}, \quad D^2(t) = \left(\frac{1}{\mu}\right)^2 \quad (236)$$

Przydatność rozkładu wykładniczego wynika z jego właściwości. Przykładowo, funkcja gęstości jest ściśle malejącą funkcją t . W związku z tym zachodzi:

$$P(0 \leq T \leq \Delta t) > P(t \leq T \leq t + \Delta t) \quad (237)$$

Dlatego jest nie tylko możliwe, ale również stosunkowo prawdopodobne, że T przyjmie małą wartość bliską 0. Jeśli T reprezentuje odstęp czasu między przybyciami, właściwość ta wyklucza sytuacje, w których potencjalni klienci, zbliżający się do systemu masowej obsługi, będą odraczać moment wejścia, jeśli zobaczą, że inny klient pojawia się przed nimi.

Przyjmijmy, że prawdopodobieństwo pojawienia się k zgłoszeń w przedziale czasu $(t; t + T)$ nie zależy ani od ilości zgłoszeń w systemie, ani od sposobu, w jaki pojawiły się one w czasie poprzedzającym ten przedział. Rozkład prawdopodobieństwa czasu pozostałego do zdarzenia (przybycia lub zakończenia usługi) jest zawsze taki sam, niezależnie od tego, ile czasu już minęło. Czas do następnego przybycia okazuje się całkowicie niezależny od tego, kiedy nastąpiło ostatnie przybycie. W efekcie, proces „zapomina” o swojej historii, ponieważ czas do następnego przybycia staje się całkowicie niezależny od tego, kiedy nastąpiło ostatnie przybycie – to tzw. proces bez pamięci.

Mówimy, że proces jest stacjonarny, gdy prawdopodobieństwo pojawienia się pewnej liczby zdarzeń w przedziale czasu zależy tylko od długości tego przedziału, a nie od jego położenia na osi czasu.

Jeżeli prawdopodobieństwo zaistnienia dwóch lub większej liczby zdarzeń w elementarnym przedziale T jest pomijalnie małe w odniesieniu do prawdopodobieństwa pojawienia się tylko jednego zdarzenia, to jest to proces zwykły. Gdy proces jest stacjonarny, bez pamięci i zwykły to jest to tzw. strumień zgłoszeń Poissona. Modele przedstawione w następnej części rozdziału będą modelami właśnie o Poissonowskim strumieniu zgłoszeń.

2. Wybrane modele masowej obsługi

2.1. Modele typu M/M/s/∞/∞

Większość podstawowych modeli masowej obsługi zakłada, że dane wejściowe i wyjściowe zachowują się zgodnie z procesem narodzin i śmierci. Przez narodzinny rozumiemy przybycie nowego zgłoszenia do systemu, śmierć oznacza opuszczenie systemu przez zgłoszenie. Proces narodzin i śmierci opisuje probabilistycznie, w jaki sposób stan systemu zmienia się wraz z upływem czasu. Narodziny i zgony następują więc losowo, a ich średnia stopa zależy tylko od aktualnego stanu systemu, nie zaś od tego, co wydarzyło się w przeszłości. Mamy więc do czynienia z procesem bez pamięci, o którym wspominaliśmy przy okazji omawiania własności rozkładu wykładniczego.

W kontekście systemów masowej obsługi zmiana stanu zachodzi wtedy, kiedy zgłoszenie pojawia się w systemie lub go opuszcza po zakończeniu obsługi. λ_n i μ_n

($n = 0, 1, \dots$) wyrażają więc odpowiednio: średnią stopę zgłoszeń i średnią stopę obsługi – gdy w systemie obecnych jest n klientów. Bardzo ważną zasadą w procesie narodzin i śmierci jest ta, mówiąca, że dla systemu znajdującego się w stanie n średnia stopa wejść równa się średniej stopie wyjść. Zasada pozwala wyznaczyć tzw. równanie bilansowe dla stanu n . Równania dla wszystkich stanów tworzą układ (wraz z równaniem zakładającym, że suma wszystkich prawdopodobieństw wynosi 1), którego rozwiązaniem są prawdopodobieństwa P_n wystąpienia stanu n . Więcej na temat procesu narodzin i śmierci znajdziemy w literaturze przedmiotu (zob. Hillier i Lieberman, 2015).

W tym i w następnym podrozdziale omówimy wybrane modele masowej obsługi, w których czas pomiędzy kolejnymi zgłoszeniami jest zmienną losową o rozkładzie wykładniczym (bądź liczba zgłoszeń w jednostce czasu ma rozkład Poissona), czas obsługi okazuje się zmienną losową o rozkładzie wykładniczym, zaś system masowej obsługi podlega procesowi narodzin i śmierci.

Zacznijmy od przypadku, w którym nieograniczona jest zarówno populacja zgłoszeń, jak i długość kolejki, oraz obowiązuje w niej dyscyplina FIFO. Zgodnie z notacją Kendalla wariant o jednym kanale obsługi ($s = 1$) zapiszemy: M/M/1/∞/∞.

Niech λ oznacza oczekiwaną stopę przybyć (zgłoszeń) w jednostce czasu, zaś $\frac{1}{\mu}$ – oczekiwany czas obsługi jednego zgłoszenia. Stała Erlanga (intensywność obsługi) wyraża się w tej sytuacji wzorem (238):

$$\rho = \frac{\lambda}{\mu} \quad (238)$$

Kiedy $\lambda \geq \mu$, system jest niestabilny i kolejka rośnie bez ograniczeń. Wprawdzie, jeśli system zacznie działać bez żadnych zgłoszeń, to serwerowi uda się dotrzymać kroku przybywającym zgłoszeniom (np. klientom pojawiającym się zaraz po otwarciu punktu obsługi) w krótkim okresie czasu, ale na dłuższą metę nie jest to możliwe. Analizy systemu obsługi w takim stanie mijają się z celem i najpierw trzeba doprowadzić do tego, by zachodziło $\lambda < \mu$. Od tej pory właśnie taki stan będziemy domyślnie zakładać.

Zgłoszenia obsługiwane są w kolejności przybywania i każde z nich pozostaje w kolejce tak długo, aż zostanie obsłużone.

Najważniejsze charakterystyki liczbowe modelu M/M/1/∞/∞

Oczekiwana liczba zgłoszeń w systemie:

$$N = \frac{\rho}{1 - \rho} \quad (239)$$

Oczekiwana długość kolejki:

$$Q = \frac{\rho^2}{1 - \rho} \quad (240)$$

Oczekiwany czas pobytu w systemie:

$$R = \frac{1}{\mu - \lambda} \quad (241)$$

Oczekiwany czas pobytu w kolejce:

$$W = \frac{\rho}{\mu - \lambda} \quad (242)$$

Prawdopodobieństwo braku zgłoszeń w systemie:

$$P_0 = 1 - \rho \quad (243)$$

Prawdopodobieństwo tego, że w systemie znajduje się n zgłoszeń:

$$P_n = \rho^n (1 - \rho) \quad (244)$$

Prawdopodobieństwo tego, że zgłoszenie spędzi w kolejce więcej niż t_0 jednostek czasowych (Jędrzejczyk, Kukuła, Skrzypek i Walkosz, 2016, s. 212):

$$P(t > t_0) = \rho e^{-t_0(\mu - \lambda)} \quad (245)$$

W literaturze przedmiotu (Miszczyński i Miszczyńska, 2002, s. 164) znajdziemy kilka dodatkowych charakterystyk dla tego modelu.

Oczekiwany czas przestoju w okresie czasu $\langle 0, T \rangle$:

$$W_T = T(1 - \rho) \quad (246)$$

Oczekiwany czas zajętości (pracy) w okresie czasu $\langle 0, T \rangle$:

$$B_T = T\rho \quad (247)$$

Oczekiwana liczba okresów przestoju kanału obsługi w okresie czasu $\langle 0, T \rangle$:

$$FP_T = T\lambda(1 - \rho) \quad (248)$$

Przykład 1¹

Kierownik działu produkcji wspólnie z kierownikiem magazynu dostali polecenie oceny efektywności procesu dostarczania i obróbki dwóch detali A i B , trafiających z magazynu na tę samą maszynę. To niedawno zakupione urządzenie posiada ka-

¹ Na podstawie przykładu 5.1 w: Miszczyński i Miszczyńska, 2002.

setę mieszczącą pięć detali *A* lub pięć detali *B*, z której pobiera element, a następnie go wytłacza i zgrzewa. Po opróżnieniu kaseta uzupełniana jest automatycznie, dzięki podajnikowi zaopatrywanemu z magazynu. W magazynie znajduje się tak duża liczba detali obu typów, że można ją uznać za nieograniczoną – z punktu widzenia planowanej wielkości produkcji. Proces trafiania detali do obróbki ma charakter losowy. Kierownik działu produkcji wykorzystał raporty generowane przez komputer podłączony do urządzenia. Zarejestrowano w nich m.in. odstępy czasu między umieszczeniem elementów w maszynie celem obróbki oraz czas obróbki pojedynczego elementu. Tabela 81 przedstawia próbkę takich danych dla detalu *A*.

Po przeprowadzeniu pomiarów i użyciu odpowiednich testów statystycznych stwierdzono, że liczba przybyć w ciągu minuty jest zmienną losową o rozkładzie Poissona, a czas obróbki opisuje rozkład wykładniczy. Różne są jednak parametry tych rozkładów w zależności od typu detalu. Średnio w ciągu minuty do obróbki trafiają cztery detale *A*, zaś średni czas obróbki jednej sztuki wynosi dziesięć sekund. Z kolei dla detalu *B* średnio w ciągu minuty do obróbki trafia pięć sztuk, zaś średni czas obróbki jednego detalu wynosi siedem i pół sekundy.

Tab. 81. Proces przybycia i obróbki detalu *A* według raportu

Detal numer	Czas przybycia liczony od przybycia poprzedniego detalu [min]	Czas obróbki [min]
1	0	0,17
2	0,5	0,2
3	0,2	0,19
4	0,25	0,25
itd.	itd.	itd.

Źródło: opracowanie własne.

Zarząd firmy uznał, że nadeszła pora, aby ocenić, jak sprawuje się nowa maszyna. Obaj kierownicy mają opracować sprawozdanie, które odpowie na następujące pytania:

1. W jakim stopniu wykorzystywane jest urządzenie przy obróbce każdego z detali? Kiedy jest ono lepiej wykorzystywane?
2. Dla którego z detali łączny czas magazynowania w kasie oraz obróbki jest dłuższy?
3. Kiedy następuje większe nagromadzenie detali na stanowisku roboczym?
4. Jakie jest prawdopodobieństwo wystąpienia błędu podajnika? Dla którego detalu prawdopodobieństwo całkowitego zapełnienia kasety jest wyższe?
5. Jak duża powinna być kaseta, aby prawdopodobieństwo tego, że okaże się ona pusta, było nie wyższe niż 5%?

Intensywność ruchu (ρ) mniejsza od 1 (por. tabela 82) wskazuje na stabilność procesu w przypadku obu detali. Urządzenie jest wykorzystywane w około 67% podczas obróbki detalu A i 62,5% – jeśli chodzi o detal B. Różnica nie jest duża, ale intensywniej (lepiej) maszyna sprawuje się przy detalu A.

Aby odpowiedzieć na drugie z postawionych pytań, musimy poznać wartość czasu pobytu w systemie (R). Detal A znajduje się w kasecie, a potem w samej maszynie średnio przez trzydzieści sekund, detal B – przez dwadzieścia. Różnica dziesięciu sekund, przy odpowiednio dużej skali produkcji, może mieć znaczenie.

O nagromadzeniu detali decyduje oczekiwana liczba zgłoszeń w systemie (N), która dla detalu A wynosi średnio dwie sztuki, a dla detalu B – 1,67. Więcej detali gromadzi się na stanowisku roboczym podczas obróbki detalu A.

Odpowiedź na następne dwa pytania wymaga obliczenia prawdopodobieństw wystąpienia określonej liczby zgłoszeń (P_n), na podstawie których wyznaczona zostanie dystrybucja rozkładu.

Tab. 82. Wartości oczekiwane charakterystyk liczbowych dla przykładu 1

Nazwa miary	Detal A	Detal B
Stopa przybyć (λ , [szt./min])	4	5
Odstęp czasu ($1/\lambda$, [min/szt.])	0,25	0,2
Stopa obsługi (μ , [szt./min])	6	8
Czas obsługi ($1/\mu$, [min/szt.])	0,1667	0,125
Intensywność ruchu (ρ)	0,6667	0,625
Oczek. liczba detali w kasecie i w obróbce (N , [szt.])	2	1,6667
Oczek. liczba detali czekających na obróbkę (Q , [szt.])	1,3333	1,0417
Oczek. czas pobytu w kasecie i w obróbce (R , [min])	0,5	0,3333
Oczek. czas oczekiwania na obróbkę (W , [min])	0,3333	0,2083

Źródło: opracowanie własne.

Tab. 83. Rozkłady prawdopodobieństw i dystrybucje dla przykładu 1

k	Detal A			Detal B		
	$P(n = k)$	$P(n \leq k)$	$P(n \geq k)$	$P(n = k)$	$P(n \leq k)$	$P(n \geq k)$
0	0,3333	0,3333	1,0000	0,3750	0,3750	1,0000
1	0,2222	0,5556	0,6667	0,2344	0,6094	0,6250
2	0,1481	0,7037	0,4444	0,1465	0,7559	0,3906
3	0,0988	0,8025	0,2963	0,0916	0,8474	0,2441
4	0,0658	0,8683	0,1975	0,0572	0,9046	0,1526

	Detal A			Detal B		
5	0,0439	0,9122	0,1317	0,0358	0,9404	0,0954
6	0,0293	0,9415	0,0878	0,0224	0,9627	0,0596
7	0,0195	0,9610	0,0585	0,0140	0,9767	0,0373
8	0,0130	0,9740	0,0390	0,0087	0,9854	0,0233
9	0,0087	0,9827	0,0260	0,0055	0,9909	0,0146
10	0,0058	0,9884	0,0173	0,0034	0,9943	0,0091

Źródło: opracowanie własne.

Przypomnijmy, że kaseta przechowuje pięć sztuk detali. W sytuacji, kiedy jeden detal jest już w maszynie, a kaseta okaże się pełna – czyli liczba zgłoszeń równa się sześciu (jedna sztuka w maszynie, pięć w kasecie) – podajnik nie może umieścić w niej kolejnego elementu. Interesuje nas zatem prawdopodobieństwo tego, że liczba zgłoszeń wyniesie co najmniej siedem, gdyż to właśnie siódme zgłoszenie spowoduje błąd podajnika. Z tabeli 83 odczytujemy wartość $P(n \geq 7)$. W przypadku detalu A prawdopodobieństwo to wynosi 5,85%, zaś dla detalu B – 3,73%. Zatem detal A odpowiada za większą awaryjność w procesie produkcji.

Pusta kaseta oznacza przestój maszyny, generujący koszty dla przedsiębiorstwa. Dlatego zarząd chciałby ograniczyć możliwość wystąpienia takich sytuacji. W tabeli 83 wyszukujemy liczbę zgłoszeń w systemie, przy której wartość dystrybuanty $P(n \geq k)$ będzie po raz pierwszy niższa od 0,05. Dla detalu A mamy:

$$P(n \geq 8) = P(n \geq 6 + 1 + 1) = 0,039$$

Dla detalu B:

$$P(n \geq 7) = P(n \geq 5 + 1 + 1) = 0,0373$$

Uwzględniając fakt, że jeden detal powinien znajdować się wewnątrz maszyny w celu obróbki oraz że podajnik dosyła kolejną sztukę, kaseta w przypadku detalu A powinna mieć pojemność minimum sześciu, a w przypadku detalu B minimum pięciu detali. Z tego płynie wniosek, że docelowa pojemność kasety, która zapewni uniknięcie przestoju, winna równać się co najmniej $\max\{5, 6\} = 6$ [detali].

Kierownik działu produkcji oraz kierownik magazynu postanowili zbadać jeszcze jeden aspekt działania nowej maszyny, a mianowicie: stopień jej faktycznego wykorzystania w czasie ośmiogodzinnego dnia pracy. W tym celu obliczyli dodatkowe miary (zob. tabela 84).

Tab. 84. Ocena wykorzystania maszyny w ciągu dnia pracy ($T = 480$)

Miernik	Detal A	Detal B
W_T [min]	160	180
B_T [min]	320	300
FP_T [szt.]	640	900

Źródło: opracowanie własne.

Wyniki wskazują, że łączny oczekiwany czas faktycznej pracy maszyny jest o dwadzieścia minut dłuższy w przypadku detalu A – co jest konsekwencją wyższej intensywności obsługi. W tej sytuacji oczekiwany łączny czas przestoju (rozumiany jako suma przestojów częściowych z całego dnia pracy) okazuje się również o dwadzieścia minut krótszy. Z kolei wartość FP_T , wyrażająca liczbę przerw podczas dnia pracy, pozwala określić średnią liczbę detali, jakie zostały poddane operacjom w ciągu ośmiu godzin. Po każdej operacji następuje bowiem przerwa na dosłanie kolejnej sztuki do obróbki. Widzimy, że liczba przetworzonych detali B znacznie przewyższa liczbę detali A. Analizy podobnego typu znajdziemy w literaturze przedmiotu (Kamberi, 2017), gdzie autor analizuje funkcjonowanie laboratorium w jednym ze szpitali w Macedonii.

Zwiększenie liczby kanałów, przy niezmienionych pozostałych parametrach, prowadzi nas do modelu $M/M/s/\infty/\infty$ gdzie $s \geq 2$. Stała Erlanga wyrażona jest wówczas wzorem (230). Poniżej prezentujemy pozostałe charakterystyki liczbowe tego modelu.

Najważniejsze charakterystyki liczbowe modelu $M/M/s/\infty/\infty$

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = \frac{1}{\sum_{j=0}^{s-1} \frac{(\lambda/\mu)^j}{j!} + \frac{(\lambda/\mu)^s}{s!(1-\rho)}} \quad (249)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = \frac{\lambda}{\mu} + P_0 = \frac{(\lambda/\mu)^{s+1}}{[s - (\lambda/\mu)]^2 (s-1)!} \quad (250)$$

Oczekiwana długość kolejki:

$$Q = N - \frac{\lambda}{\mu} \quad (251)$$

Oczekiwany czas pobytu w systemie:

$$R = \frac{N}{\lambda} \quad (252)$$

Oczekiwany czas pobytu w kolejce:

$$W = \frac{Q}{\lambda} \quad (253)$$

Prawdopodobieństwo tego, że w systemie znajduje się $n \geq 1$ zgłoszeń:

$$P_n = \begin{cases} P_0 \frac{(\lambda/\mu)^n}{n!}, & \text{dla } 1 \leq n \leq s \\ P_0 \frac{(\lambda/\mu)^n}{s! s^{n-s}}, & \text{dla } n > s \end{cases} \quad (254)$$

Prawdopodobieństwo tego, że czas oczekiwania w kolejce jest dłuższy od t_0 (Jędrzejczyk, Kukula, Skrzypek i Walkosz, 2016, s. 214):

$$P(t > t_0) = P(n > s - 1) e^{-\mu t_0 (s - \rho)} \quad (255)$$

Przykład 2²

Właściciel warsztatu samochodowego chce rozszerzyć działalność o usługi wulkanizatorskie. Na początek postanowił zaoferować wymianę opon w cenie osiemdziesięciu złotych za komplet. Musi jednak znaleźć dodatkowych pracowników, gdyż żaden z obecnie zatrudnionych nie posiada odpowiednich kwalifikacji. Waha się, czy zatrudnić osoby z dużym doświadczeniem, ale za wyższą stawkę, czy poszukać kogoś mniej doświadczonego, kto zgodzi się pracować za mniejsze wynagrodzenie. Godzinowa stawka doświadczonego pracownika wynosi dwadzieścia pięć złotych, zaś mniej doświadczonego (dla uproszczenia nazwijmy go początkującym) – piętnaście. Z obserwacji wynika, że liczba klientów pojawiających się w celu wymiany opon jest zmienną losową o rozkładzie Poissona z wartością oczekiwaną trzech zgłoszeń na godzinę. Czas wymiany kompletu opon to z kolei zmienna losowa o rozkładzie wykładniczym. Doświadczonemu pracownikowi zajmuje ona dwadzieścia minut, początkującemu – trzydzieści.

Właściciel chce zachęcić klientów do skorzystania z usług jego warsztatu. Postanowił więc, że klient dostanie dziesięć złotych zniżki za każdą rozpoczętą godzinę oczekiwania na wymianę opon. Rozważa następujące warianty zatrudnienia: dwóch doświadczonych pracowników ($D2$), dwóch początkujących ($P2$) lub trzech początkujących ($P3$). Który pozwoli na osiągnięcie największego średniego zysku?

Decydent rozpoczął od obliczenia charakterystyk liczbowych modelu $M/M/s/\infty/\infty$, i na ich podstawie wyznaczył wielkości o ekonomicznym charakterze. Znajdziemy je w tabeli 85. Największe obciążenie pracą obserwujemy w przypadku modelu $P2$, dla którego intensywność ruchu wynosi 75%. Jednocześnie jest to wariant o najmniejszym prawdopodobieństwie bezczynności (P_0).

2 Na podstawie przykładu 5.4 w: Miszczyński i Miszczyńska, 2002.

Wyjaśnijmy teraz, w jaki sposób obliczone zostały średnie zyski z godziny pracy. W tym celu średnia liczba klientów czekających i obsługiwanych (N), która wyraża liczbę klientów rozpoczynających kolejną godzinę oczekiwania na wymianę opon, została pomnożona przez stawkę zniżki. Otrzymaliśmy pierwszy składnik kosztów. Dodajemy do niego koszty osobowe – będące iloczynem stawki godzinowej i liczby pracowników w danym wariantcie modelu. Przychód powstaje przez pomnożenie stopy przybyć przez cenę wymiany.

Zyski nie różnią się zbyt wiele między wariantami modeli, ale najwyższy daje zatrudnienie dwóch doświadczonych pracowników. Zauważmy też, że w wariantcie przewidującym zatrudnienie dwóch początkujących osób, klient musi zdecydowanie najdłużej czekać w kolejce. W konsekwencji rośnie jej długość. Łatwo sprawdzić, że podnoszenie zniżki za konieczność czekania powoduje szybki wzrost zysków z wariantu $D2$ i równie szybki spadek w wariantcie $P2$. Wspominamy o tym dlatego, że obniżanie wysokości zniżki doprowadzi w końcu do sytuacji, w której to właśnie model $P2$ charakteryzować się będzie najwyższym średnim godzinowym zyskiem. Jednak zbyt długie oczekiwanie może zniechęcić przyszłych klientów.

Tab. 85. Wyniki obliczeń dla przykładu 2

Miara liczbową	Wariant modelu		
	$D2$	$P2$	$P3$
	$M/M/2$	$M/M/2$	$M/M/3$
s (liczba pracowników)	2	2	3
λ stopa przybyć [klient/h]	3	3	3
$1/\lambda$ [h/klient]	0,3333	0,3333	0,3333
μ stopa obsługi [klient/h]	3,3333	2	2
$1/\mu$ [h/klient]	0,3333	0,5	0,5
ρ (stała Erlanga)	0,45	0,75	0,5
P_0	0,3793	0,1429	0,2105
N (średnia liczba klientów czekających i obsługiwanych)	1,1285	3,4286	1,7368
Q (średnia liczba klientów czekających na wymianę)	0,2285	1,9286	0,2368
W (średni czas oczekiwania w kolejce [h])	0,0762	0,6429	0,0789
R (średni czas w kolejce oraz wymiany [h])	0,3762	1,1429	0,5789
Łączne koszty osobowe [zł/h]	50	30	45
Zniżka za czekanie [zł/h]	11,2853	34,2857	17,3684
Razem koszty [zł/h]	61,2853	64,2857	62,3684
Cena usługi [zł]	80	80	80
Przychód [zł/h]	240	240	240
Zysk [zł/h]	178,7147	175,7143	177,6316

Źródło: opracowanie własne.

2.2. Modele typu M/M/s/∞/∞ z uchylonymi założeniami

Szczególnym przypadkiem modelu M/M/s/∞/∞ jest wariant, w którym obowiązuje zakaz tworzenia kolejki. Zgłoszenie opuszcza system wtedy, gdy wszystkie stanowiska obsługi są zajęte. Z tego powodu model ten bywa nazywany także modelem M/M/s ze stratą. Wyłączenie z rozważań możliwości utworzenia kolejki wpływa na charakterystyki liczbowe, które zmieniają swoją postać (Abramov, 2008). W literaturze przedmiotu (Swathi, Vasanta Kumar i Hanumantha Rao, 2019) znajdziemy opis modelu M/M/1 uwzględniający m.in. klientów rezygnujących z obsługi i serwery wyłączone z obsługi z różnych powodów. W takim systemie masowej obsługi również rozpatrujemy straty, ale w szerszym aspekcie.

Najważniejsze charakterystyki liczbowe modelu M/M/s ze stratą

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = \frac{1}{\sum_{j=0}^s \frac{(\lambda/\mu)^j}{j!}} \quad (256)$$

Prawdopodobieństwo tego, że w systemie znajduje się $n \geq 1$ zgłoszeń:

$$P_n = \frac{(\lambda/\mu)^n}{n!} P_0 \quad (257)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = \sum_{j=1}^s \frac{\rho^j}{(j-1)!} P_0 \quad (258)$$

Oczekiwany czas pobytu w systemie:

$$R = \frac{N}{\lambda \left(1 - \frac{\rho^s}{s!} P_0 \right)} = \frac{N}{\lambda (1 - P_s)} \quad (259)$$

Złagodźmy teraz założenie dotyczące długości kolejki i przyjmijmy, że ma ona skończoną długość. Oznacza to, że liczba zgłoszeń, które zostaną obsłużone, jest ograniczona. Jeśli oznaczmy ją przez K , wtedy kolejka pomieści $(K - n)$ zgłoszeń. Zgłoszenie pojawiające się w momencie, gdy kolejka jest pełna, nie wchodzi do systemu. Stąd stopa zgłoszeń ma postać:

$$\lambda_n = \begin{cases} \lambda, & \text{dla } n = 0, 1, 2, \dots, K-1 \\ 0, & \text{dla } n \geq K \end{cases} \quad (260)$$

Ponieważ stopa przybyć w niektórych przypadkach równa się 0, to system może osiągnąć równowagę nawet, jeśli $\rho > 1$. Przybliżmy charakterystyki liczbowe jednokanałowego modelu M/M/1/K/ ∞ wersji z wieloma kanałami obsługi M/M/s/K/ ∞ .

Najważniejsze charakterystyki liczbowe modelu M/M/1/K/ ∞

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = \frac{1-\rho}{1-\rho^{K+1}} \quad (261)$$

Prawdopodobieństwo tego, że w systemie znajduje się $n \geq 1$ zgłoszeń:

$$P_n = \frac{1-\rho}{1-\rho^{K+1}} \rho^n \quad (262)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = \frac{\rho}{1-\rho} = \frac{(K+1)\rho^{K+1}}{1-\rho^{K+1}} \quad (263)$$

Oczekiwana długość kolejki:

$$Q = N - (1 - P_0) \quad (264)$$

Oczekiwany czas pobytu w systemie:

$$R = \frac{N}{\lambda} \quad (265)$$

Oczekiwany czas pobytu w kolejce:

$$W = \frac{Q}{\lambda} \quad (266)$$

gdzie:

$$\bar{\lambda} = \lambda(1 - P_K) \quad (267)$$

Charakterystyki liczbowe modelu wielokanałowego M/M/s/K/ ∞ , $s > 1$

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = \frac{1}{\sum_{n=0}^s \frac{(\lambda/\mu)^n}{n!} + \frac{(\lambda/\mu)^s}{s!} \sum_{n=s+1}^K \left(\frac{\lambda}{s\mu} \right)^{n-s}} \quad (268)$$

Prawdopodobieństwo tego, że w systemie znajduje się $n \geq 1$ zgłoszeń:

$$P_n = \begin{cases} \frac{(\lambda/\mu)^n}{n!} P_0, & \text{dla } n = 1, 2, \dots, s-1 \\ \frac{(\lambda/\mu)^n}{s! s^{n-s}} P_0, & \text{dla } n = s, s+1, \dots, K \\ 0, & \text{dla } n > K \end{cases} \quad (269)$$

Oczekiwana długość kolejki:

$$Q = \frac{P_0 (\lambda/\mu)^s}{s! (1-\rho)^2} [1 - \rho^{K-s} - (K-s)(1-\rho)\rho^{K-s}] \quad (270)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = \sum_{n=0}^{s-1} n P_n + Q + s \left(1 - \sum_{n=0}^{s-1} P_n \right) \quad (271)$$

Wartości R i W można obliczyć ze wzorów dla modelu $M/M/1/K/\infty$.

Przykład 3

Poczekalnia w zakładzie fryzjerskim mieści maksymalnie pięciu klientów. Fryzjer obsługuje jedną osobę naraz. Jeśli klient trafi na pełną poczekalnię, udaje się do innego zakładu. Klienci zjawiają się w losowych odstępach czasu ze średnią stopą sześciu osób na godzinę. Odstępy między pojawieniami się klientów opisuje zmienna losowa o rozkładzie wykładniczym. Czas obsługi jednego klienta wynosi piętnaście minut – i również opisuje go rozkład wykładniczy. Spróbujemy scharakteryzować proces obsługi w tym zakładzie. Jakie zmiany zajądą w procesie obsługi, gdyby zatrudnić drugiego fryzjera o takiej samej stopie obsługi?

Tab. 86. Wyniki obliczeń dla przykładu 3

Charakterystyka liczbowa	M/M/1/5/∞	M/M/2/5/∞
s (liczba fryzjerów)	1	2
K (pojemność kolejki)	5	5
λ (stopa przybyć [klient/h])	6	6
$1/\lambda$ [h/klient]	0,1667	0,1667
μ (stopa obsługi [klient/h])	4	4
$1/\mu$ [h/klient]	0,25	0,25
ρ (stała Erlanga)	1,5	0,75

Tab. 86 (cd.)

Charakterystyka liczbowa	M/M/1/5/∞	M/M/2/5/∞
ρ (skoryg. stała Erlanga)	0,9519	0,6862
P_0	0,0481	0,1793
P_5	0,3654	0,0851
N (średnia liczba klientów czekających i obsługiwanych)	3,5774	2,006
Q (średnia liczba klientów czekających na usługę)	2,6256	0,6336
W (średni czas oczekiwania w kolejce [h])	0,6896	0,1154
R (średni czas w kolejce oraz usługi [h])	0,9396	0,3654

Źródło: opracowanie własne.

