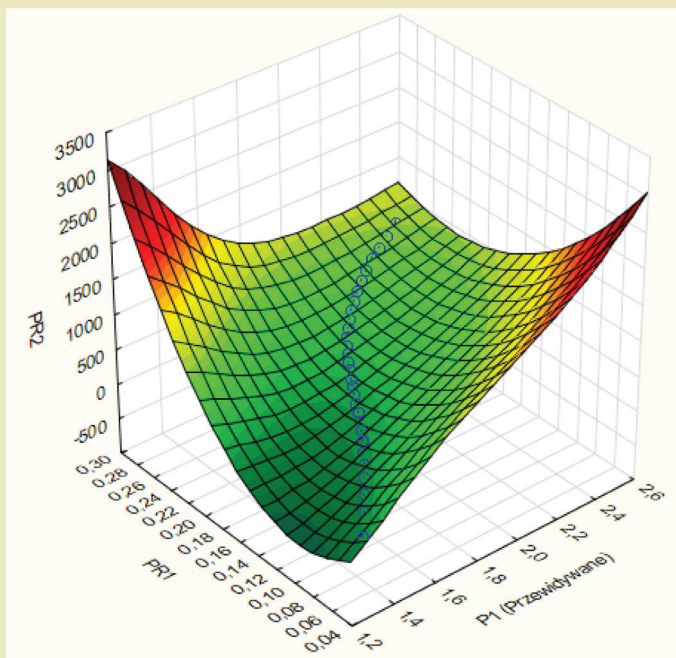




*Magdalena Rogalska*

# Wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych



Lublin 2016

# Wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych

# Monografie – Politechnika Lubelska



Politechnika Lubelska  
Wydział Budownictwa i Architektury  
ul. Nadbystrzycka 40  
20-618 Lublin

Magdalena Rogalska

# Wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych



Politechnika Lubelska  
Lublin 2016



Recenzenci:

dr hab. Mieczysław Wodecki, prof. nadzw.

dr hab. inż. Ryszard Dachowski, prof. nadzw.

Publikacja wydana za zgodą Rektora Politechniki Lubelskiej

© Copyright by Politechnika Lubelska 2016

ISBN: 978-83-7947-186-7

Wydawca: Politechnika Lubelska

ul. Nadbystrzycka 38D, 20-618 Lublin

Realizacja: Biblioteka Politechniki Lubelskiej

Ośrodek ds. Wydawnictw i Biblioteki Cyfrowej

ul. Nadbystrzycka 36A, 20-618 Lublin

tel. (81) 538-46-59, email: wydawca@pollub.pl

[www.biblioteka.pollub.pl](http://www.biblioteka.pollub.pl)

Druk: TOP Agencja Reklamowa Agnieszka Łuczak

[www.agencjatop.pl](http://www.agencjatop.pl)

---

Elektroniczna wersja książki dostępna w Bibliotece Cyfrowej PL [www.bc.pollub.pl](http://www.bc.pollub.pl)

Nakład: 100 egz.

## SPIS TREŚCI

<b>1. Wstęp</b> .....	<b>7</b>
<b>2. Opis metody</b> .....	<b>11</b>
<b>3. Wstęp do prognozowania</b> .....	<b>15</b>
3.1. Definiowanie problemu i zebranie danych empirycznych .....	15
3.2. Zmienne w arkuszu kalkulacyjnym programu STATISTICA .....	16
3.3. Analiza zmiennych .....	23
3.4. Korelacje liniowe zmiennych.....	29
3.5. Funkcja autokorelacji reszt i autokorelacji częściowej reszt.....	31
3.6. Ocena dobroci dopasowania modeli i prawidłowości predykcji, obliczenie błędów .....	33
3.7. Oznaczenia .....	37
3.7.1. Oznaczenia metod i modeli.....	37
3.7.2. Oznaczenia osi.....	38
<b>4. Prognozowanie</b> .....	<b>39</b>
4.1. Prognozowanie metodą regresji wielorakiej .....	39
4.2. Prognozowanie metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych .....	50
4.3. Prognozowanie metodą uogólnionych modeli addytywnych .....	57
4.4. Prognozowanie metodą sztucznych sieci neuronowych.....	62
4.6. Prognozowanie metodą autoregresji zintegrowanej.....	79
4.6.1. Proces autoregresji.....	80
4.6.2. Proces średniej ruchomej MA (moving average).....	84
4.6.3. Proces autoregresji zintegrowanej ARIMA .....	85
4.6.4. Przykład obliczeniowy.....	85
4.7. Analiza prawidłowości i wybór modelu prognostycznego .....	101
<b>5. Zastosowanie wieloczynnikowych modeli w prognozowaniu czasu     procesów budowlanych</b> .....	<b>103</b>
5.1. Roboty ziemne z wykorzystaniem koparek .....	103
<b>6. Podsumowanie</b> .....	<b>141</b>
<b>Bibliografia</b> .....	<b>143</b>



# 1. Wstęp

Harmonogramowanie prac jest bardzo istotną częścią realizacji przedsięwzięć budowlanych. Od prawidłowej organizacji przebiegu procesów budowlanych w dużej mierze zależy powodzenie projektu: ukończenie zadania w planowanym terminie i kosztach oraz uzyskanie zysku. Istnieje wiele metod wykonywania harmonogramów, naukowcy wciąż pracują nad udoskonalaniem istniejących rozwiązań, powstają również nowe metody.

Można wyróżnić pięć zasadniczych kierunków rozwoju harmonogramowania: pierwszy obejmujący wprowadzanie danych o charakterze rozmytym, drugi z uwzględnieniem buforów czasu, trzeci wprowadzający czynniki ryzyka, czwarty dotyczący szeregowania zadań z zastosowaniem narzędzi sztucznej inteligencji i piąty obejmujący potokowe szeregowanie zadań.

Wprowadzaniem do harmonogramowania danych o charakterze rozmytym zajmowali się Chanas, Kamburowski (1981), Buckley (1989), Chen, Hwang (1992), Jang (1992), Chang, Tsujimura, Gen, Tozawa (1993, 1995), Shipley, Korvin, Omer (1997), Chanas, Zieliński (2001), Chen (2007), Kulejewski (2010) i inni.

Analizą czasu trwania procesów budowlanych i przebiegiem ścieżki krytycznej po wprowadzeniu buforów czasu zajmowali się: Goldratt (1992,1997), Hamerlink (1995, 1998), Radovilsky (1998), Harris, Ionnu (1998), Herroelen (2001), Steyn (2002), Yang (2003), Scavino (2003), Rogalska, Hejducki (2007) i inni.

Powstało wiele prac dotyczących aplikacji czynników ryzyka do harmonogramów. Swoje osiągnięcia prezentowali Soroush (1993), Gong i Rowings (1995), Gong (1997), Dawson i Dawson (1998), Premachandra (2001), Rene van Drop (2005), Kister i Hawkins (2006), Skorupka (2007), Azaron i zespół (2008), Huang i Wang (2009), Li i Wang (2009), Yunning i Xixi (2010).

Problemem szeregowania zadań w harmonogramowaniu z zastosowaniem narzędzi sztucznej inteligencji zajmowali się: Lin, Goodman i Punch (1997), Yamada (1998), Reeves i Yamada (1998), Leu, Yang, Huang (2000), Chang, Christensen, Zhang (2001), Toklu (2002), Ugwu i Tah (2002), Borne (2004), Rogalska, Bożejko, Hejducki (2005, 2008), Nassar (2005), Li i Chen (2008), Prakash, Felix, Chan, Deshmukh (2011).

Piąty kierunek rozwoju to metoda Flow Shop (potokowego szeregowania zadań). Metoda zapoczątkowana została przez Afanasjewa (1957) i kontynuowana przez wielu badaczy. w ciągu ostatnich 25 lat między innymi przez: Arditi, Abulak (1986), Russell, Wong (1993), Adeli, Karim (1997), Yamada (1998), Harris, Ioannou (1998), Afanasjew, Afanasjew (2000), Hejducki (2000, 2003, 2004), El-Rayes, Moselhi (2001), Marcinkowski (2011), Rogalska, Hejducki (2004), Bożejko, Wodecki (2005), Podolski (2008).

Cechą wspólną wszystkich kierunków jest wykorzystywanie wartości nakładów pracy, jako elementu niezbędnego do utworzenia harmonogramu. w budownictwie wielkość nakładów pracy (Kowalczyk, Zabielski (2010)) podaje się w postaci łącznego nakładu czasu pracy wszystkich robotników wykonujących określone zadanie. Usta-

lenie normy nakładów jest trudne, gdyż nakłady na taką samą jednostkę produkcji nie zawsze są jednakowe. Zależą od wielu czynników, takich jak: metoda wykonania i rodzaj konstrukcji, rodzaj i gatunek użytych materiałów, kwalifikacje oraz zdolności robotników, warunki pracy, technika i technologia wykonania, wymagania jakościowe, ilość robót do wykonania, wielkość zespołu roboczego. Normy nakładów określa się jako czas niezbędny wykonania zadania przez przeciętnego wykonawcę w przeciętnych warunkach. Stosowanymi metodami ustalania norm pracy są:

- metody sumaryczne (szacunkowa, statystyczna, porównawcza),
- metody analityczne (analityczno-badawcza uwzględniająca chronometraż, obserwację wykorzystania dnia pracy, obserwację migawkową oraz analityczno-obliczeniowa oparta na normach scalonych i normatywach elementarnych),
- metoda parametryczna (Hoła, Mrozowicz (2003)).

Uzyskiwane w ten sposób normy nakładów są wartościami uśrednionymi i mogą w sposób bardzo znaczny różnić się od czasu realizacji pracy w określonych warunkach.

W monografii przedstawiono metodę prognozowania czasów procesów budowlanych z uwzględnieniem czynników mających wpływ na ich realizację. **Celem pracy jest wykazanie, że czas wykonania danego procesu budowlanego zależy od wielu czynników: technicznych, technologicznych, organizacyjnych i powinien być prognozowany w określonych warunkach realizacyjnych.**

Znajomość czasu przeznaczanego na wykonanie zaplanowanych robót jest niezbędna do projektowania organizacji i technologii robót, umożliwia kontrolę i sterowanie przebiegiem prac oraz czuwanie nad jego efektywnością. Nieterminowość inwestycji budowlanych jest powszechnym zjawiskiem w Polsce i na świecie. Czarnigowska, Sobotka (2010) wykazują, że 58% robót nie jest wykonywanych w zaplanowanym terminie (28% przed terminem, 30% po terminie). Istnieje bardzo wiele czynników powodujących nieterminowe ukończenie inwestycji. Jednym z nich jest przyjmowanie uśrednionych wartości nakładów pracy.

Podjęto próbę opracowania metodyki umożliwiającej wyznaczanie rzeczywistych czasów procesów budowlanych, uwzględniających warunki wykonania w odniesieniu do analizowanej budowy. Przedstawiona metodologia prognozowania czasu realizacji procesów budowlanych umożliwia sporządzenie harmonogramów uwzględniających warunki rzeczywiste a nieuśrednione. Jest to nowe podejście do wyznaczania czasów procesów budowlanych. Opracowana metoda bazuje na analizie obliczeniowej metod prognostycznych: regresji wielorakiej, wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych, uogólnionych modeli addytywnych, sztucznych sieci neuronowych, wektorów nośnych i autoregresji zintegrowanej.

W wyniku obliczeń różnymi metodami otrzymuje się wiele modeli prognostycznych, z których wybiera się do stosowania model charakteryzujący się najmniejszym błędem prognostycznym. Istnieje możliwość wprowadzania do obliczeń bardzo dużej liczby czynników mających wpływ na czas realizacji procesu w określonych warunkach. Mogą to być wartości liczbowe lub lingwistyczne. Zaproponowana metoda jest uni-

wersalna, istnieje możliwość zastosowania jej w odniesieniu do wszystkich procesów budowlanych: robót liniowych, kubaturowych, procesów prostych i złożonych. Zakres pracy obejmuje definiowanie problemu prognostycznego, zasady przygotowywania danych do obliczeń, metodologię obliczeniową, obliczenia prognostyczne z analizą i oceną porównawczą uzyskanych modeli z wykorzystaniem średniego absolutnego błędu procentowego oraz autokorelacji i autokorelacji cząstkowej szeregu resztowego.

Praca zawiera nowe ujęcie problemu wyznaczania czasu procesów budowlanych w warunkach rzeczywistych. Na podstawie badań stwierdzono, że czasy wykonania procesu budowlanego w szczegółowych uwarunkowaniach rzeczywistych różnią się w sposób znaczny od czasów uśrednionych zawartych w bazie normatywnej. Proponowana metoda umożliwia obliczenie czasu trwania procesów z uwzględnieniem czynników wpływających na czas realizacji.

Celem szeroko zakrojonych dalszych prac jest opracowanie wzorów regresyjnych wyznaczających czasy procesów budowlanych z uwzględnieniem czynników wpływu i alternatywne zastąpienie nimi bazy normatywnej w odniesieniu do pojedynczych przedsiębiorstw.



## 2. Opis metody

Istotą harmonizacji produkcji budowlanej jest ustalenie takiego scenariusza wykonania zadań produkcyjnych, aby zapewnić jednostce produkcyjnej uzyskiwanie (w czasie) przychodów z pracy własnego potencjału produkcyjnego (Marcinkowski (2007), Adamiecki (1985)). o sprawności wykonania zadań budowlanych decydują w zdecydowanej większości sytuacji planistycznych niezbędne nakłady pracy zasobów czynnych na wykonanie zadań i liczba przydzielonych zasobów, które te nakłady mają ponieść (Marcinkowski (2013)).

Planowanie produkcji budowlanej odnosi się do wielu aspektów działania:

- Projektowania systemów kompleksowej mechanizacji (Jaworski (1980, 1999), Kapiński (2008)).
- Ustalania optymalnych rozwiązań organizacyjnych współpracujących ze sobą środków pracy (Marcinkowski, Koper (2011)).
- Harmonizowania wykonywania zadań z bilansowaniem potrzeb i dostępności zasobów (Marcinkowski (2002)).
- Oceny kosztu wykonania zadań (Kowalczyk, Zabielski (2005)).
- Analizy ryzyka (Skorupka (2004, 2006, 2007, 2008)).
- Szeregowania zadań (Bożejko i in. (2011), Hejducki (2011), Rogalska (2011)).
- Logistyki w przedsiębiorstwie (Sobotka (2009,2012), Jaśkowski (2014)).

Podstawą wymienionych działań jest ustalenie norm pracy lub też norm technicznie uzasadnionych. Norma technicznie uzasadniona to forma pracy ustalona w takich warunkach, które pod względem organizacji czynności, zabiegów uznane są za najlepsze (w pełni wykorzystane są środki techniczne, zastosowane metody i sposoby wykonywania pracy są najbardziej racjonalne, uwzględnione są fizjologiczne uwarunkowania człowieka w procesie pracy). Wyznaczanie norm technicznie uzasadnionych ma na celu ustalenie podstaw racjonalnego planowania.

Bazę tę stanowią opracowania ogólne, dotyczące robót budowlanych w tym: zestawienia nakładów rzeczowych, robocizny, materiałów i sprzętu lub nakładów scalonych cenniki i zestawienia stawek, informacje o poziomie wskaźników i narzutów. Istnieje bardzo wiele grup katalogów :

- KNR: Katalogi Nakładów Rzeczowych,
- KNR-W: Katalogi Nakładów Rzeczowych wyd. WACETOB,
- NNRNKB: Normy Nakładów Rzeczowych Na Konstrukcje Budowlane – uzupełnienia do KNR,
- KNR 9-XX: Katalogi Nakładów Rzeczowych wyd. ORGBUD-SERWIS,
- KNR AT: Katalogi Nakładów Rzeczowych wyd. ATHENASOFT,
- KNR BISTYP: Katalogi Nakładów Rzeczowych wyd. BISTYP-CONSULTING,
- KNR SEK: Katalogi Nakładów Rzeczowych wyd. SEKOCENBUD,
- KNR K: Katalogi Nakładów Rzeczowych wyd. KOPRIN,
- KSNR: Katalogi Scalonych Nakładów Rzeczowych,



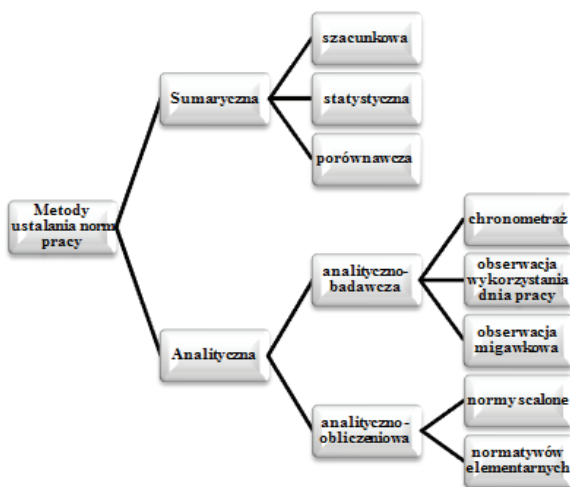
- KNNR: Kosztorysowe Normy Nakładów Rzeczowych,
- KNNR-W: Kosztorysowe Normy Nakładów Rzeczowych wyd. WACETOB,
- KNP: Katalogi Norm Pracy,
- KNKRB: Katalogi Norm Kosztorysowych Robót Budowlanych,
- KJNZMB: Katalogi Jednostkowych Norm Zużycia Materiałów Budowlanych,
- NZCPB: Normy Zużycia Czynników Produkcji Budowlanej,
- KPRR: Katalogi Pracochłonności Robót Remontowych,
- TZKNBK: Tymczasowe Zakładowe Katalogi Norm Budowlano-Konserwatorskich,
- KNCK: Katalogi Norm i Cen Kosztorysowych,
- ZNR: Zestaw Nakładów Rzeczowych.

Każdy z katalogów może być podstawą do sporządzania kosztorysu i harmonogramu robót budowlanych.

Normowanie czasów pracy i prawidłowe zarządzanie czasami planowanymi w systemie produkcyjnym daje następujące korzyści:

- lepsze planowanie i poziomowanie produkcji,
- poprawę efektywnego wykorzystania zasobów,
- planowanie zatrudnienia,
- szczegółową kalkulację kosztów wytwarzania,
- redukcję kosztów wytwarzania,
- skracanie czasów realizacji zleceń,
- możliwość pomiaru i monitoringu efektywności czasu pracy,
- definicję i pomiar elementarnych czynności dodających wartość.

Normowanie czasów pracy może być przeprowadzane w różny sposób. Metody te można podzielić wg Kowalczyk, Zabielski (2012) na metody sumaryczne i analityczne rysunek 2.1.



Rys. 2.1. Metody ustalania norm pracy

Metoda sumaryczna polega na określeniu łącznego czasu niezbędnego do wykonania operacji roboczej, bez podziału na elementy składowe. Metoda analityczna polega na określeniu czasu niezbędnego na wykonanie operacji roboczej z uwzględnieniem etapów: podziału operacji na elementy składowe, analizy możliwości produkcyjnych maszyn i urządzeń, organizacji robót, doświadczalnym sprawdzeniu poprawności zaprojektowanej technologii i organizacji robót i opracowanie normy technicznie uzasadnionej.

Sumaryczne metody ustalania norm pracy można podzielić na trzy rodzaje:

- szacunkowa – polega na ocenie czasu niezbędnego do wykonania danej operacji roboczej przez osobę ustalającą normę,
- statystyczna – polega na ustaleniu normy czasu na podstawie danych statystycznych o zużyciu czasu podczas wykonywania takiej samej pracy w przeszłości. Metoda nie uwzględnia możliwości różnicowania czasów ze względu na istniejące w przeszłości uwarunkowania,
- porównawcza – polega na określeniu czasu niezbędnego na wykonanie operacji roboczej na podstawie czasu trwania zbliżonej operacji, ustalonego także metodą sumaryczną.

Wszystkie metody sumaryczne są proste, szybkie i tanie, nie dają jednak pewności prawidłowego szacowania norm pracy.

Metody analityczne, charakteryzujące się wyższym stopniem dokładności, można je podzielić na (Kowalczyk, Zabielski (2012)):

- analityczno-badawcza, w której czas wykonania wszystkich elementów operacji roboczej określa się bezpośrednio na stanowisku roboczym. Pomiarów dokonuje się, uwzględniając zaprojektowaną metodę wykonania pracy i przyjęte warunki techniczno-organizacyjne.
- analityczno-obliczeniowej, w jakiej ustalenie zużycia czasu polega na zsumowaniu znanych czasów wykonania czynności składowych.

W 2003 roku Hoła i Mrozowicz zaproponowali zastosowanie metody parametrycznej. Normę czasu, w metodzie tej, określa się na podstawie zależności funkcyjnej między pracochłonnością zadania, a czynnikami odniesienia wyrażającymi wielkość zadania. Modele norm czasu mogą być opisane za pomocą wzorów algebraicznych lub tablic.

Proponuje się wprowadzenie nowej metody ustalania norm pracy. Metodę nazwano parametryczno-regresyjną (rys.2.2).

Proponowana metoda bazuje na obliczeniach statystycznych. Prognozuje się czas procesu budowlanego bazując na pomiarach i danych zebranych w warunkach rzeczywistych. Na tej podstawie wyznacza się równanie regresji, dzięki któremu można obliczyć czas trwania procesu realizowanego w nowych warunkach, uwzględniając w obliczeniach czynniki wpływu występujące w równaniu regresji. Na przykład mierzony jest czas malowania agregatem 100 m<sup>2</sup> powierzchni. Jako czynniki wpływu wyznacza się: rodzaj powierzchni, liczbę pracowników w brygadzie, liczbę godzin pracy,

wydajność agregatu i inne. Przeprowadzając obliczenia prognostyczne otrzymuje się równanie regresji w postaci:

czas malowania agregatem 100 m<sup>2</sup> powierzchni =  $a_0 + a_1(\text{rodzaj powierzchni}) + a_2(\text{liczbę pracowników w brygadzie}) + a_3(\text{liczbę godzin pracy}) + a_4(\text{wydajność agregatu}) + \dots + \dots$

gdzie  $a_1$  może być wartością liczbową lub uzyskaną w toku obliczeń funkcją. Zależność regresyjna może też być przedstawiona w postaci sieci neuronowej.



Rys. 2.2. Metody ustalania norm pracy z uwzględnieniem metody parametrycznej

Znając lub przewidując warunki panujące na nowej budowie można obliczyć czas procesu i zastosować go w harmonogramie budowy.

Metoda składa się z kolejnych etapów obejmujących:

- identyfikację problemu (wybór procesu budowlanego),
- sformułowanie problemu (zdefiniowanie, zebranie i analiza danych),
- dobór metod obliczeniowych z analizą możliwości zastosowania,
- obliczenie wielu modeli regresyjnych z zastosowaniem wybranych metod,
- weryfikację wyników (wybór najlepszego modelu regresyjnego na podstawie analizy błędów).

Metoda jest pracochłonna, wymaga stosowania programów komputerowych do obliczeń statystycznych np. STATISTICA firmy StatSoft. Późniejsze stosowanie równania regresji do obliczania czasu procesu z uwzględnieniem czynników wpływu jest niezwykle proste. Jest to nowa metoda dotychczas nie stosowana.

## 3. Wstęp do prognozowania

### 3.1. Definiowanie problemu i zebranie danych empirycznych

Etap definiowania problemu prognostycznego wymaga wiedzy specjalistycznej. Na tym etapie należy określić zmienną zależną, czyli wartość prognozowaną, oraz zmienne niezależne - potencjalne czynniki wpływu. Zmienne niezależne, które występują w równaniu regresyjnym nazywane są predyktorami. Nie każdy potencjalny czynnik wpływu, czyli zmienna niezależna, zostanie wykorzystany w równaniu regresyjnym jako predyktor. Zależy to od metody obliczeniowej oraz wybranego modelu obliczeniowego. w niektórych metodach np. regresji wielorakiej eliminowane są automatycznie z równania regresyjnego zmienne niezależne wzajemnie skorelowane liniowo. Wybór zmiennych niezależnych musi być racjonalny, poparty wiedzą i doświadczeniami z przeszłości. Przykładowo rozpatrując czynniki wpływu na czas betonowania, można brać pod uwagę rodzaj zastosowanego sprzętu, płynność dostaw, warunki pogodowe, doświadczenie i wydajność brygady roboczej oraz inne. Jednak istnieje wiele czynników, które z racjonalnego punktu widzenia nie są istotne w odniesieniu do analizowanego czasu betonowania np. wydajność maszyn stosowanych w procesach poprzedzających. Może zdarzyć się tak, że znajdziemy zależności regresyjne pomiędzy tymi zmiennymi i wydajność maszyn do robót ziemnych okaże się znakomitym predyktorem czasu procesu betonowania. Będzie to jednak jedynie zależność statystyczna, niemająca odzwierciedlenia w warunkach rzeczywistych. Dobór zmiennych niezależnych jest kluczowym etapem, szczególnie w przypadkach nowych, niewykonywanych przedtem prognoz. Może również wystąpić zjawisko odwrotne. Ogólnie przyjęty i akceptowany predyktor danego zjawiska nie będzie modelował w sposób zadawalający prognozy. Błędy *ex ante* lub *ex post* będą zbyt duże. Jeśli zależność jest ewidentna, wówczas należy dążyć do takiego przekształcania danych, wyboru metod i modeli prognostycznych, aby uzyskać zadawalający efekt obliczeniowy.

Etap zebrania danych i ich analiza to najbardziej pracochłonny etap prognozowania, mający na celu określenie predyktorów. Zbieranie danych zwykle zajmuje około 80% czasu potrzebnego do tworzenia prognozy. Znacznym ułatwieniem jest wykorzystywanie danych gromadzonych przez przedsiębiorstwa o procesach przez nich realizowanych. Tradycyjnym źródłem informacyjnym jest baza danych Głównego Urzędu Statystycznego. Pozyskanie danych opracowanych zgodnie z potrzebami prognostycznymi jest niezwykle kosztowne, można jednak nieodpłatnie korzystać z danych ogólnie dostępnych. Oczywiście podstawową metodą tworzenia bazy danych jest metoda doświadczalna pomiaru czasu trwania procesów bezpośrednio na budowie, przy określonych uwarunkowaniach wewnętrznych i zewnętrznych. Coraz częściej w warunkach polskich prowadzi się bardzo dokładne pomiary dotyczące czasu realizacji poszczególnych procesów. Ma to na celu określenie czasów realizacji i tworzenie harmonogramów

prac w sposób jak najbardziej zbliżony do rzeczywistego, uwzględniający warunki realne. Daje to możliwość wprowadzania doraźnych działań naprawczych, korzystanie z uprzednio założonych buforów czasu, określanie konieczności wprowadzania dodatkowych zasobów, czy też obliczania potencjalnych strat lub zysków.

### 3.2. Zmienne w arkuszu kalkulacyjnym programu STATISTICA

Arkusz kalkulacyjny w programie STATISTICA to tabela, w której zmiennym odpowiadają kolumny, natomiast wiersze to kolejne przypadki, czyli pomierzone, nadane lub obliczone wartości zmiennych. Na rysunku 3.1 widoczny jest zrzut ekranu arkusza kalkulacyjnego programu STATISTICA. Dane zawarte w arkuszu posłużą jako przykład obliczeniowy do prezentacji różnych metod obliczeniowych. Są to dane wygenerowane w celu objaśnienia toku obliczeń predykcyjnych (przykład 3.1).

	1 P1	2 PR1	3 PR2	4 Zmn4	5 Zmn5	6 Zmn6	7 Zmn7	8 Zmn8	9 Zmn9
1	1,33	0,05	10						
2	1,36	0,070711	13						
3	1,42	0,086603	18						
4	1,48	0,1	25						
5	1,53	0,111803	34						
6	1,59	0,122474	45						
7	1,63	0,132288	58						
8	1,68	0,141421	73						
9	1,74	0,15	90						
10	1,76	0,158114	109						
11	1,79	0,165831	130						
12	1,83	0,173205	153						
13	1,86	0,180278	178						
14	1,89	0,187083	205						
15	1,92	0,193649	234						
16	1,95	0,2	265						
17	1,97	0,206155	298						
18	2,00	0,212132	333						
19	2,12	0,217945	370						
20	2,05	0,223607	409						
21	2,07	0,229129	450						
22	2,09	0,234521	493						
23	2,12	0,239792	538						
24	2,14	0,244949	585						
25	2,16	0,25	634						
26	2,28	0,254951	685						
27	2,20	0,259808	738						
28	2,21	0,264575	793						
29	2,28	0,269258	850						
30	2,25	0,273861	909						

Rys.3.1. Przykład 3.1.: Zrzut ekranu arkusza kalkulacyjnego programu STATISTICA

W przykładzie 3.1 zmienną zależną jest zmienna 1 ( $v_1$ ) oznaczona P1 (proces 1), a zmiennymi niezależnymi, czyli potencjalnymi predyktorami równań regresyjnych są zmienne 2 i 3 ( $v_2, v_3$ ) oznaczone jako PR1 i PR2 (predyktor 1, predyktor 2). Liczba przypadków jest dziesięciokrotnie większa niż liczba zmiennych. Jest to standardowa proporcja umożliwiająca korzystanie z wielu metod obliczeniowych.

Dane surowe empiryczne wymagają najczęściej opracowania. w przypadku tworzenia szeregu czasowego jednym z predyktorów i zarazem zmienną niezależną jest okres  $t$ . Okresem może być rok, kwartał, miesiąc, tydzień, dzień, godzina, minuta, a nawet sekunda lub część sekundy (okresy poniżej minuty są mało użyteczne w budownictwie). Tworząc bazę danych musimy posiadać wszystkie dane odnoszące się do poszczególnych okresów, tak by istniała możliwość utworzenia szeregu. Dane można agregować, np. mając dane dzienne możemy stworzyć dane miesięczne. Problemem obliczeniowym w wielu programach (np. STATISTICA) jest występowanie wartości zerowych. Są wówczas 2 możliwości: wyeliminować daną lub wprowadzić zamiast zera wartość relatywnie bardzo małą, na przykład jeśli średni czas oczekiwania na rozładunek betonowozu na budowie wynosi 10 minut, a w niektórych przypadkach rozpoczęto rozładunek natychmiast (czyli czas oczekiwania na rozładunek wynosił 0 minut), to można czasowi oczekiwania nadać wartość  $0,0001 \cdot 10 \text{ minut} = 0,001 \text{ minuty}$ , co nie zmienia sensu fizycznego, natomiast umożliwia przeprowadzenie obliczeń.

Przygotowanie zmiennych niezależnych do obliczeń może polegać również na ich przekształceniu. Tworzymy wówczas nowe zmienne, które są powiązane z podstawową zmienną ( $v_i$ ) funkcyjnie np.  $(v_i)^2$ ,  $(v_i)^3$ ,  $(1/v_i)$ ,  $(v_i^{0,5})$ ,  $\ln(v_i)$ ,  $\log(v_i)$ . Szczególnie korzystne jest logarytmowanie zmiennej, w przypadku gdy wartości jednej zmiennej są o kilka rzędów większe niż drugiej. W ten sposób prowadzimy obliczenia dla nowych zmiennych uzyskując często wielokrotnie lepszą prognozę. Wadą metody z wykorzystaniem zmiennych przekształconych jest zwiększona trudność interpretacji wyników. Jednak pozostawiając wyjściowe zmienne, możemy niejednokrotnie w ogóle nie uzyskać prognozy nadającej się do zastosowania, zatem i tak nie mielibyśmy czego interpretować.

Inną metodą przygotowywania danych (w odniesieniu do szeregów czasowych) jest ich opóźnianie, polegające na wpisaniu w miejsce przypadku 2 wartości przypadku 1, w miejsce przypadku 3 wartości przypadku 2 itd. Przypadek 1 pozostaje wolny – bez wartości. Opóźnimy określoną zmienną ( $v_i$ ) o jeden okres i tworzymy z wartości opóźnionych nową zmienną oznaczaną jako  $\text{lag}(v_i;1)$ , następnie opóźnimy pierwotną zmienną o dwa okresy i tworzymy z wartości opóźnionych nową zmienną  $\text{lag}(v_i;2)$ . Zdarza się, że w niektórych obliczeniach potrzebne są opóźnienia nawet o 36 okresów  $\text{lag}(v_i;36)$ . Opóźnione zmienne traktujemy jak zmienne niezależne. Na rysunku 3.2 przedstawiono przykładową zmienną  $v_{10}$  (skumulowany czas naprawy) wraz ze zmiennymi opóźnionymi od  $\text{lag}(v_{10};1)$  do  $\text{lag}(v_{10};12)$ .



10	11	12	13	14	15	16	17	18	19	20
skumulow wany czas naprawy	lag(v10;1)	lag(v10;2)	lag(v10;3)	lag(v10;4)	lag(v10;5)	lag(v10;6)	lag(v10;7)	lag(v10;10)	lag(v10;11)	lag(v10;12)
1	113,37									
2	167,303	113,369808								
3	315,636	167,303141	113,36981							
4	410,036	315,636475	167,30314	113,369808						
5	463,97	410,036475	315,63647	167,303141	113,36981					
6	463,97	463,969808	410,03647	315,636475	167,30314	113,36981				
7	463,97	463,969808	463,96981	410,036475	315,63647	167,30314	113,36981			
8	463,97	463,969808	463,96981	463,969808	410,03647	315,63647	167,30314	113,369808		
9	475,97	463,969808	463,96981	463,969808	463,96981	410,03647	167,30314	113,369808		
10	480,436	475,969808	463,96981	463,969808	463,96981	463,96981	410,03647	167,303141	113,369808	
11	495,436	480,436475	475,96981	463,969808	463,96981	463,96981	410,03647	315,636475	167,303141	113,369808
12	535,903	495,436475	480,43647	475,969808	463,96981	463,96981	463,96981	410,036475	315,636475	167,303141
13	583,903	535,903141	495,43647	480,436475	475,96981	463,96981	463,96981	463,969808	410,036475	315,636475
14	639,37	583,903141	535,90314	495,436475	480,43647	475,96981	463,96981	463,969808	463,969808	410,036475
15	681,37	639,369808	583,90314	535,903141	495,43647	480,43647	463,96981	463,969808	463,969808	463,969808
16	705,37	681,369808	639,36981	583,903141	535,90314	495,43647	475,96981	463,969808	463,969808	463,969808
17	705,37	705,369808	681,36981	639,369808	583,90314	535,90314	480,43647	475,969808	463,969808	463,969808
18	751,836	705,369808	681,36981	681,369808	639,36981	583,90314	495,43647	480,436475	475,969808	463,969808
19	775,836	751,836475	705,36981	705,369808	681,36981	583,90314	535,90314	495,436475	480,436475	475,969808
20	793,836	775,836475	751,83647	705,369808	705,36981	681,36981	583,90314	535,903141	495,436475	480,436475
21	816,303	793,836475	775,83647	751,836475	705,36981	681,36981	639,36981	583,903141	535,903141	495,436475
22	816,303	816,303141	793,83647	775,836475	751,83647	705,36981	681,36981	639,369808	583,903141	535,903141
23	862,77	816,303141	816,30314	793,836475	775,83647	751,83647	705,36981	681,369808	639,369808	583,903141
24	874,77	862,769808	816,30314	816,303141	793,83647	775,83647	705,36981	705,369808	681,369808	639,369808

Rys. 3.2. Zrzut ekranu programu STATISTICA. Arkusz ze zmiennymi opóźnionymi.

1 data	2 t	3 potencjal na temperat ura	4 Lag(v3:12)	5 styczeń	6 luty	7 marzec	8 kwiecień	9 maj	10 czerwiec	11 lipiec	12 sierpień	13 wrzesień	14 październik	15 listopad	16 grudzień
1	1.2000	1	271,497	1	0	0	0	0	0	0	0	0	0	0	0
2	2.2000	2	275,393	0	1	0	0	0	0	0	0	0	0	0	0
3	3.2000	3	276,62	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
4	4.2000	4	284,25	0	0	0	1	0	0	0	0	0	0	0	0
5	5.2000	5	283,96	0	0	0	0	1	0	0	0	0	0	0	0
6	6.2000	6	286,906	0	0	0	0	0	1	0	0	0	0	0	0
7	7.2000	7	287,518	0	0	0	0	0	0	1	0	0	0	0	0
8	8.2000	8	287,58	0	0	0	0	0	0	0	1	0	0	0	0
9	9.2000	9	282,939	0	0	0	0	0	0	0	0	1	0	0	0
10	10.2000	10	282,924	0	0	0	0	0	0	0	0	0	1	0	0
11	11.2000	11	277,942	0	0	0	0	0	0	0	0	0	0	1	0
12	12.2000	12	274,33	0	0	0	0	0	0	0	0	0	0	0	1
13	1.2001	13	271,583	1	0	0	0	0	0	0	0	0	0	0	0
14	2.2001	14	272,51	275,393	0	1	0	0	0	0	0	0	0	0	0
15	3.2001	15	274,747	276,62	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
16	4.2001	16	278,45	284,25	0	0	0	1	0	0	0	0	0	0	0
17	5.2001	17	283,945	283,96	0	0	0	0	1	0	0	0	0	0	0
18	6.2001	18	285,094	286,9059	0	0	0	0	0	1	0	0	0	0	0
19	7.2001	19	289,286	287,5176	0	0	0	0	0	0	1	0	0	0	0
20	8.2001	20	288,467	287,58	0	0	0	0	0	0	0	1	0	0	0
21	9.2001	21	284,4929	282,9389	0	0	0	0	0	0	0	0	1	0	0
22	10.2001	22	281,99	282,9235	0	0	0	0	0	0	0	0	0	1	0
23	11.2001	23	274,1812	277,9421	0	0	0	0	0	0	0	0	0	0	1
24	12.2001	24	270,96	274,33	0	0	0	0	0	0	0	0	0	0	1
25	1.2002	25	274,73	271,5833	1	0	0	0	0	0	0	0	0	0	0
26	2.2002	26	276,335	272,51	0	1	0	0	0	0	0	0	0	0	0
27	3.2002	27	275,465	274,7471	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1
28	4.2002	28	281,495	278,45	0	0	0	1	0	0	0	0	0	0	0
29	5.2002	29	286,22	283,945	0	0	0	0	1	0	0	0	0	0	0
30	6.2002	30	288,0474	285,0944	0	0	0	0	0	1	0	0	0	0	0
31	7.2002	31	289,13	289,2857	0	0	0	0	0	0	1	0	0	0	0
32	8.2002	32	289,3158	288,4667	0	0	0	0	0	0	0	1	0	0	0
33	9.2002	33	281,7	284,492857	0	0	0	0	0	0	0	0	1	0	0
34	10.2002	34	279,865	281,99	0	0	0	0	0	0	0	0	0	1	0
35	11.2002	35	278,77	274,18125	0	0	0	0	0	0	0	0	0	0	1
36	12.2002	36	269,07	270,96	0	0	0	0	0	0	0	0	0	0	1

Rys. 3.3. Zrzut ekranu programu STATISTICA. Arkusz ze zmiennymi – dekompozycja sezonowa



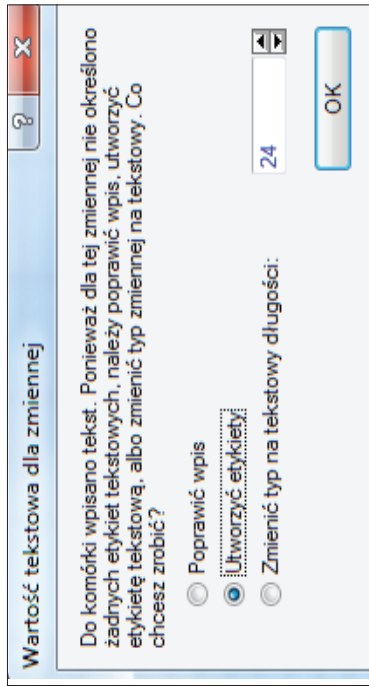
Kolejną metodą przygotowania danych jest dekompozycja sezonowa. Polega ona na tym, że tworzymy nowe zmienne, które wykorzystywane będą do obliczeń. Arkusz kalkulacyjny ze zmiennymi używanymi do dekompozycji sezonowej przedstawiono na rysunku 3.3. Dekompozycję sezonową zaprezentowano na przykładzie miesięcy (mogą to być również godziny dnia, dni tygodnia, kwartały). Należy wybrać jeden z miesięcy, który nie będzie służył jako przyszły potencjalny predyktor i nadać mu wszystkie wartości równe -1. W analizowanym przykładzie wybrano miesiąc marzec. Następnie kolejno nadawać wartości równe 1 dla jednej zmiennej w przypadku (wiersz arkusza) i zera dla pozostałych. Np. dla przypadku 1 zmiennej v5 styczeń – przypisano wartość 1, a pozostałym v6 do v16 wartości zero. Dla przypadku 2 zmiennej v6 luty – przypisano wartość 1, a pozostałym v5 oraz od v7 do v12 wartości zero. Działanie to należy wykonać dla wszystkich przypadków. Nowe zmienne zdekompozycjonowane sezonowo używa się jako predyktory w metodach obliczeniowych.

Proces przygotowywania danych przedstawiono w sposób uproszczony. W rzeczywistości jest to jednak zajęcie żmudne, wymagające wielkiej staranności i dokładności, gdyż nawet kilka błędnych danych z wielu tysięcy może spowodować błąd prognozy, lub powstanie prognozy o zbyt małej dokładności. Ponadto w przypadku kiedy późniejsze obliczenia nie są zadawalające, trzeba wrócić do danych wyjściowych, zmienić ich przekształcenia, zwiększyć liczbę przypadków lub wprowadzić nowe dane. Przygotowując dane do obliczeń nigdy nie mamy pewności, że dokonaliśmy właściwego wyboru. Zawsze istnieje możliwość, że pomimo poprawnych założeń nie uzyskamy oczekiwanych rezultatów. Nie należy się tym zrażać, należy podjąć działania naprawcze i wykonać obliczenia ponownie, z nadzieją na lepsze wyniki. Typowe zmienne niezależne w analizie czasu realizacji procesów budowlanych to: rodzaj sprzętu, warunki grunto-wo-wodne, skład brygady, doświadczenie pracowników, przyjęta technologia wykonania, miesiąc wykonywania robót, kondygnacja, doświadczenie kierownika budowy, firma podwykonawcza itp. Zmienne niezależne muszą być mierzalne lub możliwe do określenia w sposób lingwistyczny. Danym lingwistycznym można nadać wartości liczbowe (kody) i zastosować je do obliczeń statystycznych. Zwykle wartościom zmiennej w programie STATISTICA przypisuje się kody trzycyfrowe 101, 102, 103.... Jeśli mamy wiele zmiennych lingwistycznych to można przypisać im te same kody. Wtedy unikniemy automatycznego doboru kodów (przez program obliczeniowy) o wyższej wartości jako ważniejszych. Jeśli kolejnym przypadkiem zmiennej v1 nadamy kody 701, 702, 703..., to zwykle uznana ona zostanie za ważniejszą (o większym wpływie na zmienną zależną) niż zmienna v2 z nadanymi kodami 101, 102, 103.... Kodowanie zmiennych lingwistycznych przedstawiono na przykładzie 2. Na rysunku 3.4 przedstawiono arkusz kalkulacyjny przykładu 3.2. Zmienne od v1 do v5 ( w arkuszu oznaczane automatycznie jako 1, 2, 3...) to pomierzone czasy trwania procesów budowlanych P1, P2, P3, P4, P5. Danym lingwistycznym można nadać wartości liczbowe (kody) używając ich do obliczeń statystycznych. Zmienna v6 to zmienna lingwistyczna – przypisany numer brygady. Numer brygady jest w rzeczywistości nazwą brygady, dlatego nie należy

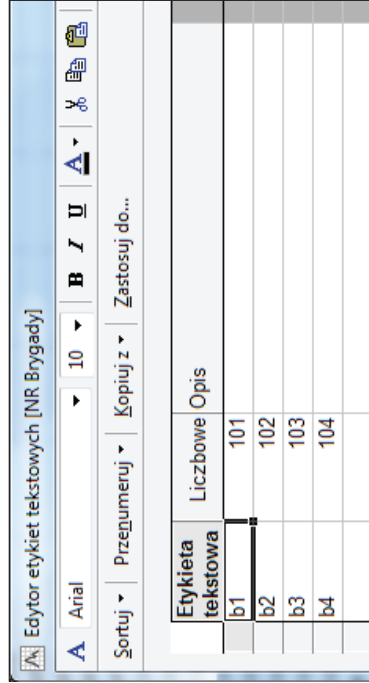
wpisywać wartości liczbowej np. 1 tylko słownie „pierwsza”. Zmienne v7 i v8 to odpowiednio: oznaczenie miesiąca realizacji prac oraz numer (nazwa) kondygnacji realizowanego budynku. Podobnie jak poprzednio oznaczenia liczbowe są nazwami miesięcy i kondygnacji, dlatego zaklasyfikowano je, jako zmienne lingwistyczne. Program STATISTICA po wpisaniu danej lingwistycznej (etykiety tekstowej) do pierwszego wiersza danej zmiennej, zada pytanie o nadanie wartości tekstowej dla zmiennej (rys.3.5). Należy wówczas zaznaczyć opcję „utworzyć etykiety” i w ten sposób otrzymamy kody zmiennych lingwistycznych rys. 3.6–3.8.