W tabeli 86 znalazło się podsumowanie charakterystyk liczbowych przykładu 3. Dla modelu jednokanałowego intensywność ruchu jest większa niż 1. Może to sprawiać kłopot z interpretacją, skoro nie należy ona do przedziału $\langle 0, 1 \rangle$. Rozwiązaniem będzie obliczenie skorygowanej wartości o prawdopodobieństwo tego, że klient opuści system nieobsłużony, według wzoru:

$$\bar{\rho} = \rho(1 - P_K) \quad (272)$$

Zatrudnienie drugiego fryzjera dwukrotnie poprawia działanie systemu (mierzona stałą Erlanga). Dwukrotnie krótsza jest też kolejka oczekujących na usługę. Znacząco wzrasta jednak prawdopodobieństwo braku zgłoszeń, czyli fryzjerzy dłużej będą bezczynni. Zatrudnienie drugiego fryzjera powoduje też, że prawdopodobieństwo całkowitego zapełnienia kolejki nie przekracza 10%. Średni czas oczekiwania, gdy strzyżeniem zajmuje się jeden fryzjer, wynosi około czterdziestu dwóch minut, a gdy jest ich dwóch – siedem. Wydaje się więc, że – mimo wszystko – warto zatrudnić w zakładzie dwóch fryzjerów, gdyż bardzo podnosi to komfort korzystania z usług zakładu.

Uchylmy teraz założenie o nieograniczonej liczebności populacji zgłoszeń. Niech L oznacza skończoną liczebność populacji tworzonej przez zgłoszenia. Każdy członek populacji zgłoszeń może znajdować się wewnątrz lub poza systemem kolejkowania. Innymi słowy, jego stan ulega zmianom w czasie, gdyż opuszczenie systemu nie jest definitywne i istnieje możliwość powrotu. Model M/M/s/∞/L zakłada, że czas, jaki upłynął od wyjścia z systemu do powrotu do niego po raz kolejny, ma rozkład wykładniczy z parametrem λ . Całkowita populacja składa się z L zgłoszeń, zaś n oznacza liczbę zgłoszeń już obecnych w systemie. Każde kolejne przybycie musi pochodzić spośród $L - n$ zgłoszeń, których nie ma jeszcze w systemie.

Najważniejszym zastosowaniem tego modelu jest problem naprawy maszyn lub urządzeń, w którym jedna osoba zajmująca się konserwacją (lub więcej) utrzymuje w gotowości do działania grupę L maszyn, naprawiając je w razie awarii. Każda z nich – gdy nie działa i czeka na naprawę – jest zgłoszeniem w systemie. Jeżeli sprawnie funkcjonuje, znajduje się poza systemem kolejkowym. W literaturze przedmiotu (Kramer, 1989) występuje ograniczona populacja zgłoszeń – tu jednak autor korzystał z innego modelu, a mianowicie $M/G/1$, o którym piszemy w dalszej części rozdziału. Dlatego tworzące populację zgłoszenia na przemian znajdują się wewnątrz i na zewnątrz systemu masowej obsługi. Przedstawmy najpierw charakterystyki modelu jednokanałowego $M/M/1/\infty/L$.

Najważniejsze charakterystyki liczbowe modelu $M/M/1/\infty/L$

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = \frac{1}{\sum_{n=0}^L \left(\frac{L!}{(L-n)!} \left(\frac{\lambda}{\mu} \right)^n \right)} \quad (273)$$

Prawdopodobieństwo tego, że w systemie znajduje się $n \geq 1$ zgłoszeń:

$$P_n = \frac{L!}{(L-n)!} \left(\frac{\lambda}{\mu} \right)^n P_0 \quad (274)$$

Oczekiwana długość kolejki:

$$Q = \sum_{n=1}^L (n-1)P_n = L - \frac{\lambda + \mu}{\lambda} (1 - P_0) \quad (275)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = L - \frac{\mu}{\lambda} (1 - P_0) \quad (276)$$

Oczekiwany czas pobytu w kolejce:

$$W = \frac{Q}{\lambda} \quad (277)$$

gdzie:

$$\bar{\lambda} = \lambda(L - N) \quad (278)$$

Charakterystyki liczbowe modelu wielokanałowego $M/M/s/\infty/L$, $s > 1$

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = \frac{1}{\sum_{n=0}^{s-1} \frac{L!}{(L-n)!n!} \left(\frac{\lambda}{\mu}\right)^n + \sum_{n=s}^L \frac{L!}{(L-n)!s!s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n} \quad (279)$$

Prawdopodobieństwo tego, że w systemie znajduje się $n \geq 1$ zgłoszeń:

$$P_n = \begin{cases} \frac{L!}{(L-n)!n!} \left(\frac{\lambda}{\mu}\right)^n P_0, & \text{dla } 0 \leq n \leq s \\ \frac{L!}{(L-n)!s!s^{n-s}} \left(\frac{\lambda}{\mu}\right)^n P_0, & \text{dla } s \leq n \leq L \\ 0, & \text{dla } n > L \end{cases} \quad (280)$$

Oczekiwana długość kolejki:

$$Q = \sum_{n=s}^L (n-s)P_n \quad (281)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = \sum_{n=0}^{s-1} nP_n + Q + s \left(1 - \sum_{n=0}^{s-1} P_n\right) \quad (282)$$

Przykład 4

W pewnej rozlewni wody mineralnej funkcjonują cztery linie technologiczne do napełniania butelek, które średnio co pięć godzin wymagają interwencji mechanika. Wiadomo, że czas pomiędzy naprawami ma rozkład wykładniczy. Średni czas naprawy (o tym samym rozkładzie) wynosi jedną godzinę. Ile średnio linii będzie naprawianych? Ile wyniesie koszt przestoju i naprawy, jeżeli wiadomo, że niepracująca linia generuje koszt równy trzystu złotym na godzinę, a stawka mechanika za naprawę wynosi sto pięćdziesiąt złotych?

Aby odpowiedzieć na postawione pytania, należy obliczyć charakterystyki liczbowe modelu $M/M/1/\infty/4$. Znajdziemy je w tabeli 87. W modelu tego typu λ oznacza liczbę zgłoszeń do obsługi każdego obiektu w jednostce czasu – i dlatego wynosi ona w naszym przypadku $1/5$. Prawdopodobieństwo tego, że mechanik nie będzie zajmował się żadną z linii technologicznych wynosi około 40%. Śred-

nio jedna linia nie pracuje i wymaga naprawy (informuje o tym wartość N), zaś na samą naprawę oczekuje ona niecałe czterdzieści minut (W). Łączny koszt przestoju (przy założeniu ośmiogodzinnego dnia pracy) i naprawy jednej linii wynosi: $8 \times 1 \times 300 + 150 = 2550$ [zł].

Tab. 87. Wyniki obliczeń dla przykładu 4

Charakterystyka liczbową	M/M/1/∞/4
s (liczba mechaników)	1
L (liczebność populacji)	4
λ (stopa przybyć)	0,2
$1/\lambda$	5
μ (stopa obsługi)	1
$1/\mu$	1
ρ (stała Erlanga)	0,2
P_0	0,3983
N (średnia liczba zgłoszeń)	0,9917
Q (średnia długość kolejki)	0,39
W (średni czas oczekiwania w kolejce [h])	0,6483

Źródło: opracowanie własne.

Kolejnym szczególnym przypadkiem modelu z rozkładem wykładniczym okazuje się ten, w którym zgłoszenie przechodzi kolejno przez m stanowisk obsługi, ułożonych szeregowo. Za przykład niech posłuży sytuacja, kiedy dokument wystawiony w jednej z komórek przedsiębiorstwa jest przekazywany drogą służbową do kolejnych jego działów (prawny, księgowość itd.). W każdym z nich zostanie on przeanalizowany, zaakceptowany i przesłany dalej – aż do uzyskania ostatecznej wersji. Przejścia pomiędzy stanowiskami obsługi wyznaczają podział całego procesu na etapy. Inny przykład wykorzystania tego modelu zaprezentowano w literaturze przedmiotu (zob. Langaris, 1993). W związku z tym analizę efektywności działania systemu masowej obsługi przeprowadza się na każdym z etapów, wykorzystując do tego celu wzory z modelu M/M/1/∞/∞.

Przykład 5

Do magazynu pewnego sklepu internetowego spływają średnio dwa zamówienia na godzinę. Odstępy czasu pomiędzy kolejnymi zgłoszeniami mają rozkład wykładniczy. Pracownicy magazynu kompletują średnio trzy zamówienia na godzinę, a następnie przekazują je do działu obsługi klienta, gdzie sprawdza się zgodność z koszykiem wypełnionym przez klienta na stronie sklepu oraz dołącza

fakturę. Pracownicy tego działu obsługują przeciętnie 2,5 zamówienia w ciągu godziny. Ostatnim etapem, przez który musi przejść zamówienie, jest dział zajmujący się pakowaniem i wysyłką. W ciągu godziny obsługuje on średnio 2,8 zgłoszenia. Czas obsługi w każdym z działów podlega rozkładowi wykładniczemu. Co należy zrobić?

1. Ocenić efektywność pracy na poszczególnych etapach obsługi zamówienia.
2. Odpowiedzieć na pytanie, w którym dziale zamówienie klienta spędza najwięcej czasu.
3. Obliczyć prawdopodobieństwo tego, że w systemie przebywa dokładnie jedno zgłoszenie.

Zamówienie złożone przez klienta przechodzi przez $m = 3$ [etapy]: kompletację w magazynie, sprawdzenie w dziale obsługi oraz pakowanie i wysyłkę. Każdy z nich zostanie potraktowany jako jednokanałowy system masowej obsługi typu M/M/1/∞/∞. Tabela 88 zawiera podsumowanie charakterystyk liczbowych tego modelu, oddzielnie dla każdego etapu.

Tab. 88. Wartości oczekiwane charakterystyk liczbowych dla przykładu 5

Nazwa miary	Etap 1	Etap 2	Etap 3
Stopa przybyć w etapie i (λ_i [szt./h])	2	2	2
Odstęp czasu w etapie i ($1/\lambda_i$ [h/szt.])	0,5	0,5	0,5
Stopa obsługi w etapie i (μ_i [szt./h])	3	2,5	2,8
Czas obsługi w etapie i ($1/\mu_i$ [h/szt.])	0,3333	0,4	0,3571
Intensywność ruchu w etapie i (ρ_i)	0,6667	0,8	0,7143
Oczek. liczba zgłoszeń w etapie i (N_i [szt.])	2	4	2,5
Oczek. długość kolejki w etapie i (Q_i [szt.])	1,3333	3,2	1,7857
Oczek. czas pobytu w etapie i (R_i [h])	1	2	1,25
Oczek. czas w kolejce w etapie i (W_i [h])	0,6667	1,6	0,8929
Prawdopodobieństwa braku zgłoszeń (P_n^i)	0,3333	0,2	0,2857

Źródło: opracowanie własne.

Największą intensywność ruchu obserwujemy na drugim etapie, gdzie wynosi ona 80%. Dział obsługi klienta jest więc najmocniej (bo najintensywniej) wykorzystywaną komórką firmy. Prawdopodobieństwo braku zgłoszeń (czyli przestoju) okazuje się dla tego działu również najmniejsze (20% przy około 33% dla magazynu i 29% dla działu wysyłkowego).

Jednak to właśnie tutaj zamówienie spędza najwięcej czasu ($R = 2$ [h] przy średniej wynoszącej $(1 + 2 + 1,25)/3 = 1,42$ [h]). Jest to o tyle ciekawe, że średni czas obsługi na poszczególnych etapach nie różni się znacząco.

Odpowiedź na ostatnie z postawionych pytań pomoże znaleźć prawdopodobieństwo tego, że w systemie przebywa k zgłoszeń, dane wzorem:

$$P_k = \sum_{j_1 + \dots + j_m = k} P_{j_1}^k \cdot P_{j_2}^k \cdot \dots \cdot P_{j_m}^k \quad (283)$$

Przy czym prawdopodobieństwo pojawienia się n zgłoszeń na etapie i (przez analogię do modelu M/M/1/∞/∞) wynosi:

$$P_n^i = \rho_i^n (1 - \rho_i) \quad (284)$$

Spełnienie warunku z trzeciego polecenia ma miejsce, gdy $j_1 + j_2 + j_3 = 1$; to zaś nastąpi, kiedy zajdzie jedna z następujących sytuacji:

- $j_1 = 1, j_2 = 0, j_3 = 0$;
- $j_1 = 0, j_2 = 1, j_3 = 0$;
- $j_1 = 0, j_2 = 0, j_3 = 1$.

Z tego płynie wniosek, że prawdopodobieństwo tego, iż w systemie przebywa dokładnie jedno zgłoszenie wynosi:

$$\begin{aligned} P_1 &= P_1^1 P_0^2 P_0^3 + P_0^1 P_1^2 P_0^3 + P_0^1 P_0^2 P_1^3 = \\ &= 0,2222 \cdot 0,2 \cdot 0,2857 + 0,3333 \cdot 0,16 \cdot 0,2857 + 0,3333 \cdot 0,2 \cdot 0,2041 = 0,0415 \end{aligned}$$

2.3. Model M/G/1/∞/∞

Do tej pory zakładaliśmy jedynie rozkład wykładniczy lub Poissona dla zmiennych losowych opisujących zgłoszenia i obsługę. Założenie można jednak oczywiście uchylić. Staje się to konieczne, kiedy zgłoszenia są z góry zaplanowane lub uregulowane. Ponadto rzeczywisty rozkład czasu obsługi również nie zawsze da się opisać przy pomocy rozkładu wykładniczego – szczególnie, gdy wymagania klientów są dość podobne. W konsekwencji każdorazowa obsługa zgłoszenia jest praktycznie identyczna i wariancja czasu jej trwania staje się bardzo niewielka. Przypomnijmy, że wariancja dla rozkładu wykładniczego okazuje się spora w porównaniu do wartości oczekiwanej. Niestety uchylenie tych założeń zazwyczaj czyni analizę dużo bardziej skomplikowaną i wykracza poza ramy tej książki. Przedstawimy jednak model, który da się względnie łatwo wykorzystać.

Model M/G/1/∞/∞, bo o nim mowa, zakłada jeden kanał obsługi – zgłoszenia podlegające rozkładowi Poissona (odstęp między zgłoszeniami podlegają rozkładowi wykładniczemu). Każde zgłoszenie obsługiwane jest zgodnie z tym samym, dowolnym rozkładem, ale tym razem nie zakładamy z góry jego postaci. Jedynie, czego wymagamy, to znajomość wartości oczekiwanej ($1/\mu$) oraz wariancji σ^2 tego rozkładu. Poniżej przedstawiamy charakterystyki liczbowe tego modelu.

Oczekiwana długość kolejki³:

$$Q = \frac{\lambda^2 \sigma^2 + \rho^2}{2(1 - \rho)} \quad (285)$$

Oczekiwana liczba zgłoszeń w systemie:

$$N = \rho + Q \quad (286)$$

Oczekiwany czas pobytu w kolejce:

$$W = \frac{Q}{\lambda} \quad (287)$$

Oczekiwany czas pobytu w systemie:

$$R = W + \frac{1}{\mu} \quad (288)$$

Prawdopodobieństwo tego, że w systemie brak zgłoszeń:

$$P_0 = 1 - \rho \quad (289)$$

Przykład 6

Wróćmy do sytuacji opisanej w przykładzie 1. Dyrekcja zapoznała się z wynikami pochodzącymi z obu modeli typu M/M/1/∞/∞ i uznała, że należy poprawić przynajmniej niektóre właściwości procesu produkcji. W związku z tym, kierownik produkcji otrzymał polecenie zmodyfikowania ustawień maszyny w taki sposób, aby odchylenie standardowe czasu obsługi każdego z detali wyniosło dwie sekundy. Pozostałe parametry wejściowe pozostają bez zmian. Kierownik wykorzystał więc model M/G/1/∞/∞ i dokonał stosownych obliczeń, których wyniki zestawiał w tabeli 89.

Tab. 89. Wartości oczekiwane charakterystyk liczbowych dla przykładu 6

Nazwa miary	Detal A	Detal B
Stopa przybyć (λ , [szt./min])	4	5
Odstęp czasu ($1/\lambda$, [min/szt.])	0,25	0,2
Stopa obsługi (μ , [szt./min])	6	8
Czas obsługi ($1/\mu$, [min/szt.])	0,1667	0,125
Odchylenie stand. (σ , [min])	0,0333	0,0333
Intensywność ruchu (ρ)	0,6667	0,625

3 Jest to tzw. wzór Chinczyna–Pollaczka, opracowany przez dwóch badaczy (niezależnie od siebie) w latach trzydziestych XX wieku.

Nazwa miary	Detal A	Detal B
Oczek. liczba detali w kasecie i w obróbce (N , [szt.])	1,36	1,1829
Oczek. liczba detali czekających na obróbkę (Q , [szt.])	0,6933	0,5579
Oczek. czas pobytu w kasecie i w obróbce (R , [min])	0,34	0,2366
Oczek. czas oczekiwania na obróbkę (W , [min])	0,1733	0,1116

Źródło: opracowanie własne.

Uwagę zwracają charakterystyki opisujące długość kolejki i spędzany w niej czas. Liczba detali oczekujących w kolejce na obróbkę w maszynie spadła o połowę. To spore zwiększenie wydajności produkcji, które ma też dalej idące konsekwencje – można w tym samym czasie realizować większe zamówienia. Mniejszą, ale nadal znaczącą poprawę (nieco ponad 30%) obserwujemy także, gdy chodzi o średnią liczbę detali znajdujących się w kasecie i w obróbce. Spadnie więc prawdopodobieństwo wystąpienia błędu podajnika. Biorąc powyższe pod uwagę, dyrekcja zaleciła pozostać przy nowych ustawieniach urządzenia.

Szczególnym przypadkiem modelu $M/G/1/\infty/\infty$ okazuje się $M/D/1/\infty/\infty$, w którym rozkład czasów obsługi jest deterministyczny, tzn. czas obsługi jest stały dla każdego zgłoszenia, zatem $\sigma = 0$. Zmianie ulega wtedy formuła na oczekiwaną długość kolejki, przyjmując postać (291):

$$Q = \frac{\rho^2}{2(1-\rho)} \quad (290)$$

Pozostałe charakterystyki obliczamy jak wcześniej. Przykład wykorzystania modelu $M/G/1$ z losowo generowaną kolejką znajdziemy w literaturze przedmiotu (Rietman i Resing, 2004). Z kolei inny autorzy (Sherzer i Kerner, 2018) rozszerzają ten model o rezygnację z obsługi i łączą go z równowagą Nasha, znaną z teorii gier konfliktowych.

Może zastanawiać, czemu w zaprezentowanych powyżej modelach tak często przynajmniej jeden z parametrów (wielkość populacji lub wielkość kolejki) występuje jako wielkość nieskończona. Otóż – takie modele, wbrew pozorom, łatwiej analizować. Przyjęcie przez któryś z tych dwóch parametrów skończonej wartości sprawia, że w charakterystykach liczbowych trzeba uwzględnić dodatkowy czynnik wpływający na wynik obliczeń. Nieskończone populacje pozwalają upraszczać założenia. Dodatkowe modele i formuły z nimi związane czytelnik znajdzie w literaturze przedmiotu (zob. Blumenfeld, 2001; Filipowicz, 1996).

Rozdział V

Prognozowanie popytu

Tworzenie właściwych i dobrych prognoz jest niezbędne w wielu obszarach działalności zarówno naukowej, jak i przemysłowej, handlowej, gospodarczej czy społecznej. Najczęściej konstruowane są one dla zjawisk tzw. niewymagających objaśniania – niepotrzebne są żadne dodatkowe informacje (wiedza o czynnikach w charakterze przyczyn determinujących analizowane zjawisko) poza szeregiem czasowym, w którym na podstawie przebiegu danych historycznych (w okresie *ex post*) identyfikowane są pewne prawidłowości, służące jako baza do ekstrapolacji modelu prognostycznego poza dostępną próbą badawczą¹ (Broc-kwell i Davis, 2016; Chatfield, 2000).

Prognozowanie podejmowane jest w celu sterowania decyzjami, np. przyszłej sprzedaży – w oczywisty sposób ułatwia ono planowanie produkcji (Dittmann, 2000). Podkreślając różnorodność aplikacyjną zadań prognostycznych, można wskazać główne obszary, w których wykorzystywane jest przewidywanie oraz sfery, gdzie decyzje wspomaga się uzyskiwanymi prognozami. Do najczęstszych należą (Nowakowska-Grunt i Nowakowska, 2012; Chatfield, 2000):

1. Planowanie i kontrola operacyjna – przedsiębiorstwa rutynowo prognozują np. poziom sprzedaży (bądź popytu) w celu wspomagania podejmowania decyzji w zakresie zarządzania zapasami, sprzedażą i planowaniem przyszłej produkcji, a także planowania strategicznego w odniesieniu do linii produktów czy wejścia na nowy rynek zbytu. Przedsiębiorcy wykorzystują wyniki predykcji, by decydować:
 - co produkować – jaki produkt lub który koszyk produktów powinien zostać wytworzony;
 - kiedy produkować – czy powinni dzisiaj gromadzić zapasy w oczekiwaniu na duże przyszłe zapotrzebowanie; jaki system zmianowy produkcji wprowadzić;

1 Według tego schematu tworzone są najczęściej prognozy dotyczące następujących zjawisk: planowania ekonomicznego, sprzedaży, produkcji i zdolności produkcyjnych, budżetowania, kontroli zapasów, zarządzania ryzykiem finansowym czy oceny alternatywnych strategii/scenariuszy.

- ile produkować, jaka powinna być zdolność produkcyjna przedsiębiorstwa – jakie są trendy w wielkości rynku i udziale rynkowym firmy; czy występują efekty sezonowe lub cykliczne; jak szybko i według którego schematu przebiega amortyzacja nowo wybudowanej fabryki lub nowo zainstalowanej technologii;
 - gdzie produkować – czy należy posiadać jedno miejsce produkcji czy kilka lokalizacji, a jeśli rozważa się wiele placówek, to gdzie powinny się znaleźć.
2. Marketing – prognozowanie odgrywa kluczową rolę w wielu decyzjach marketingowych, które dotyczą poziomu cen, ścieżki dystrybucji i wydatków reklamowych, opieranych na prognozach reakcji poziomów sprzedaży na różne programy marketingowe.
 3. Gospodarka, sytuacja ekonomiczna – organy państwowe, organizacje polityczne i przedsiębiorstwa prywatne, których działalność koncentruje się na prognozowaniu, rutynowo kontrolują główne zmienne ekonomiczne, tj. Produkt Krajowy Brutto (PKB), stopę bezrobocia, poziom konsumpcji, inwestycji, cen i stóp procentowych. Rządy wykorzystują wyniki prognoz tego typu kategorii ekonomicznych do kreowania polityki pieniężnej i fiskalnej, a prywatni odbiorcy do planowania strategicznego, gdyż źródło wahań gospodarczych upatrują w całej gospodarce – jako przyczynę zmian obserwowanych na poziomie branży i pojedynczego przedsiębiorstwa. Oprócz prognozowania „typowych” zmiennych, np. PKB, ekonomiści czasami przeprowadzają badania związane z:
 - analizą etapów cyklu koniunkturalnego, w którym np. dana gospodarka znajdzie się za sześć miesięcy (czy będzie to ekspansja, czy spadek);
 - stanu przyszłej aktywności na giełdzie papierów wartościowych (hossa/bessa), lub aprecjacji/deprecjacji przyszłej działalności na rynku walutowym.
 4. Zarządzanie aktywami finansowymi – zarządzający portfelami są zainteresowani prognozowaniem zwrotów z aktywów² (zwrotów z akcji, stóp procentowych, kursów walutowych i cen towarów).
 5. Zarządzanie ryzykiem finansowym – predykcja zmienności poziomów zwrotu z aktywów wiąże się z samym prognozowaniem aktywów. W ciągu ostatnich dekad opracowano i szeroko stosowano praktyczne metody prognozowania zmienności, które wykazują kluczowe znaczenie dla oceny i zabezpieczenia przed ryzykiem związanym z ustalonymi portfelami aktywów.

2 Debata na temat powodzenia prognoz zwrotu aktywów ma długą historię. Zwroty z aktywów powinny być bardzo trudne do przewidzenia – gdyby okazywały się bowiem łatwe, można by w prosty sposób zbić fortunę, a wszelkie metody pozwalające na „szybkie wzbogacenie się” byłyby już dawno zużyte. W tym przypadku proste, szeroko dostępne metody prognozowania winny przynieść niewielkie sukcesy na rynkach finansowych, aczkolwiek istnieją szanse zysku poprzez zastosowanie nowoczesnych technik do identyfikacji i wykorzystywania – wcześniej niedostrzeganych w danych finansowych – wzorców/prawdowości (przynajmniej w krótkim okresie).

Prognozy dotyczące zmienności (ryzyka) są również niezbędne dla funkcjonowania przedsiębiorstw i inwestorów, którzy muszą permanentnie wyceniać aktywa, tj. opcje i inne instrumenty pochodne.

6. Budżetowanie biznesowe i publiczne – przedsiębiorstwa i organy państwowe muszą stale planować i uzasadnić ponoszone wydatki, zaś ważnym elementem procesu budżetowania jest prognoza dotycząca potencjalnych dochodów. Duża część przychodów przedsiębiorstw pochodzi zazwyczaj ze sprzedaży towarów i świadczonych usług, a dochodów państw – z wpływów podatkowych; źródła te wykazują wahania długoterminowe.
7. Sytuacja demograficzna – populacje krajów i regionów na całym świecie często prognozowane są w zdezagregowanej strukturze, np. według wieku, płci czy aktywności zawodowej. Prognozy dotyczące ludności (narodzin, zgonów, salda migracji) są kluczowe w planowaniu wydatków publicznych m.in. na: opiekę zdrowotną, infrastrukturę, systemy zabezpieczenia społecznego czy walkę z ubóstwem. Wiele decyzji sektora prywatnego, tj. strategiczne decyzje dotyczące linii produktów, opierane są na prognozach demograficznych poszczególnych grup docelowych populacji.
8. Zarządzanie kryzysowe – w ramach zagadnień związanych z tym typem zarządzania często przewidywane są rozmaite zdarzenia odpowiadające różnego rodzaju kryzysom (np. gospodarczym czy naturalnym). Takie prognozy są najczęściej publikowane w formie oceny prawdopodobieństwa zdarzenia – np. w przypadku kredytów konsumpcyjnych i komercyjnych banki określają przewidywane prawdopodobieństwa niewykonania zobowiązania i odmawiają udzielania pożyczki, jeśli poziom tego prawdopodobieństwa zostanie uznany za zbyt wysoki.

Wszystkie z wymienionych obszarów zastosowań, według definicji, łączą się w problemach logistycznych przedsiębiorstw i łańcucha dostaw – czy chociażby determinują otoczenie zewnętrzne, w którym przedsiębiorstwo funkcjonuje (Christopher, 1996). Można zatem stwierdzić, że „dobre” przewidywanie jest podstawą wielu procesów realizowanych w ramach szeroko rozumianej działalności logistycznej – nie tylko zagadnień związanych z utrzymaniem odpowiedniego poziomu stanów magazynowych czy podaży produkcji, mających zaspokoić zgłaszany na rynku popyt.

1. Zarys teorii popytu w ekonomii

Retoryka przyjęta przez wielu ekonomistów, konsultantów i innych rozmaitych „ekspertów” podważyła ideę, że podaż i popyt określają pewne tendencje, a nie „prawo”. Dzięki temu wstępnemu wykluczeniu można stwierdzić, że doświadcze-

nia funkcjonowania rynku oparte są na notowanych, wielokrotnie powtarzających się procesach lub obserwowanych trendach. W trakcie analiz tych tendencji dostrzega się w ich przebiegach trwałe wzorce, a w konsekwencji opracowuje się zestaw propozycji determinant opisujących zachowania uczestników rynku, a co za tym idzie – czynników warunkujących procesy rynkowe (Ye i inni, 2020).

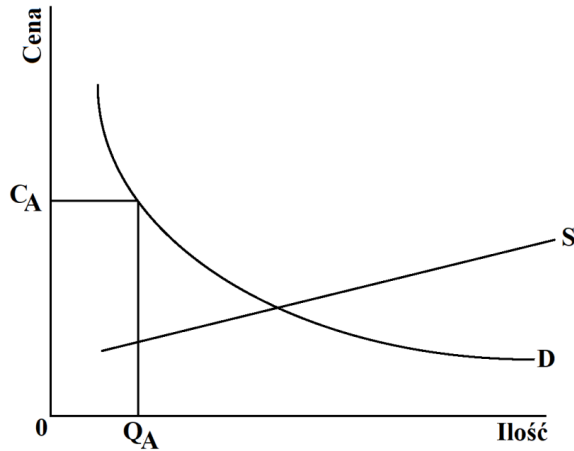
Te określenia obejmują takie stwierdzenia, jak „racjonalność zachowań” i „indywidualne zainteresowanie uczestników rynku” (Prasch, 2008; Varian, 1997; Abolghasemi i inni, 2020). Pozwoliło to na sformułowanie tezy, iż uczestnik rynku zawsze będzie szukał najlepszych produktów w najniższych cenach. Dokładne zrozumienie działania rynków wymaga jednak integracji wielu teorii, a nie jednej – dotyczącej gry podaży i popytu.

Z punktu widzenia logistyki i samego zagadnienia prognozowania sprzedaży (popytu na dobra) kategorie rynków można zawęzić do rynków towarowych, obejmujących przedmioty powszechnie dostępne lub codziennego użytku – wykorzystywane na użytek własny (Mendes, 2011). Wyłącznym powodem, dla którego konsument nabywa dany towar, jest zaspokojenie ściśle określonej potrzeby lub grupy potrzeb. Przykładami takich towarów mogą być płyn do mycia naczyń czy papier toaletowy. Towar, w przeciwieństwie do innych dóbr, tzw. doświadczalnych, których cechy ujawniają się dopiero po zakupie lub użyciu, posiada właściwości łatwo identyfikowalne poprzez bezpośrednią obserwację nawet przez niedoinformowanych nabywców (Prasch, 2008) – jest to niezwykle istotna charakterystyka, przydatna w definiowaniu np. struktury jednostki logistycznej, w skład której mogą wchodzić różne jednostki handlowe, przeznaczone do dalszego obrotu (np. detalicznego).

W analizach dotyczących zapotrzebowania należy założyć, że konsumenci mają względnie pełną wiedzę na temat właściwości kilku konkurencyjnych na danym rynku towarów oraz – nawet przy uwzględnieniu wpływu np. norm kulturowych (wynikających chociażby z lokalizacji rynku) – popyt zgłaszany przez konsumentów na towary jest zmienną niezależną (Prasch, 2008). Oznacza to, że wybory każdego człowieka i zestawy ostatecznych żądań są ustalane niezależnie od ocen dokonywanych przez innych konsumentów na rynku lub w danej populacji.

Zgodnie z definicją popyt to ilość towaru, jaką konsumenci są skłonni kupić w jednostce czasu, przy różnych poziomach cen (Milewski, 2018; Nojszewska, 2010). Zakłada się, że jeśli cena towaru zostanie obniżona, racjonalny konsument wykaże wyższy poziom zapotrzebowania na dany produkt – i analogicznie: im wyższa cena danego towaru, tym zgłaszane zapotrzebowanie będzie niższe *ceteris paribus*. Wynika z powyższego, że popyt (D) jest odwrotnie proporcjonalny do ceny (C) – gdy cena jest stosunkowo wysoka (C_A), związany z nią poziom popytu okazuje się niski (Q_A) – zob. rysunek 44.

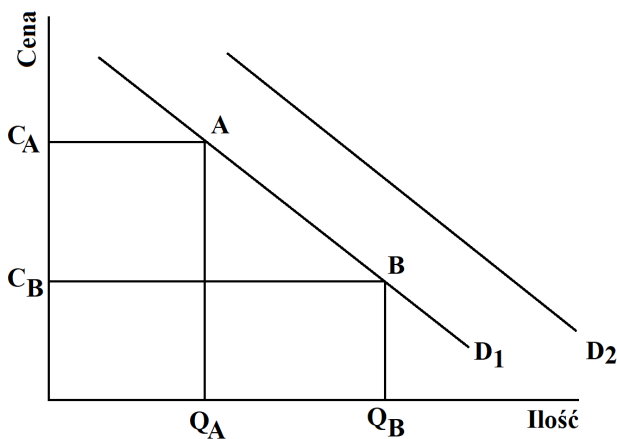
Gdy cena produktu zostaje obniżona, żądana ilość dobra (Q) wzrasta. Tendencję tę odwzorowuje nachylenie krzywej popytu (D), na której niższe poziomy ceny (C) towarzyszą wyższym poziomom ilości dobra (Q).



Rys. 44. Relacja popytu (D), podaży (S), ceny (C_A) i ilości towaru (Q_A)

Źródło: opracowanie własne.

Należy zauważyć, że ruchy cen wywołują zmiany w ilościach – i są przedstawiane przez odpowiednią zmianę popytu po stronie krzywej popytu D_1 (przejście z pkt. A do B na rysunku 45). Natomiast o wzroście popytu świadczy ruch krzywej D_1 do pozycji D_2 . Zmiany cen wywołują zmiany w żądanej ilości – i są zilustrowane ruchami wzdłuż ustalonego przebiegu krzywej popytu. Zmiany poziomu popytu z dowolnej przyczyny (innej niż cena towaru) skutkują zmianą położenia krzywej popytu (D) w prawo (wzrost popytu) lub do wewnątrz w lewo (spadek popytu).



Rys. 45. Przesunięcie popytu (D_1 na D_2) w odniesieniu do zmiany ilości (od Q_A do Q_B)

Źródło: opracowanie własne.

Podstawową zasadą leżącą u podstaw teorii popytu jest idea substytucji (Milewski, 2018). Uważa się, że konsumenci, wprost lub pośrednio, porównują cenę i cechy wielu towarów względem swoich możliwości finansowych oraz intensywności potrzeb, które mają nadzieję zaspokoić. Jeśli w wyniku tego porównania okaże się, że przy uwzględnieniu poziomów cen, towar konkurencyjny o podobnych właściwościach w pełni zaspokoi potrzeby konsumenta lub zaspokoi je po niższym koszcie, wówczas racjonalny konsument dokona zakupu na innym rynku. Ta zasada, zwana efektem substytucji, ilustrowana jest poprzez nachylenie krzywej popytu (D), przedstawiające ideę odwrotnej relacji między ceną a ilością. Efekt substytucyjny okazuje się wyjaśnieniem, dlaczego popyt na towar maleje wraz ze wzrostem jego ceny – jeśli cena rynkowa wzrośnie, konsumenci mniej przywiązani do danego produktu (jego właściwości i cech charakterystycznych) będą szukać tańszych zamienników.

W tym sensie wysokie ceny „racjonują” towary wśród nabywców, którzy wykazują zainteresowanie, by wejść w ich posiadanie. Zatem racjonowanie odbywa się raczej w zależności od poziomu gotowości (Jewczak, 2017) oraz, co jest równie ważne, zdolności do zapłaty – w tej koncepcji (Smith, 2010) wskazuje się na popyt efektywny, ponieważ towar nie zawsze może zostać wprowadzony na rynek w celu zaspokojenia potrzeb.

Jak już wspomniano, zmiany ograniczeń innych niż cena rynkowa są reprezentowane przez zmianę położenia krzywej popytu. Uwarunkowania, które powodują taką modyfikację, obejmują zmiany w (Prasch, 2008):

- poziomach dochodów obecnych i potencjalnych nabywców;
- gustach i preferencjach kupujących;
- dostępności lub ceny towarów, które są substytucyjne bądź komplementarne względem danego produktu;
- dostępności form finansowania zakupu, np. kosztów kredytu konsumentkiego;
- stanu oczekiwań konsumentów odnośnie przyszłych warunków rynkowych, w tym przyszłej ceny i dostępności tego towaru (efekt spekulacyjny).

Uważa się, że zmiany którekolwiek z tych warunków powodują przesunięcie położenia funkcji popytu; natomiast odległość i forma tej modyfikacji zależą od zakresu zmieniającego się warunku oraz reakcji konsumentów na daną zmianę.

Problem prognozowania popytu nie zawsze sprowadza się zatem do analizy jednej kategorii, np. poziomów sprzedaży – dość często staje się zjawiskiem wielowymiarowym, które uwzględnia zarówno czynniki indywidualne, tj. preferencje nabywców, jak i uwarunkowania zewnętrzne, np. przewidywaną sytuację na rynku czy potencjalną dostępność (niedobór) danego towaru w kolejnych jednostkach czasu (Chatfield, 2000; Suchecki 2006).

Postrzeganie związków przyczynowych między ceną a popytem jest często różne dla dyskursu biznesowego czy adoptowanej polityki publicznej. Oprócz odpowiedzi na pytania typu: „w którą stronę”, należy zmieniać uwarunkowania – często

wymaga się również oceny dotyczącej wymiaru („ile”). W sytuacji, gdy przedsiębiorstwo zakładało obniżenie ceny swoich towarów, istotne staje się określenie, jakiego wzrostu ilości może ono oczekiwać. Ten problem pozwala poniekąd rozwiązać koncepcja „elastyczności”, która wykazuje przydatność dla przewidywania i symulacji poziomów popytu na towary (Mendes, 2011).

Formuła elastyczności jest stosunkowo intuicyjna. Z matematycznego punktu widzenia elastyczność okazuje się miarą względnych zmian funkcji, wywołanych określonym względnym przyrostem zmiennej niezależnej (Prasch, 2008):

$$E_{xy} = \lim_{\Delta x \rightarrow 0} \left(\frac{\Delta y}{y} : \frac{\Delta x}{x} \right) = \frac{\Delta y}{y} / \frac{\Delta x}{x} \quad (291)$$

W sensie ekonomicznym elastyczność określa siłę reakcji (wrażliwość) jednej wielkości, traktowanej jako zmienna zależna, np. y (reakcja), na zmianę innej wielkości, rozumianej jako bodziec, np. x (zmienna niezależna). W przypadku funkcjonowania rynku tymi wielkościami mogą być podaż i popyt, reagujące na zmianę ceny (lub odwrotnie: cena, która reaguje na zmiany w relacji podaż/popyt). Rozróżnia się zatem elastyczności:

- podaży względem ceny, tzn. elastyczność cenowa podaży;
- popytu względem ceny, tzn. elastyczność cenowa popytu.

Odwrotność tych elastyczności określa się mianem, odpowiednio: ekspansywności ceny i giętkości ceny. Miarą elastyczności jest współczynnik elastyczności, który wskazuje, o ile procent (lub innych jednostek) zmieni się wielkość zmiennej zależnej, gdy bodziec (zmienna niezależna) zmieni się o jeden procent (o jedną jednostkę). Jeżeli zatem zmienna zależna wzrośnie o 10% przy 5% wzroście zmiennej niezależnej, to współczynnik elastyczności $E_{xy} = 2$. Oznacza to, że zmienna y jest wrażliwa (elastyczna) na zmianę zmiennej x (relatywna zmiana y jest dwukrotnie wyższa od relatywnej zmiany x).

Niska elastyczność popytu ($E_{xy} < 1$) może wystąpić, gdy konsumenci uważają, że dany produkt jest niezbędny, a jego unikatowe cechy utrudniają identyfikację odpowiednich zamienników w krótkim lub średnim okresie. Taki towar reprezentuje „potrzebę” w przeciwieństwie do „zachcianki”. Skrajnym przykładem może być zapotrzebowanie na insulinę przez osobę chorującą na cukrzycę. Niska elastyczność popytu na dany produkt dotyczy niekiedy też sytuacji, w której dany towar pochłania niewielki procent dochodów rozporządzalnych konsumenta – wtedy racjonalny nabywca rzadziej porównuje opcje rynku, jeśli wydatek w mniejszym stopniu obciąża jego możliwości dochodowe.

Wysoka elastyczność ($E_{xy} > 1$) implikuje, że ilość danego towaru jest bardzo wrażliwa na zmiany cen. Ten wysoki poziom reakcji sugeruje, że konsumenci uważają towar albo za nieistotny dla funkcjonowania, albo za mający wiele bliskich substytutów. W jednej z tych przesłanek 1% wzrost ceny tego towaru spowoduje

ponad 1% spadek sprzedawanej ilości – ponieważ konsumenci będą rezygnowali z zakupu przedmiotu bądź też zapoznają się z ofertą rynku w poszukiwaniu bliskiego substytutu. Sytuacja jest również analogiczna, gdy dobro o wysokiej elastyczności stanowi wysoki stopień obciążenia dochodu rozporządzalnego – konsumenci wówczas zbierają więcej informacji i szukają większej liczby substytutów.

Dla działającego na rynku przedsiębiorstwa podstawowe znaczenia ma znajomość popytu i jego wrażliwości na zmiany determinujących go czynności (w tym ujęciu popyt uważany jest za zmienną zależną). Za zmienne niezależne przyjmuje się czynniki kształtujące popyt, wśród których wskazywane są: wielkości dochodów, poziomy ceny – zarówno danego dobra, jak i dóbr substytucyjnych/komplementarnych. W zależności od tego, jaki z czynników determinuje wielkość popytu przedsiębiorstwa, wyróżnia się trzy rodzaje elastyczności popytu (Milewski, 2018):

- dochodową³, tj. elastyczność funkcji popytu na określone dobro (grupę dóbr) względem poziomu dochodu;
- cenową⁴, tj. elastyczność funkcji popytu względem poziomu ceny danego dobra;
- mieszaną (określaną również mianem krzyżowej), tj. elastyczność funkcji popytu na dobro i względem ceny dobra j .

Znajomość współczynników elastyczności popytu na różne dobra pozwala na dokonywanie uproszczonej predykcji (wnioskowanie w przyszłości). Proste przekształcenie elastyczności daje narzędzie prognozy popytu:

$$\frac{\Delta D}{D} = E_x \left(\frac{\Delta x}{x} \right) \quad (292)$$

3 Elastyczność dochodowa popytu kształtuje się różnie w zależności od rodzaju dobra oraz poziomu dochodów konsumentów. Na podstawie wartości liczbowych współczynnika można dokonać podziału dóbr konsumpcyjnych na trzy główne grupy:

- niższego rzędu, dla których współczynnik przyjmuje wartości ujemne ($E_d < 0$);
- pierwszej potrzeby o elastyczności dochodowej $0 < E_d < 1$ (im produkt jest bardziej niezbędny nabywcy, np. podstawowe dobra żywnościowe, tym mniej wrażliwy jest popyt na zmiany dochodów);
- dobra wyższego rzędu (luksusowe), charakteryzujące się współczynnikiem $E_d > 1$.

4 Przedsiębiorstwo może wykorzystywać informacje płynące z analizy współczynników elastyczności cenowej do projektowania swojej polityki, względem np. potencjalnego zysku. Jeśli w pewnym przedziale cenowym, o stałej tendencji, elastyczność popytu:

- wynosi $E_p < -1$ (popyt elastyczny) – wówczas należy unikać wszelkiego rodzaju podwyżek cen, gdyż będą one skutkowały obniżeniem utargu całkowitego;
- przyjmuje wartości $(-1 < E_p < 0)$ – to przy popycie nieelastycznym należy unikać obniżek cen, ponieważ przyczyniają się one do zmniejszenia utargu całkowitego;
- $E_p = -1$ – wtedy przy popycie proporcjonalnym niewielkie zmiany ceny nie wpływają na wielkość utargu.

Zakładając, że wszystkie zmienne objaśniające występujące w funkcji popytu (oprócz jednej zmiennej x) nie ulegają zmianie, można przewidzieć, że jeżeli zmienna niezależna x wzrośnie o 1%, to popyt na dane dobro (D) zmieni się o E_x [%] – wzrośnie lub zmaleje w zależności od znaku współczynnika elastyczności.

Przyпускаjąc, że spodziewany jest wzrost dochodów ludności w ciągu najbliższych trzech lat o 8%, zaś współczynnik elastyczności dochodowej popytu na towar x wynosi $E_d = 0,51$, można wywnioskować, że w ciągu trzech lat popyt na towar x wzrośnie w przybliżeniu o 4,1%. Oczywiście sytuacja zakłada, że w tym samym okresie cena towaru x i dóbr pokrewnych pozostaje na niezmiennym poziomie.

Gdyby jednak scenariusz predykcyjny zakładał, że cena towaru x w tym okresie wzrośnie prawdopodobnie o 5%, a współczynnik elastyczności cenowej popytu kształtuje się na poziomie $E_p = -0,48$, możliwe jest oszacowanie wpływu na poziom popyt zmiany obu zmiennych niezależnych. Dodając do siebie wyniki elastyczności cząstkowych notowanych zmian:

$$\frac{\Delta D}{D} = E_p \left(\frac{\Delta C}{C} \right) + E_d \left(\frac{\Delta D}{D} \right) \quad (293)$$

otrzymamy $\frac{\Delta D}{D} = (-0,48) \times 5\% + 0,51 \times 8\% = 1,7\%$. Przewiduje się że popyt

(sprzedaż) wzrośnie w analizowanych okresie w przybliżeniu o 1,7%.

Planowanie i kontrola systemów logistycznych wymaga prognoz dotyczących poziomu przyszłej działalności gospodarczej ze względu na opóźnienia – np. w dopasowaniu poziomów podaży do popytu. Typowe decyzje, które należy podjąć przed poznaniem niektórych danych, uwzględniają praktycznie każdy aspekt procesu planowania (w tym lokalizację zakładu i jego wydajności), a także produkcji, zarządzania zapasami i transportu.

Prognozowane wymagania logistyczne obejmują takie zagadnienia, jak: zapotrzebowanie klientów, ceny surowców, koszty pracy i terminy realizacji zamówień. W rzeczywistości, aby wygenerować prognozę dotyczącą poziomu zapotrzebowania, popyt musi wykazywać pewien stopień prawidłowości. Przykładowo, funkcja popytu musiałaby mieć w przyszłości podobny przebieg, co w okresach wcześniejszych – lub popyt bieżący w pewnym stopniu powinien zależeć od przeszłych wartości czynników go determinujących.

Gdy popyt charakteryzuje się przebiegiem nieregularnym, istnieje tak wiele losowości we wzorcu popytu, że dokonanie wiarygodnej prognozy może być bardzo utrudnione. Zwykle ma to miejsce w sytuacji, kiedy w strukturze popytu dominują rzadko składane, duże zamówienia klientów lub gdy wolumen każdej pozycji w zamówieniu jest niski (a stopień dywersyfikacji produktów jest wysoki). W przypadku takich przedmiotów należy badać dwie alternatywy. Jeśli popyt okazuje się niski, dokładność nie jest zwykle kluczową kwestią i można założyć

jego przeszacowanie (co z kolei doprowadzi do zwiększenia zapasów). Alternatywnie, procesy w łańcuchu dostaw (a mianowicie: wytwarzanie i transport) uelastycznia się, aby uzyskać szybką reakcję po stronie popytu (umożliwiają to systemy MTO – ang. *Make To Order*).

Prognozy popytu są konstruowane według trzech głównych horyzontów czasowych. Prognozy długoterminowe obejmują horyzont czasowy od jednego do pięciu lat, zaś predykcja w dłuższym horyzoncie czasowym staje się bardzo niewiarygodna – ze względów np. technologicznych. Prognozy długoterminowe służą podejmowaniu decyzji dotyczących wprowadzenia lub wycofania produktu z rynku, a także projektowaniu sieci logistycznej. Takie prognozy są często generowane dla grupy towarów (lub usług), a nie dla pojedynczej pozycji.

Prognozy średnioterminowe obejmują okres od kilku miesięcy do jednego roku i służą do podejmowania taktycznych decyzji logistycznych, takich jak ustalanie rocznych planów produkcyjnych i dystrybucji, zarządzania zapasami i alokacji przestrzeni w magazynach. Prognozy krótkoterminowe obejmują zaś okres od kilku dni do kilku tygodni. W działalności logistycznej wykorzystywane są do planowania i modyfikacji rozkładów zasobów w celu osiągnięcia średniookresowych celów produkcji i dystrybucji. Średnio- i długoterminowe prognozy popytu prawie nigdy nie są pozostawione logistykom – częściej zadania te przypisuje się menedżerom marketingu, próbującym stymulować popyt, np. poprzez uruchomienie kampanii reklamowej dla produktów, których sprzedaż spada. Jednak logiści często odpowiadają za krótkoterminowe prognozy popytu.

W większości przypadków klienci są rozproszeni geograficznie; warto również uwzględnić czynnik przestrzenny w predykcji, by uzyskać odpowiednie wskazania dotyczące tego kiedy i gdzie wystąpi dany poziom zapotrzebowania. Jest to dość istotne z punktu widzenia działalności logistycznej, gdyż na decyzje (takie jak lokalizacja magazynu i określenie poziomu zapasów) wpływa przestrzenna lokalizacja popytu.

W praktyce logistycznej metody prognozowania wykorzystywane są w wielu dziedzinach i na różnych poziomach struktury przedsiębiorstwa. Zainteresowania niektórych badaczy pokazują, że przedmiotem analiz i predykcji mogą być takie zjawiska, jak: poziom sprzedaży pierogów czy eksportowanego mięsa, ruch pasażerski sektora lotniczego czy bezpieczeństwo na drogach publicznych (Ghiani i inni, 2004, s. 31; Osińska, 2012, s. 103–116).

Nieustannie rosnący przepływ materiałów/towarów, środków finansowych i informacji w łańcuchach dostaw poszczególnych branż sprawia, że przedsiębiorstwa logistyczne częściej zdają sobie sprawę z korzyści stosowania progresywnych metod zarządzania. Firmy implementują odpowiednie systemy zarządzania procesami na poszczególnych poziomach funkcjonalnych, a jednocześnie – poprzez zwiększenie objętości i złożoności przepływów zasobów – wdrażają zasady logistyczne zarządzania przepływem materiałów. Z kolei w wyniku wzrostu wolu-

menu wymiany międzynarodowej zwiększa się intensywność przepływów w poszczególnych infrastrukturach logistycznych (miastach), co w efekcie końcowym podnosi znaczenie zasad zarządzania logistyką również w sferze trzeciorzędnej. Tak więc zasady lub podejścia do zarządzania logistyką w obecnym zglobalizowanym otoczeniu rynkowym są kluczowe dla zwiększenia konkurencyjności przedsiębiorstw w kontekście długoterminowego zrównoważonego wzrostu.

2. Podstawowe pojęcia, mierniki oceny jakości prognoz

Metoda prognozowania najczęściej definiowana jest jako procedura obliczania prognoz na podstawie wartości obecnych i przeszłych (Chatfield, 2000; Cieślak, 2008). W tym ujęciu może być ona rozumiana po prostu jako reguła algorytmiczna i nie musi być zależna od właściwego modelu prawdopodobieństwa. Alternatywnie, uznaje się ją za narzędzie identyfikacji konkretnego modelu dla danych, służące znalezieniu prognoz optymalnych zależnych od wykorzystywanego podejścia (Brockwell i Davis, 2016). Z tego względu należy jasno rozgraniczyć dwa przywołane tu terminy: „metoda” i „model”.

W literaturze termin „model prognostyczny” dość często stosowany jest w sposób niejednoznaczny, a czasami wręcz niewłaściwie używany do opisu metody prognostycznej. Metody prognozowania można ogólnie podzielić na trzy główne typy:

- osądowe⁵ – oparte na subiektywnej ocenie, intuicji, wiedzy „wewnętrznej” i wszelkich innych istotnych informacjach;
- jednowymiarowe – prognozy zależą tylko od bieżących i przeszłych wartości prognozowanej pojedynczej serii danych, która może zostać zwiększona o zmienną czasową, np. trend liniowy;
- wielowymiarowe – prognozy danej zmiennej zależą, przynajmniej częściowo, od wartości jednej lub większej liczby dodatkowych zmiennych, zwanych zmiennymi predykcyjnymi lub objaśniającymi; prognozy wielowymiarowe mogą zależeć od wykorzystywanego do opisu połączeń pomiędzy zmiennymi modelu wielorównaniowego, obejmującego więcej niż jedno równanie – jeżeli zmienne są współzależne.

5 Najstynniejszą metodą osądową jest prawdopodobnie technika Delphi (zob. Rowe i Wright, 1999), której celem staje się wypracowanie konsensusu opinii grupy „ekspertów” na podstawie serii kwestionariuszy wypełnianych przez poszczególne osoby, przeplatanych kontrolowaną informacją zwrotną opinii i informacji od innej grupy ekspertów.

Mówiąc ogólniej, metoda prognozowania może łączyć więcej niż jedno z powyższych podejść, np. w sytuacji gdy jedno- lub wielowymiarowe prognozy są dodatkowo dostosowywane subiektywnie – by uwzględnić informacje zewnętrzne, które trudno formalnie wyrazić w modelu matematycznym (Chatfield, 2000; Diebold, 1998).

Metody osądowe czasami działają dobrze, choć często rezultaty wypracowanych opinii są nie do przyjęcia (podobna sytuacja może charakteryzować każdą z metod prognozowania). Uznaje się zatem, że to metody o podłożu statystycznym pozwalają na uzyskiwanie lepszych prognoz – pod warunkiem, że są one rozwiązaniem praktycznym (Abraham i Ledolter, 2009). Tym samym – zdarzają się sytuacje, gdy metody oparte na modelach okazują się niepraktyczne, co wynika z faktu, że niektóre istotne informacje nie są dostępne w chwili konstruowania predykcji, którą następnie i tak należy poddać ocenie trafności (De Baets i Harvey, 2020). Często dochodzi do połączenia obu podejść w celu uzyskania dokładniejszych oszacowań.

W szczególności wiele predykcji makroekonomicznych uzyskuje się poprzez dostosowanie prognozowania opartego na modelu przez dodanie lub odjęcie odpowiedniej stałej (czasami nazywanej korektą przechwytyjącą). Niestety nie zawsze jest jasne, w jaki sposób dokonuje się tego typu ingerencji w celu uzyskania ostatecznej prognozy. Jak można się spodziewać, integracja podejścia osądowego i statystycznego poprawia dokładność, gdy analityk ma dobrą wiedzę w danej dziedzinie, aczkolwiek – jak wskazują niektórzy autorzy (Armstrong i Collopy, 1998) – pogarsza niekiedy trafność oszacowań (jeśli ocena ta jest stronnicza lub nieuporządkowana).

Alternatywnym i ważnym sposobem klasyfikowania metod prognozowania jest rozróżnienie pomiędzy metodami automatycznymi, niezakładającymi interwencji człowieka, a metodami nieautomatycznymi, które jej wręcz wymagają. Jednym z przykładów tego podziału może być zestawienie kontroli zapasów z planowaniem gospodarczym. W procesie związanym z kontrolą zapasów identyfikuje się tysiące przedmiotów (podlegają one stałemu monitorowaniu) – przez co nie jest możliwe dopasowanie indywidualnych modeli do każdego z zasobów sprzedaży. W sytuacji tego typu najczęściej sięga się po prostą, automatyczną metodę, implementowaną dla całego agregatu produktów. W przeciwieństwie do kontroli zapasów, planowanie gospodarcze wymaga od analityka ostrożności w konstrukcji odpowiedniego modelu, opisującego związek między odpowiednimi zmiennymi ekonomicznymi według schematu modelu, z którego można wygenerować „optymalne” prognozy – to podejście z pewnością nie należy do automatycznych.

Kolejna klasyfikacja metod prognostycznych dzieli narzędzia według ich złożoności – na metody jednowymiarowe i zaawansowane. Pierwsze z nich są na ogół zdecydowanie prostsze w zastosowaniu niż drugie, zaś wybór danego wariantu zależy od nastawienia związanego z wysiłkiem (rozumianym jako nakłady pracy,

kapitału, czasu, niezbędnego wyposażenia technicznego itd.), jaki należy włożyć w samo prognozowanie (Cieślak, 2008; Zeliaś, Pawełek i Wanat, 2008; Chatfield, 2000). W przypadku niektórych kategorii ekonomicznych trudno jest wskazać prostszy sposób konstrukcji prognozy niż wykorzystanie najnowszych obserwacji *ex post* zmiennej do przewidzenia kolejnej obserwacji w okresie *ex ante*.

Pracując na danych szeregów deterministycznych, równie łatwo można prowadzić prace prognostyczne – jeśli znany jest odpowiedni model opisujący zależności. Przykładowo, zmieniającą się okresowo w sposób systematyczny (schematyczny) serię łatwo przewidzieć po zaobserwowaniu pełnego cyklu. Jednak seria niezależnych obserwacji okazuje się niekiedy niemożliwa do prognozowania – ponieważ w takim przypadku jedynym sensownym rozwiązaniem jest wykorzystanie wartości centralnej, np. k -elementowej średniej ruchomej jako schematu konstrukcji prognozy. W modelowaniu/prognozowaniu zmiennych logistycznych najczęściej dostępne są szeregi spomiędzy tych skrajności – nie ma możliwości ich dokładnego przewidywania, ale zawierają one strukturę, której obraz można wykorzystać do stworzenia bardziej dopasowanych (rzeczywistych) prognoz (Duda, 2017).

Metody wykorzystywane do prognozowania, np. szeregów czasowych⁶, pozwalają otrzymać prognozy przyszłych wartości, porównywalnych z tymi faktycznie zaobserwowanymi. Jeśli istnieje dobra zgodność wartości dopasowanych, to można argumentować, że zapewnia ona bardziej przekonującą weryfikację modelu niż ocena ogólnego dopasowania w próbie – np. poprzez analizę współczynnika determinacji czy błędu ogólnego modelu (Szmit, 2007; Gajda, 2001).

Typową metodą analizy serii tego typu jest próbkowanie serii w równych odstępach czasu, w celu uzyskania dyskretnych szeregów czasowych. W wyniku tego procesu praktycznie żadna z informacji nie jest tracona – pod warunkiem, że interwał czasowy próbkowania okazuje się wystarczająco wąski. Źródłem dyskretnych szeregów czasowych może być:

- pobieranie próbek z serii ciągłej (np. chwilowe zużycie energii elektrycznej przez maszynę, mierzone w odstępach godzinnych – dane uzyskiwane są wówczas poprzez pobieranie próbek ciągłego przebiegu zużycia lub z uwzględnieniem wskazanego interwału pomiaru);
- poprzez agregację danych w czasie (np. łączenie dniowych poziomów sprzedaży w kolejnych tygodniach, miesiącach roku);
- seria dyskretna z natury (np. dywidendy wypłacane przez spółkę w kolejnych latach).

6 Szeregi czasowe to zestawy obserwacji mierzone sekwencyjnie wraz z upływem czasu. Innymi słowy, np. dla dyskretnych szeregów czasowych charakter dyskretny wprowadzany jest poprzez oś czasu. Dla ciągłych szeregów czasowych obserwowana zmienna to zazwyczaj zmienna ciągła – np. sposób odpowiadający wynikowi pomiaru aktywności mózgu, zarejestrowanemu na urządzeniu do EEG.

Dla wszystkich typów dyskretnych szeregów czasowych dane są zwykle rejestrowane w równych odstępach (jednostkach) czasu, stąd postępowanie z szeregami dyskretnymi jest zwykle dość zbieżne – pomimo np. różnych interwałów próbkowania dla serii ciągłych lub różnych okresów agregacji dla danych zagregowanych (Abraham i Ledolter, 2009).

Agregacja danych może przebiegać zarówno w czasie, jak i w szeregu (agregacja jednoczesna), np. analizie mają zostać poddane dane dotyczące sprzedaży dla kilku produktów różnych marek, oferowanych w różnych rozmiarach, w kolejnych tygodniach. Takie dane (wielowymiarowe) mogą być dość zmienne i trudne do prognozowania bez odpowiedniej formy agregacji – zarówno w czasie (np. w kolejnych czterotygodniowych okresach), jak i dla poszczególnych produktów (np. sumy wszystkich rozmiarów dla danej marki).

Badacze (Montgomery i inni, 1998) wskazują, że możliwość agregowania danych rodzi wiele problemów, np. wyboru najlepszego poziomu agregacji do sporządzania prognoz czy wykorzystania danych miesięcznych dotyczących sprzedaży, w celu poprawy dokładności prognoz kwartalnych. W przypadku kontroli zapasów kłopotliwym okazuje się dylemat, czy należy opracować prognozę podsumowującą dla agregatu określonej grupy towarów, a następnie przydzielić tę prognozę do poszczególnych pozycji na podstawie ich historycznej częstotliwości względnej (podejście *top-down*), czy utworzyć indywidualne prognozy dla każdego elementu (podejście *bottom-up*)⁷.