	1 P1	2 P2	3 P3	4 P4	5 P5	6 NR Brygady	7 miesiąc	8 kondygnacja
1	5	9	5,75	2,975	5,9	b1	czerwiec	czwarta
2	6	7	6,5	3,45	5,7	b1	czerwiec	czwarta
3	6	8	6,5	3,45	5,8	b1	czerwiec	czwarta
4	6	8	6,5	3,45	5,8	b1	czerwiec	czwarta
5	7	8	7,25	3,925	5,8	b1	czerwiec	czwarta
6	7	9	7,25	3,925	5,9	b1	czerwiec	czwarta
7	6	9	6,5	3,45	5,9	b1	czerwiec	czwarta
8	7	9	7,25	3,925	5,9	b1	czerwiec	trzecia
9	7	9	7,25	3,925	5,9	b1	czerwiec	trzecia
10	7	9	7,25	3,925	5,9	b1	czerwiec	trzecia
11	7	9	7,25	3,925	5,9	b2	maj	trzecia
12	7	9	7,25	3,925	5,9	b2	maj	trzecia
13	7	9	7,25	3,925	5,9	b2	maj	trzecia
14	7	9	7,25	3,925	5,9	b2	maj	trzecia
15	7	9	7,25	3,925	5,9	b2	maj	trzecia
16	7	10	7,25	3,925	6	b2	maj	druga
17	8	10	8	4,4	6	b2	maj	druga
18	8	10	8	4,4	6	b2	maj	druga
19	8	10	8	4,4	6	b2	maj	druga
20	8	10	8	4,4	6	b2	maj	druga
21	8	10	8	4,4	6	b2	kwiecień	druga
22	8	10	8	4,4	6	b2	kwiecień	druga
23	8	10	8	4,4	6	b2	kwiecień	druga
24	8	10	8	4,4	6	b3	kwiecień	druga
25	8	10	8	4,4	6	b3	kwiecień	druga
26	8	10	8	4,4	6	b3	kwiecień	druga
27	8	10	8	4,4	6	b3	kwiecień	druga
28	9	10	8,75	4,875	6	b3	marzec	druga
29	9	10	8,75	4,875	6	b3	marzec	druga
30	9	10	8,75	4,875	6	b3	marzec	druga
31	9	10	8,75	4,875	6	b4	marzec	pierwsza
32	8	11	8	4,4	6,1	b4	marzec	pierwsza
33	9	11	8,75	4,875	6,1	b4	marzec	pierwsza
34	9	11	8,75	4,875	6,1	b4	marzec	pierwsza
35	9	11	8,75	4,875	6,1	b4	marzec	pierwsza
36	9	11	8,75	4,875	6,1	b4	marzec	pierwsza
37	9	11	8,75	4,875	6,1	b4	marzec	pierwsza
38	10	12	9,5	5,35	6,2	b4	luty	pierwsza
39	10	12	9,5	5,35	6,2	b4	luty	pierwsza
40	11	12	10,25	5,825	6,2	b4	luty	pierwsza

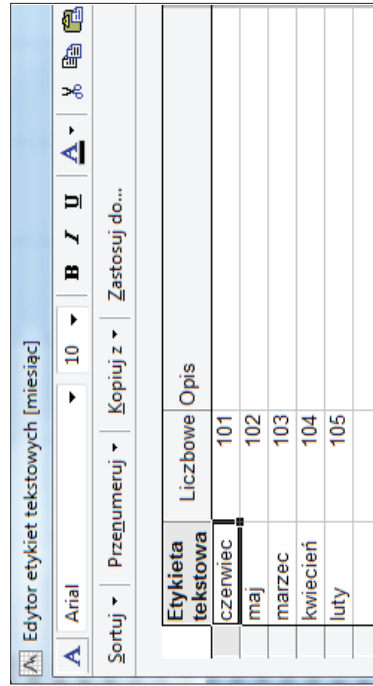
Rys.3.4. Przykład 3.2: Arkusz kalkulacyjny programu STATISTICA



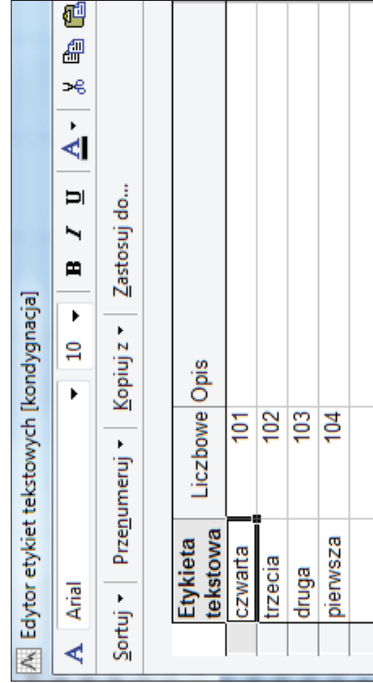
Rys. 3.5. Okno dialogowe programu STATISTICA nadanie wartości tekstowej dla zmiennej



Rys. 3.6. Okno dialogowe programu STATISTICA edycja etykiet tekstowych, zmienna numer brygady



Rys. 3.7. Okno dialogowe programu STATISTICA edycja etykiet tekstowych, zmienna miesiąc



Rys. 3.8. Okno dialogowe programu STATISTICA edycja etykiet tekstowych, zmienna kondygnacja

Po wykonaniu obliczeń statystycznych otrzymane równanie regresji będzie zawierało etykiety tekstowe, pod które należy podstawić wartości liczbowe. w równaniu 3.1 (model regresyjny metody wieloimiennej regresji adaptacyjnej z użyciem funkcji skleja-nych MARS) pod etykietę tekstową miesiąc luty podstawić wartość 105, miesiąc marzec wartość 103 itd.

$$P1 = 7,000000000000000e+000 + 2,55952380952381e+000 * \max(0; \text{miesiąc\_luty} - 0,000000000000000e+000) + 1,08333333333333e+000 * \max(0; \text{miesiąc\_marzec} - 0,000000000000000e+000) + 9,16666666666668e-001 * \max(0; \text{kondygnacja\_druga} - 0,000000000000000e+000) - 8,57142857142854e-001 * \max(0; \text{kondygnacja\_czwarta} - 0,000000000000000e+000) + 7,73809523809524e-001 * \max(0; \text{NR Brygady\_b4} - 0,000000000000000e+000) \quad (3.1)$$

Wykorzystywanie zmiennych lingwistycznych do prognozowania znacznie poszerza możliwości obliczeniowe. Wiele czynników zostałoby pominiętych, gdyby istniała konieczność wykorzystywania tylko zmiennych, którym można przypisać pomierzone wartości liczbowe.

W prezentowanym arkuszu kalkulacyjnym do przykładu 3.2 (rys.3.4), zmiennymi zależnymi są czasy trwania procesów od P1 do P5, natomiast zmiennymi niezależnymi: numer brygady, miesiąc i kondygnacja. Każda ze zmiennych zależnych wymaga przeprowadzenia indywidualnego toku obliczeniowego. Zatem wyznaczamy jedną zmienną zależną i wiele zmiennych niezależnych. Jeśli w równaniu regresji występują zmienne niezależne to nazywamy je wówczas predyktorami.

Arkusze kalkulacyjne programu STATISTICA mogą być ogromne. Mogą zawierać miliony danych. Zebranie bardzo dużej liczby danych w budownictwie zwykle okazuje się problematyczne. Zatem należałoby określić minimalną liczbę danych do obliczeń. Zwykle przyjmuje się, że liczba przypadków powinna być dziesięciokrotnie większa od liczby predyktorów. Czyli jeśli w prezentowanym arkuszu z przykładu 2, mamy 3 zmienne niezależne i naszą intencją jest by jednocześnie były one predyktorami zmiennej zależnej, to powinniśmy dysponować co najmniej 30 przypadkami do analizy. Nie jest to jednak warunek konieczny a jedynie zalecenie. Można przeprowadzić obliczenia dysponując mniejszą liczbą danych, dobierając metody obliczeniowe. Zależne to jest również od jednorodności danych i trafności wyboru zmiennych niezależnych.

### 3.3. Analiza zmiennych

Podstawowym i najbardziej korzystnym, z obliczeniowego punktu widzenia, rozkładem zmiennej jest rozkład normalny. Podczas toku obliczeniowego wyznaczania równania regresji wiele metod wymaga podania rodzaju rozkładu zmiennej. Należy określić czy zmienne zależna i niezależne mają rozkład normalny. Bardzo wygodną metodą jest zastosowanie testu w Shapiro-Wilka (Shapiro, Wilk (1965)). Formułujemy hipotezę zerową  $H_0$ , że zmienna ma rozkład zbliżony do normalnego, przeciwko hipotezie alternatywnej  $H_1$ :  $\sim H_0$ , że zmienna nie

ma rozkładu zbliżonego do normalnego. Wynikiem przeprowadzonej analizy będzie stwierdzenie, że nie ma podstaw do odrzucenia hipotezy zerowej, lub że taka podstawa istnieje. Obliczamy wartość statystyki w testu Shapiro-Wilka zgodnie ze wzorem (3.2):

$$W = \frac{\left( \sum_{i=1}^{\frac{n}{2}} a_{n-i+1} (x_{(n-i+1)} - x_{(i)}) \right)^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (3.2)$$

gdzie:

$n/2$  – część całkowita liczby  $n/2$ ,

$x_{(i)}$  – zmienna przyjmująca  $i$ -tą co do wielkości wartość w próbie ( $i$ -ta statystyka pozycyjna),

$a_{n-i+1}$  – stabilizowane współczynniki Shapiro-Wilka.

Obszar odrzucenia testu jest postaci  $[0, W_\alpha]$ , gdzie  $W_\alpha$  jest kwantylem rzędu  $\alpha$  rozkładu Shapiro-Wilka. Wartości  $W_\alpha$  zestawione są w tablicach kwantyli rozkładu Shapiro-Wilka (tab. 3.1.).

Tab. 3.1. Wartości krytyczne  $W_{kr}$  testu Shapiro-Wilka dla  $\alpha=0,05$

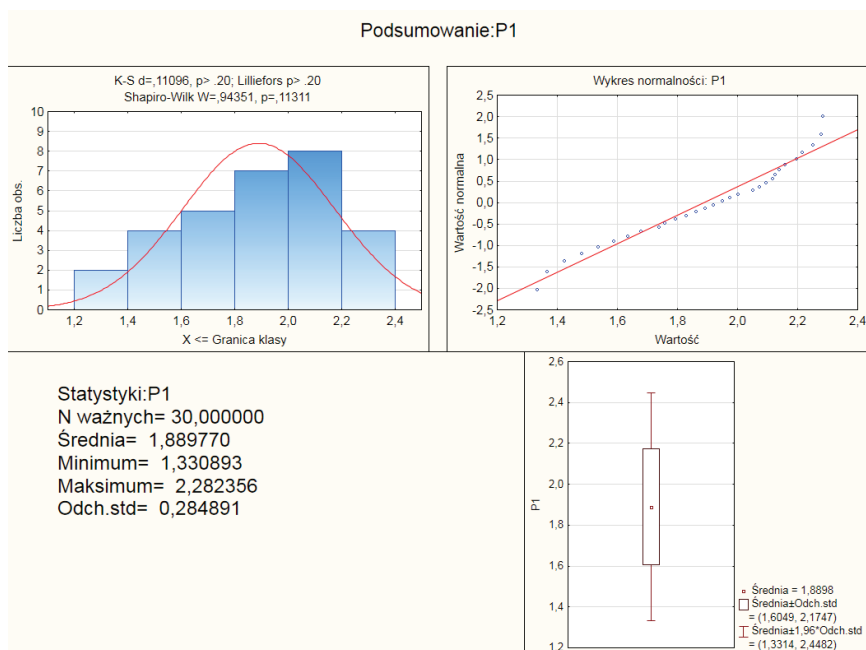
n	$W_{kr}$	n	$W_{kr}$	n	$W_{kr}$	n	$W_{kr}$
1	-	26	0,920	51	0,949	76	0,959
2	-	27	0,923	52	0,950	77	0,959
3	0,767	28	0,924	53	0,950	78	0,959
4	0,748	29	0,926	54	0,951	79	0,960
5	0,762	<b>30</b>	<b>0,927</b>	55	0,951	80	0,960
6	0,788	31	0,929	56	0,952	81	0,960
7	0,803	32	0,930	57	0,952	82	0,961
8	0,818	33	0,931	58	0,953	83	0,961
9	0,829	34	0,933	59	0,953	84	0,961
10	0,842	35	0,934	60	0,954	85	0,961
11	0,850	36	0,935	61	0,954	86	0,962
12	0,859	37	0,936	62	0,955	87	0,962
13	0,866	38	0,938	63	0,955	88	0,962
14	0,874	39	0,939	64	0,955	89	0,962
15	0,881	40	0,940	65	0,956	90	0,962
16	0,887	41	0,941	66	0,956	91	0,963
17	0,892	42	0,942	67	0,956	92	0,963
18	0,897	43	0,943	68	0,957	93	0,963
19	0,901	44	0,944	69	0,957	94	0,963
20	0,905	45	0,945	70	0,957	95	0,963
21	0,908	46	0,945	71	0,958	96	0,963
22	0,911	47	0,946	72	0,958	97	0,964
23	0,914	48	0,947	73	0,958	98	0,964
24	0,916	49	0,947	74	0,959	99	0,964
25	0,918	50	0,947	75	0,959	100	0,964

Analizując dane z przykładu 3.1 (arkusz kalkulacyjny, rys. 3.1.) przyjęto kolejno dla hipotezy, że zmienne mają rozkład normalny. Zweryfikowano tę hipotezę tworząc test Shapiro-Wilka na poziomie istotności  $\alpha = 0,05$ . Sformułowano hipotezy zerowe –  $H_0$ ; proces P1 oraz zmienne niezależne PR1 i PR2 podlegają rozkładowi normalnemu, przeciwko hipotezie alternatywnej  $H_1: \sim H_0$ . Dla  $n = 30$  przypadków i poziomu istotności  $\alpha = 0,05$  wartość  $W_{\alpha} = 0,927$  (oznaczona w tabeli 3.1 czcionką bold). Obliczono wartość statystyki W testu Shapiro-Wilka i otrzymano wartości zestawione w tabeli 3.2.

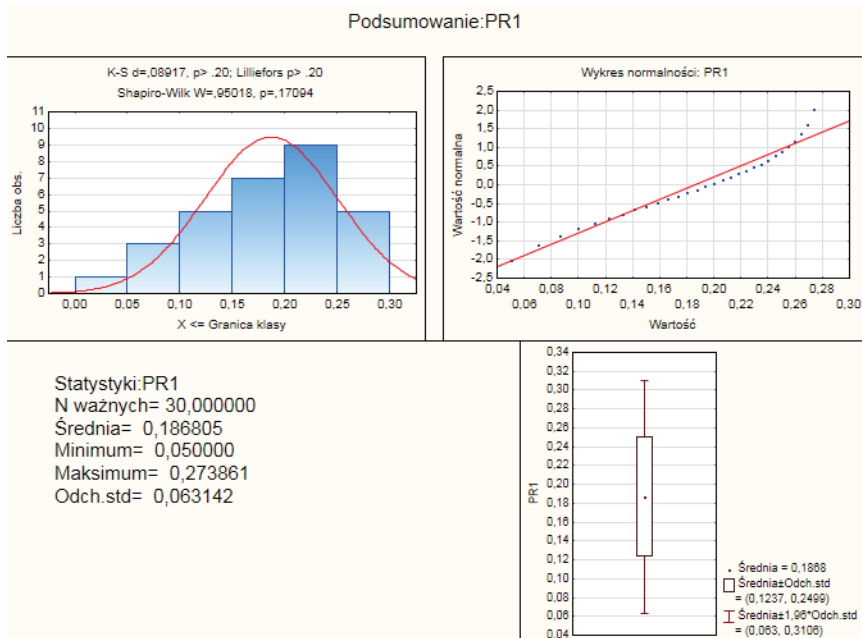
Tab. 3.2. Wartości statystyk w testu normalności Shapiro-Wilka dla zmiennych z przykładu 3.1

LP	Nazwa zmiennej	$W_{0,05}, n = 30$	W	Rozkład normalny
1	P1	<b>0,927</b>	0,94351	TAK $W_{0,05} < W$
2	PR1	<b>0,927</b>	0,95018	TAK $W_{0,05} < W$
3	PR2	<b>0,927</b>	0,90142	NIE $W_{0,05} > W$

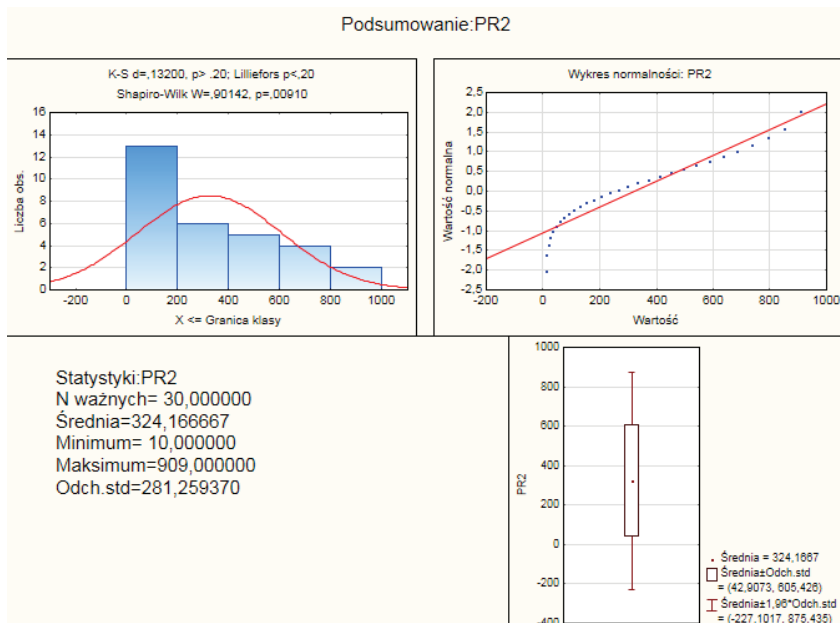
Podsumowania zmiennych P1, PR1 i PR2 przedstawiono na rysunkach 3.9–3.11.



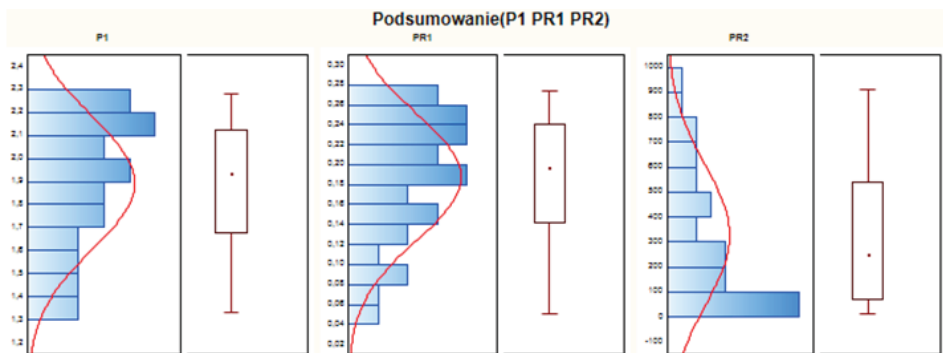
Rys. 3.9. Podsumowanie statystyk podstawowych zmiennej P1 z obliczoną wartością W testu normalności Shapiro-Wilka



Rys. 3.10. Podsumowanie statystyk podstawowych zmiennej PR1 z obliczoną wartością w testu normalności Shapiro-Wilka



Rys. 3.11. Podsumowanie statystyk podstawowych zmiennej PR2 z obliczoną wartością w testu normalności Shapiro-Wilka



Rys. 3.12. Zestawienie histogramów zmiennych P1, PR1 i PR2

Zadanie prognostyczne to wyznaczenie wzoru regresyjnego do obliczania wartości P1 ze znanych wartości PR1 i PR2. Histogramy zmiennej P1 i PR1 są podobne (wykazują rozkłady zbliżone do normalnego) i nie powinny wystąpić komplikacje obliczeniowe. Histogram zmiennej PR2 w znaczny sposób różni się od histogramu zmiennej P1 (rys. 3.12.). Może okazać się, że nie we wszystkich proponowanych metodach obliczeniowych PR2 jest predyktorem równania regresyjnego do wyznaczania wartości P1.

Zmienna PR2 nie wykazuje rozkładu zbliżonego do normalnego, wartość współczynnika w Shapiro-Wilka wynosi 0,90142. W wielu zadaniach modelowania statystycznego ogólnym wymaganiem jest normalność rozkładu zmiennych. Można przekształcić zmienną tak by wykazywała rozkład zbliżony do normalnego. Do tego celu służy między innymi przekształcenie Boxa-Coxa (Box, Cox (1964)), które transformuje zmienne tak aby ich układ był zbliżony do rozkładu normalnego. Wzór przekształcenia Boxa-Coxa ma następującą postać (3.3):

$$X_{\text{BOXCOC}} = \begin{cases} \frac{(X + \sigma)^\lambda - 1}{\lambda} & \text{dla } \lambda \neq 0 \\ \log(X + \sigma) & \text{dla } \lambda = 0 \end{cases} \quad (3.3)$$

gdzie:

$X$  – oryginalna zmienna,

$X_{\text{BOXCOC}}$  – przekształcona zmienna,

$\sigma$  – odchylenie standardowe oryginalnej zmiennej,

$\lambda$  – główny parametr przekształcenia Boxa-Coxa.

Program STATISTICA celem obliczenia parametru  $\lambda$  używa podejścia, określane jako maksymalizacja wiarygodności, gdzie tworzy się funkcję wiarygodności, a właściwie minus logarytm tej funkcji (3.4):

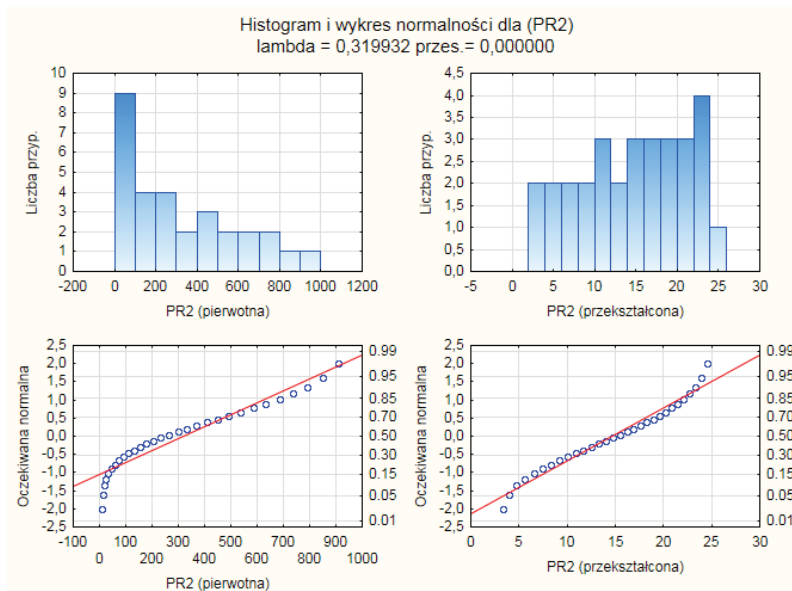
$$E(\lambda) = -\frac{N}{2} \ln(\sigma_{\text{BOXCOC}}^2) + (1 - \lambda) \sum_{i=1}^N \ln(x_{i\text{COCBOX}}) \quad (3.4)$$

gdzie:

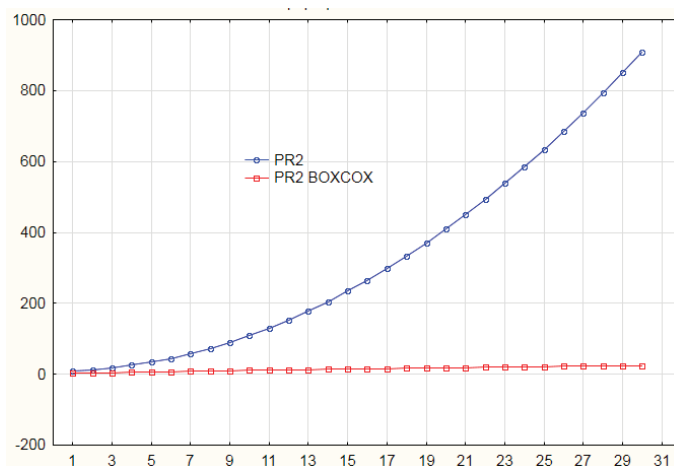
$\sigma_{\text{BOXCOC}}$  – odchylenie standardowe przekształconej zmiennej,

$x_{i\text{COCBOX}}$  – i-ty przypadek przekształconej zmiennej.





Rys. 3.13. Histogramy i wykresy normalności zmiennych PR2 i PR2<sub>BOXCOX</sub>



Rys. 3.14. Zestawienie wartości zmiennych PR2 i zmiennej przekształconej metodą Boxa-Coxa PR2<sub>BOXCOX</sub>

Poszukuje się takiej wartości  $\lambda$ , która minimalizuje E. Stosowana jest metoda poszukiwania Goldena. w algorytmie tym zaczyna się obliczenia od najmniejszej wartości  $\lambda$ , stosuje się przekształcenie Boxa-Coxa przy tej wartości parametru i oblicza się wartość E. Proces ten jest powtarzany aż do osiągnięcia największego założonego  $\lambda$ , a wynikiem algorytmu jest lambda, dla której E osiągnęło najmniejszą wartość.

Zmienną PR2 poddano przekształceniu Boxa-Coxa. Przekształconą zmienną nazwano PR2<sub>BOXCOX</sub>. Uzyskano wartość lambda równą 0,319932. Obliczona wartość

w testu normalności Shapiro-Wilka wynosi  $W_{0,05} = 0,95126$ , zatem nie ma już podstaw do odrzucenia hipotezy o normalności rozkładu. Histogramy i wykresy normalności zmiennych PR2 i  $PR2_{\text{BOXCOX}}$  przedstawiono na rys. 3.13. Wykres porównawczy zmiennych PR2 i  $PR2_{\text{BOXCOX}}$  przedstawiono na rysunku 3.14.

### 3.4. Korelacje liniowe zmiennych

Jeśli zmienne niezależne są skorelowane liniowo to podczas wykonywania obliczeń metodą regresji wielorakiej z wykorzystaniem programu STATISTICA, jedna ze zmiennych niezależnych zostanie wyeliminowana z równania regresji. Na przykład, jeśli analizujemy czas wykonania autostrady i między innymi zmiennymi niezależnymi są koszt wykonania i liczba mostów to jedna z tych zmiennych zostanie wyeliminowana z równania regresji. Otrzymany wynik obliczeń, wykazywał będzie, że koszt inwestycji nie zależy od liczby wybudowanych mostów, co mogłoby wydawać się nielogiczne. Jednak w rzeczywistości liczba mostów zawierać się będzie w zmiennej koszt. Celem uniknięcia błędnej interpretacji równania regresji, należy przeprowadzić każdorazowo analizę korelacji zmiennych. Analizie podlegać będą zmienne nieliniowe o ciągłych rozkładach.

Oblicza się estymatory współczynnika korelacji liniowej Pearsona  $r_{xy}$  (Pearson (1905), Cohen (1988), Buda, Jarynowski (2010)), zgodnie z wzorem (3.5):

$$r_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (3.5)$$

gdzie:  $x_i$  i  $y_i$  oznaczają wartości prób losowych, natomiast  $\bar{x}$  i  $\bar{y}$  wartości średnie z tych prób.

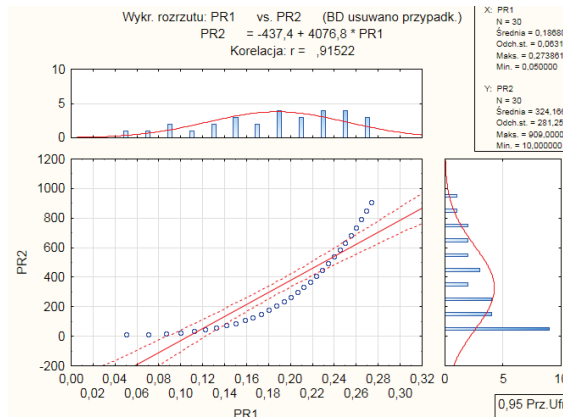
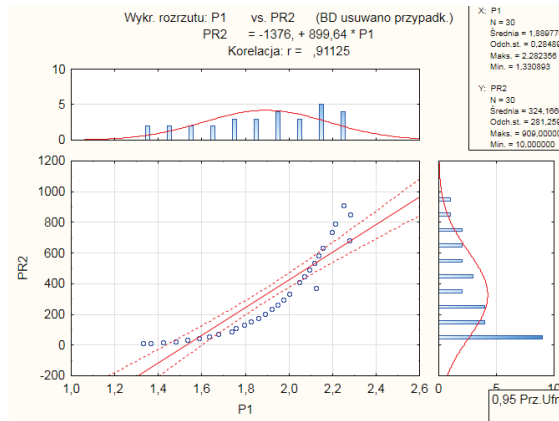
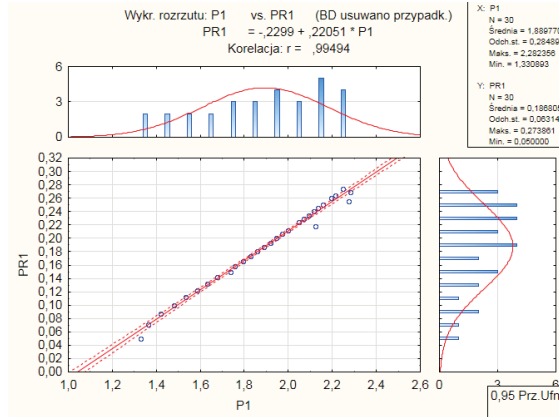
Współczynnik korelacji liniowej dwóch zmiennych X i Y można również zapisać jako iloraz kowariancji i iloczynu odchyłeń standardowych tych zmiennych (3.6):

$$r_{xy} = \frac{\text{cov}(X, Y)}{\sigma_x \sigma_y} \quad (3.6)$$

gdzie:  $\text{cov}(X, Y)$  to kowariancja zmiennych X i Y, natomiast  $\sigma_x$  i  $\sigma_y$  oznaczają wartości odchyłeń standardowych tych zmiennych.

Tab. 3.3. Wartości współczynników korelacji liniowej Paersona  $r_{xy}$  analizowanych zmiennych P1, PR1 i PR2

Zmienna	Korelacje (p1pr1pr2)				
	Oznaczone współczynniki korelacji są istotne z $p < 0,5000$ N = 30 (Braki danych usuwano przypadkami)				
	Średnia	Odch. std	P1	PR1	PR2
P1	1,8898	0,2849	1,000000	0,994941	0,911252
PR1	0,1868	0,0631	0,994941	1,000000	0,915222
PR2	324,1667	281,2594	0,911252	0,915222	1,000000



Rys.3.15 Obliczone współczynniki korelacji liniowej  $r_{xy}$  Pearsona

Obliczone estymatory współczynnika korelacji liniowej Pearsona  $r_{xy}$  dla zmiennych P1, PR1 i PR2 z przykładu 3.1, zestawiono w tablicy 3.3 i zobrazowano na rys. 3.15. Korelacje uznaje za silne gdy  $r_{xy}$  należy do przedziału  $r_{xy} \in [-1,0;-0,8] \cup [0,8;1,0]$ . W analizowanym przykładzie stwierdzono istnienie silnych korelacji liniowych pomiędzy wszystkimi zmiennymi. Jedna ze zmiennych niezależnych zostanie prawdopodobnie wyeliminowana z równania regresji w metodzie regresji wielorakiej MR. przypadku stosowania innych metod obliczeniowych jest to mniej prawdopodobne.

### 3.5. Funkcja autokorelacji reszt i autokorelacji cząstkowej reszt

Zgodnie z definicją regresji, szereg resztowy, który powstaje jako różnica pomiędzy wartościami danymi (pomierzonymi) a wartościami prognozowanymi, powinien mieć średnią bliską zera, chociaż jednocześnie jego odchylenie standardowe nie musi być wartością zerową. Z tego względu warunkiem koniecznym w procesie prognozowania jest analiza szeregów resztowych wygenerowanych modeli. Służy do tego analiza funkcji autokorelacji reszt (składnika losowego) oraz funkcja autokorelacji cząstkowej reszt.

Autokorelacja składnika losowego to korelacja między składnikami losowymi  $e_1, e_2, \dots, e_{t-k}, \dots, e_t$  modelu (G. S. Maddala Introduction to Econometrics, Wiley 2002). Korelacja między  $e_t$  i  $e_{t-k}$  jest określana mianem autokorelacji rzędu  $k$  i oznaczana jest przez  $\rho_k$  (3.7):

$$\rho_k = \frac{\text{Cov}(e_t, e_{t-k})}{\sigma_t \cdot \sigma_t} = \frac{\sum_{k=1}^t [(e_t - \bar{e}) \cdot (e_{t-k} - \bar{e})]}{t-1} \cdot \frac{1}{\sqrt{\frac{\sum_{k=1}^t (e_t - \bar{e})^2}{t-1}}} = \frac{\sum_{k=1}^t [(e_t - \bar{e}) \cdot (e_{t-k} - \bar{e})]}{\sum_{k=1}^t (e_t - \bar{e})^2} \quad (3.7)$$

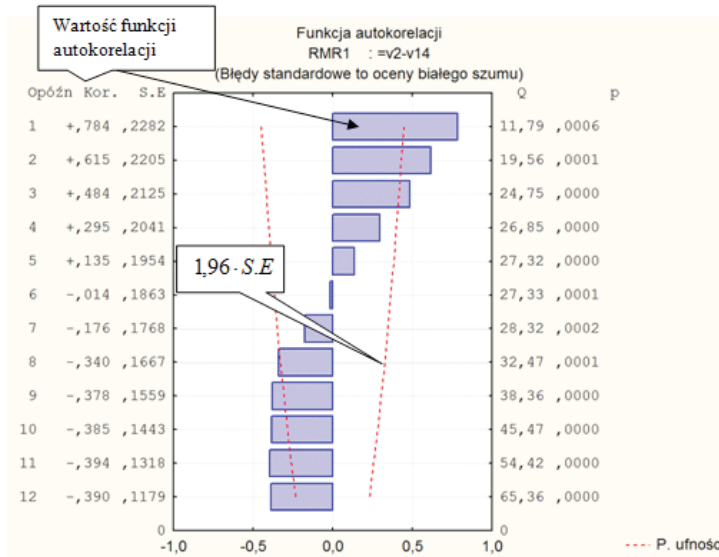
Jeśli model jest prawidłowy nie powinna zachodzić autokorelacja składnika losowego.

Skutki autokorelacji składnika losowego dla estymatorów metody najmniejszych kwadratów MNK są następujące: jeśli wśród zmiennych objaśniających:

nie występują opóźnienia zmiennej objaśnianej, estymatory są nieobciążone, ale nieefektywne, natomiast estymatory wariancji są obciążone,

występują opóźnienia zmiennej objaśnianej, estymatory MNK nie są zgodne.

Na rysunku 3.16 przedstawiono zestawienie wartości funkcji autokorelacji dla 12 opóźnień oraz wartości graniczne, których nie powinna przekraczać. Wartości funkcji są przekroczone dla następujących opóźnień 1, 2, 3, 8, 9, 10, 11 i 12. W prezentowanym przypadku należy stwierdzić, że hipoteza o występowaniu autokorelacji potwierdziła się. Nie można zatem uznać, że równanie, z którego pochodzi szereg resztowy, jest równaniem regresyjnym.



Rys. 3.16. Przedstawienie graficzne funkcji autokorelacji przykładowego modelu 1 metody MR. Na wykresie wartości liczbowe w kolumnach od lewej to: opóźnienie szeregu, wartości funkcji autokorelacji dla kolejnych opóźnień, błędy standardowe funkcji autokorelacji, Q statystyka Ljunga-Boxa, p – prawdopodobieństwo testowe statystyki Q

Statystyką badającą istotność kolejnych współczynników korelacji jest statystyka Ljunga-Boxa postaci (3.8) (Ljung, Box (1978)):

$$Q = t \cdot (t + 2) \cdot \sum_{k=1}^h \frac{\rho_k^2}{t - k} \quad (3.8)$$

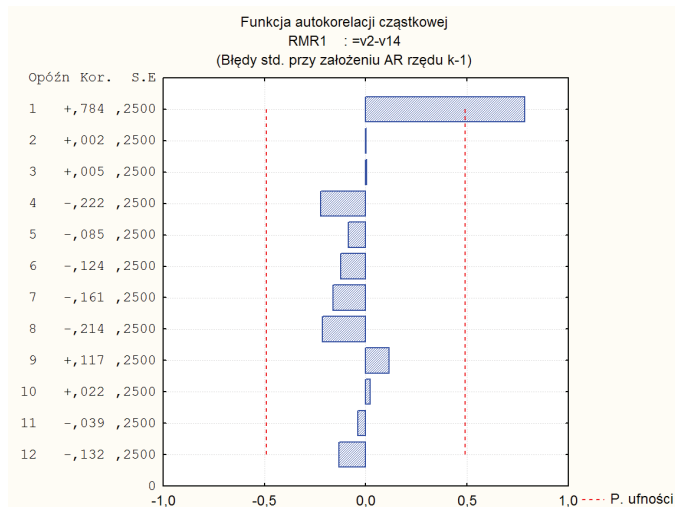
Statystyka Q ma rozkład  $\chi^2$  z k stopniami swobody. Wartości sprawdzianu większe od wartości krytycznych pozwalają na odrzucenie hipotezy zerowej zakładającej nieistotność autokorelacji rzędu k. W przeciwnym przypadku nie ma podstaw do odrzucenia

hipotezy zerowej. Jeśli  $Q > \chi^2_{1-\alpha, k}$  to autokorelacja rzędu k jest istotna. Przykładowo dla opóźnienia  $k = 1$  i  $\alpha = 0,05$ ,  $\chi^2_{kr}$  obliczone przy użyciu opcji Kalkulator Prawdopodobieństwa programu STATISTICA wynosi 0,003932. Otrzymana wartość jest mniejsza od  $Q = 11,79$ . Nie ma podstaw do odrzucenia hipotezy zerowej zakładającej nieistotność autokorelacji rzędu pierwszego.

Następnie należy sprawdzić hipotezę o występowaniu w szeregu resztowym autokorelacji cząstkowych. Funkcja autokorelacji cząstkowej ACF obliczana jest zgodnie z (3.9):

$$ACF(k) = \frac{\frac{1}{n} \sum (e_t - \bar{e})(e_{t-k} - \bar{e})}{\sigma_e^2 + (\bar{e})^2} \quad (3.9)$$

Poziom ufności oznaczony linią przerywaną na rys. 3.17. wyznaczany jest jako iloraz liczby 1,96 i odchylenia standardowego badanego szeregu. Jeśli wartość funkcji autokorelacji cząstkowej przekracza linię przerywaną oznacza to, że należy przyjąć hipotezę o występowaniu w szeregu resztowym autokorelacji cząstkowych. Równanie otrzymane z obliczeń nie może być uznane za równanie regresji (Pierce, Box (1970)).



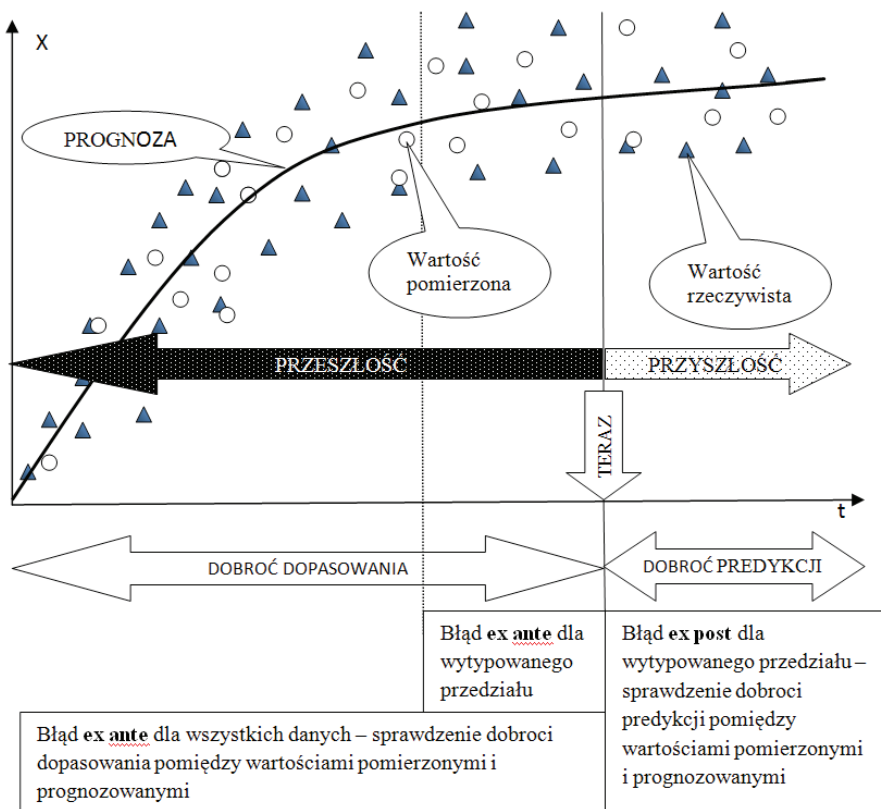
Rys. 3.17. Przedstawienie graficzne funkcji autokorelacji cząstkowej przykładowego modelu 1 metody MR

Z praktyki obliczeniowej wynika, że jeśli szereg resztowy równania regresji ma wysoką wartość funkcji autokorelacji cząstkowej lub/i autokorelacji cząstkowej reszt, przekraczającej dozwoloną wartość to błąd prognozy jest wysoki. Mniej istotny wpływ na błąd prognozy ma występowanie autokorelacji dla opóźnienia powyżej piątego.

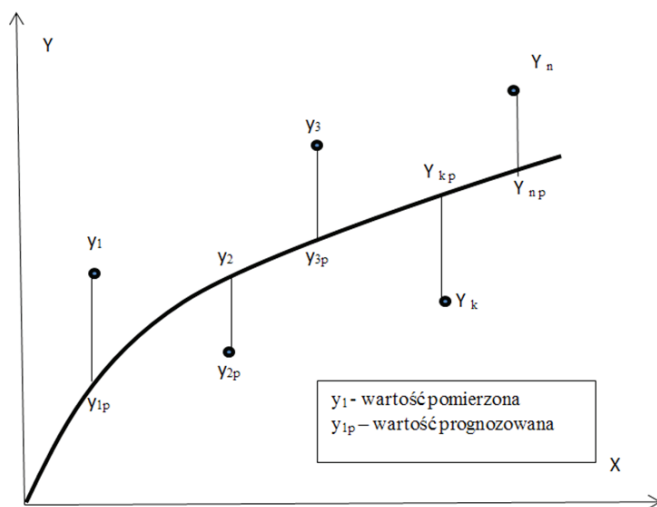
### 3.6. Ocena dobroci dopasowania modeli i prawidłowości predykcji, obliczenie błędów

Predykcja to szacowanie wartości jednej zmiennej na podstawie wyników innej zmiennej, lub zmiennych, wtedy gdy zmienne te są ze sobą skorelowane. Korelacja pomiędzy zmiennymi nie musi mieć charakteru liniowego. Im wyższa korelacja między tymi zmiennymi tym lepsza predykcja.

Predykcja na podstawie linii regresji to poszukiwanie linii regresji najlepiej dopasowanej do wartości zmiennej zależnej (przy jej predykcji) na podstawie zmiennych niezależnych. Linia regresji opisywana jest za pomocą równania regresji. Metoda ta wykorzystuje wszystkie wyniki w próbie badawczej. Linia wyznaczona na podstawie danych z próby może różnić się od linii najlepiej dopasowanej do wyników populacji. Wartości pomierzone również mogą różnić się od wartości rzeczywistych z populacji (rys. 3.18), np. ze względu na błędy pomiarowe.



Rys. 3.18. Schemat predykcji na podstawie linii regresji



Rys. 3.19. Wartości pomierzone i prognozowane wraz z linią regresji

Najlepiej dopasowana linia regresji (rys. 3.19) to taka linia, która minimalizuje sumę kwadratów odchyleń między wartościami pomierzonymi a ich wartościami przewidywanymi (tzw. kryterium najmniejszych kwadratów).

Linia regresji jest opisywana równaniem regresji. Postać tego równania jest różna w zależności od korelacji między zmiennymi oraz przyjętej metody obliczeniowej. Bardzo rzadko zdarza się, że przyjmuje ona postać liniową. Prognozując wykorzystuje się zwykle wiele metod obliczeniowych, a w metodach tych wiele modeli predykcyjnych. Otrzymuje się wiele możliwości potencjalnych równań regresji. Warunkiem podstawowym doboru jednego z nich jako właściwego, jest nie występowanie w jego szeregu resztowym autokorelacji i autokorelacji cząstkowych (szczególnie w pierwszych 5 opóźnieniach). Wybór najlepszego rozwiązania następuje poprzez uwzględnienie najmniejszych błędów predykcyjnych. Jest to element wspólny wszystkich zastosowanych do obliczeń metod. Można określać błędy *ex ante* i *ex post*. Mierniki *ex ante* to mierniki, które podają spodziewany rząd odchyleń rzeczywistych realizacji zmiennej prognozowanej od prognoz; mierniki te oblicza się przed realizacją. Mierniki *ex post* podają wielkość rzeczywistego odchylenia wartości zmiennych od prognoz; mierniki te oblicza się po zrealizowaniu wydarzeń prognozowanych. Zobrazowano to na rysunku 3.18.

Celem analitycznej oceny poprawności predykcji oblicza się błędy prognoz (King, Minium 2009): *ME*, *MAE*, *MPE* i *MAPE* dane wzorami (3.10), (3.11), (3.12) i (3.13).

$$ME = \frac{1}{T-n} \sum_{i=T-n}^T (Y_i - Y_{ip}) \quad (3.10)$$

$$MAE = \frac{1}{T-n} \sum_{i=T-n}^T |Y_i - Y_{ip}| \quad (3.11)$$

$$MPE = \frac{1}{T-n} \sum_{i=T-n}^T \frac{Y_i - Y_{ip}}{Y_i} \quad (3.12)$$

$$MAPE = \frac{1}{T-n} \sum_{i=T-n}^T \frac{|Y_i - Y_{ip}|}{Y_i} \quad (3.13)$$

gdzie:

*ME* – średni błąd (mean error),

*MAE* – średni osiągnięty błąd (mean average error),

*MPE* – średni błąd procentowy (mean percentage error),

*MAPE* – średni absolutny procentowy błąd (mean absolute percentage error),

*T* – suma liczby okresów obliczeniowych i prognozowanych,

*n* – liczba okresów prognozowanych,

$Y_i$  – wartość rzeczywista zmiennej w okresie *i*,

$Y_{ip}$  – wartość prognozowana zmiennej w okresie *i*.