Szczególną cechą danych szeregów czasowych jest to, że obserwacje zwykle nie są niezależne, dlatego analiza musi uwzględniać kolejność gromadzenia obserwacji. W efekcie każda obserwacja mierzonej zmiennej staje się obserwacją dwuwymiarową – z uwzględnieniem czynnika czasu jako drugiej składowej. Za główne cele analizy szeregów czasowych przyjmuje się (Brockwell i Davis, 2016):

- opis – przy użyciu statystyk podsumowujących i/lub metod graficznych; sporządzenie np. wykresu szeregu czasowego jest szczególnie cenne przy dekompozycji jego składowych;
- modelowanie – odnalezienie odpowiedniego modelu opisującego proces generowania danych; model jednowymiarowy opiera się wyłącznie na wartościach *ex post* jednej zmiennej, podczas gdy podejście wielowymiarowe może zakładać wykorzystanie nie tylko przeszłych wartości analizowanej zmiennej, ale również bieżących i przeszłych wartości innych zmiennych (predyktorów) – w drugim przypadku zmiana w jednej serii danych pomaga w wyjaśnieniu zmiany zaobserwowanej w innej serii;
- prognozowanie – oszacowanie przyszłych wartości serii danych; istnieje wyraźne rozróżnienie między prognozowaniem w stanie ustalonym, gdzie

7 Byron J. Dangerfield i John S. Morris (Dangerfield i Morris, 1992) sugerują, że podejście odolne (*bottom-up*) jest zwykle bardziej dokładne.

przyjmuje się, że przyszłość będzie bardzo podobna do przeszłości, a sytuacją „co, jeśli”, w której do badania wpływu zmiany implementowany jest scenariusz wielowymiarowy;

- kontrolę – dobre prognozy umożliwiają analitykowi podejmowanie działań w celu kontrolowania danego procesu, niezależnie od tego, czy jest to proces przemysłowy, czy gospodarka (w tym sensie proces kontrolny okazuje się powiązany głównie z prognozowaniem wielowymiarowym).

3. Rodzaje metod prognostycznych wykorzystywanych w prognozowaniu popytu

O wyborze narzędzia prognostycznego (zob. dodatek 1) decydują przesłanki prognostyczne, które obejmują zarówno hipotezy badawcze (dotyczące kształtowania się zjawiska), jak i dostępne o nim informacje. Należy zatem przyjąć odpowiednią postawę (Cieślak, 2008):

- pasywną – według której podmiot będzie postrzegał przyszłość przez pryzmat przeszłości, zakładając, że zidentyfikowane w okresie historycznym mechanizmy są powielane w przebiegu danej zmiennej w okresach przyszłych;
- aktywną – która przyszłość nakazuje traktować jako stosunkowo niezależną od przeszłości.

Jakość uzyskiwanych prognoz zależy również od założeń (identyfikowanych bądź niemożliwych do identyfikacji) aplikowanych metod lub subiektywnych osądów. Zatem prognoza może zostać zdefiniowana ogólnie jako pewne warunkowe stwierdzenie w formie zweryfikowanej przy pomocy metod naukowych, o pewnym poziomie akceptowalnej niepewności – np. że jeśli zdarzenia danego typu będą się utrzymywać w przyszłości, to wówczas poziom zapasów powinien wzrosnąć. Z powyższej interpretacji wynika, że każdorazowo należy zakładać konieczność modyfikacji prognoz w sytuacji jakichkolwiek pojawiających się dodatkowych informacji lub opracować na starcie szereg różnych wariantów prognoz (a nie tylko pojedynczej prognozy), z których każdy opiera się na znanym zestawie jasno określonych założeń (scenariuszy)⁸. Ernst Friedrich Schumacher (Schumacher, 1986) wskazuje również, że prognozy długoterminowe są pewnymi przypuszczeniami – nie przytacza jednak kryteriów, kiedy

8 Ernst F. Schumacher (Schumacher 1986) nazywa to prognozowaniem scenariuszowym, wskazując na sensowność prowadzenia długoterminowych studiów wykonalności, opartych na jasno określonych założeniach.

studium wykonalności staje się właściwą prognozą w rozumieniu definicyjnym. Warto przywołać przykłady błędnych prognoz związanych z założeniami scenariuszy długoterminowych, np. założyciela IBM, przewidującego w 1947 roku miejsce na światowym rynku dla około pięciu komputerów, lub prezesa Digital Equipment, twierdzącego w 1977 roku, że nikt nie ma powodów, aby posiadać w domu komputer.

Oczywiście prognozy mogą generować błędne wskazania również i w krótkim okresie, np. dla procesów ekonomicznych podlegających nagłym zmianom pod wpływem klęski żywiołowej, epidemii czy kryzysu gospodarczego lub tzw. przerwy strukturalnej⁹ w danych.

3.1. Kryteria wyboru metod prognostycznych

W trakcie każdego doświadczenia prognostycznego należy rozważyć sześć istotnych zagadnień, które pozwalają określić podstawy właściwego prognozowania. Zagadnienia te wprowadzają szereg pytań związanych z projektowaniem schematu, celu i realizacji samego badania, a należą do nich (Chatfield, 2000):

- 1) Środowisko decyzyjne i funkcja strat – w ramach tego zagadnienia należy odpowiedzieć na pytania związane z: decyzjami, do których doprowadzi prognoza; wynikającymi z zastosowania i przyjętej oceny danego modelu prognostycznego konsekwencjami dla całego projektu; określeniem sposobu, w jaki będzie rozumiane określenie „dobra” prognoza, a także kosztu lub straty związanej z błędami prognozy; zdefiniowaniem prognozy optymalnej w konkretnej sytuacji i sposobów ich wyznaczania;
- 2) Informacje o obiekcie prognozowanym – zawarte są tu informacje dotyczące: przedmiotu prognozy, rodzaju danych opisujących obiekt (czy jest to szereg czasowy, np. rejestr poziomu sprzedaży przedsiębiorstwa w czasie, czy zdarzenie, tj. dewaluacja waluty), ilości i jakości danych statystycznych (rozpiętość próby, prognozowanie dla jednego lub wielu obiektów, np. sprzedaży każdego z n oferowanych produktów albo całego wolumenu zasobów), brakujących i nietypowych (odstających) obserwacji;
- 3) Określenie prognozy, czyli sposobu, w jaki zostanie przedstawiona prognoza, co jest istotne w związku ze zdefiniowanymi na wcześniejszych etapach konstrukcji badania korzyściami i kosztami – jeśli np. przedmiotem prognozy okazuje się zmienna opisywana przez szereg czasowy, konieczne

⁹ W ekonometrii i statystyce przerwa strukturalna jest definiowana jako nieoczekiwana zmiana parametrów modeli regresji w czasie, co może prowadzić do ogromnych błędów w prognozowaniu i niewiarygodności modelu.

staje się zdefiniowanie rezultatu, czy ma być to pojedyncza najbardziej trafna prognoza (prognoza punktowa), czy może pewien dopuszczalny zakres możliwych przyszłych wartości (prognoza przedziałowa), który w swojej konstrukcji zakłada pewien poziom niepewności szacunku, związanego zarówno z problemem samego prognozowania, jak i prawdopodobieństwa rozkładu możliwych przyszłych wartości;

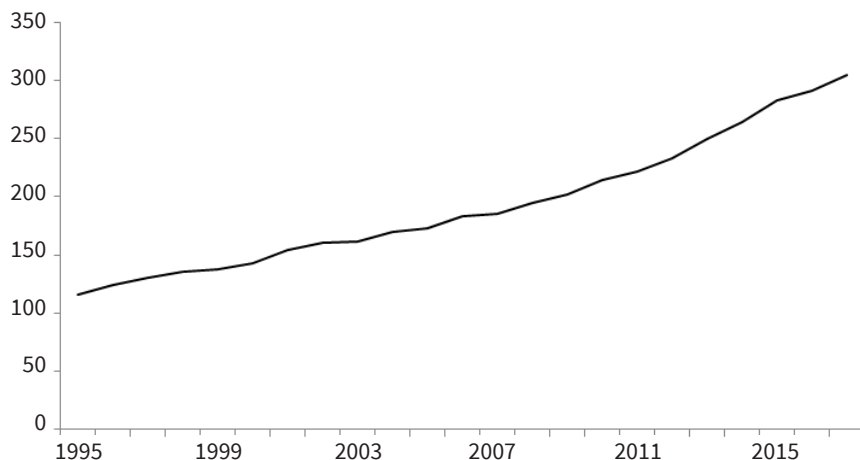
- 4) Określenie horyzontu prognozy – należy jasno określić, na jaki okres ma zostać skonstruowana prognoza (a także to, co go determinuje); czy ma być to prognoza krótko- bądź długookresowa (miesięczna, roczna, dziesięcioletnia). Warto zasygnalizować, że – w zależności od przyjętego horyzontu prognostycznego – najlepsza strategia (zarówno modelowania zjawiska, jak i prognozowania) będzie wykazywać różnicowanie;
- 5) Zdefiniowanie banku danych – na jakich danych opierać się będzie prognoza, co jest źródłem danych oraz jaka jest ich jakość, rzetelność, kompletność; czy dostępne dane to po prostu rejestr danych historycznych – np. dłużej już niegromadzonych – i czy w związku z tym dostępne są inne serie danych, powiązanych z głównym zjawiskiem prognozowanym;
- 6) Wybór metody i jej złożoność – wybór odpowiedniego narzędzia odpowiadającego potrzebom konkretnego problemu prognostycznego oraz określenie poziomu zaawansowania proponowanego rozwiązania. Wybór narzędzia powinien być podparty doświadczeniami wskazującymi, które z metod najlepiej nadają się do prognozowania w logistyce, finansach czy ekonomii, a także poddyktowany zasadą parsymonii.

3.2. Dobór odpowiedniej metody prognozowania

Klasyczna analiza szeregów czasowych ma na celu dekompozycję ich zmienności ze względu na kilka składowych, do których można zaliczyć (Brockwell i Davis, 2016; Chatfield, 2000; Cieślak, 2008; Szmit, 2007):

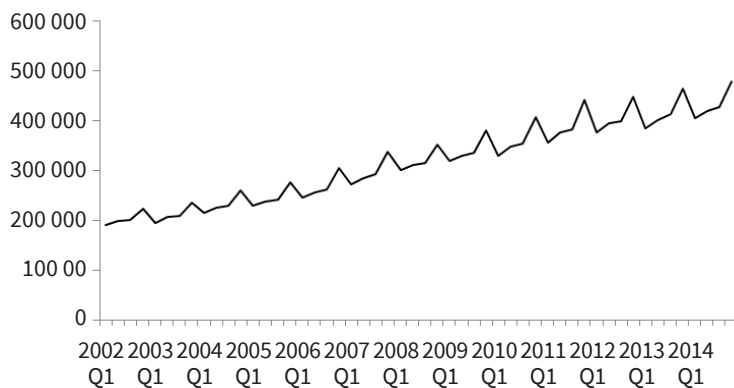
- 1) Trend (tendencję rozwojową) – ten rodzaj zmiany występuje, gdy seria danych wykazuje stały wzrost w górę lub spadek w dół (przynajmniej w kilku kolejnych okresach analizy). Trend można zdefiniować jako długoterminową zmianę średniego poziomu, zaś jego postrzeganie jest częściowo zależne od długości obserwowanych serii (zob. rysunek 46). Przeciwnościem tendencji rozwojowej będzie składowa systematyczna, wyrażająca stały, przeciętny poziom analizowanego zjawiska;
- 2) Wahania sezonowe (zmiana sezonowa) – ten rodzaj zmienności występuje zasadniczo co roku, a jego identyfikacja opiera się na obserwacji podobnych wzorców/schematów przebiegu serii danych. Należy mieć na

uwadze, że jeśli dla danego szeregu czasowego pomiar dokonywany jest tylko raz w roku, nie można wówczas stwierdzić, czy zmiany sezonowe występują. Najczęściej o sezonowości mówi się w przypadku serii danych mierzonych codziennie, co tydzień, co miesiąc lub co kwartał, oraz gdy podobne wzorce zachowań obserwuje się w określonych porach roku (np. wyższa podaż produkcji rolnej, wzrost zatrudnienia w okresie wakacyjnym) – zob. rysunek 47.



Rys. 46. Wartość produkcji przedsiębiorstwa w tys. PLN w latach 1995–2015

Źródło: opracowanie własne.



Rys. 47. Produkt Krajowy Brutto w cenach bieżących w kwartałach w latach 2002–2014

Źródło: opracowanie własne na podstawie danych GUS.

- 3) Wahania cykliczne – obejmują regularne zmiany cykliczne o okresowości dłuższej niż jeden rok, a przykładem tego typu wahań mogą być cykle koniunkturalne trwające około pięciu–siedmiu lat;
- 4) Wahania przypadkowe (wahania nieregularne) – wyrażenie „nieregularne” jest często używane do opisanie wszelkich zmian identyfikowanych w serii danych po usunięciu z niej tendencji rozwojowej, sezonowości czy innych efektów systematycznych; mogą to być zatem zmiany całkowicie losowe, które są trudne do przewidzenia, ale także zmiany wykazujące korelację krótkoterminową lub obejmujące pojedyncze nieciągłości w serii danych.

Metody klasycznego prognozowania pozwalają na uzyskanie dobrych oszacowań (o wysokiej trafności i aktualności) – w sytuacji, gdy zmienność szeregu/serii jest zdominowana przez regularny trend liniowy i/lub regularną sezonowość (Szmuksta-Zawadzka i Zawadzki, 2012). Nie wykazują jednak dużej przydatności, kiedy trendy i/lub efekty sezonowe zmieniają się w czasie w sposób nieregularny lub wykazują korelację z kolejnymi wartościami wahań przypadkowych, co określa się mianem autokorelacji¹⁰.

Przy rozwiązywaniu problemu prognostycznego pierwszym krokiem jest zatem identyfikacja składowych szeregów. Wykorzystuje się w tym celu najczęściej analizę graficzną, którą można dodatkowo wzbogacić, np. o analizę zmienności danego szeregu. Możliwe staje się więc zastosowanie współczynnika zmienności, definiowanego jako iloraz odchylenia standardowego (S_y) i średniej arytmetycznej (\bar{y}) (Sobczyk, 2000):

$$V_y = \frac{S_y}{\bar{y}} \quad (294)$$

Przyjmuje się że jeśli wartość współczynnika zmienności nie przekroczy 10%, wówczas dla analizowanej serii wpływ wahań przypadkowych jest dopuszczalny. Informacja wynikająca z powyższego spostrzeżenia daje możliwość korzystania z metod prostego prognozowania – przeznaczonych dla szeregów o niskiej zmienności. Najczęściej w tego typu seriach, zamiast tendencji rozwojowej, której występowanie podlega dodatkowemu testowaniu istotności, identyfikowany jest stały, przeciętny poziom analizowanych danych – z występującymi wahaniami przypadkowymi lub sezonowymi (zob. rysunek 48).

10 Zależność między kolejnymi wartościami tego samego szeregu czasowego. Często okazuje się również, że kolejne wartości składników resztowych modelu trendu i wahania sezonowe stają się skorelowane, gdy oddzielone są krótkim przedziałem czasowym (korelacja krótkoterminowa) – wówczas zaleca się stosowanie bardziej wyszukanych (odpornych) podejść modelowania w celu poprawy jakości prognoz.



Rys. 48. Wizualizacja przykładowych szeregów czasowych. Składowe systematyczne szeregi: a) stały przeciętny poziom zjawiska i wahania przypadkowe, b) tendencja rozwojowa i wahania przypadkowe, c) stały przeciętny poziom zjawiska, wahania sezonowe i przypadkowe

Źródło: opracowanie własne.

Wykres szeregu czasowego pozwala na wskazanie zarówno obserwacji nietypowych czy punktów zwrotnych szeregów, których występowanie w znaczącym stopniu może wpływać na końcowy rezultat prognoz, jak również wartości błędów oszacowań – szczególnie, jeśli dostępny szereg jest krótki (np. w sytuacji, kiedy dane przedsiębiorstwo funkcjonuje na rynku od kilku miesięcy). Na etapie obserwacji przebiegu serii analityk jest w stanie wychwycić obserwacje odstające, mogące wynikać np. z pomyłki przy wprowadzaniu danych, i wyeliminować je we wstępnej części prognozowania – stosując interpolację zastępuje obserwację nietypową (np. średnią arytmetyczną, wyznaczoną na podstawie wartości z dwóch sąsiadujących z obserwacją odstającą okresów czasowych).

Przed próbą modelowania i prognozowania pożądane staje się wstępne zapoznanie z dostępnymi szeregami danych, aby je „wyczuć” i dokonać identyfikacji ich głównych właściwości. Proces początkowej analizy ma istotny wpływ na późniejsze modelowanie i jakość uzyskiwanych rezultatów (trafnych prognoz). Najczęściej opisywany jest on jako wstępna analiza danych (ang. *Initial Data Analysis* – IDA), zaś wśród narzędzi powszechnie stosowanymi w jej realizacji wymienia się wizualizację szeregów danych czy ocenę statystyk podsumowujących rozkłady zmiennych. W momencie prowadzenia analizy wstępnej analityk może „wyczyścić” dane, usuwając lub dostosowując wszelkie zidentyfikowane błędy.

3.3. Błędy prognoz

Założmy, że dostępne są obserwacje pojedynczego szeregu czasowego, oznaczonego y_1, y_2, \dots, y_n , i należy skonstruować prognozę dla: y_{n+h} , gdzie $h = 1, 2, \dots, k$. Metoda prognozowania jednowymiarowego jest procedurą obliczania prognozy

punktowej, x_{n+h}^* , opartej tylko na przeszłych i obecnych wartościach danej serii (z uwzględnieniem funkcji czasu, wyrażonej w postaci np. trendu liniowego). Natychmiast nasuwa się kilka wątpliwości związanych z wyborem funkcji, która najlepiej odwzorowałaby historyczny przebieg serii: jakie właściwości będą miały uzyskiwane prognozy; czy jest to najlepsza z możliwych prognoz oraz wreszcie – co należy rozumieć przez słowo „najlepsza” (Nolte, Nolte i Pohlmeier, 2019).

Przed udzieleniem odpowiedzi na te pytania analityk musi zdecydować, w jaki sposób należy oceniać prognozy, a w szczególności – jaki wybór funkcji straty będzie właściwy. Funkcja straty definiowana jest w następujący sposób, który możemy zapisać jako:

$$q_t = y_t - y_t^*, \text{ dla } t > n \quad (295)$$

gdzie:

- y_t – wartość rzeczywista w okresie t ,
- y_t^* – wartość prognozowana w okresie t ,
- n – liczba obserwacji szeregu czasowego,
- q_t – błąd prognozy.

Błąd ten można wskazać zarówno po upływie czasu, na który prognoza była skonstruowana – w przypadku, kiedy znana jest wartość rzeczywista prognozowanej zmiennej w momencie prognozy, wówczas ocenia się ją jakość *ex post* prognozy, czyli jej trafność – jak również przed upływem tego czasu – określana jest wtedy dopuszczalność prognozy (jakość *ex ante*) (Cieślak, 2008). Następnie wyznacza się funkcję straty, powiedzmy $L(q_t)$, określającą poziom błędu prognozy. Ma ona zazwyczaj następujące właściwości (Chatfield, 2000):

- $L(0) = 0$,
- $L(q_t)$ jest funkcją ciągłą, która rośnie wraz z wartością bezwzględną q_t .

Funkcja straty może być:

- symetryczna, tak że $L(q_t) = L(-q_t)$ – w sytuacji, gdy rozważana jest strata kwadratowa, gdzie $L(q_t) = k_1 q_t^2$, a funkcja straty bezwzględnej błędu $L(q_t) = k_2 |q_t|$, gdzie k_1, k_2 oznaczają stałą;
- asymetryczna – funkcja straty powstaje w planowaniu produkcji opierając się na ten samej marży, jeśli np. strata w utracie zamówień z powodu braku zapasów przekracza stratę z powodu ich nadmiaru.

Rzeczywiste koszty związane z przeszacowaniem bądź niedoszacowaniem prognoz są czasami zupełnie inne. Jednak funkcja straty pozwala na zdefiniowanie prognozy „dobrej”, czyli takiej, która minimalizuje średnią stratę – gdy np. jest powiązana z uśrednioną wartością rozkładu prawdopodobieństwa błędów prognozy. W praktyce zapisywanie funkcji straty specyficznej dla wybranego kontekstu może być trudne, dlatego np. strata kwadratowa jest używana częściej niż jakakolwiek inna funkcja.

Bezwzględny błąd prognozy pozwala jedynie na wskazanie, czy wartość rzeczywista szeregu była wyższa od prognozowanej (dodatnia wartość błędu), czy niższa (ujemna wartość błędu). Stąd – częściej w ocenie rezultatów prognoz wykorzystywana jest wartość błędów względnych *ex post*, obliczanych dla szeregów wyrażonych na skali ilorazowej według następującej formuły (Gajda, 2001):

$$\Psi_t = \frac{y_t - y_t^*}{y_t}, \quad \text{dla } t > n \quad (296)$$

Wartość błędu względnego informuje o procentowej wielkości odchylenia prognozy od rzeczywistej wartości zmiennej, względem wartości rzeczywistej tej zmiennej. Na podstawie błędu względnego możliwe jest również zarówno określenie wielkości odchylenia względnego, jak i kierunku tego odchylenia¹¹.

Powyższe formuły błędów (odchyżeń względnych i bezwzględnych) umożliwiają konstrukcje mierników dopasowania prognoz do danych rzeczywistych. Do najczęściej stosowanych należą: absolutny błąd procentowy (ang. *Mean Absolute Percentage Error* – MAPE) oraz średniokwadratowy błąd (ang. *Root Mean Squared Error* – RMSE). W ocenie dopasowania prognoz do danych rzeczywistych wykorzystywany jest niekiedy współczynnik Theila I^2 , pozwalający na dodatkową dekompozycję, która wskazuje, jaka część błędu przypisywana jest niedopasowaniu względem średniej, wahań i tendencji rozwojowej dla prognozowanej serii (zob. dodatek 2).

3.4. Trafność i dopuszczalność prognoz

Niezależnie od tego, która miara dokładności prognozy zostanie wybrana do oceny strat predykcji, analityk stwierdza często, że uzyskiwane wyniki dla różnych metod prognozowania wydają się dość podobne, więc nie ma możliwości wskazania wyraźnego „zwycięzcy”. Metoda prowadząca do wyznaczenia „najlepszych” prognoz może być – według przyjętej miary oceny dokładności prognozy – nieznacznie lepsza niż dostępne alternatywy. Pojawia się zatem pytanie, czy w takim przypadku istnieje znacząca różnica między stosowanymi podejściami. Koncentrując uwagę na przesłankach, czy różnice między metodami prognozowania można przypisać, np. zmienności próbkowania, wciąż nie jest jasne, ani który mechanizm *samplingu* okazuje się odpowiedni, ani jakie założenia należy przyjąć w odniesieniu do różnych metod i właściwości błędów prognoz, wynikających z różnych metod.

11 Miernik ten nie należy do grupy symetrycznych i – co wynika z jego konstrukcji – mocniej wskazuje na przeszacowanie prognoz niż na ich niedoszacowanie.

Ponieważ większość własności prognoz i samej metody prognostycznej opiera się głównie na analizie *ex post*, to nadrzędność jednej metody nad innymi wynika z możliwości sporządzania prognoz poza próbą. Praktyczne znaczenie metody staje się zwykle bardziej interesujące niż teoretyczne – i często kwestią jednostkowego przypadku jest istotność ponoszenia dodatkowych nakładów na korzystanie z wyszukiwanych (bardziej kosztownych i czasochłonnych) metod, pozwalających poprawić dokładność dopasowania, o np. 0,5%.

4. Zastosowanie wybranych metod w prognozowaniu popytu i sprzedaży

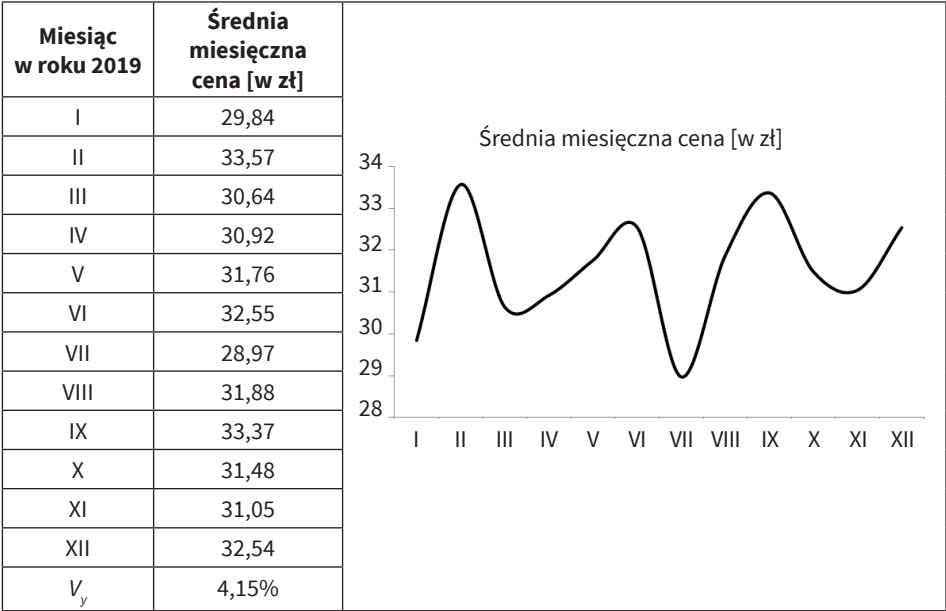
Prognozowanie szeregów czasowych, które najczęściej dostępne są w opisie procesów logistycznych, jest zasadniczo formą ich ekstrapolacji, opierającą się na dopasowaniu metody do zestawu danych historycznych (w okresie *ex post*), a następnie wykorzystaniu zidentyfikowanych prawidłowości poza zakresem dostępnych danych (w okresie *ex ante*). Ekstrapolacja okazuje się niebezpiecznie ignorowana w innych obszarach badań statystycznych, takich jak analiza regresji, jednak przy prognozowaniu zmiennych gromadzonych w postaci szeregów czasowych jest nieunikniona.

4.1. Prognozowanie szeregów o niskiej zmienności

Metody prognozowania prostego pozwalają na tworzenie prognoz krótkookresowych (Duda, 2017; Kardas, 2016) zazwyczaj na jeden okres *ex ante* – rzadziej na kilka (dla dalszych okresów dochodzi do tworzenia prognozy dynamicznej).

Założmy, że dane przedsiębiorstwo chce przewidzieć, jakiego poziomu przeciętnej ceny swojego produktu może się spodziewać w kolejnym okresie. W tym celu dział analiz przygotował odpowiednie zestawienie miesięcznych cen produktu (zob. tabela 90). Analityk, obserwując przebieg serii rocznej, dostrzega, że w poszczególnych miesiącach ceny odchylają się od poziomu przeciętnego $\bar{y} = 31,55$ [zł] w sposób różny, natomiast nie jest w stanie zidentyfikować wyraźnej tendencji (wzrostowej/spadkowej) cen. Potwierdza to również kalkulacją współczynnika zmienności $V_y = 4,15$ [%], co wskazuje na niewielkie zróżnicowania średnich cen w danych miesiącach oraz pozwala ocenić, że wpływ wahań przypadkowych kształtuje się na akceptowalnym poziomie (współczynnik zmienności mniejszy niż 10%). Seria danych jest zatem zbyt krótka, by doszukiwać się zmian sezonowych (miesięcznych), a tym bardziej cykliczności.

Tab. 90. Średnia miesięczna cena produktu w kolejnych miesiącach 2019 roku



Źródło: opracowanie własne.

Analityk postanawia zatem sięgnąć po metody przeznaczone do prognozowania szeregów czasowych o niskiej zmienności (zob. dodatek 1) – i dzięki zastosowanym narzędziom wyznacza wartości prognoz przeciętnej ceny (zob. tabela 91).

Tab. 91. Wartości prognoz wygaśłych *ex post* i *ex ante* dla przeciętnej ceny w PLN

Miesiąc w roku 2019	Naiwna	Naiwna z poprawką		Średnia ruchoma prosta		Średnia ruchoma ważona		Wygładzanie wykład. Browna
		liniową	nieliniową	k=3	k=5	k=3	k=5	
I	---	---	---	---	---	---	---	31,35
II	29,84	---	---	---	---	---	---	31,31
III	33,57	37,30	34,70	---	---	---	---	31,37
IV	30,64	27,71	31,55	31,35	---	31,36	---	31,35
V	30,92	31,20	31,93	31,71	---	31,37	---	31,34
VI	31,76	32,60	32,79	31,11	31,35	31,28	31,37	31,35
VII	32,55	33,34	33,57	31,74	31,89	31,99	31,85	31,38
VIII	28,97	25,40	29,86	31,09	30,97	30,60	30,88	31,32
IX	31,88	34,79	32,98	31,13	31,22	31,14	31,08	31,34

Miesiąc w roku 2019	Naiwna	Naiwna z poprawką		Średnia ruchoma prosta		Średnia ruchoma ważona		Wygładzanie wykład. Browna
		liniową	nieliniową	k=3	k=5	k=3	k=5	
X	33,37	34,86	34,42	31,41	31,71	32,04	31,88	31,39
XI	31,48	29,59	32,42	32,24	31,65	32,13	31,93	31,39
XII	31,05	30,61	32,03	31,97	31,35	31,64	31,62	31,38
Wartości prognoz ex ante								
I 2020	32,54	34,04	33,59	31,69	32,06	31,88	31,93	31,41

Objaśnienia: przyjęto wagi dla średniej ruchomej ważonej dla $k=3$: $w_1=0,2$, $w_2=0,3$, $w_3=0,5$; dla $k=5$: $w_1=0,05$, $w_2=0,1$, $w_3=0,2$, $w_4=0,3$, $w_5=0,35$; parametr wygładzania zoptymalizowano w dodatku SOLVER Microsoft Excel na poziomie $\alpha=0,247$.

Źródło: opracowanie własne.