W procesie oceny prognozy pomocne są również współczynniki rozbieżności Theila (Theil (1979)), które dostarczają informacji o źródłach błędów prognoz. Podstawowy współczynnik  $I^2$  zdefiniowany jest wzorem (3.14).

$$I^2 = \frac{\sum_{i=T-n}^T (Y_p - Y_i)^2}{\sum_{i=T-n}^T Y_i^2} \quad (3.14)$$

Pierwiastek kwadratowy z parametru  $I^2$  oznacza przeciętny względny błąd dopasowania  $V$ . Interpretację współczynnika Theila ułatwia jego dekompozycja na trzy składowe  $I_1^2$ ,  $I_2^2$  i  $I_3^2$  zgodnie z równaniem (3.14). Składowe opisane równaniami (3.15–3.18), pozwalają sprecyzować rząd i charakter błędów aproksymacji. Składnik  $I_1^2$  jest miarą błędu systematycznego, który wynika z obciążenia metody estymacji modelu, inaczej z obciążenia predykcji (niedostateczna zgodność przeciętnych wartości  $Y_i$  i  $Y_p$ ). Parametr  $I_2^2$  jest miernikiem stopnia odwzorowania przez model wariacji zmiennej objaśnianej i rozmiaru błędu związanego z niedostateczną elastycznością predykcji. Innymi słowy odzwierciedla błędy spowodowane faktem, że zróżnicowanie prognoz nie jest zgodne z faktycznymi wahaniami zmiennej prognozowanej. Składnik  $I_3^2$  charakteryzuje błąd wynikający z braku pełnej zgodności kierunku zmian wartości teoretycznych ze zmianami wartości empirycznych, a więc z nie dość dobrej predykcji punktów zwrotnych. Ogólnie, jeżeli odbiorca prognozy nie poda własnych kryteriów dopuszczalności prognoz, przyjmujemy, że jeżeli względny miernik dokładności predykcji  $V$  przyjmuje wartości:

- $\sqrt{I^2} \leq 3\%$ , to prognozy są wysoce dokładne,
- $3\% < \sqrt{I^2} \leq 5\%$ , to prognozy uznajemy za dobre,
- $5\% < \sqrt{I^2} \leq 10\%$ , to prognozy są mało dokładne, ale mogą być przyjęte (dopuszczalne),
- $\sqrt{I^2} > 10\%$ , to prognozy są niedokładne, nie powinny być przyjęte (niedopuszczalne).

$$I^2 = I_1^2 + I_2^2 + I_3^2 \quad (3.15)$$

$$I_1^2 = \frac{(\bar{Y}_p - \bar{Y}_i)^2}{\frac{1}{T-n} \sum_{i=T-n}^T Y_i^2} \quad (3.16)$$

$$I_2^2 = \frac{(s_p - s_i)^2}{\frac{1}{T-n} \sum_{i=T-n}^T Y_i^2} \quad (3.17)$$

$$I_3^2 = \frac{2s_{ip}s_i(1-r_{ipi})}{T-n \sum_{i=T-n}^T Y_i^2} \quad (3.18)$$

gdzie:

$I^2, I_1^2, I_2^2, I_3^2$  – współczynniki Theila,

$I$  – względny miernik dokładności predykcji,

$T$  – suma liczby okresów obliczeniowych i prognozowanych,

$n$  – liczba okresów prognozowanych,

$Y_i$  – wartość rzeczywista zmiennej w okresie  $i$ ,

$Y_{ip}$  – wartość prognozowana zmiennej w okresie  $i$ ,

$s_i$  – odchylenie standardowe zmiennej prognozowanej,

$s_{ip}$  – odchylenie standardowe prognoz,

$r_{ipi}$  – współczynnik korelacji liniowej pomiędzy wielkościami  $Y_i$  i  $Y_{ip}$ ,

$\bar{Y}_i$  – średnia arytmetyczna realizacji zmiennej prognozowanej,

$\bar{Y}_p$  – średnia arytmetyczna wartości prognoz.

Błąd *MAPE* jest najbardziej miarodajny. Jest podstawowym miernikiem prawdziwości obliczeń. Dlatego proponuje się używanie go do wyboru najlepszego modelu predykcyjnego. Jeżeli odbiorca prognozy nie poda własnych kryteriów dopuszczalności prognoz, proponuje się przyjęcie, że jeżeli błąd *MAPE* przyjmuje wartości:

- $MAPE \leq 1\%$ , to prognozy są doskonałe (celujący 6),
- $1\% < MAPE \leq 3\%$ , to prognozy są wysoce dokładne (bardzo dobry 5),
- $3\% < MAPE \leq 5\%$ , to prognozy uznajemy za dobre (dobry 4),
- $5\% < MAPE \leq 10\%$ , to prognozy uznajemy za dostateczne (dostateczny 3),
- $10\% < MAPE \leq 15\%$ , to prognozy są mało dokładne, ale mogą być przyjęte dopuszczalne (dopuszczający 2),
- $MAPE > 15\%$ , to prognozy są niedokładne, nie powinny być przyjęte (niedostateczny 1).

### 3.7. Oznaczenia

#### 3.7.1. Oznaczenia metod i modeli

Wyznaczanie najlepszego równania regresyjnego odbywa się poprzez obliczenia wieloma metodami z wykorzystaniem szeregu modeli obliczeniowych. Celem ujednoczenia wprowadzono system oznaczeń modeli prognostycznych w postaci  $M_i(y; x_1, x_2, \dots, x_k)$ , gdzie:

$M$  – nazwa metody prognostycznej,

$i$  – numer modelu metody  $M$ ,

$y$  – zmienna zależna,

$x_k$  – zmienne niezależne-predyktory,

$k$  – liczba predyktorów.

W tabeli 3.4 Zestawiono oznaczenia nazw metod prognostycznych.

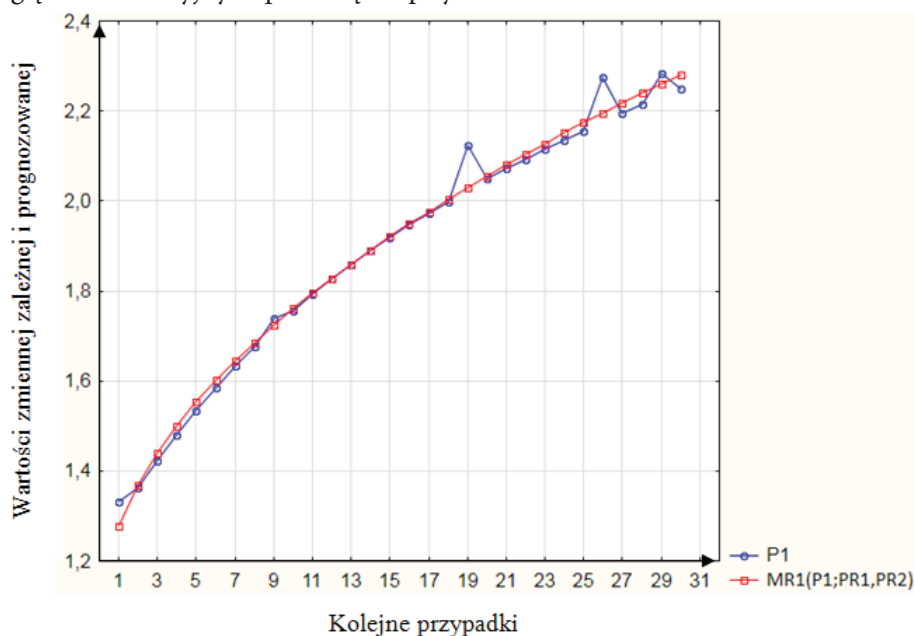
Tab. 3.4. Oznaczenia nazw metod prognostycznych

Lp.	Nazwa skrócona metody	Pełna nazwa metody w języku polskim	Pełna nazwa metody w języku angielskim
1	MR	regresja wieloraka	Multiple Regression
2	MARS	wieloimienna regresja adaptacyjna z użyciem funkcji sklepanych	Multivariate Adaptive Regression Splines
3	GAM	uogólnione modele addytywne	Generalized Additive Methods
4	NN	sztuczne sieci neuronowe	Neural Network
5	SVN	wektorów nośnych	Support Vector Machine
6	ARIMA	autoregresji zintegrowanej	Autoregressive Integrated Average Process

MR1(P1; PR1,PR2) oznaczać będzie: metoda regresji wielorakiej, model 1, zmienna zależna P1, predyktory PR1 i PR2. Szereg resztowy modelu MR1(P1; PR1,PR2) oznaczono jako RMR1.

### 3.7.2. Oznaczenia osi

W toku obliczeń wielokrotnie prezentowane będą wykresy obrazujące wartości zmiennej zależnej i prognozowanej dla kolejnych przypadków (rys. 3.20). Ze względów redakcyjnych pominięto opisy osi.



Rys.3.20 Wykres wartości zmiennej zależnej i wartości prognozowanych dla kolejnych przypadków

## 4. Prognozowanie

Prognozowanie to naukowa metoda przewidywania w jaki sposób będą kształtowały się w przyszłości procesy lub zdarzenia. Prognozowanie rozumiane jest jako racjonalne i zarazem naukowe przewidywanie przyszłych zdarzeń przez: Grenia (1984), Cieślaka (2004), Zeliasia (2003), Pawłowskiego (1982), Dittmanna, Nowaka (1998), Czerwińskiego (1992), Gajdę (2001). Prognozowanie, wg Grenia (1984), czyli predykcja, to „praktyczna działalność zmierzająca do sformułowania naukowo uzasadnionej oceny, tj. szacunku prawdopodobnego stanu w przyszłości na podstawie posiadanych informacji z przeszłości oraz merytorycznej wiedzy o danym wycinku z przeszłości będącym przedmiotem prognozowania.” W trakcie procesu prognozowania formułuje się sąd na temat przyszłych stanów zjawisk i zdarzeń. Sąd ten nazywany jest prognozą. Uzyskane w wyniku przeprowadzonych obliczeń prognozy mogą służyć jako narzędzia planistyczne; w budownictwie mogą stanowić wytyczne do określania czasu trwania procesów budowlanych oraz buforów czasu zapewniających terminową realizację. Proces prognozowania jest postępowaniem wieloetapowym. Metoda prognozowania to sposób przetwarzania informacji opisujących sytuację prognostyczną w prognozę, dostosowany do przyjętej zasady prognozowania. Znanych jest wiele metod prognozowania. Generalnie dobór metody prognozowania uzależniony jest od posiadanych danych oraz zależności występujących między nimi. Wykonując obliczenia regresyjne określoną metodą możemy wygenerować wiele modeli prognostycznych, odpowiednio dobierając i przekształcając zmienne niezależne lub też przyjmując określony rozkład zmiennych (często zdarza się, iż można uznać, że dana zmienna ma rozkład normalny, beta lub gamma). Ocena prawidłowości predykcji, a zatem wyboru najlepszej metody i modelu prognostycznego, odbywa się zgodnie z procedurami opisanymi poniżej.

### 4.1. Prognozowanie metodą regresji wielorakiej

Ogólnym celem regresji (termin ten został po raz pierwszy użyty przez Pearsona w 1908 roku) jest ilościowe ujęcie związków pomiędzy wieloma zmiennymi niezależnymi (objaśniającymi) a zmienną zależną (kryterialną, objaśnianą). Metoda pozwala na badanie związku pomiędzy wielkościami danych i przewidywanie na tej podstawie nieznanymi wartościami jednych wielkości na podstawie znanych wartości innych. Regresja, w której występuje więcej niż jedna zmienna objaśniająca, zwana jest regresją wieloraką (ang. multiple regression, MR).

Użycie regresji w praktyce sprowadza się do dwóch faz (Kornacki, Ćwik (2005)):

- Konstruowanie modelu regresyjnego, czyli funkcji opisującej, jak zależy wartość oczekiwana zmiennej objaśnianej od zmiennych objaśniających. Model konstruuje się tak, aby jak najlepiej pasował do danych z próby.
- Stosowanie modelu (tzw. scoring) – użycie wyliczonego modelu do danych, w których znamy tylko zmienne objaśniające, w celu wyznaczenia wartości oczekiwanej zmiennej objaśnianej.

Najprostszym modelem regresji jest model liniowy, w którym zmienna objaśniana jest funkcją liniową zmiennych objaśniających i składnika losowego (Kot, Jakubowski, Sokołowski (2007)). Model regresji liniowej w zbiorowości generalnej przedstawia równanie (4.1):

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \xi \quad (4.1)$$

gdzie:

$y$  – zmienna zależna (objaśniana),

$x_i$  – zmienna niezależna (objaśniająca),

$\beta_0$  – wyraz wolny regresji,

$\beta_i$  – współczynnik regresji cząstkowej (parametry strukturalne modelu regresji),

$\xi$  – składnik losowy (ksi).

Składnik losowy  $\xi$  powinien być zmienną losową podlegającą rozkładowi normalnemu o wartości przeciętnej równej zero i stałej wariancji (jest homoskedastyczny). Jeśli składnik losowy nie spełnia tych warunków to równanie (4.1) nie jest równaniem regresji. Zatem podczas obliczeń istnieje konieczność sprawdzenia reszt.

Miarą wykorzystywaną do oceny dobroci dopasowania modelu regresyjnego do danych wejściowych jest współczynnik determinacji  $R^2$ . Oznacza on, jaka część zmienności zmiennej objaśnianej została wyjaśniona przez model. Współczynnik determinacji opisuje tę część zmienności objaśnianej, która wynika z jej zależności od uwzględnionych w modelu zmiennych objaśniających. Współczynnik determinacji przyjmuje wartości z przedziału  $[0;1]$ . Dopasowanie modelu jest tym lepsze, im wartość  $R^2$  jest bliższa jedności. Wyraża się on wzorem (4.2):

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (4.2)$$

gdzie:

$R^2$  – współczynnik determinacji,

$y_i$  – zaobserwowana (empiryczna) wartość zmiennej  $y$ ; wartość jaką zmienna  $y$  przyjęła w obiekcie o numerze  $i$ ,

$\hat{y}_i$  – teoretyczna wartość zmiennej  $y$ , obliczona poprzez podstawienie do równania regresji wartości, jakie zmienne objaśniające przyjęły w obiekcie o numerze  $i$ ,

$\bar{y}$  – średnia arytmetyczna wartości zmiennej  $y$  z próby,

$n$  – liczba obserwacji,

$i$  – numer obserwacji.

Współczynnik determinacji jest wrażliwy na wprowadzanie dodatkowych zmiennych niezależnych do równania regresji. Jeżeli do modelu dodajemy nowe zmienne to  $R^2$  ulega zwiększeniu, nawet jeżeli wpływ nowej zmiennej nie jest istotny statystycznie. Celem zneutralizowania tego efektu w modelach regresji wielorakiej zaleca się (Kot, Jakubowski, Sokołowski (2007)) stosowanie skorygowanego współczynnika determinacji  $R_s^2$  (4.3):

$$R_s^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \cdot \frac{n-1}{n-k-1} \quad (4.3)$$

gdzie:

$k$  – liczba zmiennych niezależnych.

Do oceny dobroci dopasowania można wykorzystać pierwiastek kwadratowy współczynnika determinacji. Jest to współczynnik korelacji wielorakiej, który mierzy siłę łącznego wpływu wszystkich zmiennych objaśniających na zmienną objaśnianą.

Najbardziej istotne i jednocześnie decydujące o prawidłowości modelu regresji jest sprawdzenie wartości składnika losowego równania (4.1). Jeśli zostanie wykazane, że składnik losowy nie ma średniej zbliżonej do zera i jednakowej wariancji (nie jest homoskedastyczny), to nie ma podstaw do przyjęcia, że otrzymane równanie jest równaniem regresji. Możliwe jest występowanie takiej sytuacji, że skorygowany współczynnik determinacji  $R_s^2$  jest zbliżony do wartości 1 (jesteśmy wówczas zadowoleni z otrzymanego równania regresji), natomiast okazuje się po wykonaniu obliczeń, że składnik losowy nie jest białym szumem i nie można wówczas uznać modelu za prawidłowy.

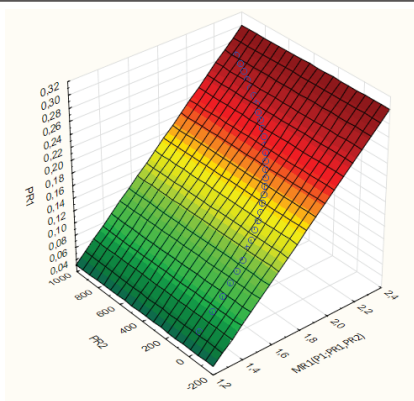
Dla danych z przykładu 3.1 (rys. 3.1) przeprowadzono obliczenia metodą regresji wielorakiej w programie STATISTICA. Wyznaczono zmienną zależną P1, przyjmując jako zmienne niezależne PR1 i PR2. Zmienne PR1 i PR2 są skorelowane liniowo (tab.3.3.) zatem jedna z nich prawdopodobnie zostanie wyeliminowana z równania regresji. Zestawienie obliczeń przedstawiono w tabeli 4.1. Wygenerowany model jest nieprawidłowy ze względu na wartość  $p = 0,933049$  zmiennej PR2. Pomimo, że:

- skorygowany współczynnik determinacji modelu jest równy 0,98916290 (zbliżony do jedności),
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd MAPE = 0,938094% (prognoza doskonała),
- stwierdzono brak autokorelacji i autokorelacji cząstkowych reszt modelu,
- model nie może być przyjęty do stosowania.

Tab. 4.1. Zestawienie obliczeń modelu MR1(P1;PR1,PR2) metodą regresji wielorakiej

METODA OBLICZEŃ O MODEL REGRESYJNY	<b>Regresja wieloraka MR1(P1;PR1,PR2)</b>																												
ZMIENNA ZALEŻNA	P1																												
ZMIENNENIEZALEŻNE	PR1,PR2																												
PODSUMOWANIE REGRESJI	Podsumowanie regresji zmiennej zależnej: P1 (p1pr1pr2) R= ,99494235 R^2= ,98991029 Popraw. R2= ,98916290 F(2,27)=1324,5 p<0,0000 Błąd std. estymacji: ,02966																												
	N=30																												
	<table border="1"> <thead> <tr> <th></th> <th>b*</th> <th>Bł. std. z b*</th> <th>b</th> <th>Bł. std. z b</th> <th>t(27)</th> <th>p</th> </tr> </thead> <tbody> <tr> <td><b>W. wolny</b></td> <td></td> <td></td> <td>1,052987</td> <td>0,027323</td> <td>38,53865</td> <td>0,000000</td> </tr> <tr> <td>PR1</td> <td>0,991218</td> <td>0,047974</td> <td>4,472303</td> <td>0,216455</td> <td>20,66160</td> <td>0,000000</td> </tr> <tr> <td>PR2</td> <td>0,004068</td> <td>0,047974</td> <td>0,000004</td> <td>0,000049</td> <td>0,08480</td> <td>0,933049</td> </tr> </tbody> </table>		b*	Bł. std. z b*	b	Bł. std. z b	t(27)	p	<b>W. wolny</b>			1,052987	0,027323	38,53865	0,000000	PR1	0,991218	0,047974	4,472303	0,216455	20,66160	0,000000	PR2	0,004068	0,047974	0,000004	0,000049	0,08480	0,933049
		b*	Bł. std. z b*	b	Bł. std. z b	t(27)	p																						
<b>W. wolny</b>			1,052987	0,027323	38,53865	0,000000																							
PR1	0,991218	0,047974	4,472303	0,216455	20,66160	0,000000																							
PR2	0,004068	0,047974	0,000004	0,000049	0,08480	0,933049																							
WZÓR REGRESYJNY	$P1 = 1,052987 + 4,472303PR1 + 0,000004PR2$																												
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH																													
BŁĄD MAPE [%]	0,938094																												
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIE PRAWIDŁOWY																												
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT																													

WYKRES  
TRÓJWYMIAROWY  
PREDYKTORÓW  
i WARTOŚCI  
PROGNOZOWANEJ



Potwierdziło się przypuszczenie, że dwie zmienne niezależne skorelowane liniowo nie będą mogły być jednocześnie predyktorami zmiennej zależnej. Należy przeprowadzić obliczenia przyjmując kolejno pojedynczo predyktory, czyli analizować modele gdzie predyktorem będzie zmienna PR1 a następnie zmienna PR2. Wyboru najlepszego modelu prognostycznego dokonuje się na podstawie wielkości błędu *MAPE* (im niższy błąd *MAPE* tym lepszy model predykcyjny).

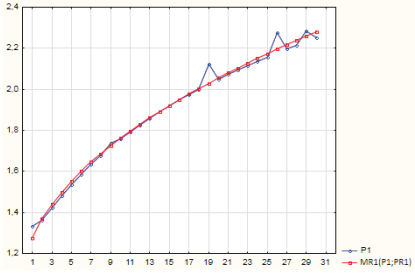
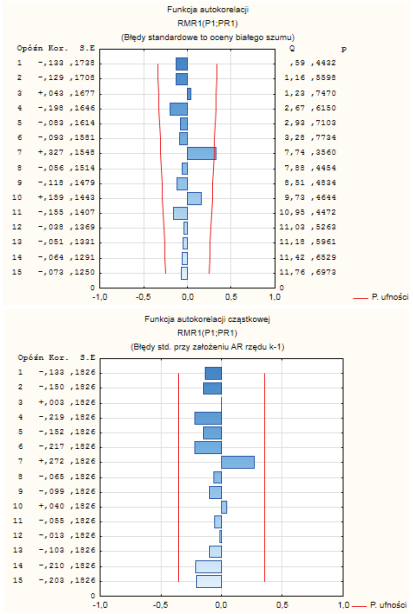
Jako pierwszy do analizy, model z pojedynczym predyktorem przyjęto MR2(P1;PR1). Zestawienie obliczeń przedstawiono w tabeli 4.2. Wygenerowany model jest prawidłowy ze względu na to, że:

- wartość *p* wyrazu wolnego predyktora PR1 zbliżona do zera,
- skorygowany współczynnik determinacji modelu jest równy 0,98954716 (zbliżony do jedności),
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd *MAPE* = 0,943469% (prognoza doskonała),
- stwierdzono brak autokorelacji i autokorelacji cząstkowych reszt modelu,
- model może być przyjęty do stosowania.

Tab. 4.2. Zestawienie obliczeń modelu MR2(P1;PR1) metodą regresji wielorakiej

METODA OBLICZEŃ i MODEL REGRESYJNY	Regresja wieloraka MR2(P1;PR1)						
ZMIENNA ZALEŻNA	P1						
ZMIENNE NIEZALEŻNE	PR1						
PODSUMOWANIE REGRESJI	Podsumowanie regresji zmiennej zależnej: P1 (Arkusz1) R= ,99494100 R <sup>2</sup> = ,98990760 Popraw. R2= ,98954716 F(1,28)=2746,4 p<0,0000 Błąd std. estymacji: ,02913						
	N=30	<b>b*</b>	Bł. std. z b*	<b>b</b>	Bł. std. z b	t(28)	<b>p</b>
	<b>W. wolny</b>			1,051184	0,016862	62,33943	0,000000
	<b>PR1</b>	0,994941	0,018985	4,489101	0,085660	52,40577	0,000000
WZÓR REGRESYJNY	P1= 1,051184 + 4,489101 PR1						



<p>WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH</p>	
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BŁĄD MAPE [%]</p>	<p>0,943469</p>
<p>OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO</p>	<p>PRAWIDŁOWY</p>

Wzór regresyjny  $P1 = 1,051184 + 4,489101 PR1$  modelu  $MR2(P1;PR1)$  otrzymany metodą regresji wielorakiej można uznać za poprawny. Istnieje jednak możliwość, że wykorzystując do predykcji zmienną niezależną  $PR2$  otrzymamy równanie regresji z mniejszym błędem  $MAPE$ . W tabeli 4.3. zestawiono obliczenia modelu  $MR3(P1;PR2)$ .

Tab. 4.3. Zestawienie obliczeń modelu MR3(P1;PR2) metodą regresji wielorakiej

METODA OBLICZEN i MODEL REGRESYJNY	<b>Regresja wieloraka MR3(P1;PR2)</b>						
ZMIENNA ZALEŻNA	P1						
ZMIENNE NIEZALEŻNE	PR2						
PODSUMOWANIE REGRESJI	Podsumowanie regresji zmiennej zależnej: P1 (Arkusz1) R= ,91125194 R^2= ,83038010 Popraw. R2= ,82432225 F(1,28)=137,07 p<.00000 Błąd std. estymacji: ,11941						
	N=30	<b>b*</b>	Bł. std. z b*	b	Bł. std. z b	t(28)	p
	<b>W. wolny</b>			1,590558	0,033592	47,34950	0,000000
	PR2	0,911252	0,077832	0,000923	0,000079	11,70790	0,000000
WZÓR REGRESYJNY	P1 = 1,590558 + 0,000923 PR2						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT							

BŁĄD MAPE [%]	5,494223
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY

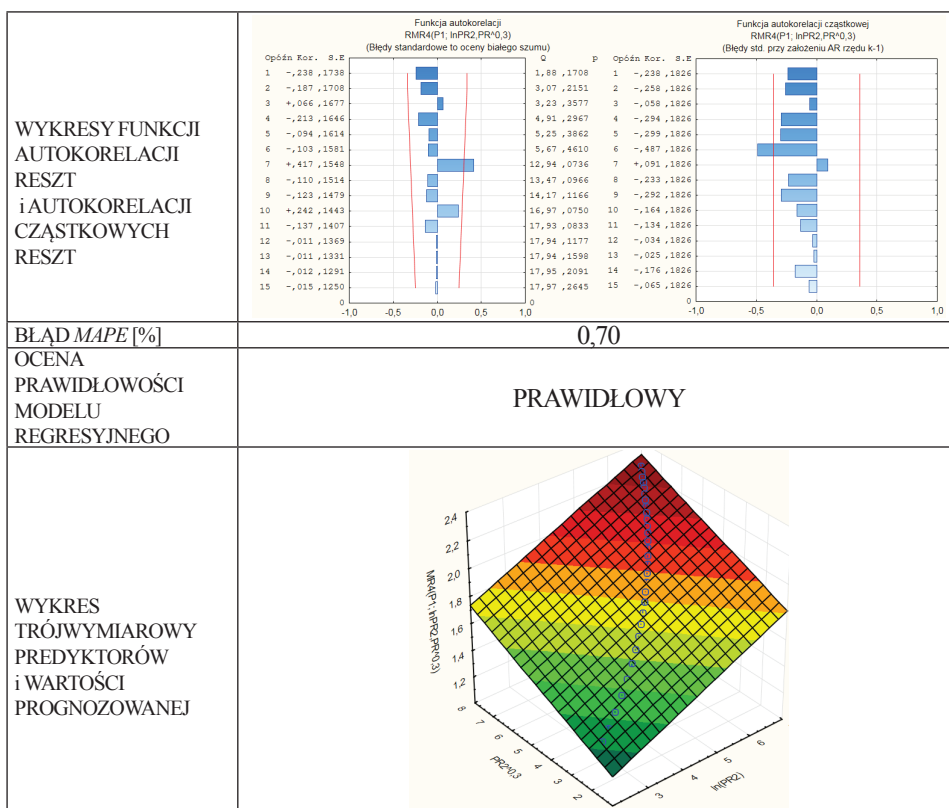
Wygenerowany model jest nieprawidłowy ze względu na to, że:

- skorygowany współczynnik determinacji modelu jest równy 0,82432225,
- wykresy wartości rzeczywistych i prognozowanych nie są zbliżone,
- błąd MAPE = 5,494223% (prognoza dostateczna),
- stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu,
- model nie może być przyjęty do stosowania.

Zmienna niezależna PR2 posiada rozkład gęstości prawdopodobieństwa, który nie jest zbliżony do normalnego. z tego względu podjęto próbę utworzenia nowej zmiennej logarytm naturalny ze zmiennej PR2 i oznaczono ją  $\ln PR2$ . Metodą prób i błędów (wykonano wiele przekształceń zmiennej PR2) wybrano do analizy także zmienną  $PR2^{0,3}$ . Obliczono model regresyjny  $MR4(P1; \ln PR2, PR2^{0,3})$  metodą regresji wielorakiej, wyniki obliczeń przedstawiono w tabeli 4.4.

Tab. 4.4. Zestawienie obliczeń modelu  $MR4(P1; \ln PR2, PR2^{0,3})$  metodą regresji wielorakiej

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>Regresja wieloraka <math>MR4(P1; \ln PR2, PR2^{0,3})</math></b>						
ZMIENNA ZALEŻNA	P1						
ZMIENNE NIEZALEŻNE	$\ln PR2, PR2^{0,3}$						
PODSUMOWANIE REGRESJI	Podsumowanie regresji zmiennej zależnej: P1 R= ,99599701 R <sup>2</sup> = ,99201005 Popraw. R <sup>2</sup> = ,99141820 F(2,27)=1676,1 p<0,0000 Błąd std. estymacji: ,02639						
	N=30	b*	Bł. std. z b*	b	Bł. std. z b	t(27)	p
	W. wolny			0,908430	0,033137	27,41419	0,000000
	$\ln(PR2)$	0,552144	0,097739	0,118368	0,020953	5,64918	0,000005
	$PR2^{0,3}$	0,447720	0,097739	0,072444	0,015815	4,58078	0,000094
WZÓR REGRESYJNY	$P1 = 0,908430 + 0,118368 \ln(PR2) + 0,072444 PR2^{0,3}$						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							



Obliczony model jest prawidłowy ze względu na to, że:

- skorygowany współczynnik determinacji modelu jest równy 0,99141820,
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd  $MAPE = 0,7\%$  (prognoza doskonała),
- stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy 7 i 6 opóźnieniu (wynik zadawalający),
- model może być przyjęty do stosowania.

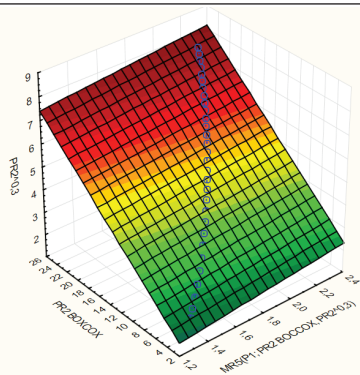
Podjęto kolejną próbę ulepszenia modelu regresyjnego poprzez wprowadzenie do obliczeń zmiennej PR2 przekształconej zgodnie z metodą Boxa-Coxa. Dodatkowo wprowadzono zmienną  $PR2^{0,3}$ , która w modelu  $MR4(P1; \ln PR2, PR2^{0,3})$  w sposób prawidłowy modelowała zmienną zależną P1. Wyniki obliczeń zestawiono w tabeli 4.5.

Tab. 4.5. Zestawienie obliczeń modelu  $MR5(P1; PR2_{\text{BOXCOX}}, PR2^{0,3})$  metodą regresji wielorakiej

<p>METODA OBLICZEŃ i MODEL REGRESYJNY</p>	<p><b>Regresja wieloraka <math>MR5(P1; PR2_{\text{BOXCOX}}, PR2^{0,3})</math></b></p>
<p>ZMIENNA ZALEŻNA</p>	<p>P1</p>
<p>ZMIENNE NIEZALEŻNE</p>	<p><math>PR2_{\text{BOXCOX}}, PR2^{0,3}</math></p>

PODSUMOWANIE REGRESJI	Podsumowanie regresji zmiennej zależnej: P1 R= ,99595363 R <sup>2</sup> = ,99192363 Popraw. R2= ,99132538 F(2,27)=1658,0 p<0,0000 Błąd std. estymacji: ,02653																																																																																																																					
	N=30	b*	Bł. std. z b*	b	Bł. std. z b	t(27)	p																																																																																																															
	W. wolny			-0,498272	0,281787	-1,76826	0,08831																																																																																																															
	PR2 BOXCOX	-8,70267	1,555965	-0,381074	0,068133	-5,59310	0,00000																																																																																																															
PR2 <sup>0,3</sup>	9,69337	1,555965	1,568453	0,251766	6,22982	0,00000																																																																																																																
WZÓR REGRESYJNY	P1 = -0,498272-0,381074 PR2 <sub>BOXCOX</sub> + 1,568453 PR2 <sup>0,3</sup>																																																																																																																					
WYKRES WARTOŚCI RZECZYWISTYCH I PROGNOZOWANYCH																																																																																																																						
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	<p>Funkcja autokorelacji RMR5(P1; PR2 BOXCOX, PR2<sup>0,3</sup>) (Błędy standardowe do oceny białego szumu)</p> <table border="1"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E</th> <th>Q</th> </tr> </thead> <tbody> <tr><td>1</td><td>-,217</td><td>,1738</td><td>1,55 ,2128</td></tr> <tr><td>2</td><td>-,181</td><td>,1708</td><td>2,67 ,2625</td></tr> <tr><td>3</td><td>+,072</td><td>,1677</td><td>2,86 ,4136</td></tr> <tr><td>4</td><td>-,199</td><td>,1646</td><td>4,32 ,3647</td></tr> <tr><td>5</td><td>-,111</td><td>,1614</td><td>4,79 ,4419</td></tr> <tr><td>6</td><td>-,120</td><td>,1581</td><td>5,37 ,4974</td></tr> <tr><td>7</td><td>+,394</td><td>,1548</td><td>11,86 ,1052</td></tr> <tr><td>8</td><td>-,126</td><td>,1514</td><td>12,55 ,1283</td></tr> <tr><td>9</td><td>-,121</td><td>,1479</td><td>13,22 ,1530</td></tr> <tr><td>10</td><td>+,251</td><td>,1443</td><td>16,24 ,0931</td></tr> <tr><td>11</td><td>-,113</td><td>,1407</td><td>16,88 ,1116</td></tr> <tr><td>12</td><td>-,013</td><td>,1369</td><td>16,89 ,1539</td></tr> <tr><td>13</td><td>-,010</td><td>,1331</td><td>16,89 ,2043</td></tr> <tr><td>14</td><td>-,010</td><td>,1291</td><td>16,90 ,2617</td></tr> <tr><td>15</td><td>-,015</td><td>,1250</td><td>16,91 ,3241</td></tr> </tbody> </table> <p>Funkcja autokorelacji cząstkowej RMR5(P1; PR2 BOXCOX, PR2<sup>0,3</sup>) (Błędy std. przy założeniu AR rzędu k-1)</p> <table border="1"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E</th> </tr> </thead> <tbody> <tr><td>1</td><td>-,217</td><td>,1826</td></tr> <tr><td>2</td><td>-,239</td><td>,1826</td></tr> <tr><td>3</td><td>-,031</td><td>,1826</td></tr> <tr><td>4</td><td>-,256</td><td>,1826</td></tr> <tr><td>5</td><td>-,261</td><td>,1826</td></tr> <tr><td>6</td><td>-,421</td><td>,1826</td></tr> <tr><td>7</td><td>+,161</td><td>,1826</td></tr> <tr><td>8</td><td>-,208</td><td>,1826</td></tr> <tr><td>9</td><td>-,228</td><td>,1826</td></tr> <tr><td>10</td><td>-,074</td><td>,1826</td></tr> <tr><td>11</td><td>-,053</td><td>,1826</td></tr> <tr><td>12</td><td>+,005</td><td>,1826</td></tr> <tr><td>13</td><td>-,019</td><td>,1826</td></tr> <tr><td>14</td><td>-,167</td><td>,1826</td></tr> <tr><td>15</td><td>-,011</td><td>,1826</td></tr> </tbody> </table>						Opóźn	Kor.	S.E	Q	1	-,217	,1738	1,55 ,2128	2	-,181	,1708	2,67 ,2625	3	+,072	,1677	2,86 ,4136	4	-,199	,1646	4,32 ,3647	5	-,111	,1614	4,79 ,4419	6	-,120	,1581	5,37 ,4974	7	+,394	,1548	11,86 ,1052	8	-,126	,1514	12,55 ,1283	9	-,121	,1479	13,22 ,1530	10	+,251	,1443	16,24 ,0931	11	-,113	,1407	16,88 ,1116	12	-,013	,1369	16,89 ,1539	13	-,010	,1331	16,89 ,2043	14	-,010	,1291	16,90 ,2617	15	-,015	,1250	16,91 ,3241	Opóźn	Kor.	S.E	1	-,217	,1826	2	-,239	,1826	3	-,031	,1826	4	-,256	,1826	5	-,261	,1826	6	-,421	,1826	7	+,161	,1826	8	-,208	,1826	9	-,228	,1826	10	-,074	,1826	11	-,053	,1826	12	+,005	,1826	13	-,019	,1826	14	-,167	,1826	15	-,011	,1826
Opóźn	Kor.	S.E	Q																																																																																																																			
1	-,217	,1738	1,55 ,2128																																																																																																																			
2	-,181	,1708	2,67 ,2625																																																																																																																			
3	+,072	,1677	2,86 ,4136																																																																																																																			
4	-,199	,1646	4,32 ,3647																																																																																																																			
5	-,111	,1614	4,79 ,4419																																																																																																																			
6	-,120	,1581	5,37 ,4974																																																																																																																			
7	+,394	,1548	11,86 ,1052																																																																																																																			
8	-,126	,1514	12,55 ,1283																																																																																																																			
9	-,121	,1479	13,22 ,1530																																																																																																																			
10	+,251	,1443	16,24 ,0931																																																																																																																			
11	-,113	,1407	16,88 ,1116																																																																																																																			
12	-,013	,1369	16,89 ,1539																																																																																																																			
13	-,010	,1331	16,89 ,2043																																																																																																																			
14	-,010	,1291	16,90 ,2617																																																																																																																			
15	-,015	,1250	16,91 ,3241																																																																																																																			
Opóźn	Kor.	S.E																																																																																																																				
1	-,217	,1826																																																																																																																				
2	-,239	,1826																																																																																																																				
3	-,031	,1826																																																																																																																				
4	-,256	,1826																																																																																																																				
5	-,261	,1826																																																																																																																				
6	-,421	,1826																																																																																																																				
7	+,161	,1826																																																																																																																				
8	-,208	,1826																																																																																																																				
9	-,228	,1826																																																																																																																				
10	-,074	,1826																																																																																																																				
11	-,053	,1826																																																																																																																				
12	+,005	,1826																																																																																																																				
13	-,019	,1826																																																																																																																				
14	-,167	,1826																																																																																																																				
15	-,011	,1826																																																																																																																				
BLĄD MAPE [%]	0,77666																																																																																																																					
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	PRAWIDŁOWY																																																																																																																					

WYKRES  
TRÓJWYMIAROWY  
PREDYKTORÓW  
i WARTOŚCI  
PROGNOZOWANEJ



- Obliczony model jest prawidłowy ze względu na to, że:
- skorygowany współczynnik determinacji modelu jest równy 0,99132538,
  - wykresy wartości rzeczywistych i prognozowanych są zbliżone,
  - błąd  $MAPE = 0,77666\%$  (prognoza doskonała),
  - stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy 7 i 6 opóźnieniu (wynik zadawalający),
  - model może być przyjęty do stosowania.

Metoda regresji wielorakiej polega na obliczaniu wielu modeli i poszukiwaniu modelu, który w sposób poprawny statystycznie wyznaczy krzywą regresji przy najmniejszym błędzie  $MAPE$ . Otrzymane wyniki obliczeń metodą regresji wielorakiej zestawiono w tabeli 4.6.

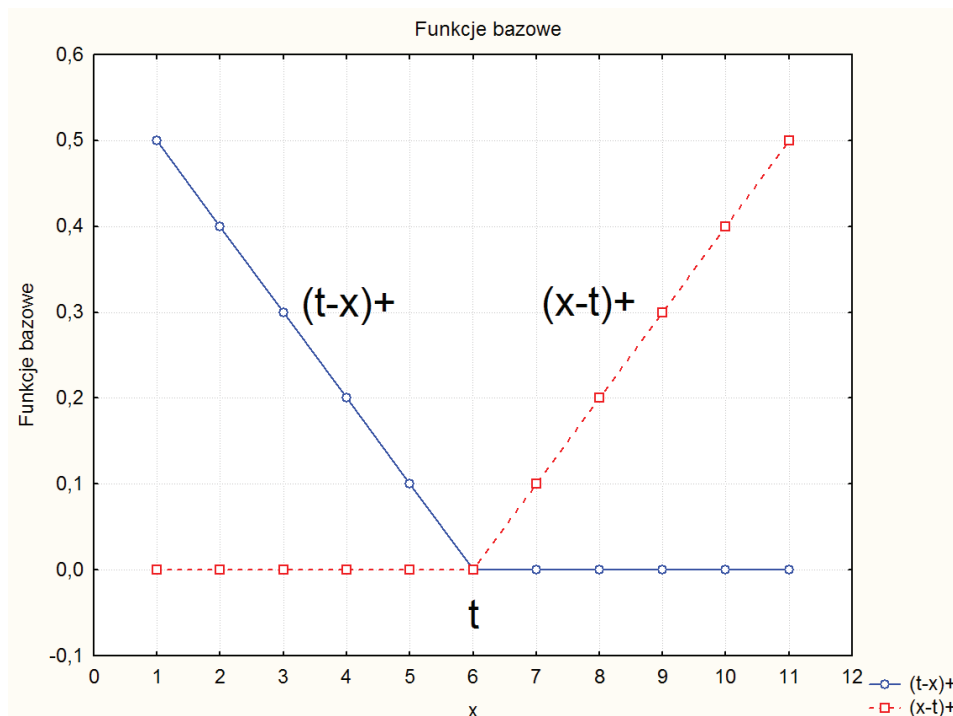
Tab. 4.6. Zestawienie przeprowadzonych obliczeń metodą regresji wielorakiej

LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszt	Wartość błędu $MAPE$ [%]	Poprawność modelu
1	MR1(P1;PR1,PR2)	$P1 = 1,052987 + 4,472303PR1 + 0,000004PR2$	NIE	0,938094	NIE
2	MR2(P1;PR1)	$P1 = 1,051184 + 4,489101 PR1$	NIE	0,943469	TAK
3	MR3(P1;PR2)	$P1 = 1,590558 + 0,000923 PR2$	TAK	5,494223	NIE
4	MR4 (P1;lnPR2,PR2^0,3)	$P1 = MR4(P1;lnPR2,PR2^0,3)$	NIE	0,700012	TAK
5	MR5 (P1; PR2 <sub>BOXCOX</sub> PR2^0,3)	$P1 = -0,498272 - 0,381074 PR2_{BOXCOX} + 1,568453 PR2^0,3$	NIE	0,776666	TAK

Najlepszym z obliczonych modeli metodą regresji wielorakiej jest model MR4(P1; lnPR2,PR2^0,3). Spośród trzech poprawnych modeli posiada najmniejszy błąd  $MAPE$  równy 0,700012%.

## 4.2. Prognozowanie metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklejanych

Multivariate Adaptive Regression Splines (MARSplines) w polskim tłumaczeniu znana jest jako wieloimienna regresja adaptacyjna z użyciem funkcji sklejanych (Fijorek (2010)) lub też składanych (Sokołowski (2004)). Metoda została opracowana przez Friedmana (1991). Wykorzystywana jest do rozwiązywania problemów regresyjnych i klasyfikacyjnych, w których celem jest znalezienie wartości zmiennych wyjściowych na podstawie zmiennych wejściowych (zależnych, predykcyjnych).



Rys.4.1. Dwie funkcje bazowe  $(t-x)_+$  i  $(x-t)_+$  z parametrem  $t$  – węzłem funkcji bazowej

W metodzie nie istnieje konieczność przyjmowania założeń dotyczących funkcjonalnych zależności pomiędzy zmiennymi zależnymi i niezależnymi. Na podstawie danych MARSplines generuje zbiór współczynników i funkcji bazowych, a następnie modeluje zależności regresyjne lub klasyfikacyjne. Przestrzeń wejściowa dzielona jest na obszary, w których określone są funkcje. Analizowany jest wpływ predyktorów oraz obserwacje danej zmiennej objaśniającej a obszar zmienności cechy dzielony jest na przedziały, w których ma ona różny wpływ na badane zjawisko. Jak opisuje Fijorek (2009) ogólny mechanizm działania MARSplines wyobrazić można sobie jako wie-

lokrotną, odcinkową regresję liniową; granice odcinków (wyznaczone na podstawie danych) określają „zakresy stosowalności” poszczególnych równań liniowych. Równania te tworzą krzywą składaną czyli zbiór zdefiniowanych lokalnie wielomianów stopnia pierwszego, które łączą się ze sobą w punktach zwanych węzłami. Funkcje te zwane są funkcjami bazowymi i służą do wyznaczania prognozy wartości zmiennej zależnej w funkcji zmiennych niezależnych. w metodzie istnieje możliwość uwzględnienia interakcji między zmiennymi. Opisywana jest ona przy pomocy iloczynu funkcji bazowych.

Przykładowe funkcje bazowe przedstawiono na rysunku 4.1 i opisano zależnością (4.4). Znaczniki „+” za wyrażeniami  $(t-x)$  i  $(x-t)$  oznaczają, że do obliczeń używane są tylko wartości dodatnie funkcji. Zamiast wartości ujemnych do obliczeń przyjmuje się wartość zero. Alternatywnie stosuje się zapis  $\max(0, t-x) = (t-x)_+$

$$(x-t)_+ = \max(0, x-t) = \begin{cases} x-t, & \text{dla } x > t \\ 0, & \text{dla } x \leq t \end{cases} \quad (4.4)$$

Ogólne równanie MARSplines to ważona suma wybranych funkcji bazowych oraz iloczynów tych funkcji (4.5). Parametry modelu znajdowane są za pomocą metody najmniejszych kwadratów.

$$y = f(x) = \beta_0 + \sum_{m=1}^M \beta_m h_m(x) \quad (4.5)$$

gdzie:

$$h_m(x) = \prod_{k=1}^q h_k(x, t) \quad (4.6)$$

$$h_k(x, t) = [t-x]_+^k \quad (4.7)$$

gdzie:

$x$  – zmienne predykcyjne,  
 $y$  – funkcja zmiennych predykcyjnych,

$\beta_0$  – wyraz wolny,  
 $M$  – liczba składników modelu,

$\beta_m$  – wagi funkcji bazowych,  
 $h_m(x)$  – funkcje bazowe,  
 $t_m^k$  – węzeł funkcji bazowej.



Przykładowe równanie regresji (4.8) otrzymane metodą MARSplines przedstawiono poniżej.

$$\begin{aligned}
 & \text{liczba pracowników w sektorze budowlanym} = 2,33561973707142e+004 \\
 & + 5,53065236560456e+002 * \max(0; 5,60000000000000e+001 - t) + 2,11982648739131e- \\
 & + 000 * \max(0; t^2 - 8,28100000000000e+003) - 8,61862212627501e-001 * \max(0; 8,281000- \\
 & 00000000e+003 - t^2) - 1,36183439099150e-001 * \max(0; 5,60000000000000e+001 - t) * \max(0; \\
 & t^2 - 3,24000000000000e+002) - 1,47469387667582e-001 * \max(0; 5,60000000000000e+001 - \\
 & - t) * \max(0; 3,24000000000000e+002 - t^2) - 5,45871124133420e-001 * \max(0; t^2 - \\
 & - 8,28100000000000e+003) * \max(0; \text{grudzień} + 1,00000000000000e+000) + 1,71780365238847e- \\
 & - 001 * \max(0; 8,28100000000000e+003 - t^2) * \max(0; \text{październik} + 1,00000000000000e+000) \\
 & - 3,28177758840381e+001 * \max(0; 5,60000000000000e+001 - t) * \max(0; \\
 & \text{styczeń} + 1,00000000000000e+000) + 1,80791590829487e-001 * \max(0; \\
 & 8,28100000000000e+003 - t^2) * \max(0; \text{listopad} + 1,00000000000000e+000) \\
 & - 1,17065832322996e-001 * \max(0; 8,28100000000000e+003 - t^2) * \max(0; \text{luty} - \\
 & + 1,00000000000000e+000) + 6,05559143250795e-001 * \max(0; t^2 - 8,28100000000000e- \\
 & + 003) * \max(0; \text{styczeń} + 1,00000000000000e+000) - 2,65065401359613e+001 * \max(0; t - \\
 & - 5,60000000000000e+001) * \max(0; \text{listopad} + 1,00000000000000e+000) \quad (4.8)
 \end{aligned}$$

W równaniu występują sumy funkcji bazowych oraz ich iloczyny (oznaczenie „\*” to iloczyn arytmetyczny zgodnie z notacją systemową programu STATISTICA i Excel), które odwzorowują interakcje pomiędzy funkcjami bazowymi. Równanie (4.8) można przenieść do arkusza kalkulacyjnego programu STATISTICA jako zmienną, dokonywać obliczeń dla nowych zmiennych bezpośrednio z arkusza (obliczenia dla nowych zmiennych w MARSplines wymagają wpisywania danych dla pojedynczych zmiennych, przy dużej liczbie predyktorów jest to bardzo pracochłonne).

MARSplines posiada własną miarę dobroci dopasowania, która uwzględni błąd resztowy i poziom skomplikowania modelu. Miara ta zwana jest błędem uogólnionego sprawdzianu krzyżowego GCV (Generalized Cross Validation) i dana jest zależnością (4.9).

$$GCV = \frac{\sum_{i=1}^N (y_i - f(x_i))^2}{\left(1 - \frac{C}{N}\right)^2}, \quad C = 1 + d \quad (4.9)$$

gdzie:

$C$  – parametr sterujący karą za dodanie kolejnej funkcji bazowej,  
 $d$  – efektywna liczba stopni swobody, równa liczbie niezależnych funkcji

bazowych najlepsze  $c$  otrzymuje się dla  $2 < d < 3$  (Hastie (2001)),

$N$  – liczba przypadków w danych,

$y_i$  – wartość rzeczywista przypadku  $i$ ,

$f(x_i)$  – wartość prognozy dla przypadku  $i$ .

Błąd uogólnionego sprawdzianu krzyżowego GCV nie jest miarą bezwzględną. Zależy od wartości liczbowych prognozowanej zmiennej i jej wartości rzeczywistej, dodatkowo jest tym większy im więcej przypadków jest branych pod uwagę. Proponuje się używanie błędu GCV jedynie do porównywania dobroci dopasowania poszczególnych modeli MARSplines. Celem sprawdzenia błędów ex ante i ex post proponuje się stosowanie tradycyjnych mierników np. *MAPE*.

Metoda obliczeniowa MARSpline wykorzystywana była przez wielu naukowców podczas realizacji zadań badawczych: Nisbet, Elder, Miner (2009), Kusiak, Li, Zhang (2010), Nisbet, Elder, Miner (2009), Vidoli (2011), Breiman (1991), Friedman (1991), Lewis, Stevens (1991), Stokes (2008), Guan (2009), Han (2009), Ligas (2015), Serinaldi (2011).

Poniżej przedstawiono obliczenia metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklejanych dla danych z przykładu 3.1. Podobnie jak w przypadku regresji wielorakiej obliczenia rozpoczęto od przyjęcia dwóch zmiennych niezależnych PR1 i PR2 do wyznaczenia zmiennej zależnej P1. Zestawienie obliczeń przedstawiono w tabeli 4.7. Model wyeliminował zmienną niezależną PR2 (jako skorelowaną z PR1).

Tab. 4.7. Zestawienie obliczeń modelu MARS1(P1;PR1) metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklejanych

METODA OBLICZEŃ i MODEL REGRESYJNY	MARS1 (P1;PR1)
ZMIENNA ZALEŻNA	P1
ZMIENNE NIEZALEŻNE	PR1, PR2
PREDYKTORY	PR1
PODSUMOWANIE REGRESJI	<p>Wyniki MARSplines:</p> <p>Zależna: P1</p> <p>Niezależne: PR1, PR2</p> <p>Liczba czynników = 3</p> <p>Liczba funkcji bazowych = 2</p> <p>Rząd interakcji = 1</p> <p>Kara = 2,000000</p> <p>Próg = 0,000500</p> <p>Błąd (GCV) = 0,001140</p> <p>Usuwanie = Tak</p>
WZÓR REGRESYJNY	$P1 = 1,92074667692681e+000 + 4,48278744930603e+000\max(0; PR1 - 1,93649167310371e-001) - 4,49281728608202e+000\max(0; 1,93649167310371e-001 - PR1)$
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	

<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BLĄD MAPE [%]</p>	<p>0,945105</p>
<p>OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO</p>	<p>PRAWIDŁOWY</p>

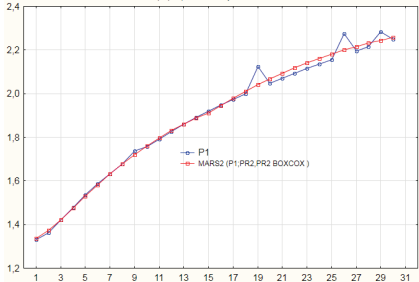
Obliczony model jest prawidłowy ze względu na to, że:

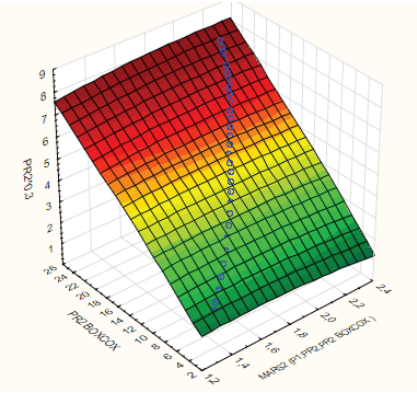
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd MAPE = 0,945105% (prognoza doskonała),
- nie stwierdzono występowania autokorelacji i autokorelacji cząstkowych reszt modelu,
- model może być przyjęty do stosowania.

Celem sprawdzenia czy model MARS1 (P1;PR1) można ulepszyć wykonano obliczenia uwzględniając kolejne zmienne niezależne, będące funkcjami zmiennych PR1 i PR2 (tab.4.8.).

Tab. 4.8. Zestawienie obliczeń modelu MARS2(P1;PR2, PR2<sub>BOXCOX</sub>) metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklejanych

<p>METODA OBLICZEŃ i MODEL REGRESYJNY</p>	<p><b>MARS2 (P1;PR2,PR2<sub>BOXCOX</sub>)</b></p>
<p>ZMIENNA ZALĘŻNA</p>	<p>P1</p>
<p>ZMIENNE NIEZALĘŻNE</p>	<p>PR1, PR2, PR2<sup>2</sup>, PR2<sub>BOXCOX</sub>, ln(PR2), PR2<sup>3</sup>, PR2<sup>0,5</sup>, PR2<sup>0,3</sup></p>
<p>PREDYKTORY</p>	<p>PR2, PR2<sub>BOXCOX</sub></p>

<p>PODSUMOWANIE REGRESJI</p>	<p>Współczynniki modelu (p1pr1pr2 startowy) Uwaga: Podświetlenie komórki wskazuje na funkcję bazową typu max(0, niezal.-węzeł), niepodśw.: max(0, węzeł-niezależna)</p> <table border="1" data-bbox="405 227 1114 336"> <thead> <tr> <th>Współczyn., węzły i funkcje bazowe</th> <th>Współcz. P1</th> <th>Węzły PR1</th> <th>Węzły PR2</th> <th>Węzły PR2^2</th> <th>Węzły PR2 BOXCOX</th> <th>Węzły ln(PR2)</th> <th>Węzły PR2^3</th> <th>Węzły PR2^0,5</th> <th>Węzły PR2^0,3</th> </tr> </thead> <tbody> <tr> <td>W. wolny</td> <td>1,127274</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Wsp.1</td> <td>-0,000482</td> <td></td> <td>234,0000</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Wsp.2</td> <td>0,000936</td> <td></td> <td>234,0000</td> <td></td> <td></td> <td></td> <td></td> <td></td> <td></td> </tr> <tr> <td>Wsp.3</td> <td>0,068931</td> <td></td> <td></td> <td></td> <td>3,403751</td> <td></td> <td></td> <td></td> <td></td> </tr> </tbody> </table>	Współczyn., węzły i funkcje bazowe	Współcz. P1	Węzły PR1	Węzły PR2	Węzły PR2^2	Węzły PR2 BOXCOX	Węzły ln(PR2)	Węzły PR2^3	Węzły PR2^0,5	Węzły PR2^0,3	W. wolny	1,127274									Wsp.1	-0,000482		234,0000							Wsp.2	0,000936		234,0000							Wsp.3	0,068931				3,403751																																																																																		
Współczyn., węzły i funkcje bazowe	Współcz. P1	Węzły PR1	Węzły PR2	Węzły PR2^2	Węzły PR2 BOXCOX	Węzły ln(PR2)	Węzły PR2^3	Węzły PR2^0,5	Węzły PR2^0,3																																																																																																																								
W. wolny	1,127274																																																																																																																																
Wsp.1	-0,000482		234,0000																																																																																																																														
Wsp.2	0,000936		234,0000																																																																																																																														
Wsp.3	0,068931				3,403751																																																																																																																												
<p>WZÓR REGRESYJNY</p>	$P1 = 1,12727391702521e+000 - 4,81924619348619e-004 * \max(0; PR2 - 2,34000000000000e+002) + 9,35605496631056e-004 * \max(0; 2,34000000000000e+002 - PR2) + 6,89306253447718e-002 * \max(0; PR2_{\text{BOXCOX}} - 3,40375079961819e+000)$																																																																																																																																
<p>WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH</p>																																																																																																																																	
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	<p>Funkcja autokorelacji RMARS2 (P1, PR2, PR2 BOXCOX) (Błędy standardowe to oceny białego szumu)</p> <table border="1" data-bbox="553 846 966 1173"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E</th> <th>Q</th> <th>P</th> </tr> </thead> <tbody> <tr><td>1</td><td>-,181</td><td>,1738</td><td>1,08</td><td>,2977</td></tr> <tr><td>2</td><td>-,171</td><td>,1708</td><td>2,09</td><td>,3524</td></tr> <tr><td>3</td><td>+,098</td><td>,1677</td><td>2,43</td><td>,4886</td></tr> <tr><td>4</td><td>-,128</td><td>,1646</td><td>3,03</td><td>,5526</td></tr> <tr><td>5</td><td>-,143</td><td>,1614</td><td>3,81</td><td>,5768</td></tr> <tr><td>6</td><td>-,171</td><td>,1581</td><td>4,98</td><td>,5463</td></tr> <tr><td>7</td><td>+,321</td><td>,1548</td><td>9,28</td><td>,2333</td></tr> <tr><td>8</td><td>-,180</td><td>,1514</td><td>10,70</td><td>,2194</td></tr> <tr><td>9</td><td>-,143</td><td>,1479</td><td>11,63</td><td>,2349</td></tr> <tr><td>10</td><td>+,269</td><td>,1443</td><td>15,10</td><td>,1284</td></tr> <tr><td>11</td><td>-,023</td><td>,1407</td><td>15,13</td><td>,1767</td></tr> <tr><td>12</td><td>-,008</td><td>,1369</td><td>15,13</td><td>,2343</td></tr> <tr><td>13</td><td>-,012</td><td>,1331</td><td>15,14</td><td>,2986</td></tr> <tr><td>14</td><td>-,001</td><td>,1291</td><td>15,14</td><td>,3686</td></tr> <tr><td>15</td><td>-,019</td><td>,1250</td><td>15,16</td><td>,4397</td></tr> </tbody> </table> <p>Funkcja autokorelacji cząstkowej RMARS2 (P1, PR2, PR2 BOXCOX) (Błędy std. przy założeniu AR rzędu k-1)</p> <table border="1" data-bbox="553 1219 966 1547"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E</th> </tr> </thead> <tbody> <tr><td>1</td><td>-,181</td><td>,1826</td></tr> <tr><td>2</td><td>-,211</td><td>,1826</td></tr> <tr><td>3</td><td>+,024</td><td>,1826</td></tr> <tr><td>4</td><td>-,149</td><td>,1826</td></tr> <tr><td>5</td><td>-,193</td><td>,1826</td></tr> <tr><td>6</td><td>-,342</td><td>,1826</td></tr> <tr><td>7</td><td>+,175</td><td>,1826</td></tr> <tr><td>8</td><td>-,239</td><td>,1826</td></tr> <tr><td>9</td><td>-,218</td><td>,1826</td></tr> <tr><td>10</td><td>-,026</td><td>,1826</td></tr> <tr><td>11</td><td>-,026</td><td>,1826</td></tr> <tr><td>12</td><td>-,009</td><td>,1826</td></tr> <tr><td>13</td><td>-,053</td><td>,1826</td></tr> <tr><td>14</td><td>-,194</td><td>,1826</td></tr> <tr><td>15</td><td>+,008</td><td>,1826</td></tr> </tbody> </table>	Opóźn	Kor.	S.E	Q	P	1	-,181	,1738	1,08	,2977	2	-,171	,1708	2,09	,3524	3	+,098	,1677	2,43	,4886	4	-,128	,1646	3,03	,5526	5	-,143	,1614	3,81	,5768	6	-,171	,1581	4,98	,5463	7	+,321	,1548	9,28	,2333	8	-,180	,1514	10,70	,2194	9	-,143	,1479	11,63	,2349	10	+,269	,1443	15,10	,1284	11	-,023	,1407	15,13	,1767	12	-,008	,1369	15,13	,2343	13	-,012	,1331	15,14	,2986	14	-,001	,1291	15,14	,3686	15	-,019	,1250	15,16	,4397	Opóźn	Kor.	S.E	1	-,181	,1826	2	-,211	,1826	3	+,024	,1826	4	-,149	,1826	5	-,193	,1826	6	-,342	,1826	7	+,175	,1826	8	-,239	,1826	9	-,218	,1826	10	-,026	,1826	11	-,026	,1826	12	-,009	,1826	13	-,053	,1826	14	-,194	,1826	15	+,008	,1826
Opóźn	Kor.	S.E	Q	P																																																																																																																													
1	-,181	,1738	1,08	,2977																																																																																																																													
2	-,171	,1708	2,09	,3524																																																																																																																													
3	+,098	,1677	2,43	,4886																																																																																																																													
4	-,128	,1646	3,03	,5526																																																																																																																													
5	-,143	,1614	3,81	,5768																																																																																																																													
6	-,171	,1581	4,98	,5463																																																																																																																													
7	+,321	,1548	9,28	,2333																																																																																																																													
8	-,180	,1514	10,70	,2194																																																																																																																													
9	-,143	,1479	11,63	,2349																																																																																																																													
10	+,269	,1443	15,10	,1284																																																																																																																													
11	-,023	,1407	15,13	,1767																																																																																																																													
12	-,008	,1369	15,13	,2343																																																																																																																													
13	-,012	,1331	15,14	,2986																																																																																																																													
14	-,001	,1291	15,14	,3686																																																																																																																													
15	-,019	,1250	15,16	,4397																																																																																																																													
Opóźn	Kor.	S.E																																																																																																																															
1	-,181	,1826																																																																																																																															
2	-,211	,1826																																																																																																																															
3	+,024	,1826																																																																																																																															
4	-,149	,1826																																																																																																																															
5	-,193	,1826																																																																																																																															
6	-,342	,1826																																																																																																																															
7	+,175	,1826																																																																																																																															
8	-,239	,1826																																																																																																																															
9	-,218	,1826																																																																																																																															
10	-,026	,1826																																																																																																																															
11	-,026	,1826																																																																																																																															
12	-,009	,1826																																																																																																																															
13	-,053	,1826																																																																																																																															
14	-,194	,1826																																																																																																																															
15	+,008	,1826																																																																																																																															
<p>BŁĄD MAPE [%]</p>	<p>0,7742</p>																																																																																																																																

OCENA PRAWDŹLIWOŚCI MODELU REGRESYJNEGO	PRAWDŹLIWY
WYKRES TRÓJWYMIAROWY PREDYKTORÓW i WARTOŚCI PROGNOZOWANEJ	

- Obliczony model MARS2 (P1;PR2,PR2<sub>BOXCOX</sub>) jest prawidłowy ze względu na to, że:
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
  - błąd  $MAPE = 0,7742\%$  (prognoza doskonała),
  - nie stwierdzono występowania autokorelacji i autokorelacji cząstkowych reszt modelu.

Model może być przyjęty do stosowania. Otrzymane wyniki przeprowadzonych obliczeń metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych zestawiono w tabeli 4.9.

Tab. 4.9. Zestawienie przeprowadzonych obliczeń metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych

LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszt	Wartość błędu $MAPE$ [%]	Poprawność modelu
1	MARS1(P1;PR1)	$P1 = 1,92074667692681e+000 + 4,48278744930603e + 000\max(0; PR1 - 1,93649167310371e-001) - 4,49281728608202e+000\max(0; 1,93649167310371e-001-PR1)$	NIE	0,945105	TAK
2	MARS2(P1;PR2, PR2 <sub>BOXCOX</sub> )	$P1 = 1,12727391702521e+000 - 4,81924619348619e-004*\max(0; PR2-2,34000000000000e+002) + 9,35605496631056e-004*\max(0; 2,34000000000000e+002-PR2) + 6,89306253447718e-002*\max(0; PR2_{BOXCOX} - 3,40375079961819e+000)$	NIE	0,774200	TAK

Lepszym z policzonych modeli metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych jest model MARS2(P1;PR2, PR2<sub>BOXCox</sub>). Oba policzone modele są poprawne, jednak w drugim przypadku błąd *MAPE* jest mniejszy i wynosi 0,774200%.

Metoda MARSplines automatycznie eliminuje zmienne niezależne z równania regresji, co jest bardzo istotnym ułatwieniem obliczeniowym. Jest jednak wrażliwa na skorelowanie zmiennych niezależnych oraz rozkład danych odbiegający od normalnego.

### 4.3. Prognozowanie metodą uogólnionych modeli addytywnych

Uogólnione modele addytywne GAM (Generalized Additive Methods) są rozwinięciem i rozszerzeniem uogólnionych modeli liniowych (Hasti, Tibshirani (1990), Schimek (2000)). W metodzie GAM wartości przekształconej zmiennej zależnej przewidywane są na podstawie liniowej kombinacji zmiennych niezależnych (predyktorów). Przekształcenie jest określone przez funkcję wiążącą. Stosuje się nieparametryczną funkcję uzyskaną przez zastosowanie wygładzania do wykresu rozrzutu reszt cząstkowych (dla przekształconych wartości zmiennej zależnej).

W uogólnionym modelu addytywnym zakłada się, że związek pomiędzy zmiennymi ma postać (4.10):

$$y = g(b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k) + e \quad (4.10)$$

gdzie:  $e$  oznacza błąd,  $g(\dots)$  jest pewną funkcją. Formalnie, funkcja odwrotna do  $g(\dots)$  oznaczona  $f(\dots)$  jest nazywana funkcją wiążącą (4.11), taką, że:

$$f(mi_y) = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k \quad (4.11)$$

gdzie:  $mi_y$  oznacza wartość oczekiwaną  $y$ .

W zależności od zakładanego rozkładu zmiennej  $y$  można użyć różnych funkcji wiążących (McCullagh, Nelder (1989)). Najprostszymi funkcjami wiążącymi są funkcje: identycznościowa, logarytmiczna i potęgowa. Metoda GAM w analizach procesów budowlanych wykazuje dużą skuteczność obliczeniową.

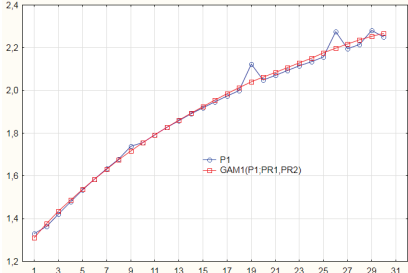
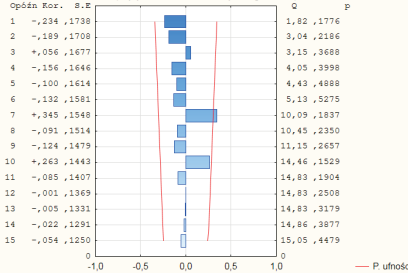
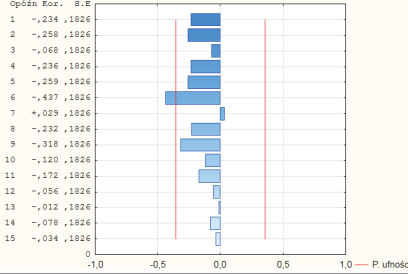
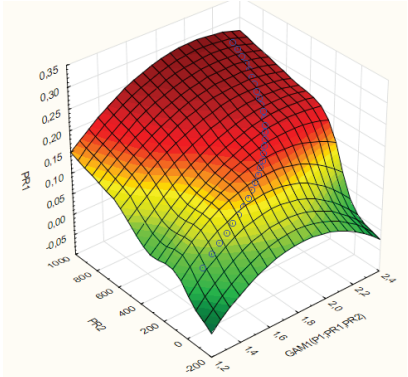
Stosowana była w pracach z różnych dziedzin nauki: Belusic (2015), Breg (2007), Coussemnet (2010), Dominici (2002), Etienne (2009), Song (2010), Tsujitani (2013), Underwood (2009), Wang (2010).

W tabeli 4.10 przedstawiono obliczenia metodą uogólnionych modeli addytywnych GAM dla danych z przykładu 3.1. Podobnie jak w przypadku regresji wielorakiej i wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych, obliczenia rozpoczęto od przyjęcia dwóch zmiennych niezależnych PR1 i PR2 do wyznaczenia zmiennej zależnej P1. Model nie wyeliminował zmiennej niezależnej PR2 pomimo jej skorelowania liniowego z PR1, jak to się działo w metodach MR i MARS. W modelu GAM1(P1; PR1,PR2) założono, że rozkład prognozowa-

nej zmiennej jest gamma, a funkcja wiążąca logarytmiczna. W punkcie 3.3. analizowano rozkłady zmiennych – zmienne P1 i PR1 posiadały rozkłady normalne, a PR2 rozkład inny niż normalny. Zatem prognozowana zmienna teoretycznie może mieć rozkład gamma lub normalny. Dlatego obliczono dwa modele GAM1 (tab. 4.10) i GAM2 (tab.4.11) W poszukiwaniu równania regresji, wykazującego mniejszy błąd MAPE, gdzie GAM2(P1; PR1,PR2) ma rozkład prognozowanej zmiennej normalny i funkcję wiążącą identycznościową.

Tab. 4.10. Zestawienie obliczeń modelu GAM1(P1; PR1,PR2) metodą uogólnionych modeli addytywnych, funkcja wiążąca logarytmiczna, rozkład gamma

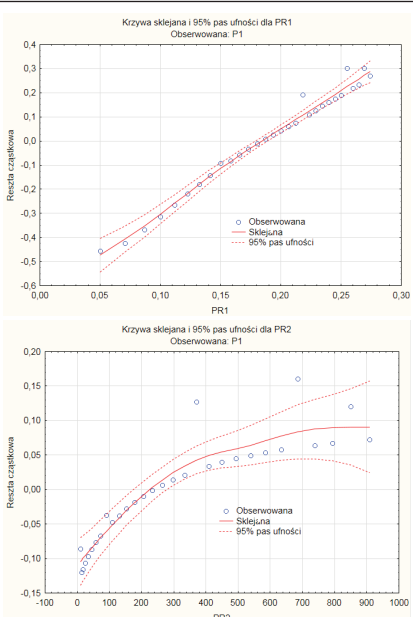
METODA OBLICZEŃ i MODEL REGRESYJNY	<b>GAM1 (P1;PR1, PR2)</b>						
ZMIENNA ZALEŻNA	P1						
ZMIENNE NIEZALEŻNE	PR1, PR2						
PREDYKTORY	PR1,PR2						
PODSUMOWANIE REGRESJI	Podsumowanie dopasowania (p1pr1pr2) Obserwowana: P1 Rozkład: Gamma; Funkcja wiążąca: Log						
		Indeks zmiennej	Stopnie swobody	GAM współcz.	Błąd standardowy	Standaryz. wynik	Nieliniowe p-wartość
	w.wolny	0	1,000000	0,123414	0,013304	9,27663	
	PR1	1	3,997435	2,827793	0,105394	26,83063	0,591318
PR2	2	4,004421	-0,000083	0,000024	-3,50572	0,820717	
WZÓR REGRESYJNY	$P1 = e^{(0,123414+2,827793PR1-0,000083PR2)}$						
KRZYWE SKLEJANE i 95% PAS UFNOŚCI	<p>The figure contains two plots. The top plot is titled 'Krzywa sklejana i 95% pas ufności dla PR1' and shows observed values for P1 on the y-axis (ranging from -0.5 to 0.4) against PR1 on the x-axis (ranging from 0.00 to 0.30). The bottom plot is titled 'Krzywa sklejana i 95% pas ufności dla PR2' and shows observed values for P1 on the y-axis (ranging from -0.07 to 0.06) against PR2 on the x-axis (ranging from -100 to 1000). Both plots include observed data points (blue circles), a fitted curve (red line), and a 95% confidence interval (dotted red lines).</p>						

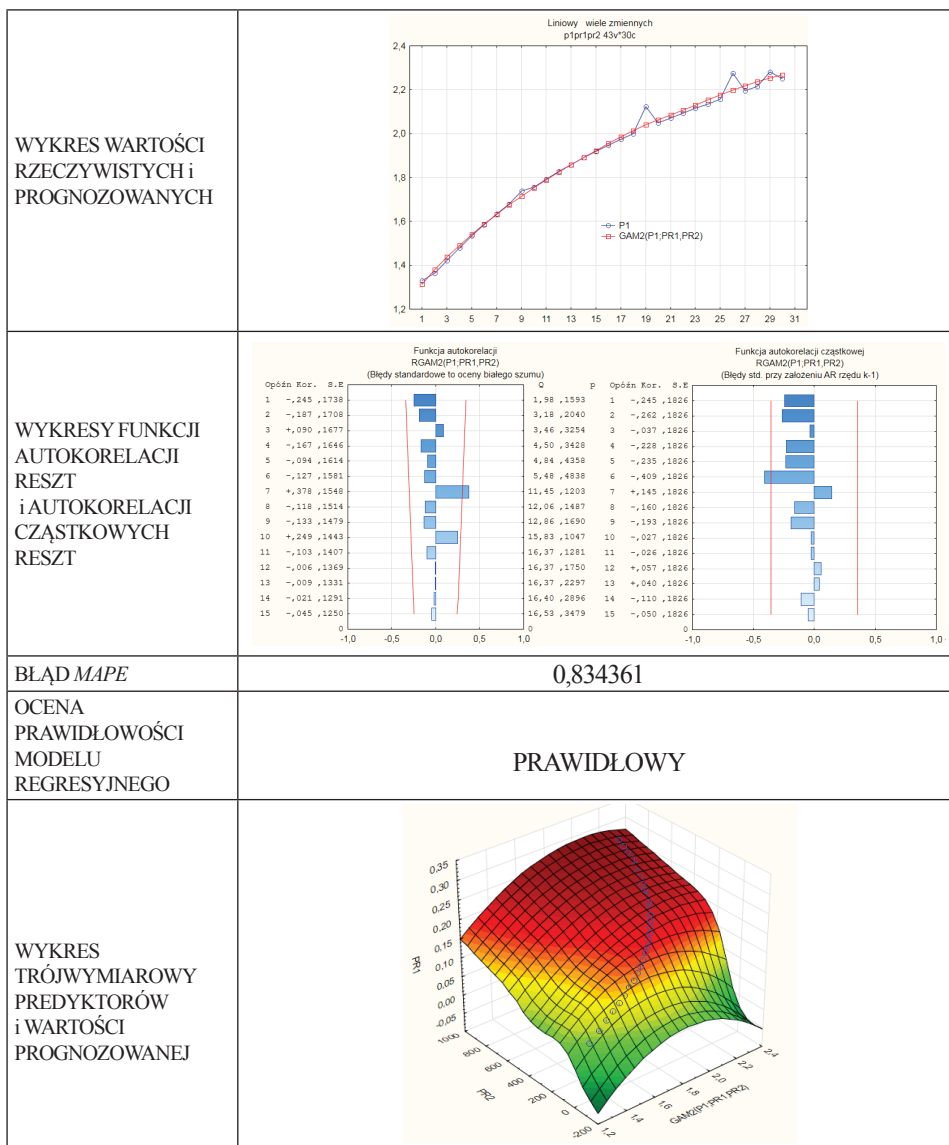
<p>WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH</p>	
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	<div style="display: flex; flex-direction: column; align-items: center;"> <div style="text-align: center;"> <p>Funkcja autokorelacji RGAM1(P1,PR1,PR2) (Błędy standardowe to oceny białego szumu)</p>  </div> <div style="text-align: center;"> <p>Funkcja autokorelacji cząstkowej RGAM1(P1,PR1,PR2) (Błędy std. przy założeniu AR rzędu k-1)</p>  </div> </div>
<p>BŁĄD MAPE [%]</p>	<p style="text-align: center;"><b>0,803304</b></p>
<p>OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO</p>	<p style="text-align: center;"><b>PRAWIDŁOWY</b></p>
<p>WYKRES TRÓJWYMIAROWY PREDYKTORÓW i WARTOŚCI PROGNOZOWANEJ</p>	



- Obliczony model GAM1(P1; PR1,PR2) jest prawidłowy ze względu na to, że:
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
  - błąd  $MAPE = 0,803304\%$  (prognoza doskonała),
  - stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy 7 i 6 opóźnieniu (wynik zadowalający),
  - model może być przyjęty do stosowania.

Tab. 4.11. Zestawienie obliczeń modelu GAM2(P1; PR1,PR2) metodą uogólnionych modeli addytywnych, funkcja wiążąca identycznościowa, rozkład normalny

METODA OBLICZEŃ i MODEL REGRESYJNY	GAM2 (P1;PR1, PR2)					
ZMIENNA ZALEŻNA	P1					
ZMIENNE NIEZALEŻNE	PR1, PR2					
PREDYKTORY	PR1,PR2					
PODSUMOWANIE REGRESJI	Podsumowanie dopasowania (p1pr1pr2) Obserwowana: P1 Rozkład: Normalny; Funkcja wiążąca: Identycznościowa					
	Indeks zmiennej	Stopnie swobody	GAM współcz.	Błąd standardowy	Standaryz. wyniku	Nieliniowe p-wartość
	w.wolny	0	1,000000	1,179935	0,027622	42,71663
PR1	1	3,997435	3,390149	0,218828	15,49233	0,441641
PR2	2	4,004421	0,000236	0,000049	4,80619	0,000468
WZÓR REGRESYJNY	$P1 = 1,1799 + 3,390149PR1 + 0,000236PR2$					
KRZYWE SKLEJANE i 95% PAS UFNOŚCI						



Obliczony model  $GAM2(P1; PR1, PR2)$  jest prawidłowy ze względu na to, że:

- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd  $MAPE = 0,834361\%$  (prognoza doskonała),
- stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy 7 i 6 opóźnieniu (wynik zadowalający).
- model może być przyjęty do stosowania.

Otrzymane wyniki przeprowadzonych obliczeń metodą uogólnionych modeli addytywnych zestawiono w tabeli 4.12.

Tab. 4.12. Zestawienie przeprowadzonych obliczeń metodą uogólnionych modeli addytywnych

LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszti	Wartość błędu MAPE [%]	Poprawność modelu
1	GAM1(P1; PR1,PR2)	$P1 = e^{(0,123414+2,827793PR1-0,000083PR2)}$	NIE	0,803304	TAK
2	GAM2(P1; PR1,PR2)	$P1 = 1,1799+3,390149PR1+0,000236PR2$	NIE	0,834361	TAK

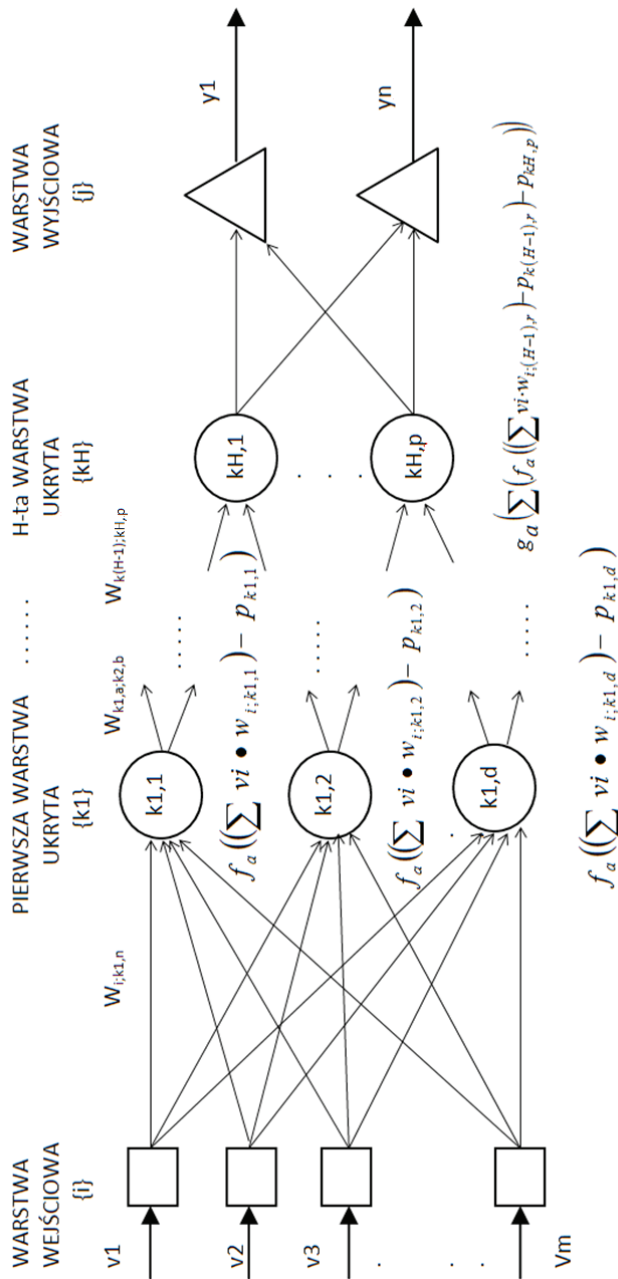
Lepszym z policzonych modeli metodą uogólnionych modeli addytywnych jest model GAM1(P1; PR1,PR2) z funkcją wiążącą logarymiczną o rozkładzie gamma. Oba policzone modele są poprawne, jednak dla GAM1 błąd MAPE jest mniejszy i wynosi 0,803304%.

#### 4.4 Prognozowanie metodą sztucznych sieci neuronowych

Sieci neuronowe są wyrafinowaną techniką modelowania, zdolną do odwzorowania nadzwyczaj złożonych funkcji (Tadeusiewicz (2010)). Sieć neuronowa to rodzaj architektury systemu komputerowego. Polega on na przetwarzaniu danych przez neurony pogrupowane w warstwy. Odpowiednie wyniki uzyskuje się dzięki procesowi uczenia, który polega na modyfikowaniu wag tych neuronów, które są odpowiedzialne za błąd. (Agnieszka Nowak-Brzezińska, Wykład „Sieci neuronowe – wprowadzenie”)

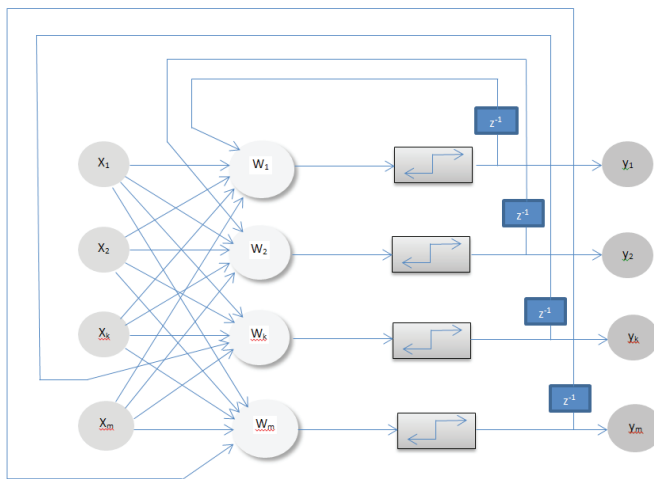
Efektywność sieci neuronowych doceniana jest ze względu na możliwość nieliniowego modelowania zjawisk i procesów. Istnieje wiele typów i rodzajów sieci neuronowych, różniących się między sobą strukturą i zasadami działania. Najpopularniejsza obecnie architektura sieciowa (Lula, Tadeusiewicz (2001)) związana jest z koncepcją wielowarstwowego perceptronu MLP. Koncepcja ta została opisana po raz pierwszy przez Rumelharta i McClellanda (1986), prace kontynuowane były przez Tadeusiewicza (1993), Bishopa (1995). Schemat działania sieci neuronowej typu MLP przedstawiono na rys. 4.2.

W perceptronie wielowarstwowym każdy z neuronów realizuje pewną transformację liniową wektora wejściowego  $v_i$  (czyli zbioru wartości wprowadzanych do neuronu). Neurony te wyznaczają ważoną sumę wartości wejściowych, a następnie odejmują od tej sumy ustaloną wartość progową. Uzyskany rezultat, będący pewną kombinacją liniową sygnałów wejściowych, jest poddawany w sieci MLP działaniu nieliniowej funkcji aktywacji  $f_a$  – najczęściej o kształcie sigmoidy. Liniowa funkcja na wejściu neuronu powoduje, że perceptron wielowarstwowo rozwiązuje stawiane mu zadanie dzieląc przestrzeń sygnałów wejściowych przy użyciu hiperpłaszczyzn na rozłączne obszary przypisywane do różnych wartości sygnałów wyjściowych  $y_j$ .



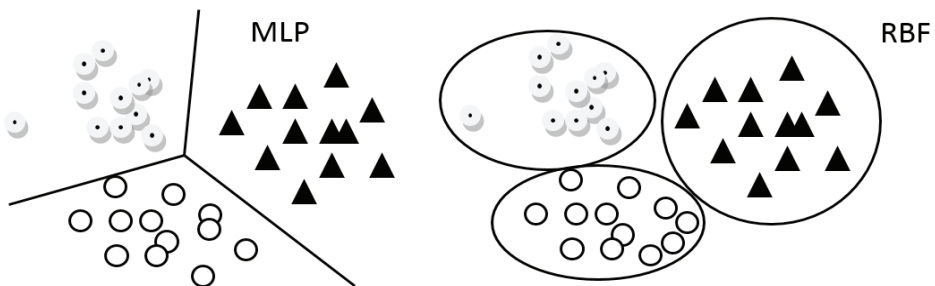
Rys. 4.2. Schemat działania sieci neuronowej typu MLP gdzie:  $v$  - zmienne niezależne wejściowe,  $m$  - liczba zmiennych niezależnych,  $w$  - wagi obliczone przy wejściu do warstwy ukrytej neuronów,  $p$  - obliczona wartość progowa,  $f$  i  $g$  - funkcje aktywacji,  $H$  - liczba warstw ukrytych w perceptronie wielowarstwowym,  $d$  i  $p$  - liczba neuronów w kolejnych warstwach ukrytych,  $y$  - zmienne zależne,  $n$  - liczba neuronów wyjściowych.

Innym popularnym modelem sieciowym jest sieć rekurencyjna RBF (Radial Basic Function). W tej sieci połączenia między neuronami stanowią graf z cyklami (obieg zamknięty) tzn. sygnały z warstwy wyjściowej sieci podawane są z powrotem do warstwy wejściowej. Wśród różnorodności modeli rekurencyjnych sztucznych sieci neuronowych wyróżnić można: sieć Hopfielda, układ gęsto połączonych ze sobą neuronów (każdy z każdym, ale bez połączeń zwrotnych) realizującą dynamikę gwarantującą zbieżność do preferowanych wzorców oraz maszynę Boltzmana będącą modyfikacją sieci Hopfielda (pozwoliła na uczenie neuronów ukrytych i likwidację wzorców pasożytniczych kosztem zwiększenia czasu symulacji). Sygnały wyjściowe stanowią jednocześnie wektor wejściowy dla następnego cyklu ( $z-1$  – jednostkowy operator opóźnienia). W sieci tej nie występuje zwykle sprzężenie zwrotne neuronu od własnego sygnału wyjściowego. Na rys. 4.3 przedstawiony został model sieci rekurencyjnej z jedną ukrytą warstwą neuronów.



Rys. 4.3. Schemat działania sieci neuronowej typu RBF z jedną ukrytą warstwą neuronów

W dużym uproszczeniu różnicę pomiędzy sieciami MLP i RBF można przedstawić jak na rysunku 4.4.



Rys. 4.4. Schemat podziału przestrzeni sygnałów wejściowych w sieciach neuronowych typu MLP i RBF

Wykonując obliczenia metodą automatycznych sieci neuronowych jako wynik otrzymujemy tabelę z zestawieniem aktywnych sieci neuronowych oraz ich parametrów. Jako wynik obliczeń wybiera się zwykle pięć sieci charakteryzujących się najmniejszymi błędami uczenia, testowania i walidacji. Im wartości błędów są bliższe zeru tym model sieciowy jest lepszy. Jakość uczenia, testowania i walidacji jest tym lepsza im jest bliższa jedności. Aby uzyskać wartość błędu *MAPE* mniejszą niż 3%, jakość uczenia, testowania i walidacji muszą wykazywać się wartościami zbliżonymi do 0,99 (oczywiście najlepiej byłoby gdyby miały wartość jeden, ale jest to bardzo mało prawdopodobne, szczególnie w odniesieniu do danych uzyskiwanych z pomiarów).

Zwykle uzyskuje się dobre wyniki obliczeń metodą sieci neuronowych, jeśli mamy do dyspozycji około sto przypadków dla każdej zmiennej. Wadą metody jest brak wygenerowanego wzoru regresji. Wykonując obliczenia prognostyczne musimy wprowadzać nowe dane (dla nowego przypadku) bezpośrednio do wytypowanej sieci. Zatem nowe obliczenia może wykonywać tylko osoba, która posiada wszystkie dane poprzednie oraz sieć neuronową. Ogranicza to uniwersalność metody oraz przekazywanie wyników w postaci wzoru regresyjnego do szerokiego stosowania.

Metoda obliczeniowa sztucznych sieci neuronowych wykorzystywana była przez wielu naukowców podczas realizacji zadań badawczych z różnych dziedzin: Moon (2015), Yalcin (2015), Hisaki (2015), Momeni (2015), Young (2015), Zhang (2015), Gab-Allah (2015), Amusan (2014), Rodger (2014), Platon (2014), Suryanita (2012), Yang (2003), Boussabaine (1996).

W tabelach 4.13–4.15 przedstawiono obliczenia metodą automatycznych sieci neuronowych dla danych z przykładu 3.1. Podobnie jak w przypadkach regresji wielorakiej, wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych i uogólnionych modeli addytywnych obliczenia przeprowadzono dla dwóch zmiennych niezależnych PR1 i PR2 celem wyznaczenia zmiennej zależnej P1.

W tabeli 4.13. zestawiono obliczenia modelu NN1 MLP 2-9-1(P1; PR1,PR2) metodą sieci neuronowych z wykładniczymi funkcjami aktywacji warstw ukrytych i wyjściowej. Przyjęto sieć typu wielowarstwowego perceptronu MLP o dziewięciu warstwach ukrytych.