Analitik wyznaczył wartości prognoz wygasłych *ex post*, na podstawie których w kolejnym kroku, dokonując odpowiedniej oceny trafności i dopuszczalności prognoz za pomocą kilku wybranych mierników (zob. dodatek 2), jest w stanie zarekomendować „najlepsze” z podejść (zob. tabela 92).

Tab. 92. Wartości błędów prognoz oraz innych mierników

Rodzaj miernika	Naiwna	Naiwna z poprawką		Średnia ruchoma prosta		Średnia ruchoma ważona		Wygładzanie wykładnicze Browna
		liniową	nieliniową	k=3	k=5	k=3	k=5	
Ψ_t	-6,84%	-11,75%	-10,28%	-4,04%	-5,27%	-4,67%	-4,83%	-3,12%
MAPE	5,87%	9,50%	5,53%	3,40%	4,20%	3,96%	4,35%	3,47%
MAPE_w	5,38%	7,79%	4,47%	2,52%	2,43%	2,95%	2,52%	3,47%
MSE	4,85	13,53	5,31	1,93	2,46	2,21	2,52	1,78
RMSE	2,20	3,68	2,30	1,39	1,57	1,49	1,59	1,34
V_{RMSE}	6,98%	11,66%	7,30%	4,40%	4,98%	4,72%	5,04%	4,23%
J	0,90	0,95	1,85	0,79	1,05	0,91	0,86	0,51

Objaśnienia: błąd względny i współczynnik Janusowy zakładają realizację prognozy w styczniu 2020 roku na poziomie: $y_{I,2020}=30,46$ [zł].

Źródło: opracowanie własne.

Dokonując analizy *ex post* wartości błędów stosowanych najczęściej (MAPE i RMSE), analitik wskazuje, że najdokładniejsze wartości poziomów średnich cen miesięcznych uzyskuje się dla metody średniej ruchomej prostej o stałej

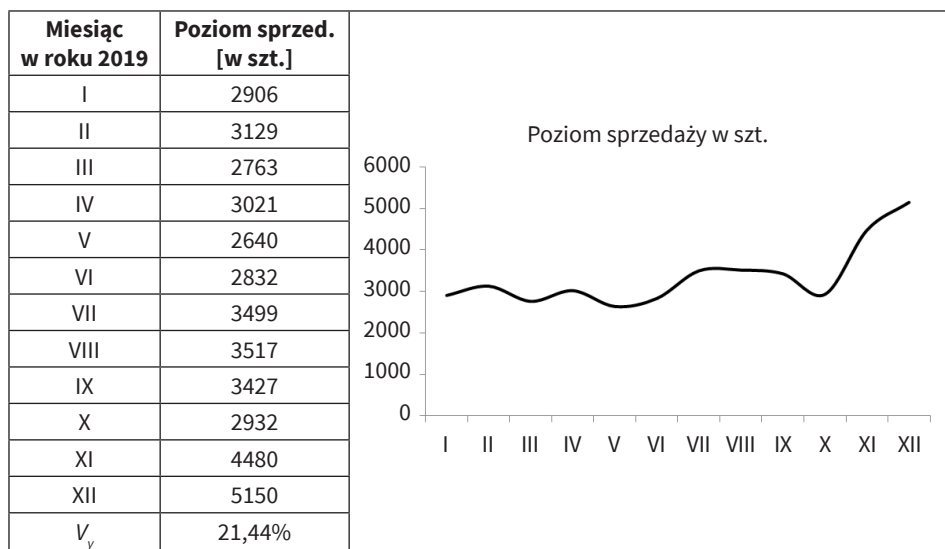
wygładzania $k = 3$ (MAPE = 3,40 [%], RMSE = 1,39 [zł]) oraz dla zoptymalizowanego wygładzania wykładniczego Browna (MAPE = 3,47 [%], RMSE = 1,34 [zł]). Ocenia zatem uzyskane wskazanymi metodami prognozy jako najbardziej trafne i jest gotów dokonać rekomendacji zarządowi, jaką należałoby przyjąć cenę średnią produktów na styczeń 2020 roku. Ponieważ nie ma jednoznacznego wskazania, która z metod daje „lepsze” rezultaty, analityk postanawia przyjrzeć się wartościom ważonych błędów $MAPE_w$ (Szmuksta-Zawadzka i Zawadzki, 2012). Na ich podstawie otrzymuje inną rekomendację – najlepsza trafność przypisywana zostaje do prognozy średniej ruchomej prostej dla stałej wygładzania $k = 5$.

Ponieważ analityk nie był w stanie dokonać wyboru, styczniowa prognoza *ex ante* zrealizowała się na poziomie $y_{1,2020} = 30,46$ [zł], co pozwoliło mu na ustalenie kryterium dopuszczalności – w ten sposób jest on w stanie dokonać rekomendacji (przy założeniu, że prognozy uważa się za dopuszczalne, jeśli wartość błędu w okresie *ex ante* nie przekracza $\pm 5\%$) i wskazać, że najbardziej dokładnym narzędziem prognostycznym, w rozważanym przypadku, okazuje się jednorównaniowy model wygładzania wykładniczego Browna, dla którego odnotowano poziom błędu względnego: $\Psi_t = -3,12$ [%] (miernik ten wskazuje również na nieznaczne przeszacowanie prognozy średniej ceny). Ponieważ analityk posiadał informacje o realizacji jednej prognozy *ex ante*, może również sprawdzić, czy rekomendowane narzędzie generuje prognozy aktualne w momencie/okresie T , na który zostały skonstruowane. Współczynnik Janusowy wskazał, że najbardziej aktualne (choć nie była to jedyna z metod, według której otrzymano aktualne prognozy) uzyskano na podstawie metod wygładzania wykładniczego Browna ($J = 0,51$).

4.2. Prognozowanie szeregów o wysokiej zmienności

W szeregach o wyższej zmienności najczęściej występującymi składowymi systematycznymi są tendencja rozwojowa oraz wahania okresowe. Dla takich serii współczynnik zmienności osiąga wartości wyższe niż $V_y = 10$ [%] – nie ma więc możliwości sięgnięcia po metody z grupy prostych. Niepodważalną przewagą metod przeznaczonych do prognozowania szeregów o wyższej zmienności jest generowanie przez nie prognoz o dalszym horyzoncie czasowym *ex ante*, co umożliwi z kolei weryfikację dopuszczalności prognoz właśnie w tym przedziale czasowym (bez konieczności posiadania realizacji prognozy w okresie konstrukcji predykcji na podstawie odpowiednich mierników bezwzględnych i względnych *ex ante*).

Założmy, że analityk otrzymał dodatkowe zadanie – przedsiębiorstwo, poza przewidywaniem przeciętnej ceny sprzedaży produktu, ma oszacować, jaki będzie popyt na ofertę przedsiębiorstwa w kolejnym kwartale 2020 roku (zob. tabela 93).

Tab. 93. Wartości poziomu sprzedaży w kolejnych miesiącach roku 2019

Źródło: opracowanie własne.

Analityk zauważa, że seria charakteryzuje się zmiennością wyższą niż $V_y = 10$ [%] oraz dostrzega rosnącą tendencję rozwojową¹². Uzyskane informacje pozwoliły na właściwe dobranie metod prognostycznych – widzi on również możliwości aplikacyjne dla funkcji trendu liniowego postaci $y_t = \alpha_0 + \alpha_1 t + \varepsilon_t$ (zob. tabela 94) oraz dwurównaniowego modelu wygładzania wykładniczego Holta – w celu określenia przyszłych poziomów sprzedaży (Halicka i Winkowski, 2013; Kardas, 2016; Wojciechowski i Wojciechowska, 2015) – zob. tabela 95.

Tab. 94. Podsumowanie modelu trendu liniowego wraz z prognozami *ex post* i *ex ante* oraz miernikami ocen trafności i dopuszczalności oszacowań

Statystyki regresji						
R^2		0,53	S_e			541,49
Parametr	Współczynniki	Błąd standardowy	t stat.	Wartość p	95% przedział ufności	
α_0	2372,68	333,26	7,12	0,00	1630,12	3115,24
α_1	151,59	45,28	3,35	0,01	50,69	252,48

12 Za pomocą odpowiedniej statystyki testowej analityk zweryfikował istotność składowej systematycznej w postaci tendencji rozwojowej. Współczynnik korelacji liniowej szeregu sprzedaży i zmiennej czasowej wyniósł $r_{xy} = 0,73$, zaś wartość statystyki testowej $t_e = 3,35$. Przy wartości krytycznej $t_{(0,05; 10)} = 2,35$ należy zatem stwierdzić istotność składowej systematycznej w postaci tendencji rozwojowej (Sobczyk, 2000).

Nr miesiąca w 2019 roku	Wartość prognoz <i>ex post</i>	Nr miesiąca w 2019 roku	Wartość prognozy <i>ex post</i>	Nr miesiąca	Wartość prognoz <i>ex ante</i>
I	2524,27	VII	3433,79		
II	2675,86	VIII	3585,38		
III	2827,44	IX	3736,97	2020; I	4343,32
IV	2979,03	X	3888,56	2020; II	4494,91
V	3130,62	XI	4040,14	2020; III	4646,49
VI	3282,21	XII	4191,73		

Okres <i>ex ante</i>	Wartości błędów <i>ex ante</i>		Współ. Theila i jego składowe		Względne składowe	
	S_{TP}	V_{TP}	I^2	0,0207		
2020; I	580,43	13,36%	I_1^2	0,0000	\hat{I}_1^2	0,00%
2020; II	600,68	13,36%	I_2^2	0,0033	\hat{I}_2^2	15,81%
2020; III	623,02	13,41%	I_3^2	0,0174	\hat{I}_3^2	84,19%

Źródło: opracowanie własne.

W wyniku modelowania ekonometrycznego (Gajda, 2001; Welfe, 2009; Szmit, 2007) model trendu oszacowano Metodą Najmniejszych Kwadratów (MNK) i otrzymano funkcję postaci $\hat{y}_t = 2372,68 + 151,59t$. Oszacowanie parametru przy zmiennej czasowej (istotne statystycznie przy przyjętym poziomie istotności $\alpha = 0,05$) wskazuje, że w roku 2019 z miesiąca na miesiąc poziom produkcji wzrastał o średnio 151,59 [szt.] *ceteris paribus*. Z powyższego podsumowania wynika, że przyjęta postać modelu trendu liniowego nie jest jakościowo dobrym narzędziem do opisu przebiegu poziomu produkcji w roku 2019. Wskazuje na to chociażby wartość współczynników determinacji $R^2 = 53$ [%] (model wyjaśnia jedynie 53% całkowitej zmienności poziomu sprzedaży w roku 2019) oraz błędu ogólnego modelu $S_e = 541,49$ (który stanowił $V_{se} = 16,13$ [%] średniej arytmetycznej odnotowanego poziomu sprzedaży w danym roku).

Prognozę zweryfikowano również pod kątem trafności – za pomocą współczynnika Theila (Cieslak, 2008) – oraz dokonano jego dekompozycji w celu uzyskania źródła błędu predykcji. W przypadku modelu trendu wartość współczynnika Theila wyniosła $I^2 = 0,0207$. Im wartość współczynnika jest większa, tym wyższe są różnice między wartościami rzeczywistymi a prognozowanymi, co stanowi poziom rozbieżności równy $I = 14,39$ [%]. Dzięki dokonanej dekompozycji współczynnika analityk zyskuje również dodatkową wiedzę – 15,81% całkowitego błędu predykcji wynika z niedostatecznej elastyczności predykcji (niezgodności zróżnicowania), a 84,19% błędu z niedostatecznej predykcji punktów zwrotnych (niezgodności kierunków zmian). Dla modelu tendencji rozwojowej nie wykaza-

no niezgodności średnich $-\hat{I}_1^2 = 0,00$ [%] nie wskazał błędu wynikającego z obciążenia predykcji.

Analityk dokonał także weryfikacji dopuszczalności prognoz w okresie *ex ante* – i w tym celu wyznaczył wartości odpowiednich błędów (S_{TP}) oraz współczynnika zmienności (V_{TP}) błędu *ex ante*. Przyjmując założenie, że prognozę uznaje się za dopuszczalną, jeśli poziom obciążenia predykcji nie przekracza 5%, uzyskane prognozy sprzedaży w kolejnych trzech miesiącach należy uznać za niedopuszczalne, ponieważ wartości współczynników błędów *ex ante* oszacowano na poziomie wyższym niż 13%.

Dla dwurównaniowego modelu wygładzania wykładniczego uzyskano wyniki zdecydowanie lepsze (zob. tabela 95). Po dokonaniu odpowiedniego porównania analityk zauważa, że drugie z proponowanych rozwiązań nie tylko generuje niższą (mówiącą o wyższej trafności) wartość współczynnika Theila $P = 0,0022$ (niż poziom rozbieżności równy $I = 4,70$ [%]), ale również wyraźnie niższe wartości błędów *ex ante* i ich współczynników zmienności. W tym przypadku, dzięki optymalizacji parametrów wygładzania, uzyskano prognozy dopuszczalne – wartości współczynników zmienności kształtowały się pomiędzy $3,30\% < V_{TP} < 3,50\%$.

Tab. 95. Wartości składowych F_t i S_t , prognoz wygładzania wykładniczego Holta wraz z prognozami *ex post* i *ex ante* oraz miernikami ocen trafności i dopuszczalności oszacowań

Rok	Miesiąc	Liczba sprzedanych sztuk	Wygładzone wart. szeregu czasowego F_t	Wygładzone wart. przyrostu trendu S_t	Wartość prognozy
2019	I	2906,00	2932,67	223,00	---
	II	3129,00	3133,63	218,10	3155,67
	III	2763,00	2865,13	109,89	3351,72
	IV	3021,00	3013,02	118,34	2975,02
	V	2640,00	2725,24	28,02	3131,36
	VI	2832,00	2818,34	42,49	2753,26
	VII	3499,00	3388,29	159,79	2860,83
	VIII	3517,00	3522,39	154,08	3548,08
	IX	3427,00	3470,28	108,23	3676,47
	X	2932,00	3044,16	-10,61	3578,50
	XI	4480,00	4229,07	255,26	3033,55
	XII	5150,00	5034,52	377,62	4484,33
2020	I				5412,14
	II				5789,75
	III				6167,37

Tab. 95 (cd.)

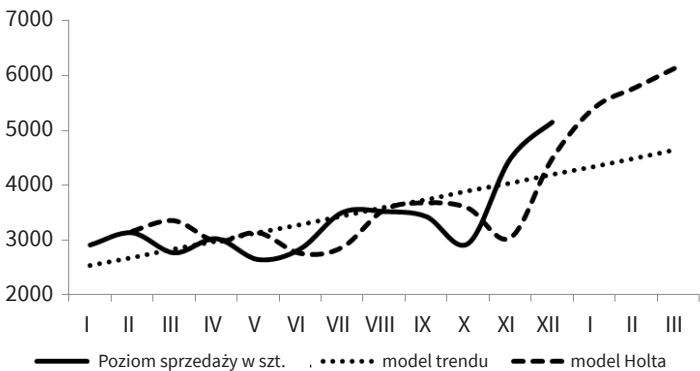
Okres <i>ex ante</i>	Wartości błędów <i>ex ante</i>		Współ. Theila i jego składowe		Względne składowe	
	S_{TP}	V_{TP}	I^2	0,0022		
2020; I	189,51	3,50%	I_1^2	0,0016	\hat{I}_1^2	73,90%
2020; II	196,12	3,39%	I_2^2	0,0000	\hat{I}_2^2	0,42%
2020; III	203,42	3,30%	I_3^2	0,0006	\hat{I}_3^2	25,67%

Objaśnienia: RMSE = 161,39, $\alpha = 0,827$, $\beta = 0,223$.

Źródło: opracowanie własne.

Dekompozycja współczynnika Theila – w przypadku modelu Holta – wskazała również na inne źródło odnotowanego błędu całkowitego: $\hat{I}_1^2 = 73,90$ [%] wynikało z błędu obciążenia predykcji (niezgodności średnich); $\hat{I}_2^2 = 0,42$ [%] przypisano niedostatecznej elastyczności predykcji (niezgodność różnicowania); część błędu spowodowana niedostateczną predykcją punktów zwrotnych (niezgodność kierunków zmian) stanowiła $\hat{I}_3^2 = 25,67$ [%].

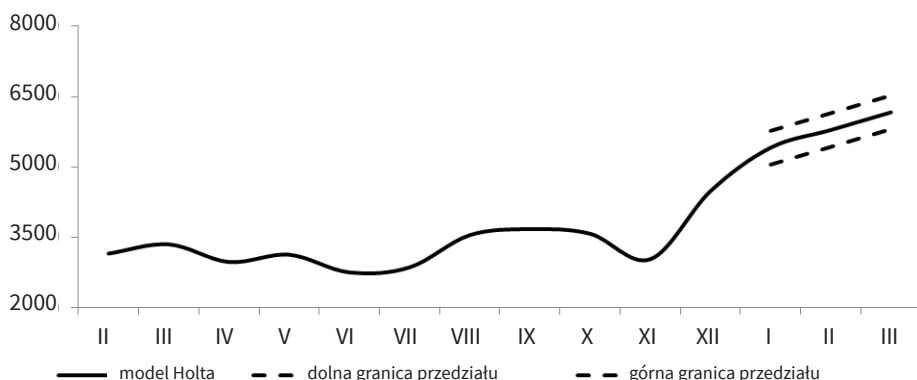
Powyższy przykład wskazuje, jak istotna jest właściwa identyfikacja składowych szeregu oraz właściwości przebiegu serii danych. Porównując oszacowania prognoz *ex post* i *ex ante*, można zauważyć, że dla szeregów o zróżnicowanym przebiegu zdecydowanie lepsze odwzorowanie w okresie *ex post* i wyższą zbieżność w okresie *ex ante* osiąga się dla bardziej elastycznych rozwiązań – dla przypomnienia, zdecydowana wyższa część rozbieżności w modelu trendu wynikała z niezgodności kierunków zmian (zob. rysunek 49).



Rys. 49. Porównanie wyników prognoz modeli Holta i trendu z wartościami rzeczywistymi

Źródło: opracowanie własne.

Ponieważ niepewność szacunku ma dość istotne znaczenie w przypadku określania przyszłych wartości, często proponowanym rozwiązaniem jest tworzenie prognoz realizujących się z pewnym przyjętym prawdopodobieństwem (Szmit, 2007). Możliwa staje się zatem konstrukcja przedziałów wartości, w których będzie znajdować się predykcja – zob. rysunek 50.



Rys. 50. Prognozy modelu Holta i prognoza przedziałowa oszacowana z 95% prawdopodobieństwem

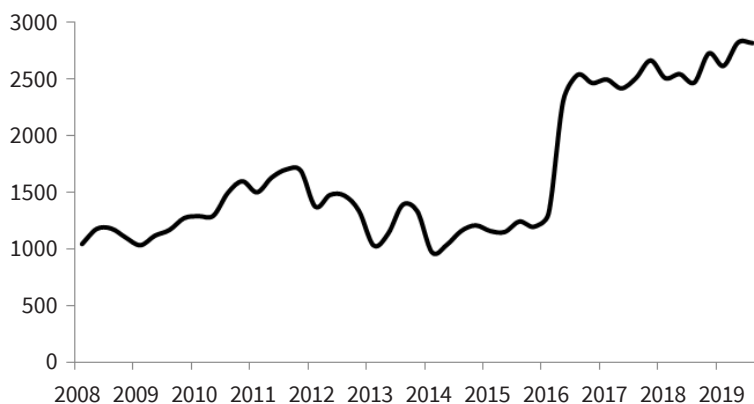
Źródło: opracowanie własne.

Zainteresowanie prognostyczne logistyków dotyczy również serii, w których mogą występować składowe systematyczne (poza identyfikowanymi już wcześniej wahaniami przypadkowymi, trendem lub stałym poziomem zjawiska) wahania okresowe, np. w postaci wahań sezonowych (Cheba, 2006; Przystupa i Juściński, 2015; Wojciechowski i Wojciechowska, 2015; Szmuksta-Zawadzka i Zawadzki, 2012).

Analitik dostał kolejne zadanie, które ma na celu określić przyszłą sytuację związaną z wartością sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa – według Polskiej Klasyfikacji Działalności. W tym celu sięgnął po odpowiednie zestawienia Głównego Urzędu Statystycznego i przygotował szeregi o najwyższej szczegółowości. Kierownictwo oczekuje prognoz do końca 2020 roku

Analitik dostrzega, że w analizowanej serii wyraźnie zauważalna staje się rosnąca tendencja rozwojowa, aczkolwiek niepokój budzi załamanie (spadek wzrostu) w latach 2012–2016. Podejrzewa (analogicznie, jak w przypadku modelu trendu liniowego, który wykorzystał do konstrukcji prognozy poziomu sprzedaży sztuk w przedsiębiorstwie), że metody prognostyczne oparte na trendzie mogą dawać

nietrafne oszacowania. Pomimo tego, dokonuje oceny istotności składowej systematycznej w postaci tendencji rozwojowej i otrzymuje potwierdzenie, że współczynnik korelacji $r_{xy} = 0,76$ i jest on istotny statystycznie przy 5% poziomie istotności ($t_{STAT} = 8,03$, przy wartości $p = 0,00$). Oszacowany współczynnik zmienności $V_y = 36,90$ [%] wskazuje na wysoką zmienność analizowanych wartości sprzedanych towarów i materiałów. Na podstawie uzyskanych rezultatów stwierdza, że w szeregu występują składowe systematyczne w postaci losowych odchyłeń oraz rosnącej (współczynnik kierunkowy prostej $\alpha_1 = 34730,11$) tendencji rozwojowej; analityk zakłada również możliwość występowania sezonowych wahań wartości sprzedaży (ze względu na kwartalność okresowości danych) – zob. rysunek 51.



Rys. 51. Wartość sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa w tys. PLN w latach 2008–2019 (dane kwartalne)

Źródło: opracowanie własne na podstawie danych GUS.

Dokonuje zatem rekomendacji, że do prognozowania serii najlepiej wykorzystać metodę wygładzania wykładniczego Wintersa bądź, aczkolwiek proponuje to niechętnie, wskaźników sezonowości.

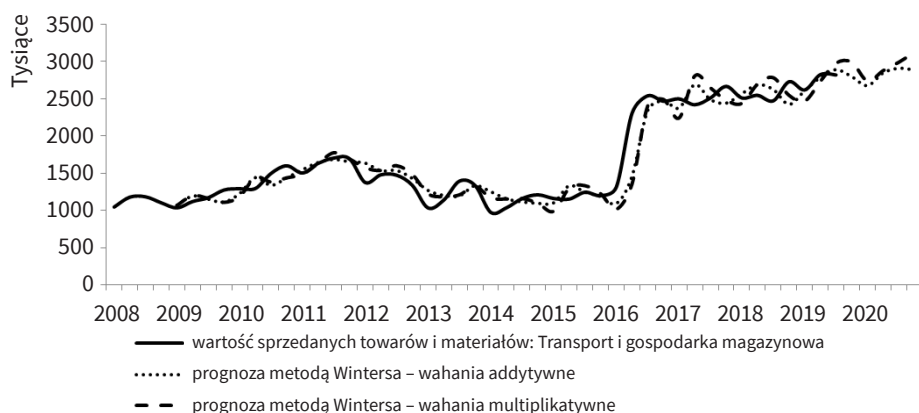
Obie metody zakładają sezonowość, która może wystąpić w postaci wahań bezwzględnie lub względnie stałych – w pierwszym przypadku wskazywana jest sezonowość addytywna¹³, w drugim zaś multiplikatywna¹⁴. Nie zawsze możliwe jest wskazanie, po obserwacji przebiegu szeregu, z jakiego rodzaju amplitudą wahań ma się do czynienia. Stąd analityk kalkuluje prognozy w obu wariantach, by

¹³ Gdy amplitudy wahań w analogicznych fazach cyklu są w przybliżeniu takiej samej wartości.

¹⁴ Gdy amplitudy wahań w analogicznych fazach cyklu zmieniają się mniej więcej w tym samym stosunku.

następnie – na podstawie odpowiedniego kryterium – dokonać końcowej rekomendacji (zob. dodatek 3, który zawiera wyniki dla alternatywnych podejść wygładzania wykładniczego oraz wskaźników sezonowości wraz z podsumowaniem dotyczącym trafności prognoz).

Dzięki uzyskanym результатам (zob. rysunek 52) potwierdzony zostaje wniosek, że metody wygładzania, dzięki swojej elastyczności, pozwoliły na interpretację zmian zachodzących w rzeczywistym przebiegu wartości sprzedanych towarów zdecydowanie lepiej niż metoda wskaźnikowa.



Rys. 52. Wartość sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa w tys. PLN w latach 2008–2020 (dane kwartalne) wraz z prognozami *ex post* i *ex ante*

Źródło: opracowanie własne na podstawie danych GUS.

Z uzyskiwanych rezultatów wynika, że do końca roku 2020 planowany jest dalszy wzrost wartości sprzedawanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa, z niewielkim spadkiem w pierwszym kwartale roku. W przypadku obu podejść (sezonowości addytywnej i multiplikatywnej), przyjmując za miernik trafności prognoz średni absolutny błąd procentowy MAPE, za generujący lepsze prognozy należy uznać wariant addytywnej sezonowości $MAPE = 7,82$ [%] – przy $MAPE = 8,60$ [%] dla podejścia multiplikatywnego. Pomimo że wartości nie należą do niskich (np. przy założeniu, że prognozę uznaje się za trafną, jeśli wartość błędu $MAPE < 5$ [%]), obie prognozy powinny zostać odrzucone. Aczkolwiek, jak pokazują obliczenia (zob. tabele 94 i 95 oraz rysunek 52), metoda alternatywna w postaci wskaźników sezonowości generuje blisko trzykrotnie wyższe wartości błędów MAPE. W przypadku wysokiej zmienności szeregów oraz częstych punktów zwrotnych w tendencji serii metody mniej

elastyczne, które opierają się modyfikacji przebiegu tendencji rozwojowej (np. o wahania okresowe), nie pozwalają na właściwe dopasowanie przebiegu do prognozowanych szeregów.

Zakłada się (często domyślnie), że uzyskanie „najlepszej” prognozy oznacza osiągnięcie najdokładniejszych wartości w wymaganym okresie czasu, przy zignorowaniu innych czynników. Precyzja nie zawsze powinna być jednak czynnikiem nadrzędnym. W praktyce preferowana może być prostsza procedura, będąca nieznacznie mniej dokładna niż podejście znacznie bardziej skomplikowane. Jak pokazują doświadczenia, w celach poprawy konstrukcji prognoz stosowane są podejścia prognozowania łączonego/kombinowanego (Shaub, 2020), które – dzięki zestawowi zintegrowanych metod/modeli – mają pozwolić na szybsze i dokładniejsze generowanie prognoz, bez konieczności spędzania dużych ilości czasu, poświęcania zasobów lub poszukiwania specjalistycznej wiedzy na temat dopasowywania modeli.

Nie zawsze jest jasne, w jaki sposób należy zmierzyć wspomnianą dokładność, a różne miary dokładności mogą dawać odmienne wskazania – co również zostało empirycznie wykazane. Jednak to precyzja prognozy okazuje się kryterium, które można rozpatrywać z ogólnego (statystycznego) punktu widzenia, zaś czynniki, jakie należy wziąć pod uwagę przy wyborze metody prognozowania, obejmują zarówno trafność prognozy, jak również koszty prowadzenia prac, dostępność odpowiedniego oprogramowania komputerowego, poprawnie określonych właściwości serii danych, a także sposób wykorzystania prognozy (i inne, wynikające z danego kontekstu istotne cechy).

Przedsiębiorstwa logistyczne potrzebują aktualnych i dość dokładnych informacji rynkowych oraz sprzedażowych, aby opracować efektywne plany całego łańcucha dostaw. Aby podejmować najlepsze decyzje prognozowane w perspektywie średnio- i długoterminowej, niezbędna jest rzetelna analiza zarówno produktu, jak i rynku. Właściwe prognozowanie popytu prowadzi do zmniejszenia zapasów, lepszego odczytywania reakcji klientów oraz krótszego czasu cyklu gotówkowego. Istotnym zagadnieniem staje się optymalizacja predykcji, która powinna pozwolić na zrozumienie pierwotnych przyczyn zmienności popytu, wzorców w okresie historycznym, a także wskazać odpowiednie algorytmy w celu zapewnienia wydajności łańcucha dostaw.

Niewłaściwe modelowanie czynników rynkowych, cen i efektów popytu prowadzi do powstawania niepotrzebnych zapasów magazynowych, nieefektywnego wykorzystania zasobów oraz ponoszenia dodatkowych kosztów, np. magazynowania. Optymalizacja prognozy ma na celu lepsze zrozumienie istoty produktu – zarówno w przeszłości, jak i w przyszłości, oraz wybór najlepszego narzędzia do jego modelowania (aby obniżyć całkowite koszty i podnieść poziom usług).

Niewątpliwie proces planowania poziomu przyszłego popytu jest z reguły złożony – z punktu widzenia ilości danych i koniecznych do dokonania analiz. Aby tego uniknąć, przedsiębiorstwa często wybierają nieoptymalną strategię

przy podejmowaniu kluczowych decyzji dotyczących analizy historycznej ścieżki przebiegu popytu. Szybko jednak skłaniają się ku stosowaniu podejść optymalizacyjnych, w celu lepszego zarządzania i usprawnienia krytycznych procesów prognozowania, które powinny uwzględniać: determinujące popyt na produkty czynniki zewnętrzne, zależność pomiędzy horyzontem prognozy a ogólną strategią dotyczącą zaopatrzenia i poziomu zapasów oraz ocenę możliwości organizacji w planowaniu popytu z perspektywy procesów i umiejętności. W zależności od problemu optymalizacyjnego prognoz, wskazuje się na dwie grupy możliwych do zastosowania algorytmów:

- 1) Konwencjonalne metody programowania matematycznego – tego typu metody są zwykle stosowane w przypadku dobrze zdefiniowanych modeli optymalizacji o pożądanых właściwościach matematycznych. W takich przypadkach często wykorzystuje się algorytmy programowania liniowego, wypukłego i ogólne metody programowania nieliniowego;
- 2) Algorytmy metaheurystyczne – ten rodzaj algorytmów stosowany jest w modelach optymalizacyjnych z niewłaściwie zdefiniowanymi funkcjami/ograniczeniami celu. W tym przypadku najczęściej są to zagadnienia optymalizacji *NP*-trudne, a ich złożoność obliczeniowa rośnie wykładniczo wraz z rozmiarem problemu. Do optymalizacji wykorzystywane są algorytmy oparte na sztucznej inteligencji (np. algorytmy genetyczne czy symulowane wyżarzanie).