Tab. 4.13. Zestawienie obliczeń modelu NN1 MLP 2-9-1(P1; PR1,PR2) metodą sieci neuronowych z wykładniczymi funkcjami aktywacji warstw ukrytych i wyjściowej

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>NN1 MLP 2-9-1(P1;PR1, PR2)</b>
ZMIENNA ZALEŻNA	P1
ZMIENNE NIEZALEŻNE	PR1, PR2
PREDYKTORY	PR1,PR2

PODSUMOWANIE REGRESJI	Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)																																																																
	1	MLP 2-9-1	0,994947	0,997438	0,999995	0,000390	0,000171																																																																
	Błąd (walidacja)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)																																																																		
	0,000048	BFGS 41	SOS	Wykładnicza																																																																			
WZÓR REGRESYJNY	NIE JEST GENEROWANY																																																																						
WYKRES WARTOŚCI RZECZYWISTYCH I PROGNOZOWANYCH																																																																							
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	<div style="display: flex; justify-content: space-around;"> <div style="width: 45%;"> <p>Funkcja autokorelacji RNN1 MLP 2-9-1(P1, PR1, PR2) (Błędy standardowe to oceny białego szumu)</p> <table border="1"> <thead> <tr> <th>Opóźn. Kor. S.B</th> <th>Q</th> </tr> </thead> <tbody> <tr><td>1</td><td>1,61 ,2046</td></tr> <tr><td>2</td><td>2,69 ,2358</td></tr> <tr><td>3</td><td>3,07 ,3806</td></tr> <tr><td>4</td><td>4,35 ,3609</td></tr> <tr><td>5</td><td>4,65 ,4345</td></tr> <tr><td>6</td><td>5,42 ,4907</td></tr> <tr><td>7</td><td>12,01 ,1003</td></tr> <tr><td>8</td><td>12,79 ,1192</td></tr> <tr><td>9</td><td>13,48 ,1421</td></tr> <tr><td>10</td><td>16,64 ,0828</td></tr> <tr><td>11</td><td>17,15 ,1035</td></tr> <tr><td>12</td><td>17,17 ,1433</td></tr> <tr><td>13</td><td>17,18 ,1914</td></tr> <tr><td>14</td><td>17,18 ,2469</td></tr> <tr><td>15</td><td>17,18 ,3061</td></tr> </tbody> </table> </div> <div style="width: 45%;"> <p>Funkcja autokorelacji cząstkowej RNN1 MLP 2-9-1(P1, PR1, PR2) (Błędy std. przy założeniu AR rzędu k-1)</p> <table border="1"> <thead> <tr> <th>Opóźn. Kor. S.B</th> <th>P</th> </tr> </thead> <tbody> <tr><td>1</td><td>-,220 ,1826</td></tr> <tr><td>2</td><td>-,254 ,1826</td></tr> <tr><td>3</td><td>-,042 ,1826</td></tr> <tr><td>4</td><td>-,254 ,1826</td></tr> <tr><td>5</td><td>-,270 ,1826</td></tr> <tr><td>6</td><td>-,440 ,1826</td></tr> <tr><td>7</td><td>*,130 ,1826</td></tr> <tr><td>8</td><td>-,243 ,1826</td></tr> <tr><td>9</td><td>-,267 ,1826</td></tr> <tr><td>10</td><td>-,121 ,1826</td></tr> <tr><td>11</td><td>-,102 ,1826</td></tr> <tr><td>12</td><td>-,049 ,1826</td></tr> <tr><td>13</td><td>-,055 ,1826</td></tr> <tr><td>14</td><td>-,194 ,1826</td></tr> <tr><td>15</td><td>*,020 ,1826</td></tr> </tbody> </table> </div> </div>							Opóźn. Kor. S.B	Q	1	1,61 ,2046	2	2,69 ,2358	3	3,07 ,3806	4	4,35 ,3609	5	4,65 ,4345	6	5,42 ,4907	7	12,01 ,1003	8	12,79 ,1192	9	13,48 ,1421	10	16,64 ,0828	11	17,15 ,1035	12	17,17 ,1433	13	17,18 ,1914	14	17,18 ,2469	15	17,18 ,3061	Opóźn. Kor. S.B	P	1	-,220 ,1826	2	-,254 ,1826	3	-,042 ,1826	4	-,254 ,1826	5	-,270 ,1826	6	-,440 ,1826	7	*,130 ,1826	8	-,243 ,1826	9	-,267 ,1826	10	-,121 ,1826	11	-,102 ,1826	12	-,049 ,1826	13	-,055 ,1826	14	-,194 ,1826	15	*,020 ,1826
Opóźn. Kor. S.B	Q																																																																						
1	1,61 ,2046																																																																						
2	2,69 ,2358																																																																						
3	3,07 ,3806																																																																						
4	4,35 ,3609																																																																						
5	4,65 ,4345																																																																						
6	5,42 ,4907																																																																						
7	12,01 ,1003																																																																						
8	12,79 ,1192																																																																						
9	13,48 ,1421																																																																						
10	16,64 ,0828																																																																						
11	17,15 ,1035																																																																						
12	17,17 ,1433																																																																						
13	17,18 ,1914																																																																						
14	17,18 ,2469																																																																						
15	17,18 ,3061																																																																						
Opóźn. Kor. S.B	P																																																																						
1	-,220 ,1826																																																																						
2	-,254 ,1826																																																																						
3	-,042 ,1826																																																																						
4	-,254 ,1826																																																																						
5	-,270 ,1826																																																																						
6	-,440 ,1826																																																																						
7	*,130 ,1826																																																																						
8	-,243 ,1826																																																																						
9	-,267 ,1826																																																																						
10	-,121 ,1826																																																																						
11	-,102 ,1826																																																																						
12	-,049 ,1826																																																																						
13	-,055 ,1826																																																																						
14	-,194 ,1826																																																																						
15	*,020 ,1826																																																																						
BŁĄD MAPE [%]	0,867906																																																																						
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	PRAWIDŁOWY																																																																						
WYKRES TRÓJWYMIAROWY PREDYKTORÓW i WARTOŚCI PROGNOZOWANEJ																																																																							

Obliczony model NN1 MLP 2-9-1(P1;PR1, PR2) jest prawidłowy ze względu na to, że:

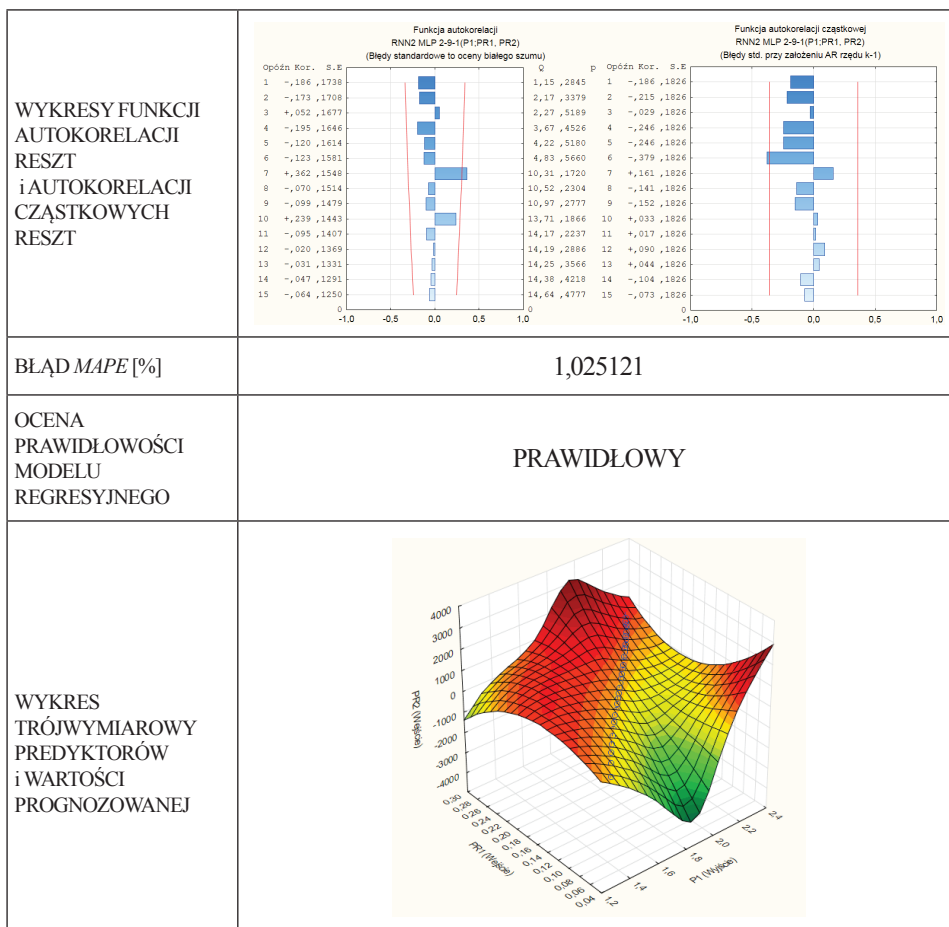
- ma wysoką jakość uczenia, testowania i walidacji (ponad 0,99)
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd  $MAPE = 0,867906\%$  (prognoza doskonała),
- stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy 7 i 6 opóźnieniu (wynik zadowalający),
- model może być przyjęty do stosowania.

W tabeli 4.14. zestawiono obliczenia modelu NN2 MLP 2-9-1(P1; PR1,PR2) metodą sieci neuronowych z wykładniczymi funkcjami aktywacji warstw ukrytych i liniową wyjściową funkcją aktywacji (liniowość ta stanowi różnicę w odniesieniu do sieci NN1). Przyjęto sieć typu wielowarstwowego perceptronu MLP o dziewięciu warstwach ukrytych.

Tab. 4.14. Zestawienie obliczeń modelu NN2 MLP 2-9-1(P1; PR1,PR2) metodą sieci neuronowych z wykładniczymi funkcjami aktywacji warstw ukrytych i liniową wyjściową funkcją aktywacji

METODA OBLICZEŃ i MODEL REGRESYJNY	NN2 MLP 2-9-1 (P1;PR1, PR2)						
ZMIENNA ZALEŻNA	P1						
ZMIENNE NIEZALEŻNE	PR1, PR2						
PREDYKTORY	PR1,PR2						
PODSUMOWANIE REGRESJI	Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
	2	MLP2-9-1	0,994448	0,997343	0,999896	0,000431	0,000155
	Błąd (walidacja)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)		
	0,000142	BFGS 10	SOS	Wykładnicza	Liniowa		
WZÓR REGRESYJNY	NIE JEST GENEROWANY						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							





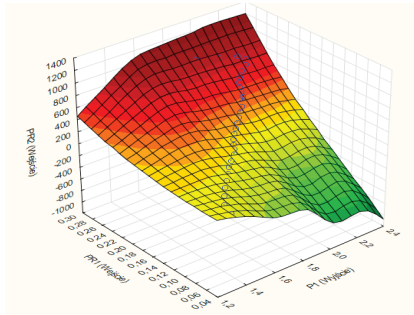
- Obliczony model NN2 MLP 2-9-1(P1;PR1, PR2) jest prawidłowy ze względu na to, że:
- ma wysoką jakość uczenia, testowania i walidacji (ponad 0,99),
  - wykresy wartości rzeczywistych i prognozowanych są zbliżone,
  - błąd MAPE = 1,025121% (prognoza doskonała),
  - stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy 7 i 6 opóźnieniu (wynik zadowalający),
  - model może być przyjęty do stosowania.

W tabeli 4.15. zestawiono obliczenia modelu NN3 RBF 2-8-1(P1; PR1,PR2) metodą sieci neuronowych z gaussowskimi funkcjami aktywacji warstw ukrytych i liniową wyjściową funkcją aktywacji. Przyjęto sieć typu rekurencyjnego RBF o ośmiu warstwach ukrytych.

Tab.4.15. Zestawienie obliczeń modelu NN3 RBF 2-8-1(P1; PR1,PR2) metodą sieci neuronowych z gaussowskimi funkcjami aktywacji warstw ukrytych i liniową wyjściową funkcją aktywacji

METODA OBLICZEŃ i MODEL REGRESYJNY	NN3 RBF 2-8-1 (P1;PR1, PR2)						
ZMIENNA ZALEŻNA	P1						
ZMIENNE NIEZALEŻNE	PR1, PR2						
PREDYKTORY	PR1,PR2						
PODSUMOWANIE REGRESJI	Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
	3	RBF 2-8-1	0,996613	0,920647	0,999960	0,000261	0,017156
	Błąd (walidacja)	Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)		Aktywacja (wyjściowe)	
	0,000258	RBFT	SOS	Gaussa		Liniowa	
WZÓR REGRESYJNY	NIE JEST GENEROWANY						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT							
BŁĄD MAPE [%]	1,38018						
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY						

WYKRES  
TRÓJWYMIAROWY  
PREDYKTORÓW  
i WARTOŚCI  
PROGNOZOWANEJ



Obliczony model NN3 RBF 2-8-1(P1;PR1, PR2) jest nieprawidłowy ze względu na to, że:

- wykresy wartości rzeczywistych i prognozowanych nie są zbliżone, wyraźne odchylenie od 25 do 30 przypadku,
- stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy pierwszym opóźnieniu (wynik eliminujący),
- jakość walidacji 0,92 – zbyt niska.

Model nie może być przyjęty do stosowania pomimo, że ma wysoką jakość uczenia i testowania (ponad 0,99) oraz niski błąd  $MAPE = 1,38018\%$ .

Otrzymane wyniki przeprowadzonych obliczeń metodą automatycznych sieci neuronowych zestawiono w tabeli 4.16.

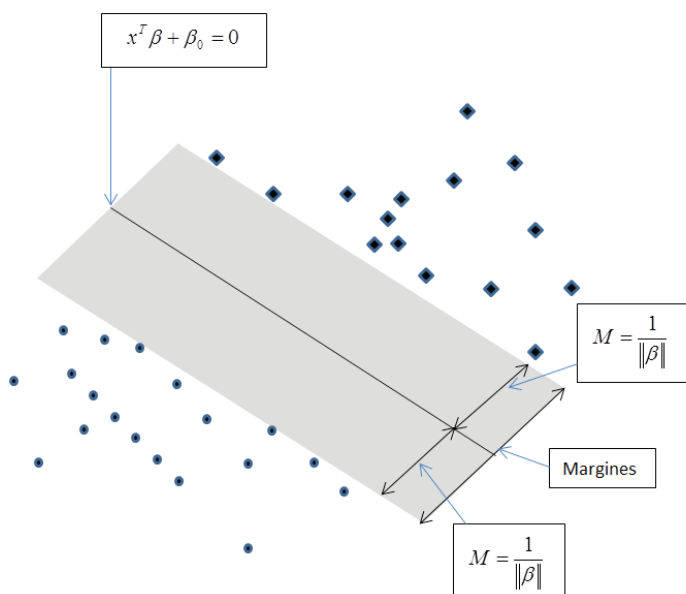
Tab. 4.16. Zestawienie przeprowadzonych obliczeń metodą automatycznych sieci neuronowych

LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszt	Wartość błędu $MAPE$ [%]	Poprawność modelu
1	NN1 MLP 2-9-1(P1;PR1, PR2)	brak	NIE	0,803304	TAK
2	NN2 MLP 2-9-1(P1;PR1, PR2)	brak	NIE	0,834361	TAK
3	NN3 RBF 2-8-1(P1;PR1, PR2)	brak	TAK	1,38018	NIE

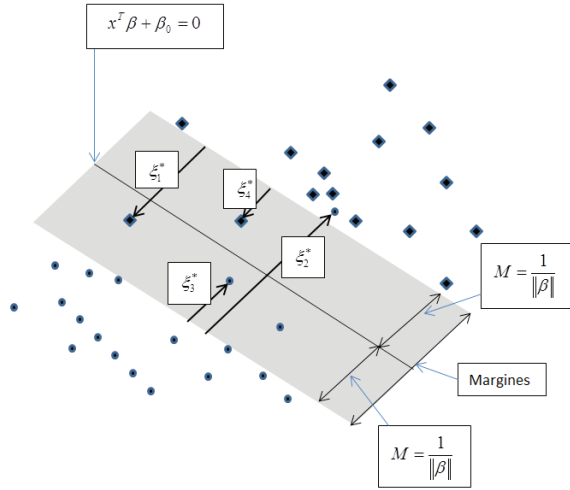
Najlepszym z policzonych modeli metodą automatycznych sieci neuronowych jest model NN1 MLP 2-9-1(P1; PR1,PR2) z wykładniczymi funkcjami aktywacji warstw ukrytych i wyjściowej. Modele NN1 i NN2 są poprawne, jednak dla NN1 błąd  $MAPE$  jest mniejszy i wynosi 0,803304%.

## 4.5. Prognozowanie metodą wektorów nośnych

Metoda wektorów nośnych SVM (Support Vector Machine) stosowana jest do rozwiązywania problemów regresyjnych i klasyfikacyjnych. Polega na budowaniu nieliniowych granic decyzyjnych, oddzielających obszary w przestrzeni predyktorów, odpowiadające różnym wartościom zmiennej zależnej. Ze względu na właściwości przestrzeni cech (predyktorów) metoda SVM wykazuje dużą elastyczność przy rozwiązywaniu zadań klasyfikacyjnych i regresyjnych o różnej złożoności. Czasami stosuje się również nazwy metoda wektorów wspierających i metoda wektorów podpierających. SVM stosowana była przez An (2007), Chen (2010), Cheng (2009, 2010, 2011), Jiang (2011), Lam (2009), Li (2009), Lute (2009), Mahfouz (2010), Pal (2008), Shin(2009), Son (2011), Tserng (2011) wraz z zespołami do rozwiązywania różnych zagadnień inżynierskich takich jak: obliczanie kosztów przedsięwzięcia, ryzyka zatrudniania podwykonawców, elementów zarządzania w budownictwie, zużycia energii elektrycznej, prekwalfikacji podwykonawców, aerodynamicznej analizy konstrukcji mostów, wyboru technologii i materiałów budowlanych i inne. w najprostszej formie działanie algorytmu SVM polega na znalezieniu hiperpłaszczyzny rozdzielającej z maksymalnym marginesem punkty należące do dwóch klas. Zobrazowano to na rysunkach 4.5 i 4.6 dla przypadków separowalnych i nie separowalnych.



Rys. 4.5. Przypadek separowalny. Maksymalny margines hiperprzestrzeni oddziela grupy przypadków.



Rys. 4.6 Przypadek nieseparowalny.  $\zeta_i^*$  – odległość punktów znajdujących się po „złej stronie” od

brzegu marginesu,  $\|\beta\|$  norma wektora  $\|\beta_1, \beta_2, \dots, \beta_n\| = \sqrt{\sum_{n=1}^n (\beta_i)^2}$

W regresji SVM poszukujemy zależności funkcyjnej zmiennej zależnej  $y$  od zbioru zmiennych niezależnych  $x$ . Przyjmuje się, że zależność ta jest typu deterministycznego (4.12), z pewnym dodatkiem losowego szumu:

$$y = f(x) + \text{szum} \tag{4.12}$$

Podstawowym zadaniem jest znalezienie postaci funkcji jądrowej  $f$ , która powinna możliwie najlepiej podawać wartość zmiennej zależnej dla nowych przypadków, których model SVM nie „widział” wcześniej. Zadanie to rozwiązuje się ucząc model SVM za pomocą próby przypadków, zwanej próbą uczącą. Podobnie jak przy klasyfikacji, proces ten polega na sekwencyjnym minimalizowaniu pewnej funkcji błędu. Stosuje się dwa typy tej funkcji i, odpowiednio do tego, są dwa typy regresji SVM typ pierwszy i drugi. Stosując regresję SVM typu 1 minimalizuje się funkcję błędu następującej postaci (4.13):

$$\frac{1}{2} w^T w + C \sum_{i=1}^N \zeta_i + C \sum_{i=1}^N \zeta_i^* \tag{4.13}$$

gdzie:

$C$  – stała, nazywana pojemnością,

$w$  – wektor współczynników,

$b$  – stała,

$N$  – liczba przypadków,

$\phi$  – funkcja jądrowa,

$\zeta_i - \zeta_i^*/M$ ,

$\zeta_i^*$  – odległość punktów znajdujących się po „złej stronie” od brzegu marginesu,

$M$  – margines hiperprzestrzeni (rys.4.6).

Stosując regresję SVM typu 2 minimalizuje się funkcję błędu następującej postaci (4.14):

$$\frac{1}{2} w^T w - C \left( \nu \varepsilon + \frac{1}{N} \sum_{i=1}^N (\zeta_i + \zeta_i^*) \right) \quad (4.14)$$

Funkcje jądrowe wywodzą się z badań liniowych przestrzeni wektorowych Hilberta, Banacha. Stosuje się kilka typów funkcji jądrowych w modelu SVM: liniową, wielomian, radialne funkcje bazowe (RBF) i funkcję sigmoidalną (4.15):

$$\phi = \begin{cases} x_i \cdot x_j \\ (\gamma x_i x_j + c)^\gamma \\ \exp(-\gamma |x_i - x_j|^2) \\ \tanh(\gamma x_i x_j + c) \end{cases} \quad (4.15)$$

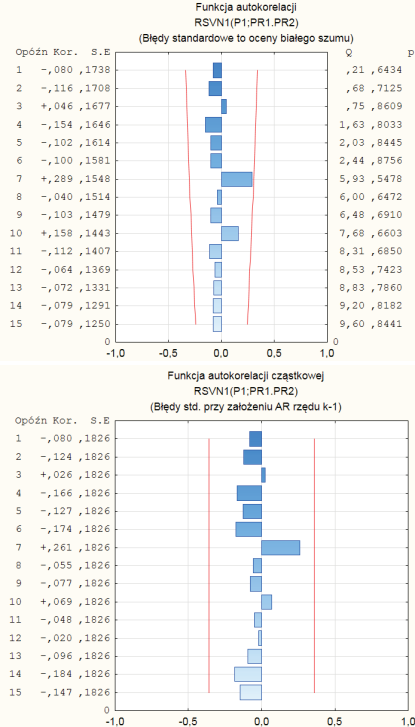
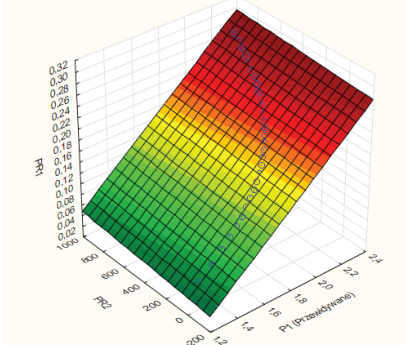
RBF jest najczęściej wybierana jako funkcja jądrowa w SVM, głównie ze względu na ograniczony zasięg w polu zmiennych  $x$ .

W tabelach 4.17–4.20 przedstawiono obliczenia metodą wektorów nośnych dla danych z przykładu 3.1. Podobnie jak w przypadkach regresji wielorakiej, wielomiennej regresji adaptacyjnej z użyciem funkcji sklepanych, uogólnionych modeli addytywnych i sieci neuronowych obliczenia przeprowadzono dla dwóch zmiennych niezależnych PR1 i PR2 celem wyznaczenia zmiennej zależnej P1. Przyjmowano kolejno funkcje jądrowe liniowe, wielomianowe, RBF i sigmoidalne.

W tabeli 4.17. zestawiono obliczenia modelu SVN1(P1; PR1,PR2) metodą wektorów nośnych z liniową funkcją jądrową.

Tab. 4.17. Zestawienie obliczeń modelu SVN1(P1; PR1,PR2) metodą wektorów nośnych z liniową funkcją jądrową

METODA OBLICZEN i MODEL REGRESYJNY	SVN1 (P1;PR1, PR2)
ZMIENNA ZALEŻNA	P1
ZMIENNE NIEZALEŻNE	PR1, PR2
PREDYKTORY	PR1,PR2
TYP JĄDRA	LINIOWY
WZÓR REGRESYJNY	NIE JEST GENEROWANY
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	

<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BLĄD MAPE [%]</p>	<p>2,328698</p>
<p>OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO</p>	<p>PRAWIDŁOWY</p>
<p>WYKRES TRÓJWYMIAROWY PREDYKTORÓW i WARTOŚCI PROGNOZOWANEJ</p>	

Obliczony model SVN1(P1;PR1, PR2) jest prawidłowy ze względu na to, że:

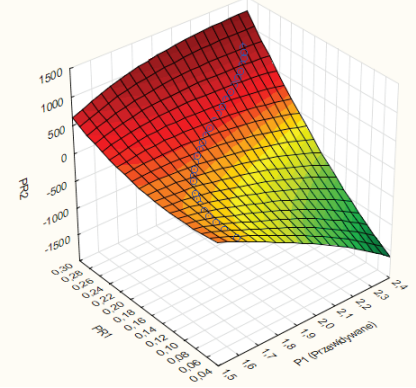
- wykresy wartości rzeczywistych i prognozowanych są zbliżone,
- błąd MAPE = 2,328698% (prognoza bardzo dobra),
- nie stwierdzono występowania autokorelacji i autokorelacji cząstkowych reszt modelu,
- model może być przyjęty do stosowania.

W tabeli 4.18. zestawiono obliczenia modelu SVN2(P1; PR1,PR2) metodą wektorów nośnych z wielomianową funkcją jądrową.

Tab. 4.18. Zestawienie obliczeń modelu SVN2(P1; PR1,PR2) metodą wektorów nośnych z wielomianową funkcją jądrową

METODA OBLICZEŃ i MODEL REGRESYJNY	SVN2 (P1;PR1, PR2)																																																																																																																																
ZMIENNA ZALEŻNA	P1																																																																																																																																
ZMIENNE NIEZALEŻNE	PR1, PR2																																																																																																																																
PREDYKTORY	PR1,PR2																																																																																																																																
TYP JĄDRA	WIELOMIANOWE																																																																																																																																
WZÓR REGRESYJNY	NIE JEST GENEROWANY																																																																																																																																
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH																																																																																																																																	
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	<div style="display: flex; flex-direction: column;"> <div style="margin-bottom: 10px;"> <p>Funkcja autokorelacji RSVN2(P1;PR1,PR2) (Błędy standardowe to oceny białego szumu)</p> <table border="1"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E</th> <th>Q</th> <th>p</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,768</td><td>,1738</td><td>19,51</td><td>,0000</td></tr> <tr><td>2</td><td>+,598</td><td>,1708</td><td>31,75</td><td>,0000</td></tr> <tr><td>3</td><td>+,449</td><td>,1677</td><td>38,93</td><td>,0000</td></tr> <tr><td>4</td><td>+,282</td><td>,1646</td><td>41,86</td><td>,0000</td></tr> <tr><td>5</td><td>+,154</td><td>,1614</td><td>42,76</td><td>,0000</td></tr> <tr><td>6</td><td>+,031</td><td>,1581</td><td>42,80</td><td>,0000</td></tr> <tr><td>7</td><td>-,041</td><td>,1548</td><td>42,87</td><td>,0000</td></tr> <tr><td>8</td><td>-,162</td><td>,1514</td><td>44,02</td><td>,0000</td></tr> <tr><td>9</td><td>-,211</td><td>,1479</td><td>46,05</td><td>,0000</td></tr> <tr><td>10</td><td>-,236</td><td>,1443</td><td>48,72</td><td>,0000</td></tr> <tr><td>11</td><td>-,292</td><td>,1407</td><td>53,04</td><td>,0000</td></tr> <tr><td>12</td><td>-,303</td><td>,1369</td><td>57,94</td><td>,0000</td></tr> <tr><td>13</td><td>-,312</td><td>,1331</td><td>63,43</td><td>,0000</td></tr> <tr><td>14</td><td>-,310</td><td>,1291</td><td>69,19</td><td>,0000</td></tr> <tr><td>15</td><td>-,298</td><td>,1250</td><td>74,88</td><td>,0000</td></tr> </tbody> </table> </div> <div> <p>Funkcja autokorelacji cząstkowej RSVN2(P1;PR1,PR2) (Błędy std. przy założeniu AR rzędu k-1)</p> <table border="1"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,768</td><td>,1826</td></tr> <tr><td>2</td><td>+,020</td><td>,1826</td></tr> <tr><td>3</td><td>-,038</td><td>,1826</td></tr> <tr><td>4</td><td>-,138</td><td>,1826</td></tr> <tr><td>5</td><td>-,038</td><td>,1826</td></tr> <tr><td>6</td><td>-,088</td><td>,1826</td></tr> <tr><td>7</td><td>+,016</td><td>,1826</td></tr> <tr><td>8</td><td>-,200</td><td>,1826</td></tr> <tr><td>9</td><td>+,033</td><td>,1826</td></tr> <tr><td>10</td><td>-,032</td><td>,1826</td></tr> <tr><td>11</td><td>-,117</td><td>,1826</td></tr> <tr><td>12</td><td>-,034</td><td>,1826</td></tr> <tr><td>13</td><td>-,052</td><td>,1826</td></tr> <tr><td>14</td><td>-,061</td><td>,1826</td></tr> <tr><td>15</td><td>-,032</td><td>,1826</td></tr> </tbody> </table> </div> </div>	Opóźn	Kor.	S.E	Q	p	1	+,768	,1738	19,51	,0000	2	+,598	,1708	31,75	,0000	3	+,449	,1677	38,93	,0000	4	+,282	,1646	41,86	,0000	5	+,154	,1614	42,76	,0000	6	+,031	,1581	42,80	,0000	7	-,041	,1548	42,87	,0000	8	-,162	,1514	44,02	,0000	9	-,211	,1479	46,05	,0000	10	-,236	,1443	48,72	,0000	11	-,292	,1407	53,04	,0000	12	-,303	,1369	57,94	,0000	13	-,312	,1331	63,43	,0000	14	-,310	,1291	69,19	,0000	15	-,298	,1250	74,88	,0000	Opóźn	Kor.	S.E	1	+,768	,1826	2	+,020	,1826	3	-,038	,1826	4	-,138	,1826	5	-,038	,1826	6	-,088	,1826	7	+,016	,1826	8	-,200	,1826	9	+,033	,1826	10	-,032	,1826	11	-,117	,1826	12	-,034	,1826	13	-,052	,1826	14	-,061	,1826	15	-,032	,1826
Opóźn	Kor.	S.E	Q	p																																																																																																																													
1	+,768	,1738	19,51	,0000																																																																																																																													
2	+,598	,1708	31,75	,0000																																																																																																																													
3	+,449	,1677	38,93	,0000																																																																																																																													
4	+,282	,1646	41,86	,0000																																																																																																																													
5	+,154	,1614	42,76	,0000																																																																																																																													
6	+,031	,1581	42,80	,0000																																																																																																																													
7	-,041	,1548	42,87	,0000																																																																																																																													
8	-,162	,1514	44,02	,0000																																																																																																																													
9	-,211	,1479	46,05	,0000																																																																																																																													
10	-,236	,1443	48,72	,0000																																																																																																																													
11	-,292	,1407	53,04	,0000																																																																																																																													
12	-,303	,1369	57,94	,0000																																																																																																																													
13	-,312	,1331	63,43	,0000																																																																																																																													
14	-,310	,1291	69,19	,0000																																																																																																																													
15	-,298	,1250	74,88	,0000																																																																																																																													
Opóźn	Kor.	S.E																																																																																																																															
1	+,768	,1826																																																																																																																															
2	+,020	,1826																																																																																																																															
3	-,038	,1826																																																																																																																															
4	-,138	,1826																																																																																																																															
5	-,038	,1826																																																																																																																															
6	-,088	,1826																																																																																																																															
7	+,016	,1826																																																																																																																															
8	-,200	,1826																																																																																																																															
9	+,033	,1826																																																																																																																															
10	-,032	,1826																																																																																																																															
11	-,117	,1826																																																																																																																															
12	-,034	,1826																																																																																																																															
13	-,052	,1826																																																																																																																															
14	-,061	,1826																																																																																																																															
15	-,032	,1826																																																																																																																															
BŁĄD MAPE [%]	4,41385																																																																																																																																



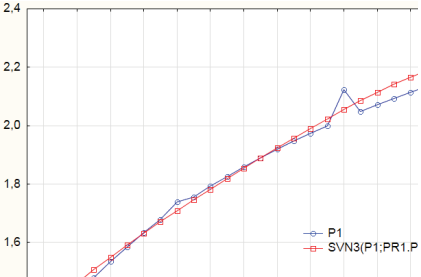
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY
WYKRES TRÓJWYMIAROWY PREDYKTORÓW i WARTOŚCI PROGNOZOWANEJ	

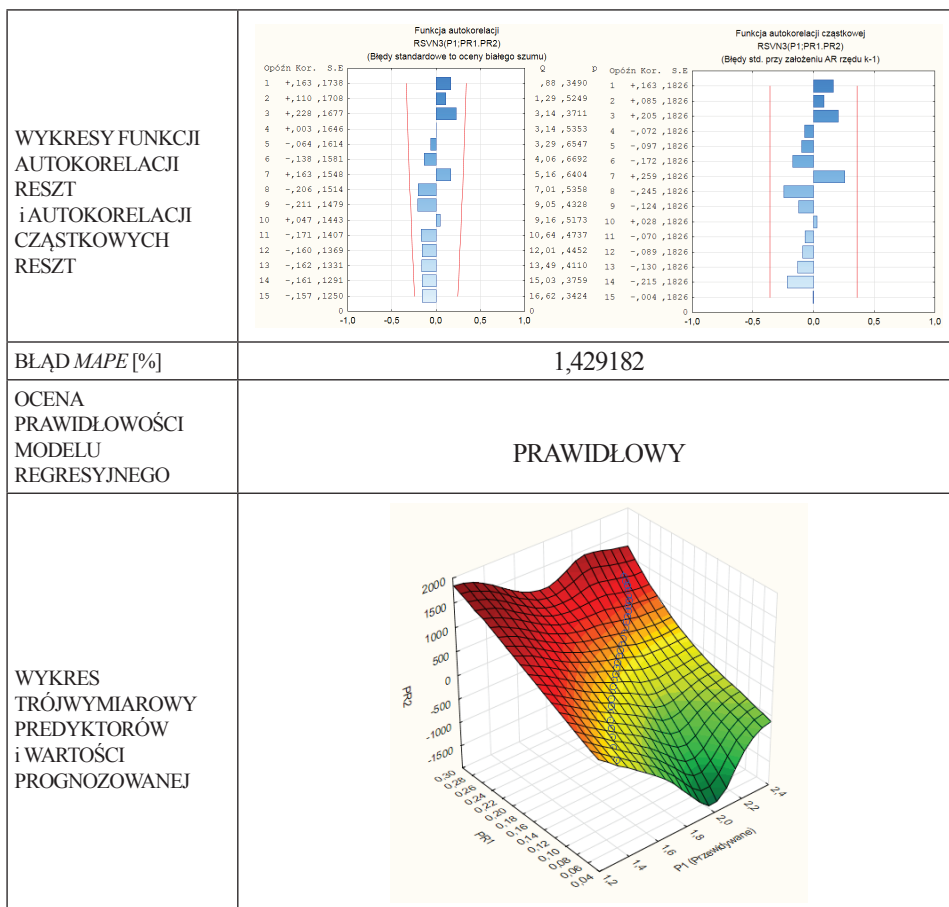
Obliczony model  $SVN2(P1;PR1, PR2)$  nie jest prawidłowy ze względu na to, że:

- wykresy wartości rzeczywistych i prognozowanych nie są zbliżone,
- stwierdzono występowanie autokorelacji i autokorelacji cząstkowych reszt modelu przy pierwszym opóźnieniu,
- model nie może być przyjęty do stosowania pomimo, że błąd *MAPE* wynosi 4,41385% (prognoza dobra).

W tabeli 4.19 zestawiono obliczenia modelu  $SVN3(P1;PR1,PR2)$  metodą wektorów nośnych z funkcją jądrową typu RBF.

Tab. 4.19. Zestawienie obliczeń modelu  $SVN3(P1; PR1,PR2)$  metodą wektorów nośnych z funkcją jądrową typu RBF

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>SVN3 (P1;PR1,PR2)</b>	
ZMIENNA ZALEŻNA	P1	
ZMIENNE NIEZALEŻNE	PR1, PR2	
PREDYKTORY	PR1,PR2	
TYP JĄDRA	RBF	
WZÓR REGRESYJNY	NIE JEST GENEROWANY	
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH		



Obliczony model SVN3(P1;PR1, PR2) jest prawidłowy ze względu na to, że:

- wykresy wartości rzeczywistych i prognozowanych są zbliżone (pewna rozbieżność występuje od przypadku 21 do 29),
- błąd MAPE = 1,429182% (prognoza bardzo dobra),
- nie stwierdzono występowania autokorelacji i autokorelacji cząstkowych reszt modelu,
- model może być przyjęty do stosowania.

W tabeli 4.20. zestawiono obliczenia modelu SVN4(P1; PR1,PR2) metodą wektorów nośnych z sigmoidalną funkcją jądrową.

Tab. 4.20. Zestawienie obliczeń modelu SVN4(P1; PR1,PR2) metodą wektorów nośnych z sigmoidalną funkcją jądrową

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>SVN4 (P1;PR1, PR2)</b>
ZMIENNA ZALEŻNA	P1
ZMIENNE NIEZALEŻNE	PR1, PR2
PREDYKTORY	PR1,PR2

TYP JĄDRA	SIGMOIDALNE																																																																																																
WZÓR REGRESYJNY	NIE JEST GENEROWANY																																																																																																
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH																																																																																																	
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	<div style="display: flex; justify-content: space-around;"> <div style="text-align: center;"> <p>Funkcja autokorelacji RSVN4(P1;PR1,PR2) (Błędy standardowe to oceny białego szumu)</p> <table border="1"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,620</td><td>,1738</td></tr> <tr><td>2</td><td>+,500</td><td>,1708</td></tr> <tr><td>3</td><td>+,345</td><td>,1677</td></tr> <tr><td>4</td><td>+,126</td><td>,1646</td></tr> <tr><td>5</td><td>+,135</td><td>,1614</td></tr> <tr><td>6</td><td>+,008</td><td>,1581</td></tr> <tr><td>7</td><td>-,035</td><td>,1548</td></tr> <tr><td>8</td><td>-,198</td><td>,1514</td></tr> <tr><td>9</td><td>-,292</td><td>,1479</td></tr> <tr><td>10</td><td>-,319</td><td>,1443</td></tr> <tr><td>11</td><td>-,431</td><td>,1407</td></tr> <tr><td>12</td><td>-,297</td><td>,1369</td></tr> <tr><td>13</td><td>-,296</td><td>,1331</td></tr> <tr><td>14</td><td>-,282</td><td>,1291</td></tr> <tr><td>15</td><td>-,255</td><td>,1250</td></tr> </tbody> </table> </div> <div style="text-align: center;"> <p>Funkcja autokorelacji częściowej RSLN4(P1;PR1,PR2) (Błędy std. przy założeniu AR rzędu k-1)</p> <table border="1"> <thead> <tr> <th>Opóźn</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,620</td><td>,1826</td></tr> <tr><td>2</td><td>+,188</td><td>,1826</td></tr> <tr><td>3</td><td>-,039</td><td>,1826</td></tr> <tr><td>4</td><td>-,218</td><td>,1826</td></tr> <tr><td>5</td><td>+,148</td><td>,1826</td></tr> <tr><td>6</td><td>-,095</td><td>,1826</td></tr> <tr><td>7</td><td>-,038</td><td>,1826</td></tr> <tr><td>8</td><td>-,294</td><td>,1826</td></tr> <tr><td>9</td><td>-,072</td><td>,1826</td></tr> <tr><td>10</td><td>-,045</td><td>,1826</td></tr> <tr><td>11</td><td>-,168</td><td>,1826</td></tr> <tr><td>12</td><td>+,100</td><td>,1826</td></tr> <tr><td>13</td><td>-,046</td><td>,1826</td></tr> <tr><td>14</td><td>-,089</td><td>,1826</td></tr> <tr><td>15</td><td>-,145</td><td>,1826</td></tr> </tbody> </table> </div> </div>	Opóźn	Kor.	S.E.	1	+,620	,1738	2	+,500	,1708	3	+,345	,1677	4	+,126	,1646	5	+,135	,1614	6	+,008	,1581	7	-,035	,1548	8	-,198	,1514	9	-,292	,1479	10	-,319	,1443	11	-,431	,1407	12	-,297	,1369	13	-,296	,1331	14	-,282	,1291	15	-,255	,1250	Opóźn	Kor.	S.E.	1	+,620	,1826	2	+,188	,1826	3	-,039	,1826	4	-,218	,1826	5	+,148	,1826	6	-,095	,1826	7	-,038	,1826	8	-,294	,1826	9	-,072	,1826	10	-,045	,1826	11	-,168	,1826	12	+,100	,1826	13	-,046	,1826	14	-,089	,1826	15	-,145	,1826
Opóźn	Kor.	S.E.																																																																																															
1	+,620	,1738																																																																																															
2	+,500	,1708																																																																																															
3	+,345	,1677																																																																																															
4	+,126	,1646																																																																																															
5	+,135	,1614																																																																																															
6	+,008	,1581																																																																																															
7	-,035	,1548																																																																																															
8	-,198	,1514																																																																																															
9	-,292	,1479																																																																																															
10	-,319	,1443																																																																																															
11	-,431	,1407																																																																																															
12	-,297	,1369																																																																																															
13	-,296	,1331																																																																																															
14	-,282	,1291																																																																																															
15	-,255	,1250																																																																																															
Opóźn	Kor.	S.E.																																																																																															
1	+,620	,1826																																																																																															
2	+,188	,1826																																																																																															
3	-,039	,1826																																																																																															
4	-,218	,1826																																																																																															
5	+,148	,1826																																																																																															
6	-,095	,1826																																																																																															
7	-,038	,1826																																																																																															
8	-,294	,1826																																																																																															
9	-,072	,1826																																																																																															
10	-,045	,1826																																																																																															
11	-,168	,1826																																																																																															
12	+,100	,1826																																																																																															
13	-,046	,1826																																																																																															
14	-,089	,1826																																																																																															
15	-,145	,1826																																																																																															
BŁĄD MAPE [%]	2,657474																																																																																																
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY																																																																																																
WYKRES TRÓJWYMIAROWY PREDYKTORÓW i WARTOŚCI PROGNOZOWANEJ																																																																																																	

- Obliczony model SVN4(P1;PR1, PR2) nie jest prawidłowy ze względu na to, że:
- wykresy wartości rzeczywistych i prognozowanych nie są zbliżone,
  - stwierdzono występowanie autokorelacji i autokorelacji częściowych reszt modelu przy pierwszym opóźnieniu,
  - model nie może być przyjęty do stosowania pomimo, że błąd MAPE wynosi 2,657474% (prognoza bardzo dobra).

Otrzymane wyniki przeprowadzonych obliczeń metodą wektorów nośnych zestawiono w tabeli 4.21.

Tab. 4.21. Zestawienie przeprowadzonych obliczeń metodą wektorów nośnych

LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszł	Wartość błędu MAPE [%]	Poprawność modelu
1	SVN1(P1;PR1,PR2)	brak	NIE	2,328698	TAK
2	SVN2(P1;PR1,PR2)	brak	TAK	4,41385	NIE
3	SVN3(P1;PR1,PR2)	brak	NIE	1,429182	TAK
4	SVN4(P1;PR1,PR2)	brak	TAK	2,657474	NIE

Najlepszym z policzonych modeli metodą wektorów nośnych jest model SVN3(P1; PR1,PR2) z funkcjami jądrowymi typu RBF. Model SVN1 z liniowymi funkcjami aktywacji jest poprawny; jednak dla SVN3 błąd MAPE jest mniejszy i wynosi 1,429182%.

#### 4.6. Prognozowanie metodą autoregresji zintegrowanej

Metoda autoregresji zintegrowanej ARIMA (Autoregressive Integrated Average Process) służy do prognozowania przyszłych wartości zmiennej na podstawie niej samej. W metodzie tej nie uwzględnia się wpływu predyktorów na wartość zmiennej zależnej. Zwykle prognozowanie dotyczy szeregów czasowych. Szeregi czasowe to sekwencja obserwacji przeprowadzona w kolejnych okresach. w szeregu czasowym wyróżnić można:

- trend (stały wzrost lub spadek średniego poziomu zjawiska w czasie),
- okresowość (np. dzienną, tygodniową, sezonową),
- autokorelację (obserwacje nie są niezależne, ale te, które leżą blisko siebie są do siebie bardziej podobne niż położone daleko od siebie).

Model autoregresyjny średniej ruchomej ARIMA to ogólny model wprowadzony przez Boxa i Jenkinsa (1976). Zawiera on zarówno parametry autoregresyjne, jak i średniej ruchomej oraz wprowadza do postaci modelu operator różnicowania. W modelu wyróżnia się trzy typy parametrów: parametry autoregresyjne (p), rząd różnicowania (d) oraz parametry średniej ruchomej (q). Wedle notacji wprowadzonej przez Boxa i Jenkinsa, modele określa się jako ARIMA (p, d, q); na przykład opisanie modelu jako (0, 1, 2) oznacza, że zawiera on 0 (zero) parametrów autoregresyjnych p i dwa parametry średniej ruchomej q, które zostały obliczone dla szeregu po jednokrotnym różnicowaniu d. W modelu ARIMA zakłada się, że można oszacować współczynniki modelu, które opisują kolejne elementy szeregu na podstawie opóźnionych w czasie poprzednich elementów (proces autoregresyjny) oraz że pozostają one pod wpływem realizacji składnika losowego w okresach przeszłych (proces średniej ruchomej). Zatem każda obserwacja składa się ze składnika losowego oraz kombinacji liniowej składników losowych z przeszłości a wartość szeregu czasowego jest sumą składnika losowego oraz kombina-

cji liniowej poprzednich obserwacji. W rozdziałach 4.6.1 i 4.6.2 opisano kolejno procesy autoregresji i średniej ruchomej. W rozdziale 4.6.3 umieszczono zapis formalny metody ARIMA, będącej połączeniem modeli autoregresji i średniej ruchomej.

#### 4.6.1. Proces autoregresji

Szeregi czasowe złożone są z obserwacji, które zależą od siebie, a dokładniej zależą od wartości obserwacji wcześniejszych. Prowadzi to do sytuacji, w której możliwe jest oszacowanie równań lub współczynników pozwalających na obliczenie kolejnych wartości szeregu czasowego przy użyciu kolejnych wartości poprzednich.

Jeśli wartość szeregu w chwili  $t$  zależy tylko od wartości w chwili poprzedniej (lub chwilach poprzednich) oraz od składnika losowego to mówimy, że mamy proces autoregresji (AR-autoregression model).

Jeśli wartość szeregu w chwili  $t$  zależy tylko od wartości w chwili poprzedniej oraz od składnika losowego to mówimy, że mamy proces autoregresji rzędu pierwszego i oznaczamy AR(1) zgodnie z równaniem (4.16).

$$y_t = c + \phi y_{t-1} + \varepsilon_t \quad (4.16)$$

gdzie:

$\varepsilon_t$  – biały szum,

$c, \phi$  – parametry modelu (stałe).

Jeśli wartość szeregu w chwili  $t$  zależy tylko od wartości w dwóch chwilach poprzednich oraz od składnika losowego to mówimy, że mamy proces autoregresji rzędu drugiego i oznaczamy AR(2) zgodnie z równaniem (4.17).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t \quad (4.17)$$

gdzie:

$\phi_1, \phi_2$  – parametry modelu (stałe).

Jeśli wartość szeregu w chwili  $t$  zależy tylko od wartości w  $p$  chwilach poprzednich oraz od składnika losowego to mówimy, że mamy proces autoregresji rzędu  $p$  i oznaczamy AR( $p$ ) zgodnie z równaniem (4.18).

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (4.18)$$

gdzie:

$\phi_1, \dots, \phi_p$  – parametry modelu (stałe).

Często wykorzystywany jest zapis z wykorzystaniem operatora przesunięcia wstecz, oznaczany jako  $L$ . Operator przesunięcia wstecz jest definiowany następująco (4.19):

$$Ly_t = y_{t-1} \quad (4.19)$$

Kolejne potęgi operatora przesunięcia definiowane są następująco (4.20, 4.21, 4.22):

$$L^0 = 1 \quad (4.20)$$

$$L^2 y_t = LLy_t = Ly_{t-1} = y_{t-2} \quad (4.21)$$

$$L^p y_t = L^{p-1} y_{t-1} = y_{t-p} \quad (4.22)$$

Procesy autoregresji zapisane przy pomocy operatora przesunięcia wyglądają zgodnie z równaniami (4.23, 4.24, 4.25) odpowiednio dla AR(1), AR(2) i AR(p). Wielomiany, które powstają przy zapisie wykorzystującym operator przesunięcia  $L$  są nazywane wielomianami charakterystycznymi procesu.

$$(1 - \phi L)y_t = c + \varepsilon_t \quad (4.23)$$

$$(1 - \phi_1 L - \phi_2 L^2)y_t = c + \varepsilon_t \quad (4.24)$$

$$(1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p)y_t = c + \varepsilon_t \quad (4.25)$$

Warunkiem koniecznym w procesie autoregresji jest stacjonarność szeregu czasowego, czyli jego stała średnia (brak trendu), stała wariancja (homoskedastyczność) i stała korelacja wartości w czasie (szereg bez zmienności sezonowej). Jeżeli dane nie spełniają warunków stacjonarności, powinny być one odpowiednio przetransformowane w następujący sposób:

- zamiast oryginalnego szeregu czasowego budujemy nowy szereg złożony z przyrostów danych (szereg różnicowy):  $Y(i) = Z(i) - Z(i-1)$ ; stopień różnicowania określany jest jako  $d$ , można różnicować szereg wielokrotnie, zwykle wystarczające jest jednokrotne różnicowanie, rzadko dwukrotne, praktycznie nie spotyka się procesów zintegrowanych wyższego rzędu,
- w przypadku występowania trendu modelujemy ten trend, a następnie analizujemy (modelujemy) odchylenia resztowe,
- w przypadku występowania zmian wariancji w czasie obliczenie pierwiastka kwadratowego lub logarytmu z obserwacji szeregu może ustabilizować wariancję; w przypadku ujemnych wartości szeregu można przed transformacją danych dodać do każdego elementu stałą, dzięki której wszystkie dane będą większe od zera.

Wyżej opisane techniki służą do tego, aby w efekcie ich stosowania uzyskać szereg czasowy ze stałym położeniem i skalą.

Formalnie na pytanie o stopień zintegrowania (czyli czy i ile razy szereg należy różnicować) odpowiadają tzw. testy pierwiastka jednostkowego. Najczęściej stosowane są w tym celu: integracyjna statystyka Durbina Watsona, test DF (Dickey'a-Fullera) oraz rozszerzony test Dickey'a-Fullera ADF.

- **Integracyjna statystyka Durбина Watsona IDW**

Model szeregu czasowego rozważanego w testach pierwiastka jednostkowego przyjmuje postać (4.26):

$$y_t = \phi y_{t-1} + \varepsilon_t \quad (4.26)$$

Rozważane będą: hipoteza zerowa  $H_0$  i hipoteza alternatywna  $H_1$  dane poniżej.

$H_0: \phi = 1$  proces jest zintegrowany w stopniu 1,  $I(1)$ ,  $d = 1$ , proces niestacjonarny,

$H_1: \phi < 1$  proces jest zintegrowany w stopniu 0,  $I(0)$ ,  $d = 0$ , proces stacjonarny.

Integracyjna statystyka Durбина Watsona wyrażona jest wzorem 4.27.

$$IDW = \frac{\sum_{t=2}^n (y_t - y_{t-1})^2}{\sum_{t=1}^n (y_t - \bar{y}_t)^2} \quad (4.27)$$

Jeżeli prawdziwa jest hipoteza zerowa  $\phi = 1$ , wtedy suma w liczniku powinna być mała w stosunku do sumy w mianowniku. Dlatego wartość statystyki IDW powinna być bliska zero. W praktyce przyjmuje się, że wartość statystyki IDW mniejsza od 0,5 świadczy o braku stacjonarności (szereg należy różnicować). Jeśli wartość statystyki IDW jest bliska wartości 2 to można twierdzić, że proces jest stacjonarny i nie zachodzi konieczność różnicowania go (jest on zintegrowany w stopniu 0).

- **Test Dickey'a-Fullera (DF)**

Przyjmujemy, że model szeregu tak jak poprzednio przyjmuje postać zgodną z równaniem (4.26). Aby zastosować test DF w praktyce wymaga on przekształcenia modelu do następującej postaci 4.30 (przekształcając 4.26 zgodnie z 4.28-4.30):

czyli: 
$$y_t - y_{t-1} = \phi y_{t-1} - y_{t-1} + \varepsilon_t \quad (4.28)$$

$$\Delta y_t = (\phi - 1)y_{t-1} + \varepsilon_t \quad (4.29)$$

$$\Delta y_t = \delta y_{t-1} + \varepsilon_t \quad (4.30)$$

gdzie:  $\delta = \phi - 1$

Wówczas można założyć odpowiedni układ hipotez:

$$H_0 : \delta = 0, \quad (Y_t \approx I(1))$$

$$H_1 : \delta < 0, \quad (Y_t \approx I(0))$$

Hipoteza zerowa jest jednoznaczna z tym, że proces jest zintegrowany pierwszego rzędu, natomiast hipoteza alternatywna wskazuje na to, że proces jest zintegrowany w stopniu zero, czyli jest stacjonarny.

Statystyka testu Dickey'a-Fullera sprawdzająca powyższy układ hipotez przedstawiona jest wzorem 4.31.

$$DF = \frac{\hat{\delta}}{S(\hat{\delta})} \quad (4.31)$$

gdzie:

$\hat{\delta}$  – ocena parametru z równania 4.15

$S(\hat{\delta})$  – średni błąd szacunku parametru  $\delta$ .

Obliczoną wartość DF należy porównać z wartością krytyczną  $D_{\alpha}$  z tablic (tab. 4.22) wartości testu Dickey'a-Fullera. Wartości krytyczne testu Dickey'a-Fullera otrzymywane są za pomocą symulacji Monte Carlo, więc są one obciążone pewnym błędem. Dlatego tablice podają nie jedną ale dwie wartości krytyczne dolną i górną. Pomiędzy nimi leży obszar braku konkluzji. Kryteria wyboru odpowiedniej hipotezy są następujące:

- Jeżeli wartość statystyki  $DF < DF_{\alpha}$  to na poziomie istotności  $\alpha$  odrzucamy  $H_0$ , na korzyść  $H_1$ , proces  $Y_t$  jest stacjonarny, czyli zintegrowany rzędu zero.
- Jeżeli wartość statystyki  $DF \geq DF_{\alpha}$  to na poziomie istotności  $\alpha$  jest brak podstaw do odrzucenia  $H_0$ , zakładającej, że proces  $Y_t$  jest zintegrowany rzędu co najmniej 1, należy kontynuować procedurę dla sprawdzenia wyższego stopnia integracji  $d \geq 1$ .

W celu wyznaczenia wyższego stopnia integracji, szacujemy następujące równanie (4.32):

$$\Delta^2 y_t = \delta \Delta y_{t-1} + \varepsilon_t \quad (4.32)$$

Wówczas można założyć odpowiedni układ hipotez:

$$H_0 : \delta = 0, \quad (Y_t \approx I(2))$$

$$H_1 : \delta < 0, \quad (Y_t \approx I(1))$$

Hipoteza zerowa zakłada, że proces jest zintegrowany rzędu drugiego, natomiast hipoteza alternatywna oznacza, że proces jest zintegrowany rzędu pierwszego. Kryteria wyboru odpowiedniej hipotezy są następujące:

- Jeżeli wartość statystyki  $DF < DF_{\alpha}$  to na poziomie istotności  $\alpha$  odrzucamy  $H_0$ , na korzyść  $H_1$ , proces  $Y_t$  jest stacjonarny, czyli zintegrowany rzędu pierwszego.
- Jeżeli wartość statystyki  $DF \geq DF_{\alpha}$  to na poziomie istotności  $\alpha$  jest brak podstaw do odrzucenia  $H_0$ , zakładającej, że proces  $Y_t$  jest zintegrowany rzędu co najmniej 2, należy kontynuować procedurę dla sprawdzenia wyższego stopnia integracji  $d \geq 2$ . w praktyce niezwykle rzadko szeregi czasowe zdarzają się być zintegrowanymi w wyższym stopniu niż 2.



Tab. 4.22. Tablica rozkładu statystyki DF testu Dickey'a Fuller

Tablica rozkładu statystyki DF testu Dickey'a Fuller $\Delta y_t = \delta y_{t-1} + \varepsilon_t$ model bez stałej i bez trendu n- liczba obserwacji, $\alpha = 0,05$								
n	Wartość krytyczna		n	Wartość krytyczna		n	Wartość krytyczna	
	dolna	górna		dolna	górna		dolna	górna
10	-2,0612	-2,0894	25	-1,9601	-1,9883	125	-1,9329	-1,9611
11	-2,0417	-2,0699	30	-1,9529	-1,9811	150	-1,9321	-1,9604
12	-2,0266	-2,0548	35	-1,9483	-1,9765	200	-1,9313	-1,9595
13	-2,0146	-2,0428	40	-1,9450	-1,9733	250	-1,9308	-1,9590
14	-2,0049	-2,0332	45	-1,9427	-1,9709	300	-1,9304	-1,9587
15	-1,9970	-2,0252	50	-1,9409	-1,9691	400	-1,9301	-1,9583
16	-1,9904	-2,0186	60	-1,9384	-1,9666	500	-1,9298	-1,9580
17	-1,9848	-2,0130	70	-1,9368	-1,9650	600	-1,9297	-1,9579
18	-1,9801	-2,0083	80	-1,9356	-1,9638	750	-1,9295	-1,9577
19	-1,9760	-2,0042	90	-1,9347	-1,9629	1000	-1,9294	-1,9576
20	-1,9724	-2,0007	100	-1,9340	-1,9622	1250	-1,9293	-1,9575

#### 4.6.2. Proces średniej ruchomej

Jeżeli wartość szeregu czasowego w chwili  $t$  zależy wyłącznie od zaburzenia losowego w chwili  $t$  oraz od kilku zaburzeń poprzednich mamy do czynienia z modelem średniej ruchomej.

Jeżeli wartość szeregu czasowego w chwili  $t$  zależy wyłącznie od zaburzenia losowego w chwili  $t$  oraz  $t-1$  mamy proces średniej ruchomej rzędu pierwszego i oznaczamy go jako MA(1) zgodnie z równaniem (4.33), a przy wykorzystaniu operatora przesunięcia zgodnie z równaniem (4.34).

$$y_t = \mu + \varepsilon_t + \theta \varepsilon_{t-1} \quad (4.33)$$

$$y_t - \mu = (1 + \theta L)\varepsilon_t \quad (4.34)$$

gdzie:

$y_t$  – wartość szeregu czasowego w chwili  $t$ ,

$\varepsilon_t, \varepsilon_{t-1}$  – biały szum w chwilach odpowiednio  $t$  i  $t-1$ ,

$\mu, \theta$  – parametry modelu (stałe),

$L$  – operator przesunięcia wstecz.

Jeżeli wartość szeregu czasowego w chwili  $t$  zależy od zaburzenia losowego w chwili  $t, t-1$  i  $t-2$  to mamy proces średniej ruchomej rzędu drugiego i oznaczamy go jako MA(2) zgodnie z równaniem (4.35), a przy wykorzystaniu operatora przesunięcia zgodnie z równaniem (4.36).

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} \quad (4.35)$$

$$y_t - \mu = (1 + \theta_1 L + \theta_2 L^2) \varepsilon_t = \theta_2(L) \varepsilon_t \quad (4.36)$$

gdzie:

$\varepsilon_{t-2}$  – biały szum w chwili  $t-2$ ,

$\theta_1, \theta_2$  – parametry modelu (stałe).

Jeżeli wartość szeregu czasowego w chwili  $t$  zależy od zaburzenia losowego w chwili  $t, t-1, \dots, t-q$  to mamy proces średniej ruchomej rzędu  $q$  i oznaczamy go jako  $MA(q)$  zgodnie z równaniem (4.37), a przy wykorzystaniu operatora przesunięcia zgodnie z równaniem (4.38).

$$y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (4.37)$$

$$y_t - \mu = (1 + \theta_1 L + \theta_2 L^2 + \dots + \theta_q L^q) \varepsilon_t = \theta_q(L) \varepsilon_t \quad (4.38)$$

gdzie:

$\varepsilon_{t-q}$  – biały szum w chwili  $t-q$ ,

$\theta_q$  – parametr modelu (stała).

#### 4.6.3. Proces autoregresji

Jeśli wartość szeregu w chwili  $t$  zależy od wartości w chwili poprzedniej (lub chwilach poprzednich), od zaburzenia losowego w chwili  $t$  oraz od kilku zaburzeń poprzednich i od składnika losowego to łącząc wszystkie składowe otrzymujemy model ARIMA z trzema parametrami ARIMA  $(p, d, q)$ , gdzie  $p$  jest parametrem autoregresji,  $q$  stopniem integracji szeregu a  $q$  parametrem średniej ruchomej.

Prace badawcze i obliczeniowe z wykorzystaniem metody ARIMA prowadzili: Box (1970), Christodoulos (2010), Dickey (1981), Ediger (2007), Faruk (2010), Gilbert (2005, 2006), Han (2010), Ho (1998), Hodge (2011), Jia (2010), Khashei (2011), Kim (2011), Lee (2011), McGough (1994, 1995), Mitchell (2011), Mohammadi (2010), Ndongo (2010), Östermark (1988), Raymond (1997), Stevenson (2007), Tan (2010), Tse (1997), Venezian (2006), Wang (2011).

#### 4.6.4. Przykład obliczeniowy

Dla danych z przykładu 3.1 przeprowadzono prognozowanie wartości przyszłych predyktorów PR1 i PR2. Arkusz kalkulacyjny programu STATISTICA przedstawiono na rysunku 4.7.

Prognozowano wartości przyszłe predyktorów PR1 i PR2. Nie są znane predyktory predyktorów PR1 i PR2, zatem nie można wyznaczyć dla nich równań regresji. Można natomiast wykorzystując metodę ARIMA wyznaczyć ich wartości przyszłe. Poszukując wartości przyszłych zmiennej zależnej P1 i mając równanie regresji

$P1 = a+g(PR1)+q(PR2)+\varepsilon$  wyznaczone inną metodą prognostyczną (np. regresji wielorakiej MR), można obliczyć wartości przyszłe zmiennej zależnej P1 na podstawie obliczonych metodą ARIMA wartości przyszyłych predyktorów.

	1 t	2 P1	3 PR1	4 PR2	5 ARIMA1 PR1(1,0,0)	6 ARIMA2 PR1(1,0,1)op4	7 ARIMA3 PR1(1,0,0) op4	8 ARIMA4 PR2(1,0,0)	9 ARIMA5 PR2(4,0,1)op4	10 ARIMA6 PR2(4,0,2)op4
1	1	1,33	0,05	10	0,050000	0,05	0,05	10	10	10
2	2	1,36	0,0707107	13	0,070711	0,0707106781	0,0707106781	13	13	13
3	3	1,42	0,0866025	18	0,086603	0,0866025404	0,0866025404	18	18	18
4	4	1,48	0,1	25	0,100000	0,1	0,1	25	25	25
5	1	1,53	0,1118034	34	0,111803	0,111803399	0,111803399	34	34	34
6	2	1,59	0,1224745	45	0,122474	0,122474487	0,122474487	45	45	45
7	3	1,63	0,1322876	58	0,132288	0,132287566	0,132287566	58	58	58
8	4	1,68	0,1414214	73	0,141421	0,141421356	0,141421356	73	73	73
9	1	1,74	0,15	90	0,150000	0,15	0,15	90	90	90
10	2	1,76	0,1581139	109	0,158114	0,158113883	0,158113883	109	109	109
11	3	1,79	0,1658312	130	0,165831	0,16583124	0,16583124	130	130	130
12	4	1,83	0,1732051	153	0,173205	0,173205081	0,173205081	153	153	153
13	1	1,86	0,1802776	178	0,180278	0,180277564	0,180277564	178	178	178
14	2	1,89	0,1870829	205	0,187083	0,187082869	0,187082869	205	205	205
15	3	1,92	0,1936492	234	0,193649	0,193649167	0,193649167	234	234	234
16	4	1,95	0,2	265	0,200000	0,2	0,2	265	265	265
17	1	1,97	0,2061553	298	0,206155	0,206155281	0,206155281	298	298	298
18	2	2,00	0,212132	333	0,212132	0,212132034	0,212132034	333	333	333
19	3	2,12	0,2179449	370	0,217945	0,217944947	0,217944947	370	370	370
20	4	2,05	0,2236068	409	0,223607	0,223606798	0,223606798	409	409	409
21	1	2,07	0,2291288	450	0,229129	0,229128785	0,229128785	450	450	450
22	2	2,09	0,2345208	493	0,234521	0,234520788	0,234520788	493	493	493
23	3	2,12	0,2397916	538	0,239792	0,239791576	0,239791576	538	538	538
24	4	2,14	0,244949	585	0,244949	0,244948974	0,244948974	585	585	585
25	1	2,16	0,25	634	0,250000	0,25	0,25	634	634	634
26	2	2,28	0,254951	685	0,254951	0,254950976	0,254950976	685	685	685
27	3	2,20	0,2598076	738	0,259808	0,259807621	0,259807621	738	738	738
28	4	2,21	0,2645751	793	0,264575	0,264575131	0,264575131	793	793	793
29	1	2,28	0,2692582	850	0,269258	0,26925824	0,26925824	850	850	850
30	2	2,25	0,2738613	909	0,273861	0,273861279	0,273861279	909	909	909
31	3				0,273861	0,004437	0,273861	908,9994	892,3461	915,8424
32	4				0,273861	0,003591	0,273861	908,9988	899,5007	930,2460
33	1				0,273861	0,003902	0,273861	908,9982	915,7672	933,1806
34	2				0,273861	0,003787	0,273861	908,9975	907,3126	929,4221
35	3				0,273861	0,003830	0,273861	908,9969	907,8730	923,1768
36	4				0,273861	0,003814	0,273861	908,9963	913,2906	918,6984
37	1				0,273860	0,003820	0,273860	908,9957	910,0632	917,5710
38	2				0,273860	0,003818	0,273860	908,9951	909,8713	919,0547

Rys. 4.7. Arkusz kalkulacyjny programu STATISTICA z danymi wyjściowymi i wynikami prognozy przeprowadzonymi metodą autoregresji zintegrowanej ARIMA

Etapem pierwszym obliczeń jest sprawdzenie stacjonarności szeregów PR1 i PR2 z zastosowaniem integracyjnej statystyki IDW Durbina Watsona. W tym celu sporządzono arkusz kalkulacyjny w programie EXCEL. Jeśli statystyka IDW jest mniejsza od 0,5 można uznać szereg za stacjonarny. w tabelach 4.23. i 4.24. zaprezentowano arkusze kalkulacyjne do obliczenia statystyki IDW Durbina Watsona szeregów PR1 i PR2.

Statystyki IDW Durbina Watsona szeregów PR1 i PR2 wynoszą odpowiednio 0,0018141 i 0,0062858 i są mniejsze od wartości 0,5; oba szeregi można uznać za stacjonarne, nie istnieje konieczność ich przekształcania.

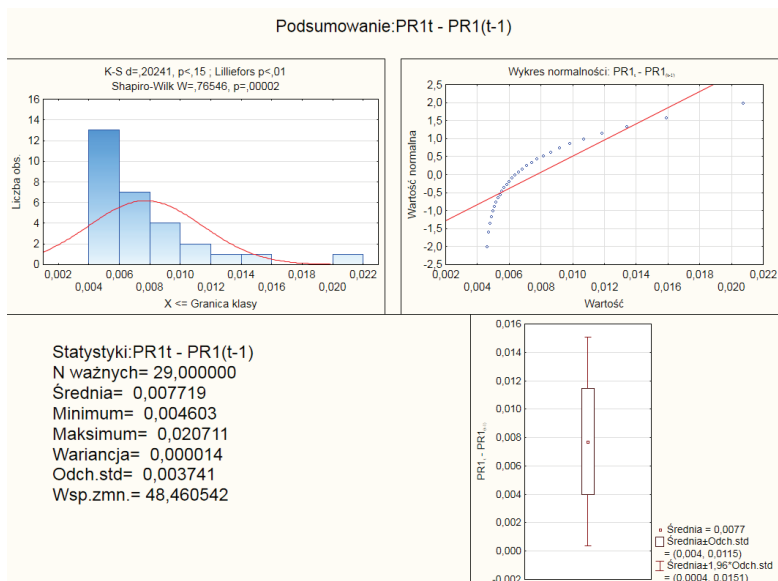
Celem potwierdzenia stacjonarności szeregów PR1 i PR2 wykonano obliczenia statystyk DF testu Dickey'a Fullera. Na podstawie analizy szeregów różnicowych rys 4.8 i 4.9 obliczono statystyki DF, porównano je z wartościami z tablicy rozkładu statystyki DF testu Dickey'a Fullera dla modelu bez stałej i bez trendu dla n = 29 obserwacji i  $\alpha = 0,05$ . Stwierdzono, że oba szeregi są stacjonarne.

Tab. 4.23. Arkusz kalkulacyjny do obliczenia statystyki IDW Durбина Watsona szeregu PR1

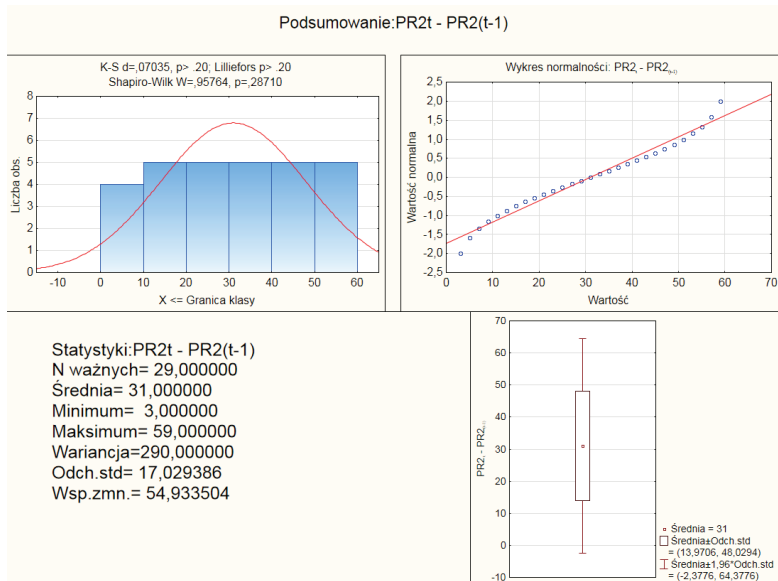
Lp.	$y_t$	$y_t - y_{(t-1)}$	$(y_t - y_{(t-1)})^2$	$y_t - \bar{y}_t$	$(y_t - \bar{y}_t)^2$
1	0,05				
2	0,0707107	0,02071068	0,000428932	-0,1160941	0,013477832
3	0,0866025	0,01589186	0,000252551	-0,1002022	0,010040481
4	0,1	0,01339746	0,000179492	-0,0868047	0,007535063
5	0,1118034	0,0118034	0,00013932	0,1118034	0,0125
6	0,1224745	0,01067109	0,000113872	0,1224745	0,015
7	0,1322876	0,00981308	9,62965E-05	0,1322876	0,0175
8	0,1414214	0,00913379	8,34261E-05	0,1414214	0,02
9	0,15	0,00857864	7,35931E-05	0,15	0,0225
10	0,1581139	0,00811388	6,58351E-05	0,1581139	0,025
11	0,1658312	0,00771736	5,95576E-05	0,1658312	0,0275
12	0,1732051	0,00737384	5,43735E-05	0,1732051	0,03
13	0,1802776	0,00707248	5,002E-05	0,1802776	0,0325
14	0,1870829	0,00680531	4,63122E-05	0,1870829	0,035
15	0,1936492	0,0065663	4,31163E-05	0,1936492	0,0375
16	0,2	0,00635083	4,03331E-05	0,2	0,04
17	0,2061553	0,00615528	3,78875E-05	0,2061553	0,0425
18	0,212132	0,00597675	3,57216E-05	0,212132	0,045
19	0,2179449	0,00581291	3,379E-05	0,2179449	0,0475
20	0,2236068	0,00566185	3,20566E-05	0,2236068	0,05
21	0,2291288	0,00552199	3,04923E-05	0,2291288	0,0525
22	0,2345208	0,005392	2,90737E-05	0,2345208	0,055
23	0,2397916	0,00527079	2,77812E-05	0,2397916	0,0575
24	0,244949	0,0051574	2,65988E-05	0,244949	0,06
25	0,25	0,00505103	2,55129E-05	0,25	0,0625
26	0,254951	0,00495098	2,45122E-05	0,254951	0,065
27	0,2598076	0,00485665	2,3587E-05	0,2598076	0,0675
28	0,2645751	0,00476751	2,27292E-05	0,2645751	0,07
29	0,2692582	0,00468311	2,19315E-05	0,2692582	0,0725
30	0,2738613	0,00460304	2,1188E-05	0,2738613	0,075
Średnia — $\bar{y}_t$	0,1868047				
suma			0,002119894		1,168553376
IDW	0,0018141 < 0,5 STACJONARNY				

Tab. 4.24. Arkusz kalkulacyjny do obliczenia statystyki IDW Durбина Watsona szeregu PR2

Lp.	$y_t$	$y_t - y_{(t-1)}$	$(y_t - y_{(t-1)})^2$	$y_t - \bar{y}_t$	$(y_t - \bar{y}_t)^2$
1	10				
2	13	3	9	-311,16667	96824,694
3	18	5	25	-306,16667	93738,028
4	25	7	49	-299,16667	89500,694
5	34	9	81	34	1156
6	45	11	121	45	2025
7	58	13	169	58	3364
8	73	15	225	73	5329
9	90	17	289	90	8100
10	109	19	361	109	11881
11	130	21	441	130	16900
12	153	23	529	153	23409
13	178	25	625	178	31684
14	205	27	729	205	42025
15	234	29	841	234	54756
16	265	31	961	265	70225
17	298	33	1089	298	88804
18	333	35	1225	333	110889
19	370	37	1369	370	136900
20	409	39	1521	409	167281
21	450	41	1681	450	202500
22	493	43	1849	493	243049
23	538	45	2025	538	289444
24	585	47	2209	585	342225
25	634	49	2401	634	401956
26	685	51	2601	685	469225
27	738	53	2809	738	544644
28	793	55	3025	793	628849
29	850	57	3249	850	722500
30	909	59	3481	909	826281
średnia	324,16667				
suma			35989		5725464,4
IDW	0,0062858 < 0,5 STACJONARNY				



Rys 4.8. Podsumowanie statystyk podstawowych szeregu różnicowego  $PR1_t - PR1_{(t-1)}$   
 $DF = 0,07719 / 0,003741 = 1,2376 < DF_\alpha (1,9529; 1,9811)$ . Proces PR1 jest stacjonarny, czyli zintegrowany rzędu pierwszego.



Rys 4.9. Podsumowanie statystyk podstawowych szeregu różnicowego  $PR2_t - PR2_{(t-1)}$   
 $DF = 31 / 17,0294 = 1,82 < DF_\alpha (1,9529; 1,9811)$ . Proces PR2 jest stacjonarny, czyli zintegrowany rzędu pierwszego.

Następnie przeprowadzono test Levene'a jednorodności wariancji. Prawidłowość obliczeń metodą ARIMA wymaga aby zmienna była homoskedastyczna (posiadała stałą wariancję). Najmocniejszym testem do sprawdzania tego założenia jest test Levene'a. Ograniczeniem testu jest to, że grupy porównywane muszą mieć tę samą liczebność. Jeśli sprawdzana jest homoskedastyczność jednej zmiennej np. PR1 o liczebności  $n = 30$  to dzielimy zmienną na grupy o tej samej liczebności i sprawdzamy jednorodność wariancji w tych grupach. W przykładzie obliczeniowym utworzono 3 nowe zmienne PR1/1, PR1/2 i PR1/3, wynikające z podziału zmiennej PR1. Grupa PR1/1 składa się z pierwszych 10 przypadków zmiennej PR1, PR1/2 z przypadków od 11 do 20, a PR1/3 z przypadków od 21 do 30. W ten sposób zmienna PR1 może być poddawana testowi Levene'a jednorodności wariancji. W teście tym obliczana jest statystyka F dana wzorem (4.39):

$$F = \frac{(n - k) \sum_{i=1}^k [n_i (\bar{x}_i - \bar{x})^2]}{(k - 1) \sum_{i=1}^k [s_i^2 (n_i - 1)]} \quad (4.39)$$

gdzie:

$n$  – liczba wszystkich obserwacji w  $k$  grupach

$k$  – liczba porównywanych grup

$\bar{x}_i$  – średnia wyników w każdej z badanych grup

$\bar{x}$  – średnia ogólna w całej badanej grupie

$s_i^2$  – wariancja wyników w  $i$ -tej próbie

$n_i$  – liczebność każdej badanej grupy

Zakłada się hipotezy: zerową  $H_0$  i alternatywną  $H_{1,0}$  następującym brzmieniu:

$H_0$ : Wariancja zmiennej zależnej w grupach czynnika jest jednakowa

$H_1$ : Wariancja zmiennej zależnej w grupach czynnika różni się.

Jeśli istotność testu (p-wartość) jest mniejsza od 0,05 to odrzucamy hipotezę zerową zakładającą równość wariancji. Przyjmujemy, że wariancje różnią się istotnie statystycznie ( $p < 0,05$ ).

Jeśli istotność testu (p-wartość) jest większa od 0,05 to nie mamy podstaw do odrzucenia hipotezy zerowej zakładającej równość wariancji.

Zmienne PR1 i PR2 z przykładu 3.1 poddano analizie jednorodności wariancji wykorzystując test Levene'a. Każdą ze zmiennych podzielono na 3 grupy, zgodnie z zasadami opisanymi powyżej i wykonano obliczenia w programie STATISTICA. Wyniki obliczeń zamieszczono w tabelach 4.25. i 4.26.

Tab. 4.25. Wyniki testu Levene'a jednorodności wariancji zmiennej PR1 podzielonej na grupy PR1/1, PR1/2 i PR1/3

Zmienna	Test Levene'a jednorodności wariancji Zaznaczone efekty są istotne z $p < ,05000$							
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
PR1/1	0,000369	3	0,000123	0,002208	6	0,000368	0,334276	0,801571
PR1/2	0,000054	3	0,000018	0,000802	6	0,000134	0,135464	0,935301
PR1/3	0,000030	3	0,000010	0,000499	6	0,000083	0,118936	0,945660

Tab. 4.26. Wyniki testu Levene'a jednorodności wariancji zmiennej PR2 podzielonej na grupy PR2/1, PR2/2 i PR2/3

Zmienna	Test Levene'a jednorodności wariancji Zaznaczone efekty są istotne z $p < ,05000$							
	SS Efekt	df Efekt	MS Efekt	SS Błąd	df Błąd	MS Błąd	F	p
PR2/1	343,230	3	114,410	1616,59	6	269,432	0,424633	0,742523
PR2/2	1980,933	3	660,311	10032,67	6	1672,111	0,394897	0,761613
PR2/3	3443,674	3	1147,891	49972,15	6	8328,691	0,137824	0,933797

W wyniku przeprowadzonych obliczeń stwierdzono, że istotność testu (p-wartość) jest większa od 0,05 w obu przypadkach i nie ma podstaw do odrzucenia hipotezy zerowej zakładającej równość wariancji.

Kolejnym krokiem jest wykonanie obliczeń metodą ARIMA. Obliczenia metodą ARIMA przeprowadza się ustalając kolejno parametry modelu. Zwykle rozpoczyna się od nadania parametrowi autoregresyjnemu p wartości 1, a pozostałym parametrom wartości zero tabela 4.27.

Tab. 27. Proces autoregresji zintegrowanej ARIMA1 PR1(1,0,0)op4 ; parametr autoregresyjny p = 1, pozostałe parametry modelu równe zero, opóźnienie sezonowe 4 zgodnie ze zmienną t, obliczenia dla 30 przypadków i 10 okresów prognozowanych

METODA OBLICZEŃ	<b>Proces autoregresji zintegrowanej ARIMA1 PR1(1,0,0) 30 przypadków</b>
ZMIENNA PROGNOZOWANA	PR1
PARAMETRY MODELU	<div style="border: 1px solid gray; padding: 5px;">                     Parametry modelu ARIMA  <input checked="" type="checkbox"/> Szacuj stałą      Opóźnienie sezonowe: 4 <input type="text"/>                      p - autoregresyjne: 1 <input type="text"/>      P - sezonowe: 0 <input type="text"/>                      q - średnia ruchoma: 0 <input type="text"/>      Q - sezonowe: 0 <input type="text"/> </div>

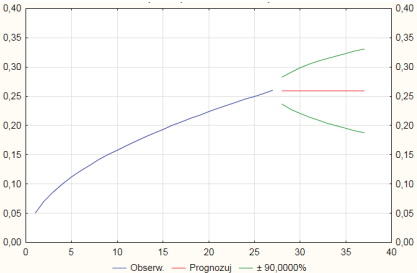
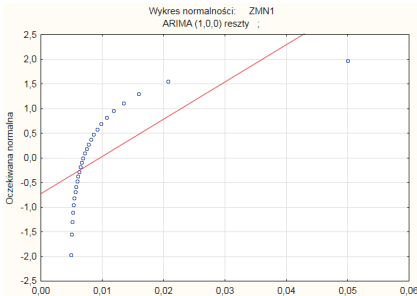


<p>WYKRES WARTOŚCI RZECZYWISTYCH I PROGNOZOWANYCH</p>	
<p>WYKRES NORMALNOŚCI RESZT</p>	
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT I AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>OCENA PRAWIDŁOWOŚCI STATYSTYCZNEJ</p>	<p>PRAWIDŁOWY</p>
<p>OCENA PRAWIDŁOWOŚCI PROGNOZYTICZNEJ</p>	<p>PRAWIDŁOWY</p>

Nie stwierdzono występowania autokorelacji i autokorelacji cząstkowych reszt nie ma zatem podstaw do odrzucenia modelu jako nieprawidłowy, zatem należy stwierdzić, że wygenerowany model jest prawidłowy statystycznie. Ocena prognozy może być również uznana za prawidłową, ale jedynie dla jednego lub dwóch okresów przyszłych. Na wykresie wartości rzeczywistych i prognozowanych wyraźnie widać, że wartości prognozowane nie zachowały trendu rosnącego wartości rzeczywistych. Dlatego prawdopodobnie tylko prognoza dla jednego lub dwóch okresów do przodu byłaby prawidłowa. Można to sprawdzić w przyszłości kiedy znane będą wartości rzeczywiste, policzyć błąd ex post *MAPE* i w ten sposób określić prawidłowość prognozy.

Istnieje też inny sposób sprawdzania prawidłowości prognozy metodą ARIMA. Z szeregu odcinamy kilka ostatnich przypadków i zakładamy, że jeszcze się nie wydarzyły. Wykonujemy prognozowanie dla szeregu skróconego, a otrzymane wartości porównujemy z wartościami odciętymi. Następnie obliczamy błąd *MAPE* pomiędzy nimi. Proces ten zobrazowano w tabeli 4.28.

Tab. 4.28. Proces autoregresji zintegrowanej ARIMA1 PR1(1,0,0)op4 27 przypadków ; parametr autoregresyjny  $p = 1$ , pozostałe parametry modelu równe zero, opóźnienie sezonowe 4 zgodnie ze zmienną  $t$ , obliczenia dla 27 przypadków i 10 okresów prognozowanych, błąd *MAPE* policzony dla przypadków od 28 do 30

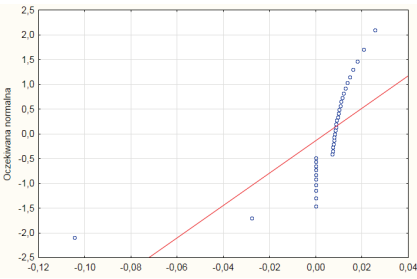
METODA OBLICZEŃ	<b>Proces autoregresji zintegrowanej ARIMA2 PR1(1,0,0) 27 przypadków</b>
ZMIENNA PROGNOZOWANA	PR1
PARAMETRY MODELU	<div data-bbox="494 560 1022 742" style="border: 1px solid gray; padding: 5px;"> <p>Parametry modelu ARIMA</p> <p><input checked="" type="checkbox"/> Szacuj stałą      Opóźnienie sezonowe: 4</p> <p>p - autoregresyjne: 1      P - sezonowe: 0</p> <p>q - średnia ruchoma: 0      Q - sezonowe: 0</p> </div>
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	
WYKRES NORMALNOŚCI RESZT	

<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>OCENA PRAWIDŁOWOŚCI STATYSTYCZNEJ</p>	<p>NIEPRAWIDŁOWY</p>
<p>BLĄD MAPE [%]</p>	<p>3,535285 prognoza dobra</p>
<p>OCENA PRAWIDŁOWOŚCI PROGNOSTYCZNEJ</p>	<p>PRAWIDŁOWY</p>

Po odcięciu trzech przypadków ze zmiennej wygenerowany model ARIMA1 PR1(1,0,0) 27 przypadków okazał się nieprawidłowy statystycznie – istnieją autokorelacje i autokorelacje cząstkowe reszt. Pomimo, że błąd prognozy jest niewielki i wynosi 3,535285% model nie może być przyjęty do stosowania. Należy poszukiwać takiego modelu ARIMA, który spełniałby wszystkie wymagania. Policzono model ARIMA3 PR1(2,0,0) 27 przypadków. Wyniki obliczeń zamieszczono w tabeli 4.29.

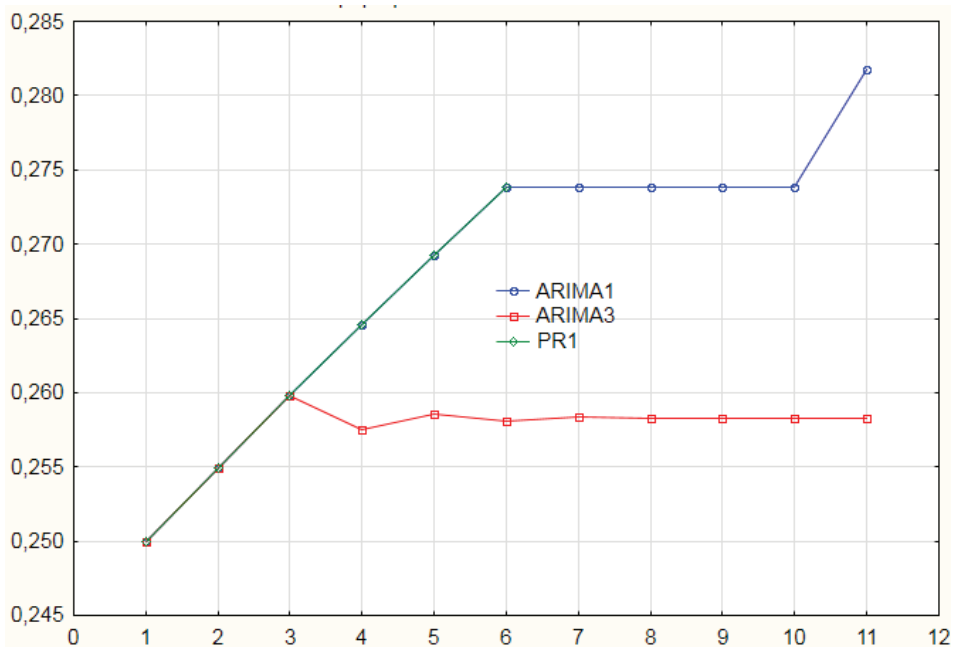
Tab. 4.29. Proces autoregresji zintegrowanej ARIMA3 PR1(2,0,0)op4 27 przypadków; parametr autoregresyjny  $p = 2$ , pozostałe parametry modelu równe zero, opóźnienie sezonowe 4 zgodnie ze zmienną  $t$ , obliczenia dla 27 przypadków i 10 okresów prognozowanych, błąd MAPE policzony dla przypadków 28–30

<p>METODA OBLICZEŃ</p>	<p><b>Proces autoregresji zintegrowanej ARIMA3 PR1(2,0,0) op4 27 przypadków</b></p>
<p>ZMIENNA PROGNOZOWANA</p>	<p>PR1</p>
<p>PARAMETRY MODELU</p>	<p>Parametry modelu ARIMA</p> <p><input checked="" type="checkbox"/> Szacuj stałą      Opóźnienie sezonowe: 4</p> <p>p - autoregresyjne: 2      P - sezonowe: 0</p> <p>q - średnia ruchoma: 0      Q - sezonowe: 0</p>
<p>WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH</p>	

<p>WYKRES NORMALNOŚCI RESZT</p>																																																																																																																																									
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	<table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th colspan="4">Funkcja autokorelacji</th> <th colspan="4">Funkcja autokorelacji cząstkowej</th> </tr> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E</th> <th>p</th> <th>Opóźn.</th> <th>Kor.</th> <th>S.E</th> <th>p</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,304</td><td>,1579</td><td>3,71 ,0542</td><td>1</td><td>+ ,304</td><td>,1644</td><td>4,05 ,1320</td></tr> <tr><td>2</td><td>- ,091</td><td>,1557</td><td>4,21 ,2394</td><td>2</td><td>- ,202</td><td>,1644</td><td>4,31 ,3655</td></tr> <tr><td>3</td><td>- ,062</td><td>,1535</td><td>4,38 ,4959</td><td>3</td><td>+ ,038</td><td>,1644</td><td>4,44 ,6176</td></tr> <tr><td>4</td><td>- ,047</td><td>,1512</td><td>4,49 ,7221</td><td>4</td><td>- ,064</td><td>,1644</td><td>4,54 ,8056</td></tr> <tr><td>5</td><td>- ,039</td><td>,1489</td><td>4,54 ,8056</td><td>5</td><td>- ,012</td><td>,1644</td><td>4,59 ,8686</td></tr> <tr><td>6</td><td>- ,035</td><td>,1466</td><td>4,64 ,9137</td><td>6</td><td>- ,033</td><td>,1644</td><td>4,70 ,9449</td></tr> <tr><td>7</td><td>- ,032</td><td>,1442</td><td>4,70 ,9449</td><td>7</td><td>- ,023</td><td>,1644</td><td>4,75 ,9657</td></tr> <tr><td>8</td><td>- ,031</td><td>,1418</td><td>4,81 ,9791</td><td>8</td><td>- ,028</td><td>,1644</td><td>4,81 ,9791</td></tr> <tr><td>9</td><td>- ,031</td><td>,1393</td><td>4,87 ,9875</td><td>9</td><td>- ,026</td><td>,1644</td><td>4,87 ,9875</td></tr> <tr><td>10</td><td>- ,032</td><td>,1368</td><td>4,94 ,9926</td><td>10</td><td>- ,028</td><td>,1644</td><td>4,94 ,9926</td></tr> <tr><td>11</td><td>- ,031</td><td>,1342</td><td></td><td>11</td><td>- ,027</td><td>,1644</td><td></td></tr> <tr><td>12</td><td>- ,031</td><td>,1316</td><td></td><td>12</td><td>- ,029</td><td>,1644</td><td></td></tr> <tr><td>13</td><td>- ,031</td><td>,1290</td><td></td><td>13</td><td>- ,030</td><td>,1644</td><td></td></tr> <tr><td>14</td><td>- ,032</td><td>,1262</td><td></td><td>14</td><td>- ,031</td><td>,1644</td><td></td></tr> <tr><td>15</td><td>- ,032</td><td>,1235</td><td></td><td>15</td><td>- ,032</td><td>,1644</td><td></td></tr> </tbody> </table>	Funkcja autokorelacji				Funkcja autokorelacji cząstkowej				Opóźn.	Kor.	S.E	p	Opóźn.	Kor.	S.E	p	1	+ ,304	,1579	3,71 ,0542	1	+ ,304	,1644	4,05 ,1320	2	- ,091	,1557	4,21 ,2394	2	- ,202	,1644	4,31 ,3655	3	- ,062	,1535	4,38 ,4959	3	+ ,038	,1644	4,44 ,6176	4	- ,047	,1512	4,49 ,7221	4	- ,064	,1644	4,54 ,8056	5	- ,039	,1489	4,54 ,8056	5	- ,012	,1644	4,59 ,8686	6	- ,035	,1466	4,64 ,9137	6	- ,033	,1644	4,70 ,9449	7	- ,032	,1442	4,70 ,9449	7	- ,023	,1644	4,75 ,9657	8	- ,031	,1418	4,81 ,9791	8	- ,028	,1644	4,81 ,9791	9	- ,031	,1393	4,87 ,9875	9	- ,026	,1644	4,87 ,9875	10	- ,032	,1368	4,94 ,9926	10	- ,028	,1644	4,94 ,9926	11	- ,031	,1342		11	- ,027	,1644		12	- ,031	,1316		12	- ,029	,1644		13	- ,031	,1290		13	- ,030	,1644		14	- ,032	,1262		14	- ,031	,1644		15	- ,032	,1235		15	- ,032	,1644	
Funkcja autokorelacji				Funkcja autokorelacji cząstkowej																																																																																																																																					
Opóźn.	Kor.	S.E	p	Opóźn.	Kor.	S.E	p																																																																																																																																		
1	+ ,304	,1579	3,71 ,0542	1	+ ,304	,1644	4,05 ,1320																																																																																																																																		
2	- ,091	,1557	4,21 ,2394	2	- ,202	,1644	4,31 ,3655																																																																																																																																		
3	- ,062	,1535	4,38 ,4959	3	+ ,038	,1644	4,44 ,6176																																																																																																																																		
4	- ,047	,1512	4,49 ,7221	4	- ,064	,1644	4,54 ,8056																																																																																																																																		
5	- ,039	,1489	4,54 ,8056	5	- ,012	,1644	4,59 ,8686																																																																																																																																		
6	- ,035	,1466	4,64 ,9137	6	- ,033	,1644	4,70 ,9449																																																																																																																																		
7	- ,032	,1442	4,70 ,9449	7	- ,023	,1644	4,75 ,9657																																																																																																																																		
8	- ,031	,1418	4,81 ,9791	8	- ,028	,1644	4,81 ,9791																																																																																																																																		
9	- ,031	,1393	4,87 ,9875	9	- ,026	,1644	4,87 ,9875																																																																																																																																		
10	- ,032	,1368	4,94 ,9926	10	- ,028	,1644	4,94 ,9926																																																																																																																																		
11	- ,031	,1342		11	- ,027	,1644																																																																																																																																			
12	- ,031	,1316		12	- ,029	,1644																																																																																																																																			
13	- ,031	,1290		13	- ,030	,1644																																																																																																																																			
14	- ,032	,1262		14	- ,031	,1644																																																																																																																																			
15	- ,032	,1235		15	- ,032	,1644																																																																																																																																			
<p>OCENA PRAWIDŁOŚCI STATYSTYCZNEJ</p>	<p>PRAWIDŁOWY</p>																																																																																																																																								
<p>BŁĄD MAPE [%]</p>	<p>4,121863 prognoza dobra5</p>																																																																																																																																								
<p>OCENA PRAWIDŁOŚCI PROGNOSTYCZNEJ</p>	<p>PRAWIDŁOWY</p>																																																																																																																																								

Otrzymano dwa modele prawidłowe ARIMA 1 i ARIMA 3, celem podjęcia decyzji o wyborze modelu analizowano wykresy w dużym zbliżeniu rys.4.10. Model ARIMA 3 zmienia trend dla 28 przypadku, model ARIMA3 dla 31. Bardziej prawdopodobny wydaje się być model ARIMA3. Zostaje zatem wybrany jako prawidłowy do dalszych obliczeń.

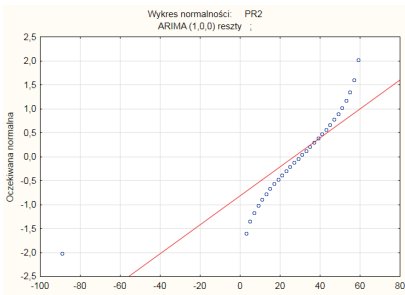
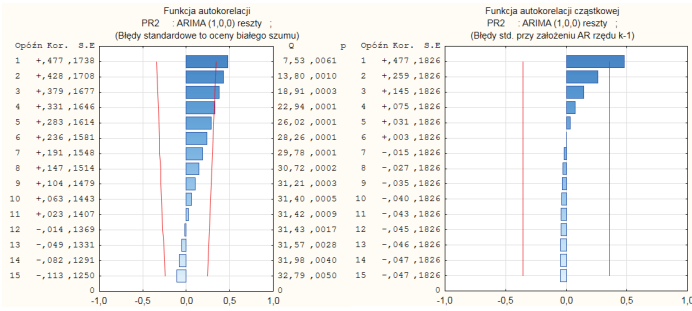
Podobnie jak w przypadku prognozowania wartości przyszłych predyktora PR1, powtarza się te same czynności poddając predykcji szereg PR2. Obliczenia tradycyjnie rozpoczyna się od nadania parametrowi autoregresyjnemu p wartości 1, a pozostałym parametrom wartości zero. Obliczenia ARIMA 4 PR2(1,0,0) op 4 dla 30 przypadków zestawiono w tabeli 4.30.