Bibliografia

- Abolghasemi M., Beh E., Tarr G., Gerlach R. (2020), *Demand forecasting in supply chain: The impact of demand volatility in the presence of promotion*, „Computers & Industrial Engineering”, nr 106380.
- Abraham B., Ledolter J. (2009), *Statistical methods for forecasting*, John Wiley & Sons Inc., New Jersey.
- Abramov V.M. (2008), *Continuity theorems for the M/M/1/n queueing system*, „Queueing Systems”, nr 59.
- Ansari A.H. (2019), *Easy Simplex (AHA Simplex) Algorithm*, „Journal of Applied Mathematics and Physics”, nr 7.
- Armstrong J.S., Collopy F. (1998), *Integration of statistical methods and judgment for time series forecasting: Principles from empirical research*, [w:] G. Wright, I.P. Goodwin (red.), *Forecasting with Judgment*, John Wiley & Sons Inc., New York.
- Baets S. de, Harvey N. (2020), *Using judgment to select and adjust forecasts from statistical models*, „European Journal of Operational Research”, t. 31, nr 284.
- Biecek P. (2017), *Przewodnik po pakiecie R*, Oficyna Wydawnicza GiS, Wrocław.
- Bładowski S. (1970), *Metody sieciowe w planowaniu i organizacji pracy*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Blumenfeld D. (2001), *Operations research calculations handbook*, Boca Raton, London–New York.
- Bogdan M. (2018), *Multiple solutions in linear programming problem*, „Procedia Manufacturing”, nr 22.
- Brockwell P.J., Davis R.A. (2016), *Introduction to time series and forecasting*, Springer-Verlag, Berlin–Heidelberg.
- Burkard R., Cela E., Pardalos P.M., Pitsoulis L. (1998), *The quadratic assignment problem*, [w:] D.Z. Du, P. Pardalos (red.), *Handbook of combinatorial optimization*, Kluwer Academic Publishers, Berlin–Heidelberg.
- Burkard R.E., Dell’Amico M., Martello S. (2009), *Assignment Problems*, SIAM, Philadelphia.
- Burton C., Michael N. (1999), *Zarządzanie projektem*, Wydawnictwo Astrum, Wrocław.
- Chatfield C. (2000), *Time-series forecasting*, CRC Press Inc., Cleveland.
- Cheba K. (2006), *Zastosowanie metody wskaźników sezonowości w prognozowaniu dla danych dekadowych*, „Prace Naukowe Akademii Ekonomicznej we Wrocławiu”, nr 1112.

- Christopher M. (1996), *Strategia zarządzania dystrybucją: praktyka logistyki biznesu*, Wydawnictwo Placet, Warszawa.
- Cieślak M. (red.), (2008), *Prognozowanie gospodarcze: metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa.
- Dangerfield B.J., Morris J.S. (1992), *Top-down or bottom-up: Aggregate versus disaggregate extrapolations*, „International Journal of Forecasting”, t. 8, nr 2.
- Dantzig G. (1951), *Maximization of a linear function of variables subject to linear*, [w:] T.C. Koopmans (red.), *Activity Analysis of Production and Allocation*, John Wiley & Sons Inc., New York.
- Dantzig G. (1963), *Linear Programming and Extensions*, Princeton University Press, Princeton.
- Dantzig G., Orden A. (1953), *Duality Theorems*, The Rand Corporation, Santa Monica.
- Diebold F.X. (1998), *Elements of forecasting*, South-Western College Publishing, Cambridge.
- Dittmann P. (2000), *Metody prognozowania sprzedaży w przedsiębiorstwie*, Wydawnictwo Akademii Ekonomicznej im. Oskara Langego, Wrocław.
- Duda A. (2017), *Prognozowanie w logistyce na przykładzie prognozy metodą średnich ruchomych*, „Zeszyty Naukowe Akademii Sztuki Wojennej”, nr 3.
- Duncan W.R. (2013), *A Guide to the Project Management Body of Knowledge*, Project Management Institute Inc., Pennsylvania.
- Egervary J. (1931), *Matrixok kombinatorius tulajdonságairól (On combinatorial properties of matrices)*, „Matematikai és Fizikai Lapok”, nr 38.
- Engel R., Hassin R. (2017), *Customer equilibrium in a single-server system with virtual and system queues*, „Queueing Systems”, nr 87.
- Filipowicz B. (1996), *Modele stochastyczne w badaniach operacyjnych*, Wydawnictwo Naukowo-Techniczne, Warszawa.
- Filipowicz B., Kwiecień J. (2008), *Zastosowanie teorii kolejek do modelowania struktur administracyjnych w szkolnictwie*, „Automatyka”, t. 3, nr 12.
- Ford L., Fulkerson D. (1969), *Przepływy w sieciach*, PWN, Warszawa.
- Frame J.D. (1999), *Zarządzanie projektami w organizacjach*, WIG-Press, Warszawa.
- Gajda J.B. (2001), *Prognozowanie i symulacja a decyzje gospodarcze*, Wydawnictwo C.H. Beck, Warszawa.
- Gantt H.L. (1913), *Work, Wages, and Profits*, The Engineering Magazine Co., New York.
- Gągolewski M. (2014), *Programowanie w języku R*, Wydawnictwo Naukowe PWN, Warszawa.
- Ghiani G., Laporte G., Musmanno R. (2004), *Introduction to logistics systems planning and control*, John Wiley & Sons Inc., New Jersey.
- Halicka K., Winkowski C. (2013), *Wykorzystanie metod wygładzania wykładniczego do prognozowania kursu sprzedaży EUR*, „Ekonomia i Zarządzanie”, nr 2.
- He L., Bai H., Jiang Y., Ouyang D., Jiang S. (2018), *Revised simplex algorithm for linear programming on GPUs with CUDA*, „Multimedia Tools and Applications”, t. 22, nr 77.
- He Q., Lee J. (2017), *Another pedagogy for pure-integer Gomory*, „RAIRO-Operation Research”, t. 1, nr 51.
- Hillier F.S., Lieberman G.J. (2015), *Introduction to operations research*, McGraw-Hill Education, New York.

- Hoffmann T. (2016), *Rola i znaczenie ewaluacji projektów współfinansowanych ze środków Unii Europejskiej w procesie rozwoju regionalnego*, „Studia i Prace WNEiZ US. Problemy Współczesnej Ekonomii”, t. 3, nr 44.
- Jacyna M., Żak J. (2012), *Zastosowanie teorii kolejek do analizy i oceny procesu transportowego w centrum logistycznym*, „Logistyka”, nr 4.
- Jędrzejczyk Z., Kukuła K., Skrzypek J., Walkosz A. (2016), *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa.
- Jewczak M. (2017), *Willingness-to-pay for Additional Health Benefits in Poland – Analysis of Spatial Diversity*, „Przedsiębiorczość i Zarządzanie”, nr 18.
- Kalichman I.L. (1971), *Algebra liniowa i programowanie*, PWN, Warszawa.
- Kamberi L. (2017), *Using Queueing Theory-M/M/1 in Hospital Tetovo*, „Journal of Mathematics and System Science”, nr 7.
- Kardas E. (2016), *Wykorzystanie metod analitycznych do prognozowania wielkości produkcji stali na świecie i w Polsce w latach 2016–2020*, „Hutnik – Wiadomości Hutnicze”, t. 11, nr 83.
- Kasparek M. (2011), *Zarządzanie projektem*, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice.
- Kendall D.G. (1951), *Some problems in the theory of queues*, „Journal of the Royal Statistical Society. Series B”, t. 2, nr 13.
- Kőnig D. (1931), *Gráfok és mátrixok*, „Matematikai és Fizikai Lapok”, nr 38.
- Koopmans T., Beckmann M. (1957), *Assignment problems and the location of economic activities*, „Econometrica”, nr 25.
- Kotarbiński T. (1970), *Sprawność i błęd*, Wydawnictwo PZWS, Warszawa.
- Kozarkiewicz A. (2012), *Zarządzanie portfelami projektów*, Wydawnictwo Naukowe PWN, Warszawa.
- Kozubski J.J. (2000), *Wprowadzenie do badań operacyjnych*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
- Koźmiński A.K., Piotrowski W. (red.), (2004), *Zarządzanie. Teoria i praktyka*, Wydawnictwo Naukowe PWN, Warszawa.
- Koźniewska I., Włodarczyk M. (1978), *Modele odnowy, niezawodności i masowej obsługi*, PWN, Warszawa.
- Kramer M. (1989), *Stationary Distributions in a Queueing System with Vacation Times and Limited Service*, „Queueing Systems”, nr 4.
- Krawczyk S. (2001), *Metody ilościowe w logistyce (przedsiębiorstwa)*, Wydawnictwo C.H. Beck, Warszawa.
- Kuhn H. (1955), *The Hungarian Method for the assignment problem*, „Naval Research Logistics Quarterly”, nr 2.
- Kundakcioglu O., Alizamir S. (2008), *Generalized assignment problem*, [w:] C. Floudas, P. Pardalos (red.), *Encyclopedia of optimization*, Springer, Boston.
- Langaris C. (1993), *Waiting time analysis of a two-stage queueing system with priorities*, „Queueing Systems”, nr 14.
- Lawler E. (1963), *The quadratic assignment problem*, „Management Science”, nr 9.
- Little J. (1961), *A proof for the queueing formula: $L = \lambda W$* , „Operations Research”, t. 9, nr 3.
- Lock D. (2009), *Podstawy zarządzania projektami*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

- Łada M., Kozarkiewicz A. (2010), *Zarządzanie wartością projektów*, Wydawnictwo C.H. Beck, Warszawa.
- Matuszak Z., Bundz S., Homik W. (2014), *Próba opisu załadunku na statek elementami teorii masowej obsługi*, „Logistyka”, nr 6.
- Mei Lee S., Adi S., Mohd Rizal B., Effendi M. (2018), *Comparison between linear programming and integer linear programming: a review*, „International Journal of Mathematics in Operational Research”, t. 1, nr 13.
- Mendes P. (2011), *Demand driven supply chain: A structured and practical roadmap to increase profitability*, Springer Science & Business Media, Berlin–Heidelberg.
- Milewski R. (red.), (2018), *Podstawy ekonomii*, Wydawnictwo Naukowe PWN, Warszawa.
- Miszczyńska D., Miszczyński M. (2002), *Wybrane metody badań operacyjnych*, Wydawnictwo Wyższej Szkoły Ekonomiczno-Humanistycznej w Skierniewicach, Skierniewice.
- Montgomery A.L., Zarnowitz V., Tsay R.S., Tiao G.C. (1998), *Forecasting the US unemployment rate*, „Journal of the American Statistical Association”, t. 93, nr 442.
- Murthy P.R. (2007), *Operations research*, New Age International Publishers Ltd., New Delhi.
- Nogalski B., Czerska J., Klimek A. (2010), *Wykorzystanie metod teorii masowej obsługi i lean management w usprawnianiu procesów w zarządzaniu organizacją publiczną*, „Współczesne Zarządzanie”, nr 1.
- Nojszewska E. (2010), *Podstawy ekonomii*, Wydawnictwa Szkolne i Pedagogiczne, Warszawa.
- Nolte I., Nolte S., Pohlmeier W. (2019), *What determines forecasters' forecasting errors?*, „International Journal of Forecasting”, t. 35, nr 1.
- Nowakowska-Grunt J., Nowakowska A. (2012), *Selected Tools of Information Flow Management in Logistics*, „Studia Ekonomiczne Uniwersytetu Ekonomicznego w Katowicach”, nr 121.
- Oberlander G.D. (2000), *Project Management for Engineering and Construction*, McGraw-Hill, Boston.
- Osińska M. (2012), *Prognozowanie w logistyce*, Wydawnictwo Uczelniane Wyższej Szkoły Gospodarki, Bydgoszcz.
- Pawlak M. (2011), *Zarządzanie projektami*, Wydawnictwo Naukowe PWN, Warszawa.
- Pisz I., Łapuńska I. (2015), *Zarządzanie projektami w logistyce*, Wydawnictwo Difin, Warszawa.
- Povh J. (2008), *Assignment Problems in Logistics*, „Logistics & Sustainable Transport”, t. 1, nr 3.
- Prasch R.E. (2008), *How markets work: supply, demand and the 'real world'*, Edward Elgar Publishing, Cheltenham.
- Przystupa W., Juściński S. (2015), *Analiza możliwości wykorzystania modelu Holta–Wintersa do oceny popytu na części zamienne do pojazdów rolniczych*, „Logistyka”, nr 5.
- Rietman R., Resing J. (2004), *An M/G/1 Queueing Model with Gated Random Order of Service*, „Queueing Systems”, nr 48.
- Rogalska D., Klepacz H., Porazińska E., Hausman H., Mikulski H., Kwiatkowska M., Strzelec K. (1991), *Programowanie liniowe. Algorytmy i zadania*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź.
- Ross G., Soland R. (1975), *A branch and bound algorithm for the generalized assignment problem*, „Mathematical Programming”, nr 8.
- Roszkowski H., Wiatrak A.P. (2005), *Zarządzanie projektem. Istoty i procedury i ich zastosowanie przy korzystaniu z środków Unii Europejskiej*, Wydawnictwo SGGW, Warszawa.

- Rowe G., Wright G. (1999), *The Delphi technique as a forecasting tool: issues and analysis*, „International Journal of Forecasting”, t. 15, nr 4.
- Sakarovitch M. (1983), *Computational Aspects of the Simplex Method: Revised Simplex Algorithm; Bounded Variables*, [w:] J.B. Thomas (red.), *Linear Programming*, Springer, New York.
- Salan J.M. (2015), *Modeling and solving linear programming with R*, OmniaScience, Catalunya.
- Schumacher E.F. (1986), *Small is beautiful*, Rowohlt-Verlag, Berlin.
- Shaub D. (2020), *Fast and accurate yearly time series forecasting with forecast combinations*, „International Journal of Forecasting”, t. 36, nr 1.
- Sherzer E., Kerner Y. (2018), *Customers' abandonment strategy in an M/G/1 queue*, „Queueing Systems”, nr 90.
- Sikora W., Anholcer M., Gaspars-Wieloch H., Godlewski M., Marcinkowski J., Stodulny P. (2008), *Badania operacyjne*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Smith A. (2010), *The Wealth of Nations: An inquiry into the nature and causes of the Wealth of Nations*, Harriman House Limited, Hampshire.
- Sobczyk M. (2000), *Statystyka*, Wydawnictwo Naukowe PWN, Warszawa.
- Srinivasan V., Thompson G. (1973), *An algorithm for assigning uses to sources in a special class of transportation problems*, „Operations Research”, t. 21, nr 1.
- Stabryła A. (2006), *Zarządzanie projektami ekonomicznymi i organizacyjnymi*, Wydawnictwo Naukowe PWN, Warszawa.
- Strategor (2001), *Zarządzanie firmą*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Sucheck B. (2006), *Kompletne modele popytu*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Swathi C., Vasanta Kumar V., Hanumantha Rao S. (2019), *M/M/1 Queueing system with Customer Balking and Reneging*, „International Journal of Innovative Technology and Exploring Engineering”, nr 8.
- Szapiro T. (2000), *Decyzje menedżerskie z Excelem*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Szmit A. (red.), (2007), *Zbiór zadań z podstaw ekonometrii i prognozowania*, Wydawnictwo Politechniki Łódzkiej, Łódź.
- Szmuksta-Zawadzka M., Zawadzki J. (2012), *O miernikach dokładności prognoz ex post w prognozowaniu zmiennych o silnym natężeniu sezonowości*, „Metody Ilościowe w Badaniach Ekonomicznych”, t. 13, nr 1.
- Szmuksta-Zawadzka M., Zawadzki J. (2012), *Z badań nad metodami prognozowania na podstawie niekompletnych szeregów czasowych z wahaniami okresowymi (sezonowymi)*, „Przegląd Statystyczny”, t. 59, nr 1.
- Taha H. (1992), *Operations Research. An Introduction*, MacMillan Publishing Co., New York.
- Taylor F. (1911), *The Principles of Scientific Management*, Harper & Brothers, New York-London.
- Tiwari S., Kumar A. (2018), *Comparison between Goal Programming and other Linear Programming Methods*, „International Journal for Research in Applied Science & Engineering Technology”, nr 6.
- Trocki M. (red.), (2012), *Nowoczesne zarządzanie firmą*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Trocki M. (red.), (2015), *Zarządzanie projektem europejskim*, Polskie Wydawnictwo Ekonomiczne, Warszawa.

- Trocki M., Juchniewicz M. (2013), *Ocena projektów – koncepcje i metody*, Oficyna Wydawnicza Szkoła Główna Handlowa, Warszawa.
- Trocki M., Grucza B., Ogonek K. (2009), *Zarządzanie projektami*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Trocki M., Grucza B., Ogonek K. (2011), *Zarządzanie projektami*, [w:] J.R. Meredith, S.J. Mantel (red.), *Project Management – a Managerial Approach*, John Wiley & Sons Inc., New York.
- Trocki M., Bukhała E., Grucza B., Juchniewicz M., Metelski W., Wyrozębski P. (2012), *Nowoczesne zarządzanie projektami*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Truetsch U. (2014), *A semidefinite programming based branch-and-bound framework for the quadratic assignment problem*, Center for Economic Research, Tilburg.
- Trzaskalik T. (2008), *Wprowadzenie do badań operacyjnych z komputerem*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Vaidya N. (2017), *Application of Quick Simplex Method on the Dual Simplex Method (A New Approach)*, „Journal of Advances in Mathematics and Computer Science”, t. 24, nr 5.
- Varian H.R. (1997), *Mikroekonomia*, Wydawnictwo Naukowe PWN, Warszawa.
- Vitor F., Easton T. (2018), *The double pivot simplex method*, „Mathematical Methods of Operations Research”, nr 87.
- Welfe A. (2009), *Ekonometria: metody i ich zastosowanie*, Polskie Wydawnictwo Ekonomiczne, Warszawa.
- Witkowski J., Rodawski B. (2007), *Pojęcie i typologia projektów logistycznych*, „Gospodarka Materialowa i Logistyka”, nr 3.
- Wojciechowski A., Wojciechowska N. (2015), *Zastosowanie modeli Holta i Wintersa w zarządzaniu paletami w kanale dystrybucyjnym firmy Quatra*, „Logistyka”, nr 6.
- Ye W., Guo R., Deschamps B., Jiang Y., Liu X. (2020), *Macroeconomic forecasts and commodity futures volatility*, „Economic Modelling”, nr 1.
- Zawiła-Niedźwiecki J., Rostek K., Gąsiorkiewicz A. (red.), (2010), *Informatyka gospodarcza*, Wydawnictwo C.H. Beck, Warszawa.
- Zeliaś A., Pawełek B., Wanat S. (2008), *Prognozowanie ekonomiczne: teoria, przykłady, zadania*, Wydawnictwo Naukowe PWN, Warszawa.

Dodatek 1

Sposoby konstrukcji prognoz dla wybranych metod

Wyszczególnienie	Konstrukcja prognozy, uwagi dodatkowe
<i>Metody naiwne</i>	
prosta	$\dot{y}_{(t+1)} = y_t$
z poprawką liniową	$\dot{y}_{(t+1)} = y_t + y_t - y_{t-1} = 2y_t - y_{t-1}$
z poprawką nieliniową	$\dot{y}_{t+1} = y_t + \frac{y_t}{t_{t-1}}$
z poprawką sezonową	$\dot{y}_{(t+1)} = y_{t-r}$ gdzie: r – określa długość okresu sezonowego
<i>Metody średniej ruchomej</i>	
prosta	$\dot{y}_{t+1} = \frac{1}{k} \sum_{i=t-k}^{t-1} y_t$ gdzie: k – określa stałą wygładzania, czyli liczbę elementów uwzględnianych w kalkulacji prognozy
ważona	$\dot{y}_{t+1} = \frac{1}{k} \sum_{i=t-k}^{t-1} w_i y_t$ gdzie: k – określa liczbę elementów uwzględnianych w kalkulacji prognozy, a w_i wagi poszczególnych obserwacji w prognozie, przy czym: $\sum_{i=1}^k w_i = 1$ oraz zachowana jest własność „postarzania informacji” $0 \leq w_1 \leq w_2 \leq \dots \leq w_k \leq 1$
<i>Wygładzanie wykładnicze</i>	
Browna	$\dot{y}_{(t+1)} = \alpha \cdot y_t + (1 - \alpha) \cdot \dot{y}_t$ gdzie: α – określa parametr wygładzania przyjmujący wartości z przedziału $(0; 1)$ Jeśli $\alpha \rightarrow 1$, prognoza będzie uwzględniać w bardzo wysokim stopniu błędy <i>ex post</i> prognoz poprzednich, i odwrotnie. Metoda wymaga przyjęcia wartości startowych: $y_1 = y_1$ lub $\dot{y}_1 = \frac{1}{3} \sum_{t=1}^3 (y_t)$

Holta	<p>Model dwurównaniowy</p> <p>równanie I – służy do wyznaczenia wygładzonych wartości szeregu czasowego (prognozowanej zmiennej) w momencie/okresie t:</p> $F_t = \alpha \cdot y_t + (1 - \alpha) \cdot (F_{t-1} + S_{t-1})$ <p>równanie II – służy do wyznaczenia wygładzonych wartości przyrostu trendu na moment/okres t:</p> $S_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot S_{t-1}$ <p>gdzie: α, β – określa parametr wygładzania przyjmujący wartości z przedziału $\langle 0; 1 \rangle$:</p> $y_T^* = F_t + (T - t) \cdot S_t$ <p>gdzie: T – określa kolejne numery okresów <i>ex ante</i>.</p> <p>Metoda wymaga przyjęcia wartości startowych:</p> $F_1 = y_1 \text{ lub } F_1 = \frac{1}{3} \sum_{t=1}^3 (y_t), \text{ zaś } S_1 = y_2 - y_1$
Wintersa dla sezonowości addytywnej	<p>Model trzyrównaniowy</p> <p>równanie I – służy do wyznaczenia wygładzonych wartości szeregu czasowego (prognozowanej zmiennej) w momencie/okresie t:</p> $F_t = \alpha \cdot (y_t - C_{t-r}) + (1 - \alpha) \cdot (F_{t-1} + S_{t-1})$ <p>równanie II – służy do wyznaczenia wygładzonych wartości przyrostu trendu na moment/okres t:</p> $S_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot S_{t-1}$ <p>równanie III – służy do wygładzania wartości wahań okresowych:</p> $C_t = \gamma \cdot (y_t - F_t) + (1 - \gamma) \cdot C_{t-r}$ <p>gdzie: α, β, γ – określa parametr wygładzania przyjmujący wartości z przedziału $\langle 0; 1 \rangle$:</p> $y_T^* = F_t + (T - t) \cdot S_t + C_{t-r}$ <p>gdzie: T – określa kolejne numery okresów <i>ex ante</i>, r – określa długość okresu sezonowego.</p> <p>Metoda wymaga przyjęcia wartości startowych:</p> $F_1 = y_1 \text{ lub } F_1 = \frac{1}{r} \sum_{t=1}^r (y_t), \text{ zaś } S_1 = y_2 - y_1$ <p>lub jest różnicą średnich arytmetycznych z drugiego i pierwszego cyklu:</p> $C_t = y_t - \frac{1}{r} \sum_{t=1}^r (y_t)$

<p>Wintersa dla sezonowości multiplikatywnej</p>	<p>Model trzyrównaniowy równanie I – służy do wyznaczenia wygładzonych wartości szeregu czasowego (prognozowanej zmiennej) w momencie/okresie t:</p> $F_t = \alpha \cdot \left(\frac{y_t}{C_{t-r}} \right) + (1 - \alpha) \cdot (F_{t-1} + S_{t-1})$ <p>równanie II – służy do wyznaczenia wygładzonych wartości przyrostu trendu na moment/okres t:</p> $S_t = \beta \cdot (F_t - F_{t-1}) + (1 - \beta) \cdot S_{t-1}$ <p>równanie III – służy do wygładzania wartości wahań okresowych:</p> $C_t = \gamma \cdot \left(\frac{y_t}{F_t} \right) + (1 - \gamma) \cdot C_{t-r}$ <p>gdzie: α, β, γ – określa parametr wygładzania przyjmujący wartości z przedziału $(0; 1)$:</p> $y_t^* = [F_t + (T - t) \cdot S_t] \cdot C_{t-r}$ <p>gdzie: T – określa kolejne numery okresów <i>ex ante</i>, r – określa długość okresu sezonowego.</p> <p>Metoda wymaga przyjęcia wartości startowych:</p> $F_1 = y_1 \text{ lub } F_1 = \frac{1}{r} \sum_{t=1}^r (y_t), \text{ zaś } S_1 = y_2 - y_1$ <p>lub jest różnicą średnich arytmetycznych z drugiego i pierwszego cyklu:</p> $C_t = \frac{y_t}{\frac{1}{r} \sum_{t=1}^r (y_t)}$
<p><i>Modele tendencji rozwojowej</i></p>	
<p>Liniowy</p>	$y_t = a_0 + a_0 t + \varepsilon_t$ <p>gdzie: t – oznacza zmienną czasową</p>
<p><i>Wskaźniki sezonowości</i></p>	
<p>Addytywny</p>	$y_{ti}^* = y_t^* + c_i$ <p>gdzie: y_{ti}^* – określa wartość prognozy wynikającą z modelu tendencji rozwojowej, c_i – czysty wskaźnik sezonowości $c_i = z_i - \frac{1}{r} \sum_{i=1}^r z_i$, gdzie $z_i = \frac{1}{k} \sum_{i,j=1}^k z_{i+jxr,j}$ – surowe wskaźniki sezonowości – wielkości średnie wyznaczone na podstawie wielkości z_{ti}, dotyczących tej samej fazy cyklu, z_{ti} – wartości szeregu oczyszczone o tendencję rozwojową $z_{ti} = y_{ti} - y_{ti}^*$</p>

Multiplikatywny	$y_{ti}^* = y_t^* \cdot c_i$ <p>gdzie: y_t^* – określa wartość prognozy wynikającą z modelu tendencji rozwojowej, c_i – czysty wskaźnik sezonowości $c_i = \frac{z_i}{\frac{1}{r} \sum_{i=1}^r z_i}$, gdzie $z_i = \frac{1}{k} \sum_{j=1}^k z_{i+jr,j}$ – surowe wskaźniki sezonowości – wielkości średnie wyznaczone na podstawie wielkości z_{ij}, dotyczących tej samej fazy cyklu, z_{ij} – wartości szeregu oczyszczone o tendencję rozwojową</p> $z_{ij} = \frac{y_{tij}}{y_{ti}^*}$
-----------------	--

Źródło: opracowanie własne na podstawie Chatfield, 2000; Cieślak, 2008; Diebold, 1998; Gajda, 2001; Szmit, 2007; Zeliaś, Pawełek i Wanat, 2008; Welfe, 2009.

Dodatek 2

Mierniki trafności prognoz

Miernik	Interpretacja
Średni błąd (ang. <i>Mean Error</i>) $ME = \frac{1}{n} \sum_{t=1}^n (y_t - y_t^*)$	Przeciętne odchylenie prognoz od wartości rzeczywistych
Średni błąd absolutny (ang. <i>Mean Absolute Error</i>) $MAE = \frac{1}{n} \sum_{t=1}^n y_t - y_t^* $	Przeciętne bezwzględne odchylenie prognoz od wartości rzeczywistych
Średni błąd procentowy (ang. <i>Mean Percentage Error</i>) $MPE = \frac{1}{n} \sum_{t=1}^n \left(\frac{y_t - y_t^*}{y_t} \right)$	Przeciętne względne odchylenie prognoz od wartości rzeczywistych – procent, jaki w rzeczywistej wartości zmiennej y stanowiło przeciętne względne odchylenie prognoz od danych rzeczywistych
Średni absolutny błąd procentowy (ang. <i>Mean Absolute Percentage Error</i>) $MAPE = \frac{1}{n} \sum_{t=1}^n \left \frac{y_t - y_t^*}{y_t} \right $	Przeciętne bezwzględne odchylenie prognoz od wartości rzeczywistych
Średni ważony absolutny błąd procentowy (ang. <i>Mean Weighted Absolute Percentage Error</i>) $MAPE_w = \frac{1}{n} \sum_{t=1}^n \left \frac{y_t - y_t^*}{y_t} \right \cdot w_t$	Ważony błąd <i>MAPE</i> , gdzie indywidualne wartości absolutnych błędów prognoz ważone są np. udziałami poszczególnych obserwacji w sumie całkowitej $w_t = \frac{y_t}{\sum_{t=1}^n y_t}$
Średni błąd kwadratowy (ang. <i>Mean Squared Error</i>) $MSE = \frac{1}{n} \sum_{t=1}^n (y_t - y_t^*)^2$	Przeciętny błąd kwadratowy prognoz

<p>Pierwiastek średniego błędu kwadratowego (ang. <i>Root Mean Squared Error</i>)</p> $RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^n (y_t - \hat{y}_t)^2}$	Przeciętne odchylenia prognoz od wartości rzeczywistych
<p>Średni procentowy błąd kwadratowy (ang. <i>Mean Squared Percentage Error</i>)</p> $MSPE = \frac{1}{n} \sum_{t=1}^n \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2$	Przeciętny procentowy błąd kwadratowy prognoz
<p>Pierwiastek średniego procentowego błędu kwadratowego (ang. <i>Mean Squared Percentage Error</i>)</p> $RMSPE = \sqrt{\frac{1}{n} \sum_{t=1}^n \left(\frac{y_t - \hat{y}_t}{y_t} \right)^2}$	Przeciętne procentowe odchylenia prognoz od wartości rzeczywistych
<p>Współczynnik zmienności błędu średniokwadratowego <i>RMSE</i></p> $V_{RSME} = \frac{RMSE}{\bar{y}}$	Miernik względny będący ilorazem miernika <i>RMSE</i> oraz średniej wartości zmiennej prognozowanej
<p>Współczynnik Theila – I^2 – i jego dekompozycja</p> $I^2 = \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T y_t^2}$ $I^2 = I_1^2 + I_2^2 + I_3^2$	Miernik przyjmuje wartość „0”, gdy prognozy są idealnie trafne – im większe są różnice pomiędzy wartościami prognostycznymi a rzeczywistymi badanej zmiennej, tym większa jest jego wartość
$I_1^2 = \frac{\sum_{t=1}^T (y_t - \hat{y}_t)^2}{\sum_{t=1}^T y_t^2}; \quad \hat{I}_1^2 = \frac{I_1^2}{I^2} \cdot 100\%$	Część całkowitego błędu predykcji stanowiąca błąd wynikający z obciążenia predykcji (niezgodność średnich)
$I_2^2 = \frac{[S_y - \hat{S}_y]^2}{\frac{1}{T} \sum_{t=1}^T y_t^2}; \quad \hat{I}_2^2 = \frac{I_2^2}{I^2} \cdot 100\%$	Część całkowitego błędu predykcji wynikająca z niedostatecznej elastyczności predykcji (niezgodność zróżnicowania)
$I_3^2 = \frac{S_y \hat{S}_y [1 - R_{y, \hat{y}}]}{\frac{1}{T} \sum_{t=1}^T y_t^2}; \quad \hat{I}_3^2 = \frac{I_3^2}{I^2} \cdot 100\%$	Część całkowitego błędu predykcji wynikająca z niedostatecznej predykcji punktów zwrotnych (niezgodność kierunków zmian)
<p>Współczynnik Janusowy</p> $J = \frac{\frac{1}{m} \sum_{T=1}^m (y_T - \hat{y}_T)^2}{\frac{1}{n} \sum_{T=1}^n (y_T - \hat{y}_T)^2}$	Miernik informujący o aktualności prognoz wyznaczonych na podstawie modelu/metody – jeśli współczynnik Janusowy $0 < J \leq 1$, wówczas metoda/model jest aktualna/y i może być stosowana/y do konstrukcji prognoz

<p>Błąd <i>ex ante</i></p> $S_{TP} = S_e \sqrt{1 + \frac{1}{n} + \frac{(T - \bar{t})^2}{\sum_{t=1}^n (t - \bar{t})^2}}$	<p>Wartość błędu prognoz w okresie <i>ex ante</i>, gdzie: S_e – błąd standardowy modelu, T – okres <i>ex ante</i>, na który sporządzono prognozę</p>
<p>Współczynnika błędu <i>ex ante</i></p> $V_{TP} = \frac{S_{TP}}{y_T} \cdot 100\%$	<p>Informuje o względnym obciążeniu prognozy błędem <i>ex ante</i>, pozwala na ocenę dopuszczalno- ści prognozy</p>
$P\{y_T^* - t_{\alpha, n-k} \cdot S_{TP} < y_T < y_T^* + t_{\alpha, n-k} \cdot S_{TP}\} = 1 - \alpha$	<p>Pozwala na określenie – z przyjętym poziomem prawdopodobieństwa (poziomem istotności) – przedział prognozy dla przyszłej wartości</p>

Objaśnienia: proponowane mierniki nie wyczerpują listy dostępnych miar dobroci dopasowania/trafności prognoz.