Rys. 4.10. Zestawienie graficzne wyników obliczeń metodą ARIMA, modele ARIMA 1 i ARIMA 3 dla przypadków od 25 do 35 oraz wartości zmiennej prognozowanej PR1 od 25 do 30

Tab. 4.30. Proces autoregresji zintegrowanej ARIMA 4 PR2(1,0,0)op4; parametr autoregresyjny  $p = 1$ , pozostałe parametry modelu równe zero, opóźnienie sezonowe 4 zgodnie ze zmienną  $t$ , obliczenia dla 30 przypadków i 10 okresów prognozowanych

METODA OBLICZEŃ	<b>Proces autoregresji zintegrowanej ARIMA4 PR2(1,0,0)op4</b>
ZMIENNA PROGNOZOWANA	PR2
PARAMETRY MODELU	<div style="border: 1px solid gray; padding: 5px;"> <p>Parametry modelu ARIMA</p> <p><input checked="" type="checkbox"/> Szacuj stałą      Opóźnienie sezonowe: 4</p> <p>p - autoregresyjne: 1      P - sezonowe: 0</p> <p>q - średnia ruchoma: 0      Q - sezonowe: 0</p> </div>
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	<p>Prognostyka: Model: (1,0,0) Opóź. sezon: 4 Resztowy MS=1568,3 Dane: PR2 Początek bazy: 1      Koniec bazy: 30</p> <p>— Obserw. — Prognozuj — ± 90,0000%</p>

<p>WYKRES NORMALNOŚCI RESZT</p>																																																																																																																																							
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	 <table border="1" data-bbox="413 496 1105 806"> <thead> <tr> <th colspan="3">Funkcja autokorelacji PR2 - ARIMA (1,0,0) reszty ; (Błędy standardowe to oceny białego szumu)</th> <th colspan="3">Funkcja autokorelacji cząstkowej PR2 - ARIMA (1,0,0) reszty ; (Błędy std. przy założeniu AR rzędu k-1)</th> </tr> <tr> <th>Opóźn</th> <th>Kor.</th> <th>s.e</th> <th>Q</th> <th>p</th> <th>Opóźn</th> <th>Kor.</th> <th>s.e</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,477</td><td>,1738</td><td>7,53</td><td>,0061</td><td>1</td><td>+,477</td><td>,1826</td></tr> <tr><td>2</td><td>+,428</td><td>,1708</td><td>13,80</td><td>,0010</td><td>2</td><td>+,259</td><td>,1826</td></tr> <tr><td>3</td><td>+,379</td><td>,1677</td><td>18,91</td><td>,0003</td><td>3</td><td>+,145</td><td>,1826</td></tr> <tr><td>4</td><td>+,331</td><td>,1646</td><td>22,94</td><td>,0001</td><td>4</td><td>+,075</td><td>,1826</td></tr> <tr><td>5</td><td>+,283</td><td>,1614</td><td>26,02</td><td>,0001</td><td>5</td><td>+,031</td><td>,1826</td></tr> <tr><td>6</td><td>+,236</td><td>,1581</td><td>28,26</td><td>,0001</td><td>6</td><td>+,003</td><td>,1826</td></tr> <tr><td>7</td><td>+,191</td><td>,1548</td><td>29,78</td><td>,0001</td><td>7</td><td>-,015</td><td>,1826</td></tr> <tr><td>8</td><td>+,147</td><td>,1514</td><td>30,72</td><td>,0002</td><td>8</td><td>-,027</td><td>,1826</td></tr> <tr><td>9</td><td>+,104</td><td>,1479</td><td>31,21</td><td>,0003</td><td>9</td><td>-,035</td><td>,1826</td></tr> <tr><td>10</td><td>+,063</td><td>,1443</td><td>31,40</td><td>,0005</td><td>10</td><td>-,040</td><td>,1826</td></tr> <tr><td>11</td><td>+,023</td><td>,1407</td><td>31,42</td><td>,0009</td><td>11</td><td>-,043</td><td>,1826</td></tr> <tr><td>12</td><td>-,014</td><td>,1369</td><td>31,43</td><td>,0017</td><td>12</td><td>-,045</td><td>,1826</td></tr> <tr><td>13</td><td>-,049</td><td>,1331</td><td>31,57</td><td>,0028</td><td>13</td><td>-,046</td><td>,1826</td></tr> <tr><td>14</td><td>-,082</td><td>,1291</td><td>31,98</td><td>,0040</td><td>14</td><td>-,047</td><td>,1826</td></tr> <tr><td>15</td><td>-,113</td><td>,1250</td><td>32,79</td><td>,0050</td><td>15</td><td>-,047</td><td>,1826</td></tr> </tbody> </table>	Funkcja autokorelacji PR2 - ARIMA (1,0,0) reszty ; (Błędy standardowe to oceny białego szumu)			Funkcja autokorelacji cząstkowej PR2 - ARIMA (1,0,0) reszty ; (Błędy std. przy założeniu AR rzędu k-1)			Opóźn	Kor.	s.e	Q	p	Opóźn	Kor.	s.e	1	+,477	,1738	7,53	,0061	1	+,477	,1826	2	+,428	,1708	13,80	,0010	2	+,259	,1826	3	+,379	,1677	18,91	,0003	3	+,145	,1826	4	+,331	,1646	22,94	,0001	4	+,075	,1826	5	+,283	,1614	26,02	,0001	5	+,031	,1826	6	+,236	,1581	28,26	,0001	6	+,003	,1826	7	+,191	,1548	29,78	,0001	7	-,015	,1826	8	+,147	,1514	30,72	,0002	8	-,027	,1826	9	+,104	,1479	31,21	,0003	9	-,035	,1826	10	+,063	,1443	31,40	,0005	10	-,040	,1826	11	+,023	,1407	31,42	,0009	11	-,043	,1826	12	-,014	,1369	31,43	,0017	12	-,045	,1826	13	-,049	,1331	31,57	,0028	13	-,046	,1826	14	-,082	,1291	31,98	,0040	14	-,047	,1826	15	-,113	,1250	32,79	,0050	15	-,047	,1826
Funkcja autokorelacji PR2 - ARIMA (1,0,0) reszty ; (Błędy standardowe to oceny białego szumu)			Funkcja autokorelacji cząstkowej PR2 - ARIMA (1,0,0) reszty ; (Błędy std. przy założeniu AR rzędu k-1)																																																																																																																																				
Opóźn	Kor.	s.e	Q	p	Opóźn	Kor.	s.e																																																																																																																																
1	+,477	,1738	7,53	,0061	1	+,477	,1826																																																																																																																																
2	+,428	,1708	13,80	,0010	2	+,259	,1826																																																																																																																																
3	+,379	,1677	18,91	,0003	3	+,145	,1826																																																																																																																																
4	+,331	,1646	22,94	,0001	4	+,075	,1826																																																																																																																																
5	+,283	,1614	26,02	,0001	5	+,031	,1826																																																																																																																																
6	+,236	,1581	28,26	,0001	6	+,003	,1826																																																																																																																																
7	+,191	,1548	29,78	,0001	7	-,015	,1826																																																																																																																																
8	+,147	,1514	30,72	,0002	8	-,027	,1826																																																																																																																																
9	+,104	,1479	31,21	,0003	9	-,035	,1826																																																																																																																																
10	+,063	,1443	31,40	,0005	10	-,040	,1826																																																																																																																																
11	+,023	,1407	31,42	,0009	11	-,043	,1826																																																																																																																																
12	-,014	,1369	31,43	,0017	12	-,045	,1826																																																																																																																																
13	-,049	,1331	31,57	,0028	13	-,046	,1826																																																																																																																																
14	-,082	,1291	31,98	,0040	14	-,047	,1826																																																																																																																																
15	-,113	,1250	32,79	,0050	15	-,047	,1826																																																																																																																																
<p>OCENA PRAWIDŁOŚCI STATYSTYCZNEJ</p>	<p>NIEPRAWIDŁOWY</p>																																																																																																																																						
<p>OCENA PRAWIDŁOŚCI PROGNOSTYCZNEJ</p>	<p>NIEPRAWIDŁOWY</p>																																																																																																																																						

Stwierdzono występowanie autokorelacji i autokorelacji cząstkowej reszt, są zatem podstawy do odrzucenia modelu jako nieprawidłowego statystycznie. Na wykresie wartości rzeczywistych i prognozowanych wyraźnie widać, że wartości prognozowane nie zachowały trendu rosnącego wartości rzeczywistych. Model ARIMA 4 PR2(1,0,0) op 4 dla 30 przypadków należy odrzucić jako całkowicie nieprawidłowy.

Należy wykonać kolejne obliczenia zwiększając wartości parametrów p autoregresyjnego, q średniej ruchomej, P opóźnienia sezonowego oraz Q opóźnienia sezonowego. Przykład obliczeń zestawiono w tabeli 4.31. Przedstawiono obliczenia dla procesu autoregresji zintegrowanej ARIMA 5 PR2(4,0,1)op 4 dla 30 przypadków.

Tab. 4.31. Proces autoregresji zintegrowanej ARIMA 5 PR2(4,0,1) op4; parametr autoregresyjny  $p = 4$ , parametr średniej ruchomej  $q = 1$ , pozostałe parametry modelu równe zero, opóźnienie sezonowe 4 zgodnie ze zmienną  $t$ , obliczenia dla 30 przypadków i 10 okresów prognozowanych

METODA OBLICZEŃ	<b>Proces autoregresji zintegrowanej ARIMA5 PR2(4,0,1) op4</b>																																																																																																		
ZMIENNA PROGNOZOWANA	PR2																																																																																																		
PARAMETRY MODELU	<div style="border: 1px solid gray; padding: 5px;"> <p>Parametry modelu ARIMA</p> <p><input checked="" type="checkbox"/> Szacuj stałą      Opóźnienie sezonowe: 4</p> <p>p - autoregresyjne: 4      P - sezonowe: 0</p> <p>q - średnia ruchoma: 1      Q - sezonowe: 0</p> </div>																																																																																																		
WYKRES WARTOŚCI RZECZYWISTYCH I PROGNOZOWANYCH																																																																																																			
WYKRES NORMALNOŚCI RESZT																																																																																																			
WYKRESY FUNKCJI AUTOKORELACJI RESZT I AUTOKORELACJI CZĄSTKOWYCH RESZT	<table border="0" style="width: 100%; font-size: small;"> <tr> <td style="width: 50%; vertical-align: top;"> <p>Funkcja autokorelacji PR2 : ARIMA (4,0,1) reszty ; (Błędy standardowe to oceny białego szumu)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,411</td><td>,1738</td></tr> <tr><td>2</td><td>+ ,252</td><td>,1708</td></tr> <tr><td>3</td><td>- ,005</td><td>,1677</td></tr> <tr><td>4</td><td>+ ,112</td><td>,1646</td></tr> <tr><td>5</td><td>+ ,111</td><td>,1614</td></tr> <tr><td>6</td><td>+ ,095</td><td>,1581</td></tr> <tr><td>7</td><td>+ ,078</td><td>,1548</td></tr> <tr><td>8</td><td>+ ,061</td><td>,1514</td></tr> <tr><td>9</td><td>+ ,044</td><td>,1479</td></tr> <tr><td>10</td><td>+ ,028</td><td>,1443</td></tr> <tr><td>11</td><td>+ ,012</td><td>,1407</td></tr> <tr><td>12</td><td>- ,004</td><td>,1369</td></tr> <tr><td>13</td><td>- ,018</td><td>,1331</td></tr> <tr><td>14</td><td>- ,033</td><td>,1291</td></tr> <tr><td>15</td><td>- ,046</td><td>,1250</td></tr> </tbody> </table> </td> <td style="width: 50%; vertical-align: top;"> <p>Funkcja autokorelacji cząstkowej PR2 : ARIMA (4,0,1) reszty ; (Błędy std przy założeniu AR rzędu k-1)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,411</td><td>,1826</td></tr> <tr><td>2</td><td>+ ,100</td><td>,1826</td></tr> <tr><td>3</td><td>- ,168</td><td>,1826</td></tr> <tr><td>4</td><td>+ ,178</td><td>,1826</td></tr> <tr><td>5</td><td>+ ,067</td><td>,1826</td></tr> <tr><td>6</td><td>- ,042</td><td>,1826</td></tr> <tr><td>7</td><td>+ ,062</td><td>,1826</td></tr> <tr><td>8</td><td>+ ,022</td><td>,1826</td></tr> <tr><td>9</td><td>- ,023</td><td>,1826</td></tr> <tr><td>10</td><td>+ ,008</td><td>,1826</td></tr> <tr><td>11</td><td>- ,007</td><td>,1826</td></tr> <tr><td>12</td><td>- ,026</td><td>,1826</td></tr> <tr><td>13</td><td>- ,018</td><td>,1826</td></tr> <tr><td>14</td><td>- ,025</td><td>,1826</td></tr> <tr><td>15</td><td>- ,033</td><td>,1826</td></tr> </tbody> </table> </td> </tr> </table>	<p>Funkcja autokorelacji PR2 : ARIMA (4,0,1) reszty ; (Błędy standardowe to oceny białego szumu)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,411</td><td>,1738</td></tr> <tr><td>2</td><td>+ ,252</td><td>,1708</td></tr> <tr><td>3</td><td>- ,005</td><td>,1677</td></tr> <tr><td>4</td><td>+ ,112</td><td>,1646</td></tr> <tr><td>5</td><td>+ ,111</td><td>,1614</td></tr> <tr><td>6</td><td>+ ,095</td><td>,1581</td></tr> <tr><td>7</td><td>+ ,078</td><td>,1548</td></tr> <tr><td>8</td><td>+ ,061</td><td>,1514</td></tr> <tr><td>9</td><td>+ ,044</td><td>,1479</td></tr> <tr><td>10</td><td>+ ,028</td><td>,1443</td></tr> <tr><td>11</td><td>+ ,012</td><td>,1407</td></tr> <tr><td>12</td><td>- ,004</td><td>,1369</td></tr> <tr><td>13</td><td>- ,018</td><td>,1331</td></tr> <tr><td>14</td><td>- ,033</td><td>,1291</td></tr> <tr><td>15</td><td>- ,046</td><td>,1250</td></tr> </tbody> </table>	Opóźn.	Kor.	S.E.	1	+ ,411	,1738	2	+ ,252	,1708	3	- ,005	,1677	4	+ ,112	,1646	5	+ ,111	,1614	6	+ ,095	,1581	7	+ ,078	,1548	8	+ ,061	,1514	9	+ ,044	,1479	10	+ ,028	,1443	11	+ ,012	,1407	12	- ,004	,1369	13	- ,018	,1331	14	- ,033	,1291	15	- ,046	,1250	<p>Funkcja autokorelacji cząstkowej PR2 : ARIMA (4,0,1) reszty ; (Błędy std przy założeniu AR rzędu k-1)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,411</td><td>,1826</td></tr> <tr><td>2</td><td>+ ,100</td><td>,1826</td></tr> <tr><td>3</td><td>- ,168</td><td>,1826</td></tr> <tr><td>4</td><td>+ ,178</td><td>,1826</td></tr> <tr><td>5</td><td>+ ,067</td><td>,1826</td></tr> <tr><td>6</td><td>- ,042</td><td>,1826</td></tr> <tr><td>7</td><td>+ ,062</td><td>,1826</td></tr> <tr><td>8</td><td>+ ,022</td><td>,1826</td></tr> <tr><td>9</td><td>- ,023</td><td>,1826</td></tr> <tr><td>10</td><td>+ ,008</td><td>,1826</td></tr> <tr><td>11</td><td>- ,007</td><td>,1826</td></tr> <tr><td>12</td><td>- ,026</td><td>,1826</td></tr> <tr><td>13</td><td>- ,018</td><td>,1826</td></tr> <tr><td>14</td><td>- ,025</td><td>,1826</td></tr> <tr><td>15</td><td>- ,033</td><td>,1826</td></tr> </tbody> </table>	Opóźn.	Kor.	S.E.	1	+ ,411	,1826	2	+ ,100	,1826	3	- ,168	,1826	4	+ ,178	,1826	5	+ ,067	,1826	6	- ,042	,1826	7	+ ,062	,1826	8	+ ,022	,1826	9	- ,023	,1826	10	+ ,008	,1826	11	- ,007	,1826	12	- ,026	,1826	13	- ,018	,1826	14	- ,025	,1826	15	- ,033	,1826
<p>Funkcja autokorelacji PR2 : ARIMA (4,0,1) reszty ; (Błędy standardowe to oceny białego szumu)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,411</td><td>,1738</td></tr> <tr><td>2</td><td>+ ,252</td><td>,1708</td></tr> <tr><td>3</td><td>- ,005</td><td>,1677</td></tr> <tr><td>4</td><td>+ ,112</td><td>,1646</td></tr> <tr><td>5</td><td>+ ,111</td><td>,1614</td></tr> <tr><td>6</td><td>+ ,095</td><td>,1581</td></tr> <tr><td>7</td><td>+ ,078</td><td>,1548</td></tr> <tr><td>8</td><td>+ ,061</td><td>,1514</td></tr> <tr><td>9</td><td>+ ,044</td><td>,1479</td></tr> <tr><td>10</td><td>+ ,028</td><td>,1443</td></tr> <tr><td>11</td><td>+ ,012</td><td>,1407</td></tr> <tr><td>12</td><td>- ,004</td><td>,1369</td></tr> <tr><td>13</td><td>- ,018</td><td>,1331</td></tr> <tr><td>14</td><td>- ,033</td><td>,1291</td></tr> <tr><td>15</td><td>- ,046</td><td>,1250</td></tr> </tbody> </table>	Opóźn.	Kor.	S.E.	1	+ ,411	,1738	2	+ ,252	,1708	3	- ,005	,1677	4	+ ,112	,1646	5	+ ,111	,1614	6	+ ,095	,1581	7	+ ,078	,1548	8	+ ,061	,1514	9	+ ,044	,1479	10	+ ,028	,1443	11	+ ,012	,1407	12	- ,004	,1369	13	- ,018	,1331	14	- ,033	,1291	15	- ,046	,1250	<p>Funkcja autokorelacji cząstkowej PR2 : ARIMA (4,0,1) reszty ; (Błędy std przy założeniu AR rzędu k-1)</p> <table border="1" style="width: 100%; border-collapse: collapse;"> <thead> <tr> <th>Opóźn.</th> <th>Kor.</th> <th>S.E.</th> </tr> </thead> <tbody> <tr><td>1</td><td>+ ,411</td><td>,1826</td></tr> <tr><td>2</td><td>+ ,100</td><td>,1826</td></tr> <tr><td>3</td><td>- ,168</td><td>,1826</td></tr> <tr><td>4</td><td>+ ,178</td><td>,1826</td></tr> <tr><td>5</td><td>+ ,067</td><td>,1826</td></tr> <tr><td>6</td><td>- ,042</td><td>,1826</td></tr> <tr><td>7</td><td>+ ,062</td><td>,1826</td></tr> <tr><td>8</td><td>+ ,022</td><td>,1826</td></tr> <tr><td>9</td><td>- ,023</td><td>,1826</td></tr> <tr><td>10</td><td>+ ,008</td><td>,1826</td></tr> <tr><td>11</td><td>- ,007</td><td>,1826</td></tr> <tr><td>12</td><td>- ,026</td><td>,1826</td></tr> <tr><td>13</td><td>- ,018</td><td>,1826</td></tr> <tr><td>14</td><td>- ,025</td><td>,1826</td></tr> <tr><td>15</td><td>- ,033</td><td>,1826</td></tr> </tbody> </table>	Opóźn.	Kor.	S.E.	1	+ ,411	,1826	2	+ ,100	,1826	3	- ,168	,1826	4	+ ,178	,1826	5	+ ,067	,1826	6	- ,042	,1826	7	+ ,062	,1826	8	+ ,022	,1826	9	- ,023	,1826	10	+ ,008	,1826	11	- ,007	,1826	12	- ,026	,1826	13	- ,018	,1826	14	- ,025	,1826	15	- ,033	,1826		
Opóźn.	Kor.	S.E.																																																																																																	
1	+ ,411	,1738																																																																																																	
2	+ ,252	,1708																																																																																																	
3	- ,005	,1677																																																																																																	
4	+ ,112	,1646																																																																																																	
5	+ ,111	,1614																																																																																																	
6	+ ,095	,1581																																																																																																	
7	+ ,078	,1548																																																																																																	
8	+ ,061	,1514																																																																																																	
9	+ ,044	,1479																																																																																																	
10	+ ,028	,1443																																																																																																	
11	+ ,012	,1407																																																																																																	
12	- ,004	,1369																																																																																																	
13	- ,018	,1331																																																																																																	
14	- ,033	,1291																																																																																																	
15	- ,046	,1250																																																																																																	
Opóźn.	Kor.	S.E.																																																																																																	
1	+ ,411	,1826																																																																																																	
2	+ ,100	,1826																																																																																																	
3	- ,168	,1826																																																																																																	
4	+ ,178	,1826																																																																																																	
5	+ ,067	,1826																																																																																																	
6	- ,042	,1826																																																																																																	
7	+ ,062	,1826																																																																																																	
8	+ ,022	,1826																																																																																																	
9	- ,023	,1826																																																																																																	
10	+ ,008	,1826																																																																																																	
11	- ,007	,1826																																																																																																	
12	- ,026	,1826																																																																																																	
13	- ,018	,1826																																																																																																	
14	- ,025	,1826																																																																																																	
15	- ,033	,1826																																																																																																	
OCENA PRAWIDŁOŚCI STATYSTYCZNEJ	NIEPRAWIDŁOWY																																																																																																		

OCENA PRAWIDŁOWOŚCI PROGNOSTYCZNEJ	NIEPRAWIDŁOWY
--	---------------

Podobnie jak w przypadku ARIMA 4 PR2(1,0,0) op 4 dla 30 przypadków, stwierdzono występowanie autokorelacji i autokorelacji cząstkowej reszt, są zatem podstawy do odrzucenia modelu jako nieprawidłowego statystycznie. Należy podejmować kolejne próby obliczeń aż do uzyskania efektu zadawalającego.

Metodą prób i błędów znaleziono model poprawny statystycznie. Jest to model ARIMA6 PR2(4,0,2)op4 liczony dla 30 przypadków. Wyniki obliczeń zestawiono w tabeli 4.32.

Tab. 4.32. Proces autoregresji zintegrowanej ARIMA 6 PR2(4,0,2)op4 ; parametr autoregresyjny  $p=4$ , parametr średniej ruchomej  $q=2$ , pozostałe parametry modelu równe zero, opóźnienie sezonowe 4 zgodnie ze zmienną  $t$ , obliczenia dla 30 przypadków i 10 okresów prognozowanych

METODA OBLICZEŃ	<b>Proces autoregresji zintegrowanej ARIMA6 PR2(4,0,2)op4</b>
ZMIENNA PROGNOZOWANA	PR2
PARAMETRY MODELU	<div style="border: 1px solid gray; padding: 5px;"> <p style="text-align: center;">Parametry modelu ARIMA</p> <p><input checked="" type="checkbox"/> Szacuj stałą      Opóźnienie sezonowe: 4</p> <p>p - autoregresyjne: 4      P - sezonowe: 0</p> <p>q - średnia ruchoma: 2      Q - sezonowe: 0</p> </div>
WYKRES WARTOŚCI RZECZYWISTYCH I PROGNOZOWANYCH	<p style="font-size: small;">Progniza, Model: (4,0,2) Opóź. sezon: 4 Resztowy MS=7282,0 Dane: PR2 Początek bazy: 1      Koniec bazy: 30</p>
WYKRES NORMALNOŚCI RESZT	<p style="font-size: small;">Wykres normalności: PR2 ARIMA (4,0,2) reszty</p>

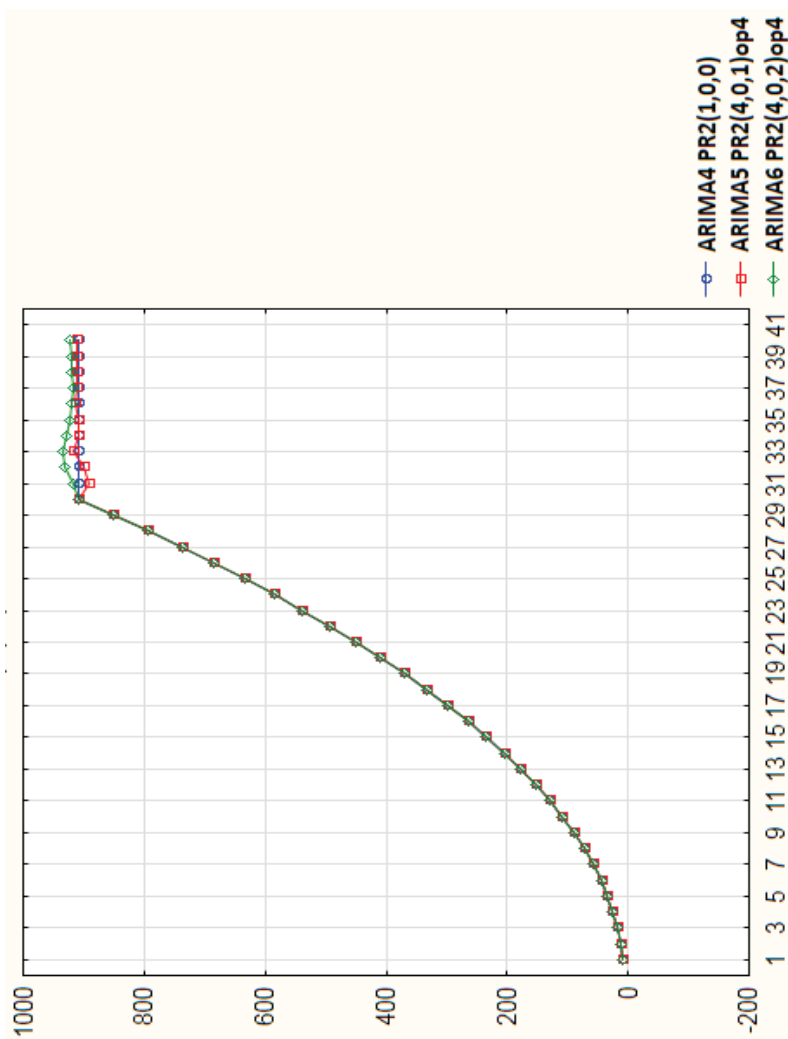


<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>OCENA PRAWIDŁOWOŚCI STATYSTYCZNEJ</p>	<p>PRAWIDŁOWY</p>
<p>OCENA PRAWIDŁOWOŚCI PROGNOSTYCZNEJ</p>	<p>PRAWIDŁOWY</p>

Nie stwierdzono występowania autokorelacji i autokorelacji cząstkowych reszt nie ma zatem podstaw do odrzucenia modelu jako nieprawidłowy, zatem należy stwierdzić, że wygenerowany model jest prawidłowy statystycznie. Ocena prognozy może być również uznana za prawidłową, ale jedynie dla jednego lub dwóch okresów przyszłych ze względu na zachowanie trendu rosnącego. Na rysunku 4.11 przedstawiono wartości szeregu PR2 wraz z prognozami wartości przyszłych obliczonych przy zastosowaniu modeli ARIMA 4, ARIMA 5 i ARIMA 6.

Statystycznie poprawny jest jedynie model ARIMA 6 i to on może być przyjęty do dalszych obliczeń, czyli do szacowania przyszłych wartości zmiennej zależnej P1 (zwykle o jeden lub dwa okresy do przodu). Mając wyznaczone wartości przeszłe PR1 i PR2 można korzystając z wzoru regresyjnego wyznaczonego w punktach 4.1 do 4.5 (wybiera się równanie regresji z najmniejszym błędem MAPE p.4.7). Istnieje również możliwość wyznaczania wartości przyszłych szeregu P1 bezpośrednio z niego samego, postępując zgodnie z opisaną wyżej procedurą.

Statystycznie poprawny jest jedynie model ARIMA 6 i to on może być przyjęty do dalszych obliczeń, czyli do szacowania przyszłych wartości zmiennej zależnej P1 (zwykle o jeden lub dwa okresy do przodu). Mając wyznaczone wartości przeszłe PR1 i PR2 można korzystając z wzoru regresyjnego wyznaczonego w punktach 4.1 do 4.5 (wybiera się równanie regresji z najmniejszym błędem MAPE p.4.7). Istnieje również możliwość wyznaczania wartości przyszłych szeregu P1 bezpośrednio z niego samego, postępując zgodnie z opisaną wyżej procedurą.



Rys 4.11. Zestawienie wartości kolejnych wyrazów szeregu PR2 wraz z prognozami wartości przyszłych obliczonych przy zastosowaniu modeli ARIMA 4, ARIMA 5 i ARIMA 6

## 4.7. Analiza prawidłowości i wybór modelu prognostycznego

W rozdziałach 4.1 do 4.5 wykonano obliczenia prognostyczne metodami regresji wielorakiej, wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych, uogólnionych modeli addytywnych, automatycznych sieci neuronowych i metoda wektorów nośnych. Otrzymano 16 modeli obliczeniowych. Parametrem wspólnym dla wszystkich modeli jest wartość błędu *MAPE* (średni absolutny błąd procentowy). Wybór najlepszego modelu wynika z najmniejszej wartości błędu *MAPE*. Wyniki obliczeń wszystkich modeli zestawiono w tabeli 4.33.

Tabela 4.33. Zestawienie wyników obliczeń modeli prognostycznych

Lp.	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszt	Wartość błędu MAPE [%]	Poprawność modelu
1	MR1(P1;PR1,PR2)	$P1 = 1,052987 + 4,472303PR1 + 0,000004PR2$	NIE	0,938094	NIE p = 0,933 dla PR2
2	MR2(P1;PR1)	$P1 = 1,051184 + 4,489101 PR1$		0,943469	TAK
3	MR3(P1;PR2)	$P1 = 1,590558 + 0,000923 PR2$	TAK	5,494223	NIE
4	MR4(P1;PR2, PR <sup>2</sup> )	$P1 = MR4(P1; \ln PR2, PR2^{0,3})$		3,230033	
5	MR5(P1;PR1, PR2COXBOX)	$P1 = -0,498272 - 0,381074 PR2_{\text{BOXCOX}} + 1,568453 PR2^{0,3}$	NIE	0,918350	TAK
6	MARS1(P1;PR1)	$P1 = 1,92074667692681e+000 + 4,48278744930603e+000 \max(0; PR1 - 1,93649167310371e-001) - 4,49281728608202e+000 \max(0; 1,93649167310371e-001 - PR1)$		0,945105	
7	MARS2(P1;PR2, PR2 <sub>BOXCOX</sub> )	$P1 = 1,12727391702521e+000 - 4,81924619348619e-004 \max(0; PR2 - 2,34000000000000e+002) + 9,35605496631056e-004 \max(0; 2,34000000000000e+002 - PR2) + 6,89306253447718e-002 \max(0; PR2_{\text{BOXCOX}} - 3,40375079961819e+000)$		0,774200	
8	GAM1(P1; PR1,PR2)	$P1 = e^{(0,123414 + 2,827793PR1 - 0,000083PR2)}$		0,803304	
9	GAM2(P1; PR1,PR2)	$P1 = 1,1799 + 3,390149PR1 + 0,000236PR2$		0,834361	
10	NN1 MLP 2-9-1(P1;PR1, PR2)	brak		0,803304	
11	NN2 MLP 2-9-1(P1;PR1, PR2)		0,834361		
12	NN3 RBF 2-8-1(P1;PR1, PR2)		TAK	1,380180	NIE
13	SVN1(P1;PR1,PR2)		NIE	2,328698	TAK
14	SVN2(P1;PR1,PR2)		TAK	4,413850	NIE
15	SVN3(P1;PR1,PR2)		NIE	1,429182	TAK
16	SVN4(P1;PR1,PR2)	TAK	2,657474	NIE	

Najlepszym modelem prognostycznym jest model MARS2(P1;PR2, PR2<sub>BOXCOX</sub>) uzyskany w metodzie wieloimiennej regresji adaptacyjnej z użyciem funkcji skleja-nych – posiada minimalny błąd MAPE = 0,7742%.

## 5. Zastosowanie wieloczynnikowych modeli w prognozowaniu czasu procesów budowlanych

Prezentowany w rozdziałach 3 i 4 przykład 3.1 jest jedynie ilustracją proponowanej metody. Występują tam dwa predyktory, zwykle jest to zbyt mała liczba czynników mogących opisać proces budowlany. Zdecydowano się na przyjęcie takiej konwencji prezentacji metody ze względu na jasność przekazu. W rzeczywistości obliczenia metodą wieloczynnikowych modeli w prognozowaniu czasu procesów budowlanych są o wiele bardziej skomplikowane. Poniżej przedstawiono rzeczywiste przykłady obliczeniowe wybranych procesów budowlanych.

### 5.1. Roboty ziemne z wykorzystaniem koparek

Celem obliczeń jest określenie wzoru regresyjnego do wyznaczenia prognozowanego czasu realizacji procesu polegającego na odpajaniu i załadunku mas ziemnych przy użyciu koparek przedsiębiernych, podsiębiernych, chwytakowych i zbierakowych, przy różnych uwarunkowaniach gruntowych, sprzętowych, zasobowych i awaryjności sprzętu. Jako zmienną zależną przyjęto  $v_1$ , zmiennymi niezależnymi są zmienne od  $v_2$  do  $v_{16}$ . Jako zmienna porównawczą przyjęto  $v_{17}$  – odczytaną z KNR wydajność koparki według polskich normatywów. w tablicy 5.1. zestawiono zmienne do arkusza obliczeniowego programu STATISTICA firmy StatSoft, rysunek 5.1. Dane uzyskano z pomiarów rzeczywistych. w toku obliczeń stosowano zmienne przekształcone zgodnie z p. 3.2 nadając im oznaczenia powyżej  $v_{17}$ .

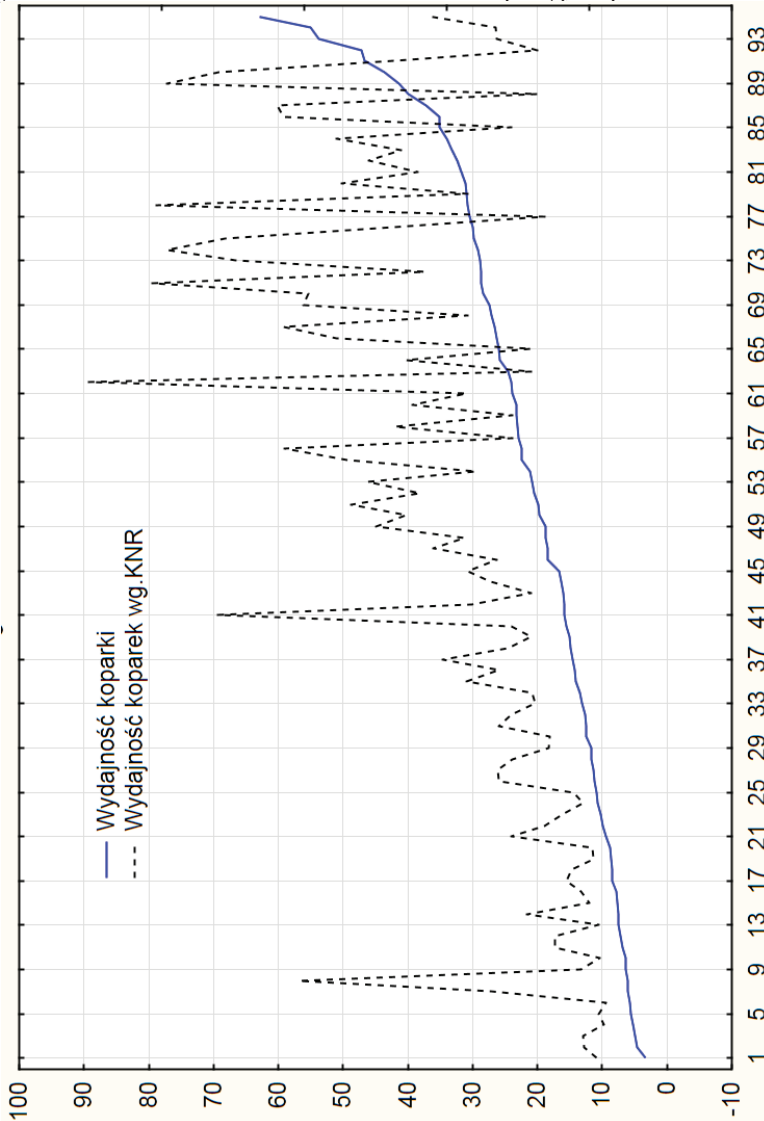
Tab. 5.1. Zestawienie zmiennych arkusza obliczeniowego w programie STATISTICA firmy Stat-Soft

Lp	Zmienna	Opis zmiennej	Uwagi
1	$v_1$	Wydajność koparki	$m^3/h$
2	$v_2$	Rodzaj łyżki koparki	zmienna lingwistyczna
3	$v_3$	Kodowany rodzaj łyżki koparki	przedsiębierna – 101; podsiębierna – 102; chwytakowa – 103; zbierakowa – 104
4	$v_4$	Pojemność łyżki koparki	$m^3$
5	$v_5$	Ładowność środka transportowego	$kG \cdot 10^3$
6	$v_6$	Rodzaj drogi	zmienna lingwistyczna
7	$v_7$	Kodowany rodzaj drogi	gruntowa – 201; utwardzona – 202
8	$v_8$	Lata pracy operatora	jako operator koparki
9	$v_9$	Liczba godzin pracy w dniu roboczym	
10	$v_{10}$	Wilgotność gruntu	
11	$v_{11}$	Kodowana wilgotność gruntu	grunt suchy – 301; grunt mokry – 302; grunt oblepiający – 303
12	$v_{12}$	Kategoria gruntu	I, II, III, IV
13	$v_{13}$	Kodowana kategoria gruntu	I – 401; II – 402; III – 403; IV – 404
14	$v_{14}$	Odległość wywozu gruntu	
15	$v_{15}$	Temperatura	
16	$v_{16}$	Awaryjność/naprawy	% czasu efektywnej pracy koparki
17	$v_{17}$	Wydajność koparki według polskich normatywów	

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Wydatność koparki [m <sup>3</sup> /h]	Rodzaj żyłki	Kod rodzaj żyłki	Pojemność bębna	Ładowność środka transportowego	Rodzaj drogi	Kod rodzaj drogi	Łata pracy operatora	Godziny pracy w danym roboczym dniu	Wygłomność gruntu	Kod wygłomności gruntu	Kategoria gruntu	Kod kategorii gruntu	Odegniość wywozu urębku	Temperatura	Awaryjność/naprawy	Wydatność koparki wg KCR
1	3.31	przedział	101	0,25	5	urządzenia	202	3	12	g. obłąpający	305	4	404	2	-5	10,8
2	4.38	podział	102	0,25	5	gruntowa	201	1	10	g. suchy	301	3	403	5	23	12,8
3	4.93	podział	102	0,25	5	urządzenia	202	1	10	g. suchy	302	2	402	5	20	13,02
4	5.29	przedział	101	0,15	5	urządzenia	201	6	8	g. suchy	302	4	404	1	10	9,8
5	5.60	podział	102	0,15	5	gruntowa	201	4	8	g. mokry	302	3	403	5	10	10,6
6	5.80	zgamakowe	103	0,25	10	urządzenia	202	3	12	g. obłąpający	302	3	403	1	20	9,42
7	6.07	zgamakowe	103	0,15	10	gruntowa	201	1	12	g. suchy	305	3	403	5	-5	2,72
8	6.11	podział	102	0,15	3	urządzenia	202	1	12	g. mokry	302	2	402	1	15	19,82
9	6.20	chwytkowe	104	0,25	3	urządzenia	202	1	10	g. suchy	301	3	403	2	10	18,87
10	6.20	chwytkowe	104	0,25	5	gruntowa	201	3	10	g. suchy	301	3	403	2	15	18,87
11	6.83	zgamakowe	103	0,25	10	urządzenia	202	3	10	g. suchy	301	3	403	2	15	18,87
12	7.13	przedział	101	0,4	5	gruntowa	201	2	8	g. suchy	302	3	403	5	15	0,81
13	7.44	przedział	101	0,15	5	gruntowa	201	1	8	g. suchy	301	3	403	10	23	17,33
14	7.56	podział	102	0,6	10	urządzenia	202	2	12	g. obłąpający	301	3	403	2	15	0,82
15	7.60	zgamakowe	103	0,25	10	urządzenia	202	1	8	g. suchy	301	2	402	2	0	0,81
16	7.77	chwytkowe	104	0,25	5	urządzenia	202	3	8	g. obłąpający	305	2	402	2	15	0,89
17	8.47	podział	102	0,6	10	gruntowa	201	5	12	g. suchy	302	4	404	1	15	0,85
18	8.49	zgamakowe	103	0,6	10	gruntowa	201	2	10	g. mokry	302	4	404	1	15	0,80
19	8.67	przedział	101	0,25	5	gruntowa	201	4	8	g. suchy	302	3	403	5	20	15,48
20	8.86	podział	101	0,15	5	gruntowa	201	2	8	g. suchy	302	2	403	5	20	11,4
21	9.48	podział	102	1,0	15	gruntowa	202	1	8	g. suchy	301	2	402	1	20	0,90
22	9.89	chwytkowe	104	0,6	10	urządzenia	202	2	8	g. suchy	302	4	404	10	0	0,86
23	10.17	podział	102	0,6	10	urządzenia	202	1	8	g. suchy	301	4	404	1	20	0,87
24	10.45	zgamakowe	103	0,25	3	gruntowa	201	3	6	g. obłąpający	305	2	404	3	35	18,87
25	10.83	przedział	101	0,25	5	gruntowa	201	3	10	g. suchy	302	4	403	2	20	11,90
26	11.16	podział	101	0,6	5	gruntowa	201	3	10	g. suchy	301	4	404	10	15	0,39
27	11.34	przedział	101	0,6	5	gruntowa	201	1	10	g. suchy	301	3	403	1	20	0,86
28	11.65	chwytkowe	104	1,0	10	urządzenia	202	1	10	g. suchy	301	3	403	2	-5	26,18
29	11.65	chwytkowe	104	0,8	10	gruntowa	201	2	10	g. suchy	301	2	402	1	-5	0,86
30	11.69	chwytkowe	104	0,8	10	gruntowa	201	6	8	g. suchy	301	4	404	2	0	0,83
31	12.49	przedział	101	0,8	10	gruntowa	201	1	10	g. suchy	302	4	404	1	5	18,06
32	12.54	przedział	101	0,6	10	gruntowa	201	1	8	g. suchy	301	3	403	1	25	25,97
33	12.60	podział	101	0,6	5	urządzenia	202	3	8	g. obłąpający	303	4	404	2	0	0,84
34	13.08	zgamakowe	103	0,6	10	gruntowa	201	5	8	g. suchy	302	2	402	2	-5	24,09
35	13.42	zgamakowe	103	0,4	10	urządzenia	202	3	6	g. suchy	301	2	402	5	10	0,84
36	14.10	przedział	101	0,6	10	gruntowa	201	3	12	g. suchy	301	2	402	2	-5	20,41
37	14.58	podział	102	0,6	10	urządzenia	202	5	8	g. suchy	301	2	402	2	-5	20,95
38	14.58	chwytkowe	104	1,2	10	urządzenia	202	6	12	g. suchy	301	2	402	2	-5	25,97
39	14.96	chwytkowe	104	0,8	10	gruntowa	201	6	8	g. suchy	302	2	402	1	10	34,6
40	14.96	chwytkowe	104	0,8	10	gruntowa	201	6	8	g. suchy	302	2	402	2	10	34,1

Rys.5.1. Fragment arkusza kalkulacyjnego programu STATISTICA

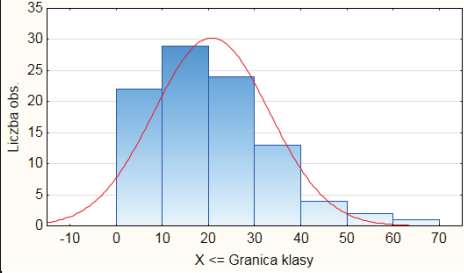
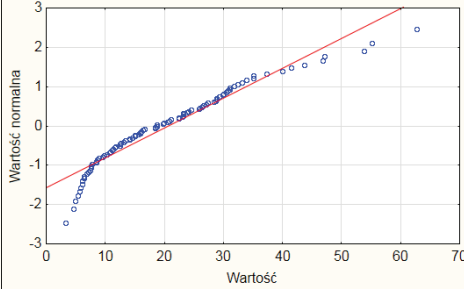
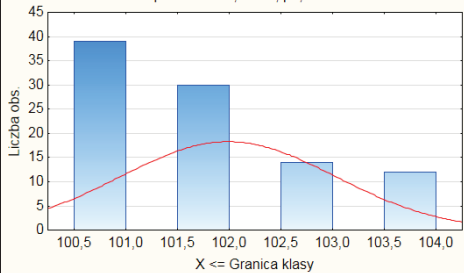
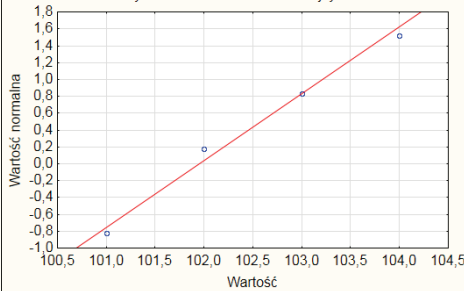
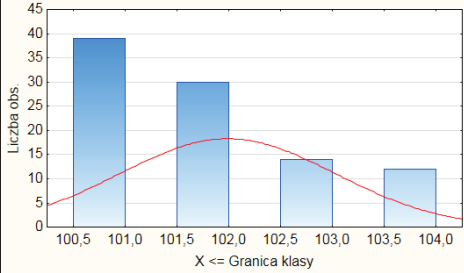
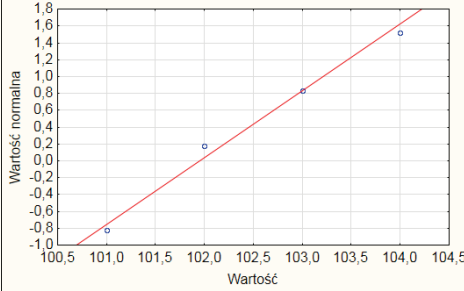
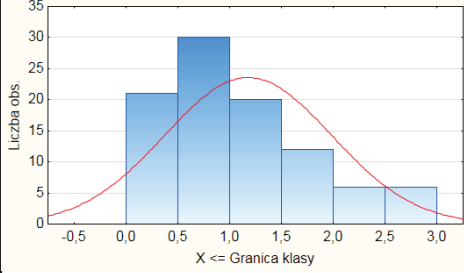
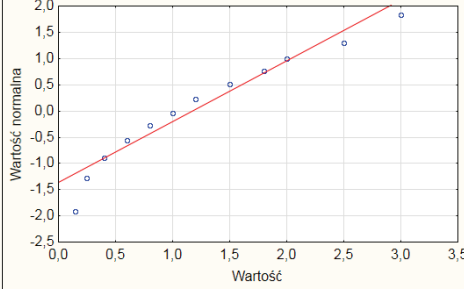
Na rysunku 5.2 przedstawiono wykres wartości zmiennych v1 (wydajność koparki) i v17 (wydajność koparki wg. KNR). Wykresy zmiennych różnią się w sposób istotny od siebie. Wydajność rzeczywista jest zdecydowanie niższa od przewidywanej w KNR. Różnica ta stanowi uzasadnienie podjęcia problemu.