Źródło: opracowanie własne na podstawie Chatfield, 2000; Cieślak, 2008; Diebold, 1998; Gajda, 2001; Szmit, 2007; Zeliaś, Pawełek i Wanat, 2008; Welfe, 2009.

Dodatek 3

Podsumowanie wyników dla metody wskaźników sezonowości addytywnej i multiplikatywnej

Tab. 96. Podsumowanie wyników dla metody wskaźników sezonowości addytywnej i multiplikatywnej

Okres, rok i kwartał		Wartość sprze- danych towarów i materia- łów y_{ti}	Wartości teoretyczne wynikające z f. trendu y_{ti}^*	z_{ti} – wartości szeregu oczysz- czone o tendencję rozwojową $z_{ti} = y_{ti} - y_{ti}^*$		Wartości prognoz y_{ti}^*		Wartość absolutnych błędów względnych – APE	
				addytywna	multi- plika- tywna	addytywna	multiplika- tywna	addy- tywna	multi- plika- tywna
2008	I	1045054.00	865284.77	179769.23	1.21	764154.03	814325.73	0.27	0.22
	II	1177822.00	900014.88	277807.12	1.31	911412.19	905319.11	0.23	0.23
	III	1182012.00	934744.99	247267.01	1.26	998031.44	969918.26	0.16	0.18
	IV	1102232.00	969475.10	132756.90	1.14	995922.08	984376.49	0.10	0.11
2009	I	1034714.00	1004205.21	30508.79	1.03	903074.47	945064.76	0.13	0.09
	II	1119030.00	1038935.32	80094.68	1.08	1050332.64	1045058.29	0.06	0.07
	III	1169460.00	1073665.43	95794.57	1.09	1136951.89	1114066.10	0.03	0.05
	IV	1273455.00	1108395.54	165059.46	1.15	1134842.52	1125432.22	0.11	0.12
2010	I	1291023.00	1143125.66	147897.34	1.13	1041994.91	1075803.79	0.19	0.17
	II	1294288.00	1177855.77	116432.23	1.10	1189253.08	1184797.46	0.08	0.08
	III	1496504.00	1212585.88	283918.12	1.23	1275872.33	1258213.94	0.15	0.16
	IV	1598712.00	1247315.99	351396.01	1.28	1273762.97	1266487.95	0.20	0.21
2011	I	1501568.00	1282046.10	219521.90	1.17	1180915.36	1206542.82	0.21	0.20
	II	1630864.00	1316776.21	314087.79	1.24	1328173.52	1324536.63	0.19	0.19
	III	1702242.00	1351506.32	350735.68	1.26	1414792.77	1402361.79	0.17	0.18
	IV	1693026.00	1386236.43	306789.57	1.22	1412683.41	1407543.68	0.17	0.17
2012	I	1374185.00	1420966.54	-46781.54	0.97	1319835.80	1337281.85	0.04	0.03
	II	1477986.00	1455696.65	22289.35	1.02	1467093.97	1464275.80	0.01	0.01
	III	1471870.00	1490426.76	-18556.76	0.99	1553713.22	1546509.63	0.06	0.05
	IV	1336526.00	1525156.88	-188630.88	0.88	1551603.85	1548599.41	0.16	0.16

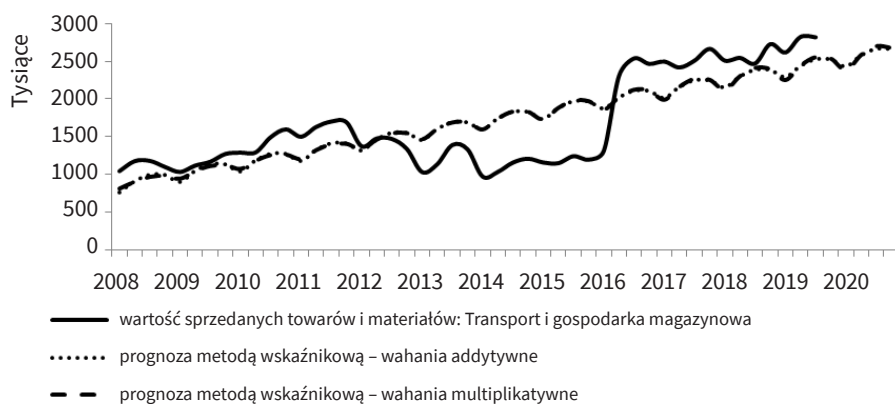
Okres, rok i kwartał		Wartość sprze- danych towarów i materia- łów y_{it}	Wartości teoretyczne wynikające z f. trendu y_{it}^*	z _{it} – wartości szeregu oczysz- czone o tendencję rozwojową $z_{it} = y_{it} - y_{it}^*$		Wartości prognoz y_t^*		Wartość absolutnych błędów względnych – APE	
				addytywna	multi- plika- tywna	addytywna	multiplika- tywna	addy- tywna	multi- plika- tywna
2013	I	1032837.00	1559886.99	-527049.99	0.66	1458756.24	1468020.88	0.41	0.42
	II	1137918.00	1594617.10	-456699.10	0.71	1606014.41	1604014.97	0.41	0.41
	III	1392573.00	1629347.21	-236774.21	0.85	1692633.66	1690657.47	0.22	0.21
	IV	1334069.00	1664077.32	-330008.32	0.80	1690524.30	1689655.14	0.27	0.27
2014	I	972855.00	1698807.43	-725952.43	0.57	1597676.69	1598759.91	0.64	0.64
	II	1035474.00	1733537.54	-698063.54	0.60	1744934.85	1743754.14	0.69	0.68
	III	1159489.00	1768267.65	-608778.65	0.66	1831554.10	1834805.31	0.58	0.58
	IV	1209548.00	1802997.76	-593449.76	0.67	1829444.74	1830710.87	0.51	0.51
2015	I	1159994.00	1837727.87	-677733.87	0.63	1736597.13	1729498.94	0.50	0.49
	II	1151885.00	1872457.98	-720572.98	0.62	1883855.30	1883493.31	0.64	0.64
	III	1243452.00	1907188.10	-663736.10	0.65	1970474.55	1978953.16	0.58	0.59
	IV	1197904.00	1941918.21	-744014.21	0.62	1968365.19	1971766.60	0.64	0.65
2016	I	1312221.00	1976648.32	-664427.32	0.66	1875517.57	1860237.97	0.43	0.42
	II	2303303.00	2011378.43	291924.57	1.15	2022775.74	2023232.48	0.12	0.12
	III	2537727.00	2046108.54	491618.46	1.24	2109394.99	2123101.00	0.17	0.16
	IV	2465351.00	2080838.65	384512.35	1.18	2107285.63	2112822.33	0.15	0.14
2017	I	2496154.00	2115568.76	380585.24	1.18	2014438.02	1990977.00	0.19	0.20
	II	2417979.00	2150298.87	267680.13	1.12	2161696.18	2162971.65	0.11	0.11
	III	2509496.00	2185028.98	324467.02	1.15	2248315.43	2267248.84	0.10	0.10
	IV	2664111.00	2219759.09	444351.91	1.20	2246206.07	2253878.06	0.16	0.15
2018	I	2508793.00	2254489.20	254303.80	1.11	2153358.46	2121716.03	0.14	0.15
	II	2543691.00	2289219.32	254471.68	1.11	2300616.63	2302710.82	0.10	0.09
	III	2469931.00	2323949.43	145981.57	1.06	2387235.88	2411396.69	0.03	0.02
	IV	2727023.00	2358679.54	368343.46	1.16	2385126.52	2394933.79	0.13	0.12
2019	I	2615952.00	2393409.65	222542.35	1.09	2292278.90	2252455.06	0.12	0.14
	II	2822208.00	2428139.76	394068.24	1.16	2439537.07	2442449.99	0.14	0.13
	III	2817123.00	2462869.87	354253.13	1.14	2526156.32	2555544.53	0.10	0.09
	IV					2524046.96	2535989.52	MAPE	
2020	I					2431199.35	2383194.09		
	II					2578457.51	2582189.16		
	III					2665076.76	2699692.37		
	IV					2662967.40	2677045.26		

Źródło: opracowanie własne na podstawie danych GUS.

Tab. 97. Wartości surowych i czystych wskaźników sezonowości

Kwartał	Surowe wskaźniki sezonowości z_i		Czyste wskaźniki sezonowości c_i	
	addytywna	multiplikatywna	addytywna	multiplikatywna
I	-100568.04	0.95	-101130.74	0.94
II	11960.01	1.02	11397.31	1.01
III	63849.15	1.05	63286.45	1.04
IV	27009.68	1.03	26446.98	1.02

Źródło: opracowanie własne na podstawie danych GUS.

**Rys. 53.** Wartość sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa w tys. PLN w latach 2008–2020 (dane kwartalne) wraz z prognozami *ex post* i *ex ante* dla metody wskaźników sezonowości

Źródło: opracowanie własne na podstawie danych GUS.

Tab. 98. Podsumowanie wyników dla modelu Wintersa dla sezonowości addytywnej

Okres, rok i kwartał		Wyglądzo- ne warto- ści szeregu czasowego F_t	Wyglądzone wartości przyrostu trendu S_t	Wyglądzone wartości wahań okre- sowych C_t	Wartości prognoz y_t^*	Wartość absolutnych błędów względnych – APE
2008	I	---	---	-81726.00	---	---
	II			51042.00		
	III			55232.00		
	IV			-24548.00		
2009	I	1118510.90	22381.68	-83796.69	1067438.75	0.03
	II	1072601.57	22374.86	46428.89	1191934.58	0.07
	III	1113009.72	22376.66	56450.16	1150208.43	0.02
	IV	1287712.24	22391.89	-14258.27	1110838.37	0.13
2010	I	1370724.34	22397.95	-79701.75	1226307.44	0.05
	II	1257051.71	22384.35	37237.21	1439551.18	0.11
	III	1429889.57	22399.39	66613.42	1335886.21	0.11
	IV	1602801.98	22414.44	-4091.00	1438030.69	0.10
2011	I	1584050.80	22410.33	-82482.52	1545514.67	0.03
	II	1594438.97	22409.12	36425.11	1643698.34	0.01
	III	1634440.11	22410.88	67801.77	1683461.51	0.01
	IV	1694568.87	22414.66	-1543.12	1652760.00	0.02
2012	I	1473140.92	22390.27	-98954.28	1634501.01	0.19
	II	1444976.26	22385.22	33010.08	1531956.30	0.04
	III	1408073.57	22379.29	63796.83	1535163.24	0.04
	IV	1343915.38	22370.63	-7388.80	1428909.74	0.07
2013	I	1146630.65	22348.67	-113792.17	1267331.74	0.23
	II	1108962.51	22342.67	28955.90	1201989.40	0.06
	III	1316279.75	22361.16	76292.00	1195102.00	0.14
	IV	1341279.54	22361.43	-7210.56	1331252.11	0.00
2014	I	1104175.98	22335.48	-131319.23	1249848.80	0.28
	II	1014111.56	22324.24	21363.20	1155467.37	0.12
	III	1080237.84	22328.62	79250.86	1112727.80	0.04
	IV	1209532.21	22339.32	15.06	1095355.90	0.09
2015	I	1287551.62	22344.89	-127558.00	1100552.30	0.05
	II	1141873.05	22328.08	10013.08	1331259.71	0.16
	III	1164201.14	22328.08	79250.86	1243452.00	0.00
	IV	1197170.07	22329.15	733.86	1186544.28	0.01

Okres, rok i kwartał		Wyglądzo- ne warto- ści szeregu czasowego F_t	Wyglądzone wartości przyrostu trendu S_t	Wyglądzone wartości wahań okre- sowych C_t	Wartości prognoz y_t^*	Wartość absolutnych błędów względnych – APE
2016	I	1425839.18	22349.78	-113619.57	1091941.22	0.17
	II	2239809.95	22428.94	63487.70	1458202.04	0.37
	III	2446057.78	22447.33	91667.97	2341489.76	0.08
	IV	2464863.18	22446.96	487.85	2469238.97	0.00
2017	I	2602023.80	22458.43	-105870.57	2373690.57	0.05
	II	2371576.96	22433.14	46403.75	2687969.93	0.11
	III	2416320.77	22435.37	93175.08	2485678.07	0.01
	IV	2649393.04	22456.44	14716.53	2439243.99	0.08
2018	I	2618282.43	22451.08	-109489.07	2565978.91	0.02
	II	2506364.86	22437.64	37327.04	2687137.26	0.06
	III	2386377.79	22423.40	83554.18	2621977.58	0.06
	IV	2693099.94	22451.83	33921.14	2423517.72	0.11
2019	I	2724815.25	22452.76	-108863.32	2606062.70	0.00
	II	2782500.72	22456.28	39707.04	2784595.05	0.01
	III	2738086.43	22449.59	79037.02	2888511.18	0.03
	IV				2794457.16	MAPE
2020	I				2674122.30	
	II				2845142.25	7.82%
	III				2906921.82	
	IV				2884255.53	

Objaśnienia: dla parametrów wygładzania: $\alpha = 0.94$, $\beta = 0.01$, $\gamma = 0.99$.

Źródło: opracowanie własne na podstawie danych GUS.

Tab. 99. Podsumowanie wyników dla modelu Wintersa dla sezonowości multiplikatywnej

Okres, rok i kwartał		Wyglądzone wartości szeregu cza- sowego F_t	Wyglądzone wartości przyrostu trendu S_t	Wyglądzone wartości wahań okresowych C_t	Wartości prognoz y'_t	Wartość absolutnych błędów względnych - APE
2008	I	---	---	0.93	---	---
	II			1.05		
	III			1.05		
	IV	1126780.00	22384.75	0.98		
2009	I	1123301.42	21454.58	0.92	1065815.17	0.03
	II	1087512.08	19395.80	1.03	1196612.29	0.07
	III	1113006.05	19615.12	1.05	1161165.80	0.01
	IV	1263116.57	24308.38	1.01	1107945.91	0.13
2010	I	1375449.00	27474.16	0.94	1185895.00	0.08
	II	1291018.40	23449.51	1.00	1443584.51	0.12
	III	1399149.51	26495.08	1.07	1381140.20	0.08
	IV	1549119.85	30935.86	1.03	1437309.01	0.10
2011	I	1595257.31	31482.58	0.94	1483068.04	0.01
	II	1626739.86	31482.58	1.00	1630864.04	0.00
	III	1606766.04	29631.96	1.06	1773600.42	0.04
	IV	1639571.02	29746.07	1.03	1688780.29	0.00
2012	I	1507820.29	23937.85	0.91	1571277.82	0.14
	II	1487402.57	22342.61	0.99	1535641.49	0.04
	III	1416859.97	19002.00	1.04	1599457.58	0.09
	IV	1326700.16	15076.00	1.01	1482675.33	0.11
2013	I	1180964.03	9292.41	0.87	1222861.06	0.18
	II	1155480.17	8041.68	0.98	1182722.12	0.04
	III	1300038.53	12951.49	1.07	1208698.34	0.13
	IV	1321680.62	13264.05	1.01	1322717.64	0.01
2014	I	1163282.94	7090.25	0.84	1167508.93	0.20
	II	1078654.32	3791.58	0.96	1152585.70	0.11
	III	1082445.90	3791.58	1.07	1159489.01	0.00
	IV	1172680.58	6900.50	1.03	1096418.79	0.09
2015	I	1339593.18	12655.33	0.87	986489.67	0.15
	II	1234759.32	8429.84	0.93	1298119.12	0.13
	III	1179667.83	6145.30	1.05	1331673.16	0.07
	IV	1166979.33	5467.94	1.03	1223090.79	0.02

2016	I	1436952.54	14980.85	0.91	1015253.86	0.23
	II	2236376.20	43193.28	1.03	1354486.77	0.41
	III	2378275.45	46743.24	1.07	2402828.81	0.05
	IV	2407037.67	46096.55	1.02	2489282.28	0.01
2017	I	2669323.73	53871.79	0.94	2240183.64	0.10
	II	2433619.09	43457.20	0.99	2804664.66	0.16
	III	2380470.56	39982.78	1.05	2643148.73	0.05
	IV	2559777.70	44993.57	1.04	2479092.23	0.07
2018	I	2664987.96	47159.26	0.94	2435783.61	0.03
	II	2594904.43	42942.63	0.98	2694726.97	0.06
	III	2410391.30	34762.20	1.02	2780826.12	0.13
	IV	2580185.06	39618.60	1.06	2544810.82	0.07
2019	I	2742448.75	44029.52	0.95	2466255.31	0.06
	II	2857856.96	46596.65	0.99	2731487.62	0.03
	III	2784715.59	42290.28	1.01	2976205.83	0.06
	IV				2987885.78	MAPE
2020	I				2736944.92	
	II				2875265.14	8,60%
	III				2988256.61	
	IV				3121976.62	

Objaśnienia: dla parametrów wygładzania: $\alpha = 0.77$, $\beta = 0.04$, $\gamma = 0.99$.

Źródło: opracowanie własne na podstawie danych GUS.

Dodatek 4

Wybrane elementy teorii prawdopodobieństwa

1. Pojęcia podstawowe

Zdarzenie elementarne E – jeden lub kilka możliwych wyników działania.

Przestrzeń zdarzeń elementarnych (przestrzeń Ω) – lista zdarzeń zawierająca wszystkie możliwe wyniki działania; może być skończona lub nieskończona.

Eksperyment losowy – działanie, którego skutkiem jest zdarzenie losowe.

Zdarzenie losowe A – podzbiór przestrzeni zdarzeń elementarnych; jeżeli podzbiór nie zawiera żadnego zdarzenia elementarnego, mówimy że A jest zdarzeniem niemożliwym; jeżeli zawiera wszystkie elementy przestrzeni Ω , mówimy że A jest zdarzeniem pewnym. Zdarzenia elementarne należące do tego podzbioru nazywamy zdarzeniami sprzyjającymi zdarzeniu losowemu A .

Zdarzenie przeciwne do zdarzenia A – nazywamy tak zdarzenie \bar{A} zawierające wszystkie zdarzenia elementarne, które nie należą do A ; $A \cup \bar{A} = \Omega$, $A \cap \bar{A} = \emptyset$.

Zdarzenia A i B wzajemnie się wykluczające – wówczas, gdy w danym czasie może mieć miejsce jedno i tylko jedno z nich; $A \cap B = \emptyset$.

2. Prawdopodobieństwo – definicje

Prawdopodobieństwo klasyczne (*a priori*) – jeżeli przestrzeń zdarzeń elementarnych Ω jest zbiorem skończonym złożonym z n jednakowo możliwych

zdarzeń, to prawdopodobieństwo zdarzenia losowego A , któremu sprzyja m zdarzeń elementarnych, jest równe:

$$P(A) = \frac{m}{n} \quad (297)$$

Prawdopodobieństwo jako względna częstość pojawiania się – obserwowana względna częstość pojawiania się zdarzenia przy dużej liczbie prób lub procent czasu, kiedy zjawisko miało miejsce w stabilnych warunkach w długim okresie czasu.

Prawdopodobieństwo subiektywne – podstawą jego określenia jest przekonanie osoby oceniającej. Stosuje się je dla zdarzeń rzadko występujących.

Prawdopodobieństwo sumy zdarzeń losowych:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (298)$$

Prawdopodobieństwo warunkowe $P(A|B)$ – Prawdopodobieństwo zdarzenia A przy założeniu, że wystąpiło zdarzenie B . Jeśli $P(B) > 0$, to obliczamy je następująco:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \quad (299)$$

Prawdopodobieństwo całkowite – jeżeli zdarzenia B_i , $i = 1, \dots, n$, takie że $P(B_i) > 0$, $i = 1, \dots, n$, wykluczają się parami oraz ich suma jest zdarzeniem pewnym¹, to zachodzi:

$$P(A) = \sum_{i=1}^n P(B_i) \cdot P(A|B_i) \quad (300)$$

Prawdopodobieństwo *a posteriori* (wzór Bayesa) – jeżeli zdarzenia B_i , $i = 1, \dots, n$, takie że $P(B_i) > 0$, $i = 1, \dots, n$, wykluczają się parami oraz ich suma jest zdarzeniem pewnym, to zachodzi:

$$P(B_i|A) = \frac{P(B_i) \cdot P(A|B_i)}{\sum_{i=1}^n P(B_i) \cdot P(A|B_i)} \quad (301)$$

W praktyce prawdopodobieństwa *a priori* $P(B_i)$ nie są często znane i zastępuje się je prawdopodobieństwami subiektywnymi.

¹ Prawdopodobieństwa $P(B_i)$, $i = 1, \dots, n$ nazywane są prawdopodobieństwami *a priori*.

3. Zmienna losowa i jej wybrane charakterystyki

Zmienna losowa – funkcja X o wartościach rzeczywistych określona na przestrzeni zdarzeń elementarnych; przyjmuje różne wartości z określonym prawdopodobieństwem $X: \Omega \rightarrow R$.

Dystrybuanta zmiennej losowej X – funkcja określona następująco:

$$F(x) = P(X < x) \quad (302)$$

Własności dystrybuanty:

1. $0 \leq F(x) \leq 1$.
2. $F(x)$ jest funkcją niemalejącą.
3. $\lim_{x \rightarrow -\infty} F(x) = 0$ oraz $\lim_{x \rightarrow \infty} F(x) = 1$.
4. Dystrybuanta jest funkcją co najmniej lewostronnie ciągłą.

Rozkład prawdopodobieństwa zmiennej losowej dyskretnej (skokowej):

$$P(X = x_i) = p_i, \quad i = 1, \dots, n; \quad \sum_{i=1}^n p_i = 1 \quad (303)$$

Funkcja gęstości zmiennej losowej ciągłej – jest to nieujemna funkcja $f(x) \geq 0$, określona na zbiorze liczb rzeczywistych, spełniająca warunek:

$$\int_{-\infty}^{\infty} f(x) dx = 1 \quad (304)$$

Dystrybuanta zmiennej losowej dyskretnej:

$$F(x) = \sum_{x_i < x} p_i \quad (305)$$

Dystrybuanta zmiennej losowej ciągłej:

$$F(x) = \int_{-\infty}^x f(x) dx \quad (306)$$

Prawdopodobieństwo przedziałowe dla zmiennej losowej ciągłej:

$$P(a \leq X \leq b) = P(a < X < b) = F(b) - F(a) \quad (307)$$

Wartość oczekiwana zmiennej losowej $E(X)$ – wartość, wokół której skupiają się wartości zmiennej losowej X przy wielu powtórzeniach eksperymentu.

- dla zmiennej dyskretnej: $E(X) = \sum_{i=1}^n x_i p_i \quad (308)$

- dla zmiennej ciągłej: $E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (309)$

Wariancja zmiennej losowej $D^2(X)$ – miara zróżnicowania (dyspersji) wartości zmiennej losowej X zdefiniowana następująco:

$$D^2(X) = E[X - E(X)]^2 = E(X^2) - [E(X)]^2 \quad (310)$$

- dla zmiennej dyskretnej: $D^2(X) = \sum_{i=1}^n [x_i - E(X)]^2 p_i \quad (311)$

- dla zmiennej ciągłej: $D^2(X) = \int_{-\infty}^{\infty} [x - E(X)]^2 f(x) dx \quad (312)$

4. Wybrane jednowymiarowe rozkłady prawdopodobieństwa

Rozkład zero-jedynkowy:

$$\begin{aligned} P(X=1) &= p \\ P(X=0) &= q \\ p+q &= 1, \quad 0 < p, q < 1 \end{aligned} \quad (313)$$

Wartość oczekiwana: $E(X) = p$

Wariancja: $D^2(X) = pq$

Rozkład Poissona:

Funkcja rozkładu prawdopodobieństwa: $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (314)$

Wartość oczekiwana: $E(X) = \lambda$

Wariancja: $D^2(X) = \lambda$

Rozkład równomierny (jednostajny) w przedziale $[a, b]$:

$$\text{Funkcja gęstości: } f(x) = \begin{cases} 0 & \text{dla } x < a \\ \frac{1}{b-a}, & \text{dla } a \leq x \leq b \\ 0 & \text{dla } x > b \end{cases} \quad (315)$$

Wartość oczekiwana: $E(X) = \frac{a+b}{2} \quad (316)$

Wariancja: $D^2(X) = \frac{1}{12} (b-a)^2 \quad (317)$

Rozkład normalny (Gaussa–Laplace’a):

$$\text{Funkcja gęstości: } f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] \quad (318)$$

Wartość oczekiwana: μ

Wariancja: σ^2

Zapis symboliczny zmiennej: $X \sim N(\mu, \sigma)$

Rozkład wykładniczy:

$$\text{Funkcja gęstości: } f(x) = \begin{cases} 0 & \text{dla } x < 0 \\ \lambda e^{-\lambda x}, & \text{dla } x \geq 0 \end{cases} \quad (319)$$

$$\text{Wartość oczekiwana: } E(X) = \frac{1}{\lambda} \quad (320)$$

$$\text{Wariancja: } D^2(X) = \frac{1}{\lambda^2} \quad (321)$$

Dodatek 5

Optymalizacja liniowa w środowisku R

Środowisko obliczeniowe R to wolne i otwarte środowisko obliczeniowe oraz język programowania przeznaczony do przeprowadzania obliczeń statystycznych i numerycznych, analizy danych i tworzenia wysokiej jakości grafiki. Wykorzystuje licencję GNU/GPL – choć istnieją także zamknięte, płatne wersje. Skupiona wokół tego oprogramowania społeczność udostępnia pakiety przydatne do analiz wszelkiego rodzaju. Ich liczba w momencie powstawiania tej książki przekracza piętnaście tysięcy.

Choć środowisko R w literaturze przedstawia się jako zbiór narzędzi przydatnych w statystyce czy prognozowaniu, to nadaje się ono również do rozwiązywania problemów optymalizacyjnych. Niektóre pakiety dysponują funkcjami pozwalającymi na wyszukanie największej lub najmniejszej wartości funkcji. Skupimy się jednak na pakietach dedykowanych, zapewniających interfejs do ogólnodostępnych bibliotek optymalizacyjnych.

W R korzysta się z linii poleceń lub skryptów. Na potrzeby niniejszej publikacji zakładamy, że czytelnik zna podstawy obsługi. Pozostałych zachęcamy do sięgnięcia do książki Przemysława Biecka *Przewodnik po pakiecie R* i Marka Gągolewskiego *Programowanie w języku R*. Polecamy również dokumentację dostępną na: <https://cran.r-project.org/> (to strona *The Comprehensive R Archive Network*, czyli sieci serwerów przechowujących najnowsze wersje zarówno samego środowiska, jak i pakietów).

Żaden z dostępnych pakietów nie oferuje jak dotąd metody graficznej. Mimo tego można przedstawić graficznie zbiór rozwiązań dopuszczalnych i np. umieścić na wykresie gradient. Zaprezentujemy kod pozwalający uzyskać taki efekt.

Środowisko R dysponuje dużymi możliwościami tworzenia zaawansowanej grafiki już w swojej podstawowej wersji. Skorzystamy jednak z pakietu `ggplot2()`, który bazuje na pomysle zwanym „gramatyką grafiki”. W tym podejściu określa się bloki konstrukcyjne wykresu (jak dane, transformacje, tzw. estetyki, skalę itd.), następnie łączy się je i uzyskuje efekt końcowy. Kolejne elementy można interpretować jak warstwy nakładane na wykres. Istnieje możliwość ustawienia właściwości poszczególnych warstw, ich dziedziczenia itd. To wygodne rozwiązanie

w sytuacji, kiedy nanosimy kolejne ograniczenia. Pakiet `ggplot2()` nie jest dostępny w standardowej instalacji R i trzeba go doinstalować.

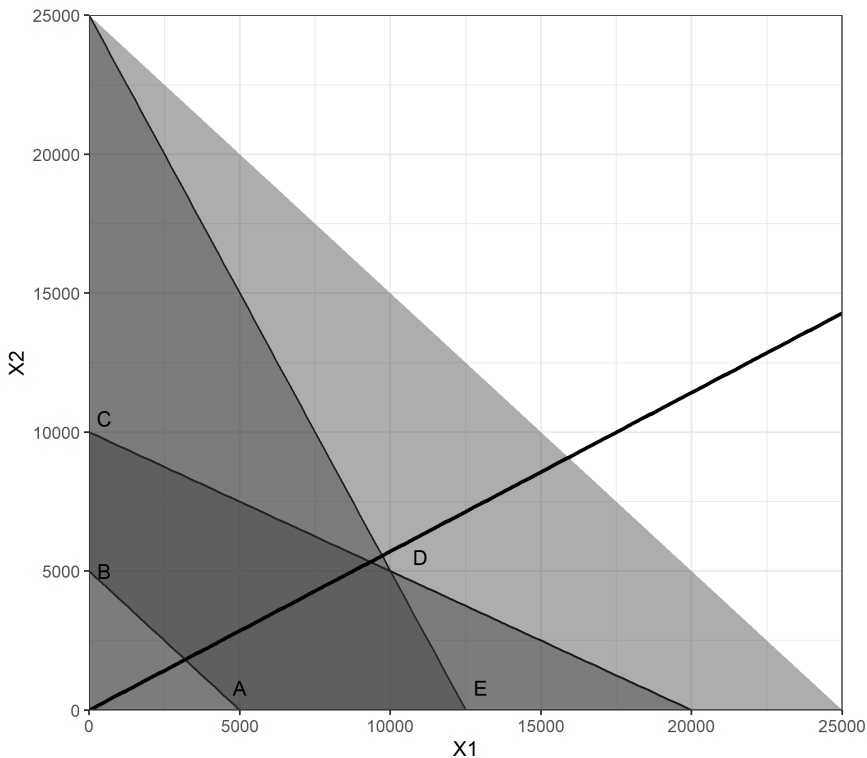
Poniżej prezentujemy kod pozwalający narysować proste wyznaczające ograniczenia oraz gradient funkcji dla przykładu 1 (z rozdziału pierwszego książki) oraz efekt działania tych poleceń. Brak gotowych funkcji automatyzujących całą procedurę powoduje, że przed przystąpieniem do tworzenia wykresu musimy się odpowiednio przygotować. Na początek należy wyznaczyć punkty przecięcia z osiami i inne, niezbędne do naszkicowania ograniczeń i gradientu współrzędne. Omówimy teraz kolejne linie składające się na wywołanie funkcji `ggplot()`.

```
library(ggplot2)
```

```
ggplot(data.frame(x = c(0, 25000)), aes(x = x)) +
  stat_function(fun = function(x) {10000 - 0.5*x}) +
  stat_function(fun = function(x) {25000 - 2*x}) +
  stat_function(fun = function(x) {5000 - x}) +
  theme_bw() +
  geom_polygon(data=data.frame(x=c(0, 0, 20000, Inf),
y = c(0, 10000, 0, 0)),
    aes(x = x, y = y),
    inherit.aes = FALSE, alpha = 0.4) +
  geom_polygon(data=data.frame(x=c(0, 0, 12500, Inf),
y = c(0, 25000, 0, 0)),
    aes(x = x, y = y),
    inherit.aes = FALSE, alpha = 0.4) +
  geom_polygon(data=data.
frame(x=c(0,0,25000,5000),y=c(5000, 25000, 0, 0)),
    aes(x = x, y = y),
    inherit.aes = FALSE, alpha = 0.4) +
  scale_x_continuous(expand = c(0, 0)) +
  scale_y_continuous(expand = c(0, 0), limits = c(0,
25000))+
  labs(x = "X1", y = "X2") +
  stat_function(fun = function(x) {(4/7)*x}, size = 1.1) +
  annotate("text", x = c(5000, 500, 500, 11000, 13000),
y = c(800, 5000, 10500, 5500, 800), label = c("A", "B",
"C", "D", "E"))
```

Funkcja `library(ggplot2)` ładuje pakiet `ggplot2()`, którego najważniejszą z kolei funkcją jest `ggplot()`. Jako pierwszą wczytujemy ramkę danych przechowującą krańce przedziału zmiennej znajdującej się na osi odciętych. W estetyce `aes` do-

myślnie oznacza się ją jako x . Kolejne warstwy dodawane są za pomocą znaku plus. Następne trzy linie kodu wywołują funkcję `stat_function()` z pakietu `ggplot()`, która oblicza wartość zmiennej y (na osi rzędnych) dla każdego z ograniczeń. Przyjęliśmy, tym samym, że zmienna x w powyższym kodzie odpowiada zmiennej decyzyjnej $X1$, zaś y zmiennej decyzyjnej $X2$. Poleceniem `theme_bw()` zmieniamy kolor tła na biały, zaś `scale_color_discrete()` tworzy legendę dodawanych prostych.



Rys. 54. Efekt działania kodu tworzącego zbiór rozwiązań dopuszczalnych

Źródło: opracowanie własne.

Po naniesieniu prostych przechodzimy do zaznaczenia zbiorów punktów spełniających odpowiednie ograniczenia. W tym celu trzeba wcześniej znać ich punkty przecięcia z osiami. Staną się one wierzchołkami wieloboków rysowanych przy pomocy funkcji `geom_polygon()`. Argument `data` tej funkcji przechowuje ramkę danych ze wspomnianymi współrzędnymi. Stopniem przeźroczystości wieloboku steruje parametr `alpha`. Kolejne wieloboki nakładają się na siebie. Obszar zaznaczony najciemniejszym kolorem to zbiór rozwiązań dopuszczalnych. Wyróżniony

został dodatkowo przez przypisanie wierzchołkom kolejnych liter alfabetu (ostatnia linia powyższego kodu).

Argument *expand* funkcji *scale_x_continuous()* i *scale_y_continuous()* spowodował, że obie osie przecinają się w punkcie (0,0). Domyślnie wszystkie wykresy w R mają osie nieco odsunięte od obszaru kreślenia, ale tym razem ważne było, aby punkty faktycznie leżały na osiach wykresu. Ponownie użyjemy funkcji *stat_function()*, żeby narysować gradient funkcji celu. Efekty działania całego kodu znajdują się na rysunku 54.

Kod przedstawiony powyżej nie służy bezpośrednio rozwiązywaniu modelu. Z tym zadaniem poradzą sobie jednak pakiety przeznaczone do optymalizacji. Na serwerach CRAN znajdziemy m.in. takie, które zapewniają interfejs do uznanych bibliotek optymalizacyjnych (zwanymi *solverami*), komercyjnych i udostępnianych na zasadach *open source*. Jedną z najpopularniejszych okazuje się darmowa biblioteka *lp_solve*, opracowana z myślą o problemach programowania liniowego i całkowitoliczbowego. Bibliotekę tę można znaleźć pod adresem: http://lp_solve.sourceforge.net/. Obsługę *solvera* zapewniają pakiety: *lpSolve()* i *lpSolveAPI()*. Pierwszy z nich zawiera zestaw predefiniowanych funkcji dla konkretnych typów zadań optymalizacyjnych, drugi pozwala na zaawansowaną kontrolę nad wprowadzaniem danych, ustawieniami parametrów itd.

Jako pierwszy przedstawimy pakiet *lpSolve()*. Poniższy kod ilustruje rozwiązanie modelu PL. Ponownie wykorzystamy przykład 1 (z pierwszego rozdziału książki).

library(lpSolve)

```
#Parametry funkcji celu
fcelu <- c(70, 40)
# Macierz współczynników lewych stron ograniczeń
A <- matrix(c(1, 2,
              2, 1,
              3, 3), nrow = 3, byrow = TRUE)
# Kierunki nierówności w ograniczeniach
ogr_kierunki <- c(„<=”, „<=”, „>=”)
# Wektor prawych stron ograniczeń
b <- c(20000, 25000, 15000)

# Rozwiązanie modelu przy pomocy funkcji lp() i zapisanie
go do zmiennej
# Ostatni parametr decyduje, czy mają zostać obliczone
składowe analizy wrażliwości
pl_rozw <- lp(„max”, fcelu, A, ogr_kierunki, b, compute.
sens = TRUE)
```

```
pl_rozw$objval # Optymalna wartość funkcji celu
## [1] 9e+05
pl_rozw$solution # Optymalne wartości zmiennych
## [1] 10000 5000
pl_rozw$duals # Optymalne wartości zmiennych dualnych oraz
przyrostów krańcowych
## [1] 3.333333 33.333333 0.000000 0.000000 0.000000
```

Następny przykład ilustruje, w jaki sposób w pakiecie *lpSolve()* rozwiązać można zadanie transportowe. Użytkownik wprowadza do funkcji *lp.transport()* macierz kosztów jednostkowych oraz wektory podaży i popytu.

```
library(lpSolve)

# Macierz kosztów jednostkowych
koszty <- matrix(c(3, 3, 6,
                  2, 4, 4,
                  5, 2, 4,
                  6, 1, 3), nrow = 4, byrow = TRUE)
# Typy ograniczeń dla podaży i popytu
wie_znaki <- rep („=", 4)
kol_znaki <- rep („=", 3)
# Wektory podaży i popytu
a <- c(11, 22, 23, 24)
b <- c(45, 13, 22)
# Rozwiązanie modelu
transp_rozw <- lp.transport(koszty, „min”, wie_znaki,
                           a, kol_znaki, b)
transp_rozw$objval # Optymalna wartość funkcji celu
## [1] 227
transp_rozw$solution # Optymalna macierz przewozów
##   [,1] [,2] [,3]
## [1,]  11  0  0
## [2,]  22  0  0
## [3,]  12  0  11
## [4,]   0 13  11
```

W pakiecie tym znajdziemy również funkcję służącą do rozwiązywania problemu optymalnego przydziału z warunkami pobocznymi (z binarnymi zmiennymi decyzyjnymi).

library(lpSolve)

```
# Macierz kosztów jednostkowych
koszty <- matrix(c(13, 16, 12, 25,
                  18, 24, 23, 19,
                  25, 14, 20, 27,
                  0, 0, 0, 0), nrow = 4, byrow = TRUE)
# Rozwiązanie modelu
przydz_rozw <- lp.assign(koszty, „min”)
przydz_rozw$objval # Optymalna wartość funkcji celu
## [1] 44
przydz_rozw$solution # Macierz optymalnego przydziału
##      [,1] [,2] [,3] [,4]
## [1,]  0  0  1  0
## [2,]  1  0  0  0
## [3,]  0  1  0  0
## [4,]  0  0  0  1
```

Pakiet *lpSolveAPI()* pozwala na rozszerzenie kontroli nad biblioteką *lp_solve* – w porównaniu do *lpSolve()*. W odróżnieniu od tego ostatniego nie zawiera wbudowanych, gotowych funkcji przeznaczonych do rozwiązywania konkretnych problemów optymalizacyjnych. Zamiast tego rozбивa cały proces przetwarzania obiektów, z których zbudowany jest model, na ciąg etapów – oddając do dyspozycji użytkownika zestaw narzędzi sterujących przebiegiem całego procesu. W pakiecie *lpSolveAPI()* istnieje np. możliwość wczytania modelu z zewnętrznego pliku, oczywiście po odpowiednim jego sformatowaniu. Zachęcamy do zapoznania się z dokumentacją, aby zobaczyć pełnię możliwości pakietu.

Kod pozwalający rozwiązać problem decyzyjny firmy Paralel S.A. z rozdziału pierwszego w pakiecie *lpSolveAPI()* wyglądać może następująco:

library(lpSolveAPI)

```
# Utworzenie obiektu: 3 ograniczenia (wiersze) i 2 zmienne (kolumny)
pl <- make.lp(3, 2)
# Ustawianie współczynników lewych stron ograniczeń
set.row(pl, 1, c(1, 2))
set.row(pl, 2, c(2, 1))
set.row(pl, 3, c(3, 3))
# Wprowadzanie parametrów funkcji celu
set.objfn(pl, c(70, 40))
```

```
# Ustawianie kierunku optymalizacji
lp.control(pl, sense = „max”)
# Ustawianie kierunków ograniczeń
set.constr.type(pl, c(„<=”, „<=”, „>=”))
# Ustawianie prawych stron ograniczeń
set.rhs(pl, c(20000, 25000, 15000))
# Nadawanie nazw elementom modelu
name.lp(pl, „Paralel”)
nazwy_zm <- c(„X1”, „X2”)
nazwy_ogr <- c(„tworzywo”, „czas”, „wkład”)
dimnames(pl) <- list(nazwy_ogr, nazwy_zm)
# Wyświetlamy gotowy model
pl
## Model name: Paralel
##      X1  X2
## Maximize  70  40
## tworzywo   1   2 <= 20000
## czas       2   1 <= 25000
## wkład      3   3 >= 15000
## Kind      Std  Std
## Type      Real Real
## Upper     Inf  Inf
## Lower     0   0
solve(pl) # Rozwiązujemy model
## [1] 0
get.objective(pl) # Optymalna wartość funkcji celu
## [1] 9e+05
get.variables(pl) # Optymalne wartości zmiennych
decyzyjnych
## [1] 10000 5000
get.constraints(pl) # Wartości lewych stron ograniczeń
w rozw. optym.
## [1] 20000 25000 45000
# Wartości zmiennych dualnych i przyrostów krańcowych
# Nie interpretujemy pierwszej z wyświetlanych wartości
get.dual.solution(pl)
## [1] 1.000000 3.333333 33.333333 0.000000 0.000000
0.000000
```

Jak widać, przygotowanie kodu wymaga zaangażowania. Pakiet *lpSolveAPI()* współpracuje też z plikami zewnętrznymi. Technicznie są to pliki tekstowe o określonej strukturze. Poniższe polecenie zapisze model przypisany do obiektu *pl* do

pliku *paralel.lp*. Rozszerzenie *.lp* to jeden z trzech dostępnych formatów zapisu, przy czym – akurat ten jest najbardziej zbliżony do sposobu, w jaki zwykle intuicyjnie zapisujemy modele.

```
write.lp(pl, "paralel.lp", type = c("lp"), use.
names = c(TRUE, TRUE))
```

Uzyskujemy efekt jak poniżej. Nazwa problemu oraz poszczególne sekcje pliku ujęte są w znaki pełniące jednocześnie funkcję ograniczników komentarzy. Poszczególne sekcje oddzielamy pustym wierszem. Funkcją *read.lp()* wczytamy model do zmiennej, by go następnie rozwiązać.

```
/* Paralel */

/* Objective function */
max: +7 X1 +4 X2;

/* Constraints */
tworzywo: +X1 +2 X2 <= 20000;
czas: +2 X1 +X2 <= 25000;
wkład: +3 X1 +3 X2 >= 15000;
```

Rozwiązywanie problemów transportowych, i im podobnych, w pakiecie *lpSolveAPI()* wymaga zapisania ich w postaci modeli decyzyjnych – czyli rozpisania macierzy kosztów jednostkowych na pojedyncze ograniczenia dla każdego dostawcy i odbiorcy.

Biblioteka *lp_solve* nie jest jedyną, do której dostęp uzyskujemy za pomocą dedykowanych pakietów. Warto zainteresować się pakietem *ROI()*, który nie wykonuje bezpośrednio optymalizacji. Zapewnia jednak ujednolicony interfejs do wielu *solverów* – darmowych i komercyjnych. Warunkiem okazuje się jedynie ich wcześniejsze zainstalowanie w środowisku obliczeniowym. Nie trzeba więc opanowywać różnorodnych składni charakterystycznych dla konkretnych bibliotek – wciąż można natomiast korzystać z ich możliwości.

Spis tabel

Tab. 1. Współrzędne wierzchołków i wartości przychodu dla przykładu 1	22
Tab. 2. Wartości ograniczeń po podstawieniu rozwiązania optymalnego	25
Tab. 3. Iteracje algorytmu <i>simplex</i> dla przykładu 2	35
Tab. 4. Alternatywne rozwiązanie dla przykładu 2	40
Tab. 5. Degeneracja rozwiązania – wariant 1	41
Tab. 6. Degeneracja rozwiązania – wariant 2	41
Tab. 7. Relacje między zmiennymi swobodnymi a dualnymi w przykładzie 2	46
Tab. 8. Tablica simpleksowa po zmianie zawartości garbnika w przykładzie 2	55
Tab. 9. Faza 1 dla przykładu 3	57
Tab. 10. Faza 1 dla przykładu 4	58
Tab. 11. Faza 2 dla przykładu 4	58
Tab. 12. Dualna metoda <i>simplex</i> dla przykładu 6	68
Tab. 13. Dualna metoda <i>simplex</i> dla przykładu 7	69
Tab. 14. Dualna metoda <i>simplex</i> dla przykładu 7 cd.	70
Tab. 15. Rozwiązanie optymalne modelu (78)–(82) dla $\lambda = 0$	73
Tab. 16. Rozwiązanie optymalne z tabeli 15 zależne od λ	73
Tab. 17. Rozwiązanie optymalne modelu (74)–(78) dla $\lambda = 20$	74
Tab. 18. Rozwiązanie optymalne z tabeli 17 zależne od λ	74
Tab. 19. Rozwiązanie optymalne modelu (78)–(82) dla $\lambda = -10/7$	75
Tab. 20. Rozwiązanie optymalne z tabeli 19 zależne od λ	76
Tab. 21. Podsumowanie analizy dla współczynników funkcji celu	76
Tab. 22. Rozwiązanie optymalne dla pierwszej iteracji	79
Tab. 23. Rozwiązanie optymalne dla drugiej iteracji	80
Tab. 24. Rozwiązanie optymalne dla trzeciej iteracji	80
Tab. 25. Rozwiązanie optymalne dla czwartej iteracji	81
Tab. 26. Podsumowanie analizy dla prawych stron ograniczeń	81
Tab. 27. Wyniki przykładu 11 dla zadania $Z_{1/-}$	89
Tab. 28. Wyniki przykładu 11 dla zadań $Z_{2/1}$ i $Z_{3/1}$	90
Tab. 29. Wyniki przykładu 11 dla zadań $Z_{4/2}$ i $Z_{5/2}$	92

Tab. 30. Rozwiązanie optymalne modelu (111)–(114)	95
Tab. 31. Dualna metoda <i>simplex</i> dla pierwszej iteracji przykładu 12	95
Tab. 32. Dualna metoda <i>simplex</i> dla drugiej iteracji przykładu 12	96
Tab. 33. Zestawienie odległości pomiędzy klientami i postojami taksówek w kilometrach w przykładzie 1	103
Tab. 34. Skorygowana macierz odległości w przykładzie 1	103
Tab. 35. Macierz kosztów utraconych korzyści i linie skreślające 0 w przykładzie 1	104
Tab. 36. Zrewidowana macierz kosztów i linie skreślające 0 w przykładzie 1	104
Tab. 37. Macierz przydziału X_1 – wariant pierwszy	105
Tab. 38. Macierz przydziału X_2 – wariant drugi	105
Tab. 39. Oczekiwana sprzedaż agentów w regionach sprzedaży	106
Tab. 40. Macierz kosztów z przykładu 2	106
Tab. 41. Skorygowana macierz kosztów z przykładu 2	107
Tab. 42. Macierz optymalnego przydziału z przykładu 2	107
Tab. 43. Oczekiwane zyski roczne z nadzoru projektów przez konsultantów w tys. PLN – przykład 3	108
Tab. 44. Utracone korzyści w tys. PLN – przykład 3a	108
Tab. 45. Macierz kosztów z przykładu 3a	109
Tab. 46. Skorygowana macierz kosztów z przykładu 3a	109
Tab. 47. Macierz optymalnego przydziału z przykładu 3a	109
Tab. 48. Utracone korzyści w tys. PLN – przykład 3b	110
Tab. 49. Macierz kosztów z przykładu 3b	110
Tab. 50. Skorygowana macierz kosztów z przykładu 3b	111
Tab. 51. Macierz optymalnego przydziału z przykładu 3b	111
Tab. 52. Porównanie wybranych definicji pojęcia „zarządzanie projektami”	120
Tab. 53. Porównanie wybranych definicji pojęcia „portfel projektów”	122
Tab. 54. Porównanie wybranych definicji pojęcia „zarządzanie portfelem projektów”	123
Tab. 55. Typologia projektów logistycznych	125
Tab. 56. Techniki określania czasu realizacji czynności projektu	134
Tab. 57. Przegląd technik planowania przebiegu projektu	134
Tab. 58. Rodzaje zapasów czasu dla czynności analizowanych w technice CPM	137
Tab. 59. Przegląd technik planowania zasobów	149
Tab. 60. Dane projektu niezbędne do wyznaczenia ścieżki krytycznej	149
Tab. 61. Wykres Gantta	151
Tab. 62. Profil zasobów w tys. PLN przy maksymalnych zasobach w każdych dwóch tygodniach wynoszących 5,5 tys. PLN	151
Tab. 63. Dane do ścieżki krytycznej po zmianach	151
Tab. 64. Wykres Gantta po zmianach	153
Tab. 65. Profil zasobów w tys. PLN przy maksymalnych zasobach w każdych dwóch tygodniach wynoszących 5,5 tys. PLN po zmianach	153
Tab. 66. Dane konieczne do sporządzenia wykresu sieciowego typu AoA oraz zapasy czasu całkowite uzyskane w wyniku zastosowania techniki CPM	158
Tab. 67. Wyniki rozwiązania modelu analizy czasowej	161

Tab. 68. Dane do analizy PERT	161
Tab. 69. Macierz obliczeń pomocniczych 1	164
Tab. 70. Macierz obliczeń pomocniczych 2	164
Tab. 71. Obliczenia zapasów czasu w projekcie	165
Tab. 72. Dane niezbędne do analizy LESS	165
Tab. 73. Koszty realizacji projektu dla czasu jego trwania równego czterdzieści dwa tygodnie	166
Tab. 74. Koszty realizacji projektu dla czasu jego trwania równego trzydzieści dziewięć tygodni	167
Tab. 75. Koszty realizacji projektu dla czasu jego trwania równego trzydzieści sześć tygodni	168
Tab. 76. Koszty realizacji projektu dla czasu jego trwania równego trzydzieści pięć tygodni	169
Tab. 77. Rozwiązanie optymalne zadania (201)–(224)	170
Tab. 78. Dane niezbędne do analizy PERT-COST	171
Tab. 79. Oczekiwane czasy graniczne, wariacje czasu i koszty wyspecyfikowanych programów realizacji projektu	175
Tab. 80. Maksymalne dopuszczalne terminy realizacji projektu zgodnie z wyspecyfikowanymi programami, w zależności od poziomu podejmowanego ryzyka	176
Tab. 81. Proces przybycia i obróbki detalu A według raportu	189
Tab. 82. Wartości oczekiwane charakterystyk liczbowych dla przykładu 1	190
Tab. 83. Rozkłady prawdopodobieństw i dystrybuanty dla przykładu 1	190
Tab. 84. Ocena wykorzystania maszyny w ciągu dnia pracy ($T = 480$)	192
Tab. 85. Wyniki obliczeń dla przykładu 2	194
Tab. 86. Wyniki obliczeń dla przykładu 3	197
Tab. 87. Wyniki obliczeń dla przykładu 4	201
Tab. 88. Wartości oczekiwane charakterystyk liczbowych dla przykładu 5	202
Tab. 89. Wartości oczekiwane charakterystyk liczbowych dla przykładu 6	204
Tab. 90. Średnia miesięczna cena produktu w kolejnych miesiącach 2019 roku	230
Tab. 91. Wartości prognoz wygaśniętych <i>ex post</i> i <i>ex ante</i> dla przeciętnej ceny w PLN	230
Tab. 92. Wartości błędów prognoz oraz innych mierników	231
Tab. 93. Wartości poziomu sprzedaży w kolejnych miesiącach roku 2019	233
Tab. 94. Podsumowanie modelu trendu liniowego wraz z prognozami <i>ex post</i> i <i>ex ante</i> oraz miernikami ocen trafności i dopuszczalności oszacowań	233
Tab. 95. Wartości składowych F_t i S_t , prognoz wykładniczego Holta wraz z prognozami <i>ex post</i> i <i>ex ante</i> oraz miernikami ocen trafności i dopuszczalności oszacowań	235
Tab. 96. Podsumowanie wyników dla metody wskaźników sezonowości addytywnej i multiplikatywnej	257
Tab. 97. Wartości surowych i czystych wskaźników sezonowości	259
Tab. 98. Podsumowanie wyników dla modelu Wintersa dla sezonowości addytywnej	260
Tab. 99. Podsumowanie wyników dla modelu Wintersa dla sezonowości multiplikatywnej	262

Spis rysunków

Rys. 1. Zbiór punktów spełniających ograniczenie (2)	21
Rys. 2. Zbiór rozwiązań dopuszczalnych dla przykładu 1	21
Rys. 3. Gradient i warstwicą dla przykładu 1	24
Rys. 4. Zmiana gradientu po zmianie współczynnika c_2	25
Rys. 5. Zmiana gradientu po zmianie współczynnika c_1	26
Rys. 6. Rozwiązanie niejednoznaczne	27
Rys. 7. Rozwiązanie nie istnieje	28
Rys. 8. Brak skończonego rozwiązania optymalnego	29
Rys. 9. Schemat postępowania w algorytmie <i>simplex</i>	32
Rys. 10. Zbiór rozwiązań modelu (98)–(101)	88
Rys. 11. Zbiory rozwiązań modeli $Z_{2/1}$ i $Z_{3/1}$	90
Rys. 12. Zbiory rozwiązań modeli $Z_{4/2}$ i $Z_{5/2}$	91
Rys. 13. Rodzaje organizacji z punktu widzenia roli projektów	121
Rys. 14. Czynność A jest bezpośrednim poprzednikiem czynności B (połączenie szere- gowe)	128
Rys. 15. Czynności A , B i C mogą być wykonywane równolegle i jednocześnie poprze- dzają czynność X ; a czynności C , D i F także mogą być wykonywane równolegle i po- przedzane są bezpośrednio przez czynność X	129
Rys. 16. Błędny zapis	129
Rys. 17. Prawidłowy zapis	129
Rys. 18. Błędny zapis	130
Rys. 19. Prawidłowy zapis	130
Rys. 20. Czynność A jest bezpośrednim poprzednikiem czynności B (połączenie szere- gowe)	130
Rys. 21. Czynności A , B i C mogą być wykonywane równolegle i jednocześnie poprze- dzają czynność X ; czynności C , D i F także mogą być wykonywane równolegle i po- przedzane są bezpośrednio przez czynność X	131
Rys. 22. Sprzężenia zwrotne w sieci typu AoN – niedopuszczalne	132
Rys. 23. Elementarna składowa sieci – czynność (i,j)	135

Rys. 24. Graficzna prezentacja oznaczeń łuków	142
Rys. 25. Rozkład β szacowanego czasu wykonania czynności	146
Rys. 26. Ścieżka krytyczna	150
Rys. 27. Ścieżka krytyczna po zmianach	152
Rys. 28. Wykres sieciowy typu AoA z przeprowadzoną analizą CPM	159
Rys. 29. Wykres sieciowy typu AoN przedstawiający analizę CPM	160
Rys. 30. Wykres sieciowy typu AoA przedstawiający analizę PERT	162
Rys. 31. Sieć typu AoN zawierająca informacje potrzebne do zastosowania techniki MPM	163
Rys. 32. Ścieżka krytyczna przy normalnym czasie trwania czynności	166
Rys. 33. Ścieżka krytyczna po skróceniu czynności E o trzy tygodnie	167
Rys. 34. Ścieżka krytyczna po skróceniu czynności K o trzy tygodnie	168
Rys. 35. Ścieżka krytyczna po skróceniu czynności I o tydzień	169
Rys. 36. Ścieżka krytyczna przy normalnym oczekiwanym czasie trwania czynności	172
Rys. 37. Ścieżka krytyczna po skróceniu czynności E	172
Rys. 38. Ścieżka krytyczna po skróceniu czynności K	173
Rys. 39. Ścieżka krytyczna po skróceniu czynności B	173
Rys. 40. Ścieżka krytyczna po skróceniu czynności I	174
Rys. 41. Ścieżka krytyczna po skróceniu czynności H	174
Rys. 42. Ścieżka krytyczna po skróceniu czynności H i F	175
Rys. 43. Schemat procesu masowej obsługi	181
Rys. 44. Relacja popytu (D), podaży (S), ceny (C_A) i ilości towaru (Q_A)	211
Rys. 45. Przesunięcie popytu (D_1 na D_2) w odniesieniu do zmiany ilości (od Q_A do Q_B)	211
Rys. 46. Wartość produkcji przedsiębiorstwa w tys. PLN w latach 1995–2015	224
Rys. 47. Produkt Krajowy Brutto w cenach bieżących w kwartałach w latach 2002–2014	224
Rys. 48. Wizualizacja przykładowych szeregów czasowych. Składowe systematyczne szeregi: a) stały przeciętny poziom zjawiska i wahania przypadkowe, b) tendencja rozwojowa i wahania przypadkowe, c) stały przeciętny poziom zjawiska, wahania sezonowe i przypadkowe	226
Rys. 49. Porównanie wyników prognoz modeli Holta i trendu z wartościami rzeczywistymi	236
Rys. 50. Prognozy modelu Holta i prognoza przedziałowa oszacowana z 95% prawdopodobieństwem	237
Rys. 51. Wartość sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa w tys. PLN w latach 2008–2019 (dane kwartalne)	238
Rys. 52. Wartość sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa w tys. PLN w latach 2008–2020 (dane kwartalne) wraz z prognozami <i>ex post</i> i <i>ex ante</i>	239
Rys. 53. Wartość sprzedanych towarów i materiałów w sekcji H: Transport i gospodarka magazynowa w tys. PLN w latach 2008–2020 (dane kwartalne) wraz z prognozami <i>ex post</i> i <i>ex ante</i> dla metody wskaźników sezonowości	259
Rys. 54. Efekt działania kodu tworzącego zbiór rozwiązań dopuszczalnych	273