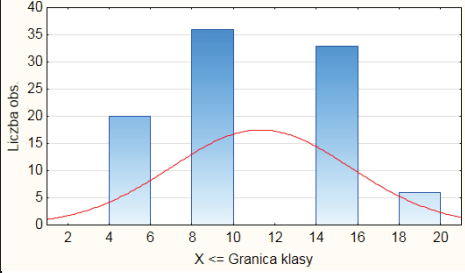
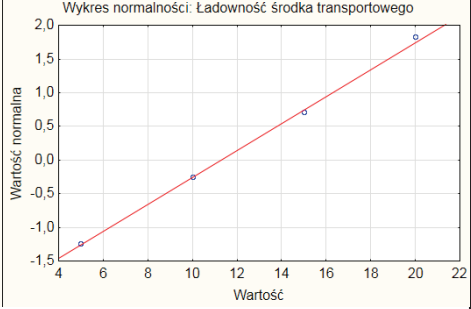
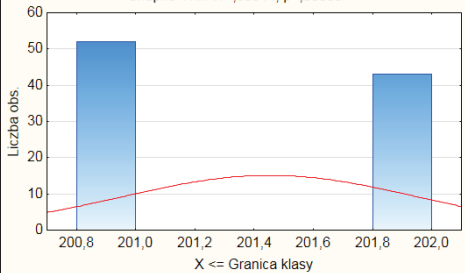
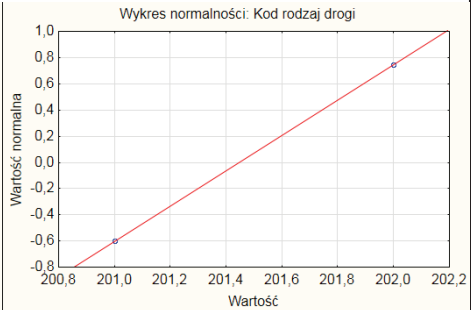
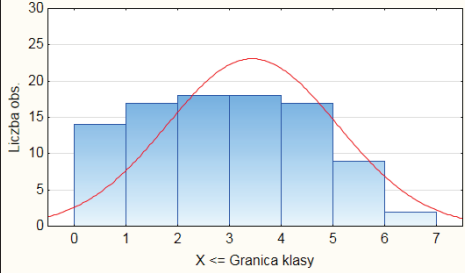
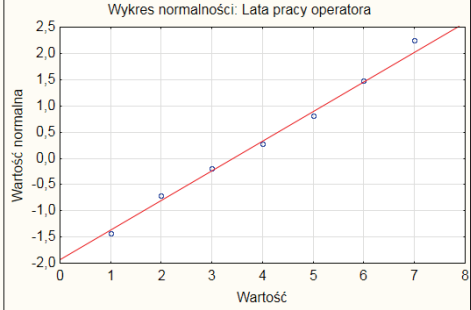
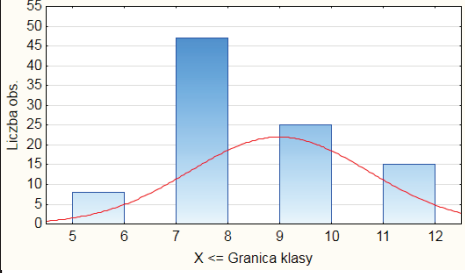
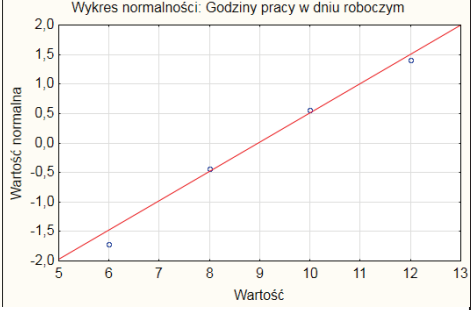


Rys. 5.2. Zestawienie wykresów wartości zmiennych v1 wydajność koparki i v17 wydajność koparki wg. KNR

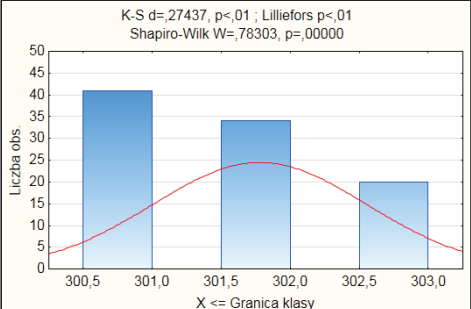
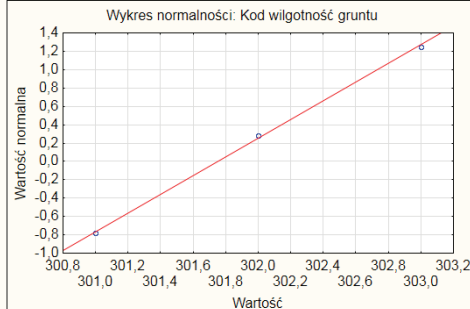
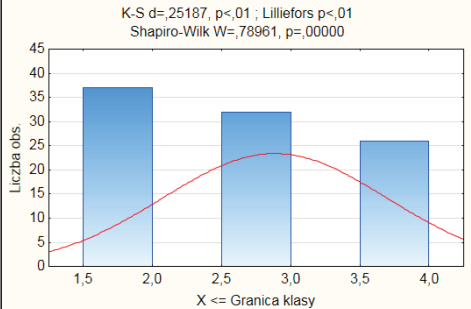
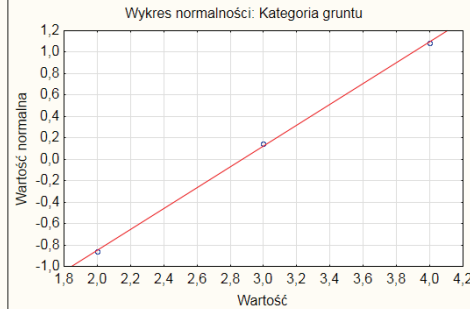
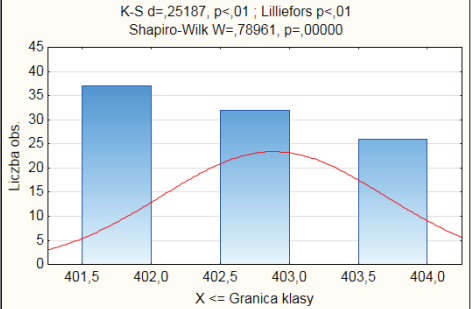
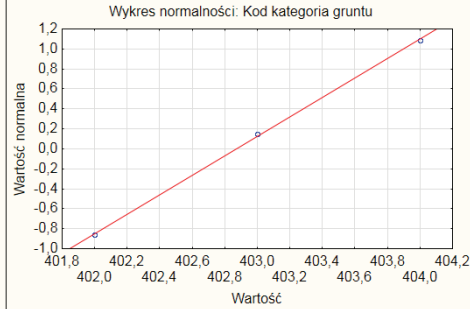
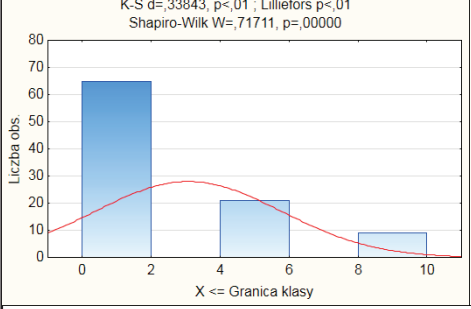
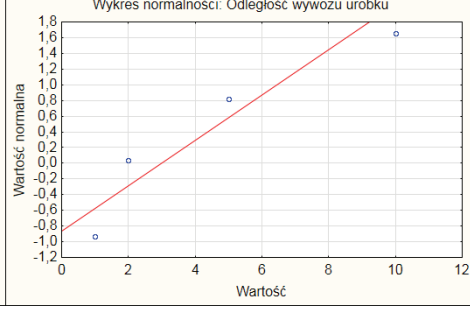
Zgodnie z punktem 3.3 opracowania wykonano analizę normalności rozkładów zmiennych. Wyniki zestawiono w tabeli 5.2.

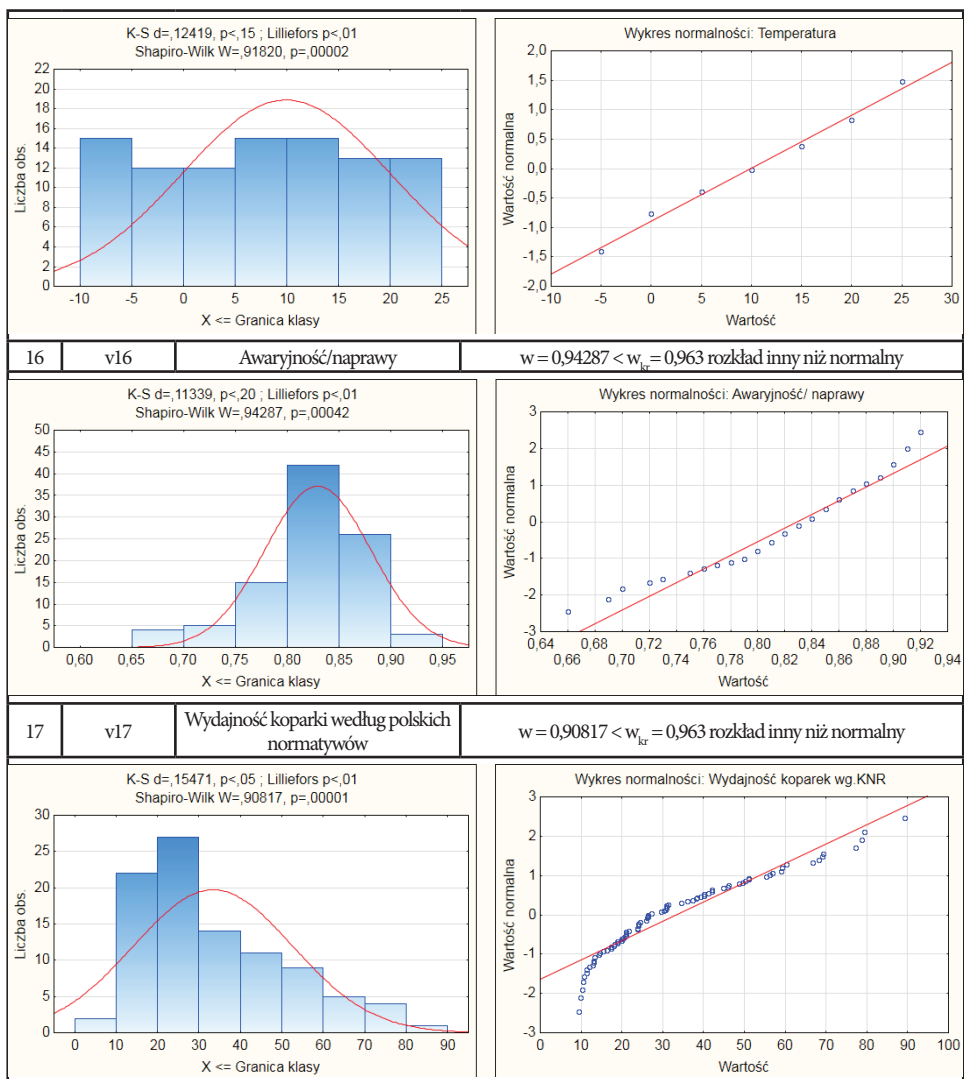
Tab. 5.2. Zestawienie analizy normalności rozkładów zmiennych

Lp	Zmienna	Opis zmiennej	Uwagi
1	v1	Wydajność koparki	$w = 0,92893 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,10188, p&gt; ,20; Lilliefors p&lt;,05 Shapiro-Wilk W=,92893, p=,00007</p> 		<p>Wykres normalności: Wydajność koparki</p> 	
2	v2	Rodzaj łyżki koparki	$w = 0,81183 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,24064, p&lt;,01; Lilliefors p&lt;,01 Shapiro-Wilk W=,81183, p=,00000</p> 		<p>Wykres normalności: Kod rodzaj łyżki</p> 	
3	v3	Kodowany rodzaj łyżki koparki	$w = 0,81183 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,24064, p&lt;,01; Lilliefors p&lt;,01 Shapiro-Wilk W=,81183, p=,00000</p> 		<p>Wykres normalności: Kod rodzaj łyżki</p> 	
4	v4	Pojemność łyżki koparki	$w = 0,91553 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,13166, p&lt;,10; Lilliefors p&lt;,01 Shapiro-Wilk W=,91553, p=,00001</p> 		<p>Wykres normalności: Pojemność łyżki</p> 	

5	v5	Ładowność środka transportowego	$w = 0,86674 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,21323, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,86674, p=,00000</p> 		<p>Wykres normalności: Ładowność środka transportowego</p> 	
6	v6	Rodzaj drogi	zmienna lingwistyczna
7	v7	Kodowany rodzaj drogi	$w = 0,63314 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,36452, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,63314, p=,00000</p> 		<p>Wykres normalności: Kod rodzaj drogi</p> 	
8	v8	Lata pracy operatora	$w = 0,93639 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,13642, p&lt;,10 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,93639, p=,00018</p> 		<p>Wykres normalności: Lata pracy operatora</p> 	
9	v9	Liczba godzin pracy w dniu roboczym	$w = 0,84331 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,29608, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,84331, p=,00000</p> 		<p>Wykres normalności: Godziny pracy w dniu roboczym</p> 	
10	v10	Wilgotność gruntu	zmienna lingwistyczna



11	v11	Kodowana wilgotność gruntu	$w = 0,78303 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,27437, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,78303, p=,00000</p> 		<p>Wykres normalności: Kod wilgotność gruntu</p> 	
12	v12	Kategoria gruntu	$w = 0,78961 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,25187, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,78961, p=,00000</p> 		<p>Wykres normalności: Kategoria gruntu</p> 	
13	v13	Kodowana kategoria gruntu	$w = 78961 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,25187, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,78961, p=,00000</p> 		<p>Wykres normalności: Kod kategoria gruntu</p> 	
14	v14	Odległość wywozu gruntu	$w = 0,71711 < w_{kr} = 0,963$ rozkład inny niż normalny
<p>K-S d=,33843, p&lt;,01 ; Lilliefors p&lt;,01 Shapiro-Wilk W=,71711, p=,00000</p> 		<p>Wykres normalności: Odległość wywozu urobku</p> 	
15	v15	Temperatura	$w = 0,91820 < w_{kr} = 0,963$ rozkład inny niż normalny



Stwierdzono, że wiele zmiennych posiada rozkład inny niż normalny. Należy spodziewać się pewnych komplikacji obliczeniowych oraz błędu *MAPE* prognozy większego niż 3%.

Zgodnie z punktem 3.4 opracowania obliczono współczynniki korelacji pomiędzy zmiennymi liczbowymi tabela 5.3. Stwierdzono występowanie korelacji liniowych pomiędzy wieloma zmiennymi.

Tab. 5.3. Zestawienie współczynników korelacji pomiędzy zmiennymi liczbowymi

Zmienna	Korelacje Oznaczone współczynnik korelacji są istotne z $p < ,05000$ N = 95 (Braki danych usuwano przypadkami)				
	Średnia	Odch.std	Wydajność koparki	Kod rodzaj łyżki	Pojemność łyżki
Wydajność koparki	20,6928	12,52053	1,000000	-0,231827	0,851309
Kod rodzaj łyżki	101,9895	1,03651	-0,231827	1,000000	-0,170066
Pojemność łyżki	1,1789	0,80393	0,851309	-0,170066	1,000000
Ładowność środka transportowego	11,3158	4,32770	0,685620	-0,151033	0,806107
Kod rodzaj drogi	201,4526	0,50039	0,101071	0,009284	0,179965
Lata pracy operatora	3,4421	1,64197	0,329519	0,077772	0,144936
Godziny pracy w dniu roboczym	8,9895	1,72278	-0,048630	-0,041766	0,062055
Kod wilgotność gruntu	301,7789	0,77431	-0,110648	-0,029440	0,079603
Kod kategoria gruntu	402,8842	0,81035	-0,172472	-0,115457	0,100729
Odległość wywozu urobku	3,0737	2,70634	0,011425	-0,026267	0,039104
Temperatura	9,9474	10,03967	0,053308	-0,097172	-0,165554
Awaryjność/ naprawy	0,8295	0,05104	0,089529	0,096417	-0,068719

Zmienna	Korelacje (drogi8 NN) Oznaczone wsp. korelacji są istotne z $p < ,05000$ N = 95 (Braki danych usuwano przypadkami)			
	Ładowność środka transportowego	Kod rodzaj drogi	Lata pracy operatora	Godziny pracy w dniu roboczym
Wydajność koparki	0,685620	0,101071	0,329519	-0,048630
Kod rodzaj łyżki	-0,151033	0,009284	0,077772	-0,041766
Pojemność łyżki	0,806107	0,179965	0,144936	0,062055
Ładowność środka transportowego	1,000000	0,115056	0,081946	0,023281
Kod rodzaj drogi	0,115056	1,000000	-0,129614	-0,080798
Lata pracy operatora	0,081946	-0,129614	1,000000	-0,020902
Godziny pracy w dniu roboczym	0,023281	-0,080798	-0,020902	1,000000
Kod wilgotność gruntu	0,008354	0,151156	-0,064561	-0,009738
Kod kategoria gruntu	0,104576	-0,000552	0,110840	0,037219

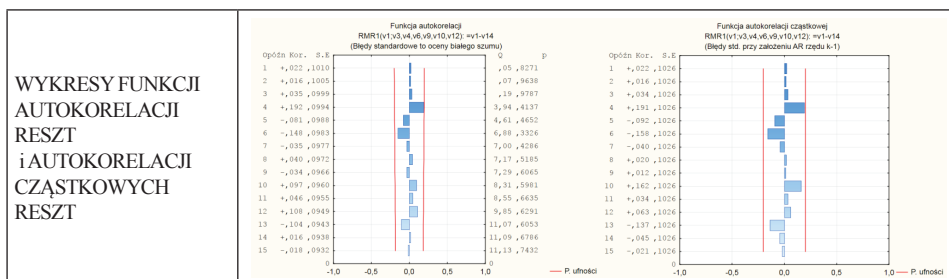
Zmienna	Korelacje (drogi8 NN) Oznaczone wsp. korelacji są istotne z $p < ,05000$ N = 95 (Braki danych usuwano przypadkami)			
	Ładowność środka transportowego	Kod rodzaj drogi	Lata pracy operatora	Godziny pracy w dniu roboczym
Odległość wywozu urobku	0,123338	0,085088	-0,148654	-0,120762
Temperatura	-0,249358	-0,079911	0,056280	-0,064614
Awaryjność/ naprawy	-0,011280	-0,057219	0,054851	-0,003693

Zmienna	Korelacje (drogi8 NN) Oznaczone wsp. korelacji są istotne z $p < ,05000$ N = 95 (Braki danych usuwano przypadkami)				
	Kod wilgotność gruntu	Kod kategoria gruntu	Odległość wywozu urobku	Temperatura	Awaryjność i naprawy
Wydajność koparki	-0,110648	-0,172472	0,011425	0,053308	0,089529
Kod rodzaj łyżki	-0,029440	-0,115457	-0,026267	-0,097172	0,096417
Pojemność łyżki	0,079603	0,100729	0,039104	-0,165554	-0,068719
Ładowność środka transportowego	0,008354	0,104576	0,123338	-0,249358	-0,011280
Kod rodzaj drogi	0,151156	-0,000552	0,085088	-0,079911	-0,057219
Lata pracy operatora	-0,064561	0,110840	-0,148654	0,056280	0,054851
Godziny pracy w dniu roboczym	-0,009738	0,037219	-0,120762	-0,064614	-0,003693
Kod wilgotność gruntu	1,000000	0,077456	0,089082	-0,145203	-0,072963
Kod kategoria gruntu	0,077456	1,000000	-0,049428	-0,098828	-0,148101
Odległość wywozu urobku	0,089082	-0,049428	1,000000	0,066705	-0,035914
Temperatura	-0,145203	-0,098828	0,066705	1,000000	0,102712
Awaryjność/ naprawy	-0,072963	-0,148101	-0,035914	0,102712	1,000000

Przeprowadzono obliczenia metodą regresji wielorakiej MR zgodnie z p. 4.1, używając modele MR1, MR2 i MR3. Wyniki obliczeń zestawiono w tabelach 5.4.–5.6.

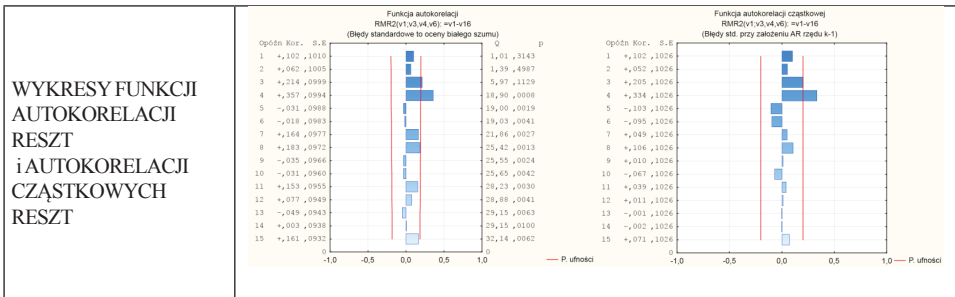
Tab. 5.4. Zestawienie obliczeń modelu  $MR1(v1;v3,v4,v6,v9,v10,v12)$  metodą regresji wielorakiej

METODA OBLICZEŃ i MODEL REGRESYJNY	Regresja wieloraka $MR1(v1;v3,v4,v6,v9,v10,v12)$						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v3,v4,v6,v9,v10,v12						
PODSUMOWANIE REGRESJI	N = 95	Podsumowanie regresji zmiennej zależnej: Wydajność koparki (drogi6) R = ,94965421 R <sup>2</sup> = ,90184312 Popraw. R2 = ,89515061 F(6,88) = 134,75 p<0,0000 Błąd std. estymacji: 4,0542					
		b*	Bł. std. z b*	b	Bł. std. z b	t(88)	p
	W. wolny			806,4193	90,52725	8,90803	0,000000
	Kod rodzaj łyżki v2	-0,12823	0,034697	-1,5490	0,41912	-3,69572	0,000380
	Pojemność łyżki v4	0,854513	0,035129	13,3083	0,54710	24,32497	0,000000
	Lata pracy operatora v8	0,231111	0,034439	1,7623	0,26260	6,71084	0,000000
	Kod kategoria gruntu v13	-0,27685	0,034220	-4,2776	0,52872	-8,09036	0,000000
	Temperatura v15	0,123337	0,034789	0,1538	0,04339	3,54531	0,000631
	Kod wilgotność gruntu v11	-0,12817	0,033969	-2,0725	0,54927	-3,77324	0,000292
WZÓR REGRESYJNY	$MR1(v1;v3,v4,v6,v9,v10,v12) = 806,4193 - 1,5490v3 + 13,3083v4 + 1,7623v6 - 4,2776v9 + 0,1538v10 - 2,0725v12$						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
BŁĄD MAPE [%]	21,02021						
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	PRAWIDŁOWY						



Tab. 5.5. Zestawienie obliczeń modelu MR2(v1;v3,v4,v6) metodą regresji wielorakiej

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>Regresja wieloraka MR2(v1;v3,v4,v6)</b>						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v3,v4,v6,v9,v10,v12						
PODSUMOWANIE REGRESJI	N = 95 Podsumowanie regresji zmiennej zależnej: Wydajność koparki (drogi6) $R = ,88340976$ $R^2 = ,78041280$ Popraw. $R2 = ,77317366$ $F(3,91) = 107,80$ $p < 0,0000$ Błąd std. estymacji: 5,9631						
		b*	Bł. std. z b*	b	Bł. std. z b	t(91)	p
	W. wolny			139,5031	61,80810	2,25704	0,026399
	Kod rodzaj łyżki v3	-0,11309	0,050126	-1,3662	0,60550	-2,25624	0,026451
	Pojemność łyżki v4	0,799842	0,050508	12,4568	0,78661	15,83604	0,000000
Lata pracy operatora v6	0,222388	0,049923	1,6958	0,38068	4,45461	0,000024	
WZÓR REGRESYJNY	$MR2(v1;v3,v4,v6) = 139,5031 - 1,3662v3 + 12,4568v4 + 1,6958v6$						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
BŁĄD MAPE [%]	26,44492						
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	NIE PRAWIDŁOWY						



Tab. 5.6. Zestawienie obliczeń modelu MR3(v1;v4,v9,v6,v10,v29,v3,v12,v37,v22,v30,v7) metodą regresji wielorakiej

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>Regresja wieloraka MR3(v1;v4,v9,v6,v10,v29,v3,v12,v37,v22,v30,v7)</b>						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v4,v9,v6,v10,v29,v3,v12,v37,v22,v30,v7						
PODSUMOWANIE REGESJII	Podsumowanie regresji zmiennej zależnej: Wydajność koparki (drogi6) $R = ,97062941$ $R^2 = ,94212146$ Popraw. R2 = ,93445081 $F(11,83) = 122,82$ $p < 0,0000$ Błąd std. estymacji: 3,2056 N = 95						
		b*	Bł. std. z b*	b	Bł. std. z b	t(83)	p
	W. wolny			837,4718	75,71544	11,06078	0,000000
	Pojemność łyżki v4	0,812735	0,034242	12,6576	0,53328	23,73533	0,000000
	Kod kategoria gruntu v9	-0,260438	0,027388	-4,0240	0,42317	-9,50911	0,000000
	Lata pracy operatora v6	0,267405	0,033275	2,0390	0,25373	8,03634	0,000000
	Temperatura v10	0,348914	0,070656	0,4351	0,08812	4,93821	0,000004
	v29 = v6^10	-0,104093	0,033166	-0,0000	0,00000	-3,13853	0,002351
	Kod rodzaj łyżki v3	-0,166160	0,028900	-2,0071	0,34910	-5,74943	0,000000
	Kod wilgotność gruntu v12	-0,132347	0,027587	-2,1400	0,44608	-4,79741	0,000007
	v37 = v13^10	0,072914	0,027537	9,9969	3,77542	2,64789	0,009692
	v22 = (1/v4)^2	-0,120993	0,033200	-0,1459	0,04004	-3,64432	0,000466
	v30 = v10^2	-0,238035	0,071393	-0,0139	0,00416	-3,33417	0,001280
Godziny pracy v7	-0,067291	0,026762	-0,4890	0,19450	-2,51439	0,013854	

WZÓR REGRESYJNY	$MR3(v1;v4,v9,v6,v10,v29,v3,v12,v37,v22,v30,v7) = 837,4718 + 12,6576v4 - 4,0240v9 + 2,0390v6 + 0,4351v10 - 0,0000v29 - 2,0071v3 - 2,1400v12 + 9,9969v37 - 0,1459v22 - 0,0139v30 - 0,4890v7$
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	
BŁĄD MAPE [%]	17,39331
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	PRAWIDŁOWY
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	

Zgodnie z procedurą opisaną w punkcie 4.2 wykonano obliczenia metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklejących MARS. Wyniki obliczeń zestawiono w tabeli 5.7.

Tab. 5.7. Zestawienie obliczeń modelu MARS1 (v1;v3,v4,v5,v8,v9,v13,v15,v11,v16) metodą wieloimiennej regresji adaptacyjnej z użyciem funkcji sklejących

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>MARS1 (v1;v3,v4,v5,v8,v9,v13,v15,v11,v16)</b>
ZMIENNA ZALEŻNA	v1
ZMIENNE NIEZALEŻNE	od v2 do v19
PREDYKTORY	v3,v4,v5,v8,v9,v13,v15,v11,v16



PODSUMOWANIE REGRESJI	Zależnych	Liczba odwołań do każdego predyktora (drogi6) Ile razy każdy predyktor był użyty Odwołania (do funkcji bazowej)																																																																																																																																
	Kod rodzaj łyżki v3	1																																																																																																																																
	Pojemność łyżki v4	2																																																																																																																																
	Ładowność środka transportowego v5	0																																																																																																																																
	Lata pracy operatora v8	1																																																																																																																																
	Godziny pracy v9	1																																																																																																																																
	Kod kategoria gruntu v13	1																																																																																																																																
	Temperatura v15	1																																																																																																																																
	Kod wilgotność gruntu v11	1																																																																																																																																
	Awaryjność/ naprawy v16	1																																																																																																																																
WZÓR REGRESYJNY	<p>Wydajność koparki = 1,79706819408758e+001 + 1,19446716463054e+001max(0; Pojemność łyżki-1,0000000000000e+000) + 1,73912859343404e+001max(0; 1,0000000000000e+000-Pojemność łyżki) - 4,26315025699904e+000max(0; Kod kategoria gruntu-1,0200000000000e+002) + 1,75923886157349e+000max(0; Lata pracy operatora-1,0000000000000e+000) + 1,71821481315091e-001max(0; Temperatura+5,0000000000000e+000) - 1,78708386834731e+000max(0; Kod rodzaj łyżki-1,0100000000000e+002) + 3,33199437052750e+000max(0; 1,0200000000000e+002-Kod wilgotność gruntu) - 5,69661471300789e-001max(0; Godziny pracy-6,0000000000000e+000)+4,07842963164529e+001max(0; Awaryjność/ naprawy-8,3000000000000e-001)</p>																																																																																																																																	
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH																																																																																																																																		
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	<div style="display: flex; justify-content: space-around;"> <div style="width: 45%;"> <p>Funkcja autokorelacji RMARS(y1.v3,v4,v6,v7,v8,v9,v10,v11,v13) (Błędy standardowe to oceny białego szumu)</p> <table border="1"> <thead> <tr> <th>Opóźn. Kor.</th> <th>S.E.</th> <th>ρ</th> <th>p</th> </tr> </thead> <tbody> <tr><td>1</td><td>,054 ,1010</td><td>,28 ,5957</td><td></td></tr> <tr><td>2</td><td>,117 ,1005</td><td>1,63 ,4421</td><td></td></tr> <tr><td>3</td><td>,056 ,0999</td><td>1,94 ,3562</td><td></td></tr> <tr><td>4</td><td>,182 ,0994</td><td>5,31 ,2568</td><td></td></tr> <tr><td>5</td><td>,122 ,0988</td><td>6,83 ,2333</td><td></td></tr> <tr><td>6</td><td>,143 ,0983</td><td>8,95 ,1766</td><td></td></tr> <tr><td>7</td><td>,007 ,0977</td><td>8,95 ,2562</td><td></td></tr> <tr><td>8</td><td>,077 ,0972</td><td>9,38 ,2396</td><td></td></tr> <tr><td>9</td><td>,038 ,0966</td><td>9,74 ,2323</td><td></td></tr> <tr><td>10</td><td>,131 ,0960</td><td>11,60 ,3125</td><td></td></tr> <tr><td>11</td><td>,068 ,0955</td><td>12,12 ,3549</td><td></td></tr> <tr><td>12</td><td>,098 ,0949</td><td>13,19 ,3554</td><td></td></tr> <tr><td>13</td><td>,090 ,0943</td><td>14,10 ,3469</td><td></td></tr> <tr><td>14</td><td>,012 ,0938</td><td>14,12 ,4412</td><td></td></tr> <tr><td>15</td><td>,046 ,0932</td><td>14,36 ,4582</td><td></td></tr> </tbody> </table> </div> <div style="width: 45%;"> <p>Funkcja autokorelacji częściowej RMARS(y1.v3,v4,v6,v7,v8,v9,v10,v11,v13) (Błędy std. przy założeniu AR(1) rzędu k-1)</p> <table border="1"> <thead> <tr> <th>Opóźn. Kor.</th> <th>S.E.</th> <th>ρ</th> <th>p</th> </tr> </thead> <tbody> <tr><td>1</td><td>,054 ,1026</td><td></td><td></td></tr> <tr><td>2</td><td>,114 ,1026</td><td></td><td></td></tr> <tr><td>3</td><td>,049 ,1026</td><td></td><td></td></tr> <tr><td>4</td><td>,168 ,1026</td><td></td><td></td></tr> <tr><td>5</td><td>,102 ,1026</td><td></td><td></td></tr> <tr><td>6</td><td>,198 ,1026</td><td></td><td></td></tr> <tr><td>7</td><td>,038 ,1026</td><td></td><td></td></tr> <tr><td>8</td><td>,084 ,1026</td><td></td><td></td></tr> <tr><td>9</td><td>,070 ,1026</td><td></td><td></td></tr> <tr><td>10</td><td>,185 ,1026</td><td></td><td></td></tr> <tr><td>11</td><td>,030 ,1026</td><td></td><td></td></tr> <tr><td>12</td><td>,016 ,1026</td><td></td><td></td></tr> <tr><td>13</td><td>,112 ,1026</td><td></td><td></td></tr> <tr><td>14</td><td>,040 ,1026</td><td></td><td></td></tr> <tr><td>15</td><td>,043 ,1026</td><td></td><td></td></tr> </tbody> </table> </div> </div>		Opóźn. Kor.	S.E.	ρ	p	1	,054 ,1010	,28 ,5957		2	,117 ,1005	1,63 ,4421		3	,056 ,0999	1,94 ,3562		4	,182 ,0994	5,31 ,2568		5	,122 ,0988	6,83 ,2333		6	,143 ,0983	8,95 ,1766		7	,007 ,0977	8,95 ,2562		8	,077 ,0972	9,38 ,2396		9	,038 ,0966	9,74 ,2323		10	,131 ,0960	11,60 ,3125		11	,068 ,0955	12,12 ,3549		12	,098 ,0949	13,19 ,3554		13	,090 ,0943	14,10 ,3469		14	,012 ,0938	14,12 ,4412		15	,046 ,0932	14,36 ,4582		Opóźn. Kor.	S.E.	ρ	p	1	,054 ,1026			2	,114 ,1026			3	,049 ,1026			4	,168 ,1026			5	,102 ,1026			6	,198 ,1026			7	,038 ,1026			8	,084 ,1026			9	,070 ,1026			10	,185 ,1026			11	,030 ,1026			12	,016 ,1026			13	,112 ,1026			14	,040 ,1026			15	,043 ,1026		
Opóźn. Kor.	S.E.	ρ	p																																																																																																																															
1	,054 ,1010	,28 ,5957																																																																																																																																
2	,117 ,1005	1,63 ,4421																																																																																																																																
3	,056 ,0999	1,94 ,3562																																																																																																																																
4	,182 ,0994	5,31 ,2568																																																																																																																																
5	,122 ,0988	6,83 ,2333																																																																																																																																
6	,143 ,0983	8,95 ,1766																																																																																																																																
7	,007 ,0977	8,95 ,2562																																																																																																																																
8	,077 ,0972	9,38 ,2396																																																																																																																																
9	,038 ,0966	9,74 ,2323																																																																																																																																
10	,131 ,0960	11,60 ,3125																																																																																																																																
11	,068 ,0955	12,12 ,3549																																																																																																																																
12	,098 ,0949	13,19 ,3554																																																																																																																																
13	,090 ,0943	14,10 ,3469																																																																																																																																
14	,012 ,0938	14,12 ,4412																																																																																																																																
15	,046 ,0932	14,36 ,4582																																																																																																																																
Opóźn. Kor.	S.E.	ρ	p																																																																																																																															
1	,054 ,1026																																																																																																																																	
2	,114 ,1026																																																																																																																																	
3	,049 ,1026																																																																																																																																	
4	,168 ,1026																																																																																																																																	
5	,102 ,1026																																																																																																																																	
6	,198 ,1026																																																																																																																																	
7	,038 ,1026																																																																																																																																	
8	,084 ,1026																																																																																																																																	
9	,070 ,1026																																																																																																																																	
10	,185 ,1026																																																																																																																																	
11	,030 ,1026																																																																																																																																	
12	,016 ,1026																																																																																																																																	
13	,112 ,1026																																																																																																																																	
14	,040 ,1026																																																																																																																																	
15	,043 ,1026																																																																																																																																	

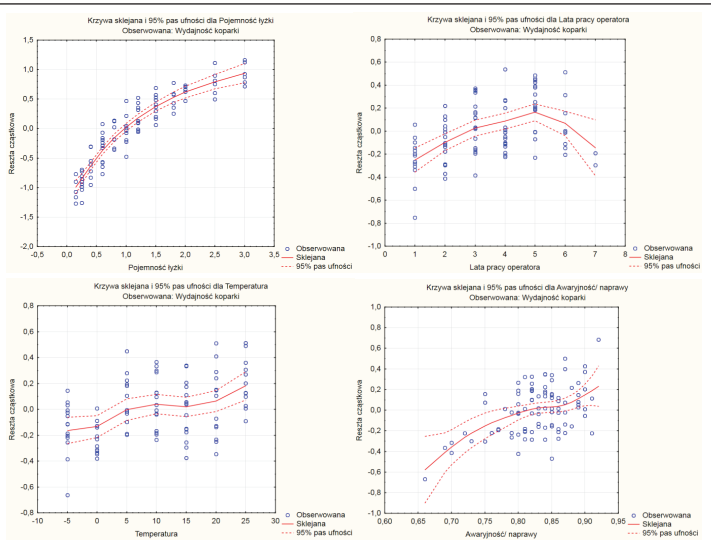
BŁĄD MAPE [%]	17,72356
OCENA PRAWDŹOWOŚCI MODELU REGRESYJNEGO	PRAWDŹOWY

Zgodnie z procedurą opisaną w punkcie 4.3 wykonano obliczenia metodą uogólnionych modeli addytywnych GAM (Generalized Additive Methods). Wyniki obliczeń zestawiono w tabeli 5.8.

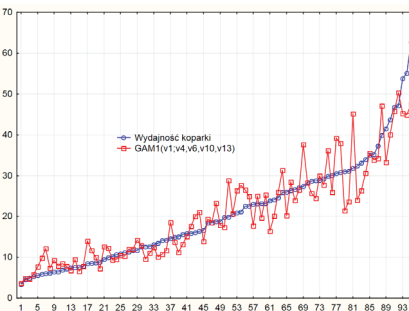
Tab. 5.8. Zestawienie obliczeń modelu GAM1(v1;v4,v8,v15,v16) metodą uogólnionych modeli addytywnych

METODA OBLICZENI i MODEL REGRESYJNY	<b>GAM1(v1;v4,v8,v15,v16)</b>						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v4,v8,v15,v16						
PREDYKTORY	v4,v8,v15,v16						
PODSUMOWANIE REGRESJI	Podsumowanie dopasowania Obserwowana: Wydajność koparki Rozkład: Gamma; Funkcja wiążąca: Log						
		Indeks zmiennej	Stopnie swobody	GAM współcz.	Błąd standardowy	Standaryz. wyniku	Nieliniowe p-wartość
	W. wolny	0	1,000000	-0,139001	0,389935	-0,35647	
	Pojemność łyżki v4	1	4,001435	0,685827	0,030217	22,69678	0,000000
	Lata pracy operatora v8	2	3,999971	0,065085	0,014611	4,45446	0,002511
	Temperatura v15	3	3,996926	0,010339	0,002403	4,30321	0,177639
Awarynosc/naprawy v16	4	4,000508	2,243254	0,466623	4,80743	0,353947	
WZÓR REGRESYJNY	$GAM1(v1;v4,v6,v10,v13) = e^{(-0,139001 + 0,685827*v4 + 0,065085*v8 + 0,010339*v15 + 2,243254*v16)}$						

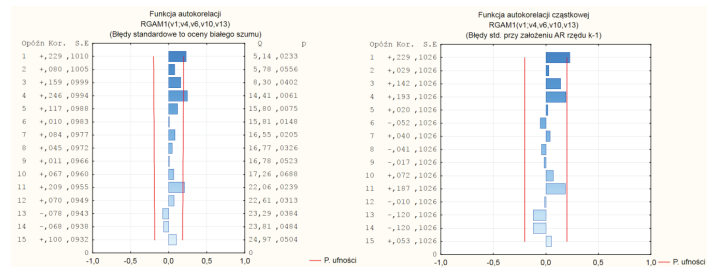
KRZYWE SKLEJANE  
i 95% PAS UFNOŚCI



WYKRES WARTOŚCI  
RZECZYWISTYCH I  
PROGNOZOWANYCH



WYKRESY FUNKCJI  
AUTOKORELACJI  
RESZT  
i AUTOKORELACJI  
CZĄSTKOWYCH  
RESZT



BŁĄD MAPE [%]

18,06137

OCENA  
PRAWIDŁOWOŚCI  
MODELU  
REGRESYJNEGO

NIEPRAWIDŁOWY

Zgodnie z procedurą opisaną w punkcie 4.4 przeprowadzono obliczenia metodą automatycznych sieci neuronowych. Obliczono 3 modele NN1, NN2 i NN3. Wyniki obliczeń zamieszczono w tabelach 5.9.– 5.11.

Tab. 5.9. Zestawienie obliczeń modelu NN1 MLP 11-9-1 (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą automatycznych sieci neuronowych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>NN1 MLP 11-9-1 (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16						
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16						
PODSUMOWANIE REGRESJI	Podsumowanie aktywnych sieci						
	Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
	1	MLP 11-9-1	0,966243	0,934685	0,924517	6,174185	6,866063
	Błąd (walidacja)		Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)	
	5,354702		BFGS 8	SOS	Wykładnicza		
WZÓR REGRESYJNY	NIE JEST GENEROWANY						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT							
BŁĄD MAPE [%]	15,83						
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY						

Tab. 5.10. Zestawienie obliczeń modelu NN2 MLP 11-13-1 (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą automatycznych sieci neuronowych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>NN2 MLP 11-13-1 (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16						
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16						
PODSUMOWANIE REGRESJI	Podsumowanie aktywnych sieci						
	Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
	2	MLP 11-13-1	0,950121	0,941306	0,951186	9,041562	5,320041
	Błąd (walidacja)		Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)	
	6,317376		BFGS 7	SOS	Tanh	Liniowa	
WZÓR REGRESYJNY	NIE JEST GENEROWANY						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT							
BŁĄD MAPE [%]	22,76						
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY						

Tab. 5.11. Zestawienie obliczeń modelu NN3 RBF 11-14-1 (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą automatycznych sieci neuronowych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>NN3 RBF 11-14-1 (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>						
ZMIENNA ZALEŻNA	v1						
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16						
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16						
PODSUMOWANIE REGRESJI	Podsumowanie aktywnych sieci						
	Id sieci	Nazwa sieci	Jakość (uczenie)	Jakość (testowanie)	Jakość (walidacja)	Błąd (uczenie)	Błąd (testowanie)
	1	RBF 11-14-1	0,717757	0,715270	0,922932	44,72542	22,09114
	Błąd (walidacja)		Algorytm uczenia	Funkcja błędu	Aktywacja (ukryte)	Aktywacja (wyjściowe)	
	10,99751		RBFT	SOS	Gaussa	Liniowa	
WZÓR REGRESYJNY	NIE JEST GENEROWANY						
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH							
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT							
BLĄD MAPE [%]	50,004						
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY						

Kolejną metodą obliczeniową zastosowaną w analizie jest metoda wektorów nośnych opisana w punkcie 4.5. Wyniki obliczeń dla funkcji jądrowych liniowych, wykładniczych, RBF i sigmoidalnych zamieszczono w tabelach 5.12.–5.15.

Tab. 5.12. Zestawienie obliczeń modelu SVM Liniowy (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą wektorów nośnych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>SVM Liniowy(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>
ZMIENNA ZALEŻNA	v1
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16
TYP JĄDRA	LINIOWY
WZÓR REGRESYJNY	NIE JEST GENEROWANY
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	
BŁĄD MAPE [%]	23,001
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY

Tab. 5.13. Zestawienie obliczeń modelu SVM Wielomian(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą wektorów nośnych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>SVM Wielomian(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>																																																																																																
ZMIENNA ZALEŻNA	v1																																																																																																
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16																																																																																																
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16																																																																																																
TYP JĄDRA	WIELOMIAN																																																																																																
WZÓR REGRESYJNY	NIE JEST GENEROWANY																																																																																																
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	<table border="1"> <thead> <tr> <th>Opóśn</th> <th>Koz.</th> <th>S.E</th> <th>Opóśn</th> <th>Koz.</th> <th>S.E</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,607</td><td>,1010</td><td>1</td><td>+,607</td><td>,1026</td></tr> <tr><td>2</td><td>+,396</td><td>,1005</td><td>2</td><td>+,043</td><td>,1026</td></tr> <tr><td>3</td><td>+,427</td><td>,0999</td><td>3</td><td>+,270</td><td>,1026</td></tr> <tr><td>4</td><td>+,507</td><td>,0994</td><td>4</td><td>+,248</td><td>,1026</td></tr> <tr><td>5</td><td>+,448</td><td>,0988</td><td>5</td><td>+,051</td><td>,1026</td></tr> <tr><td>6</td><td>+,306</td><td>,0983</td><td>6</td><td>+,078</td><td>,1026</td></tr> <tr><td>7</td><td>+,272</td><td>,0977</td><td>7</td><td>+,004</td><td>,1026</td></tr> <tr><td>8</td><td>+,255</td><td>,0972</td><td>8</td><td>+,070</td><td>,1026</td></tr> <tr><td>9</td><td>+,342</td><td>,0966</td><td>9</td><td>+,191</td><td>,1026</td></tr> <tr><td>10</td><td>+,359</td><td>,0960</td><td>10</td><td>+,110</td><td>,1026</td></tr> <tr><td>11</td><td>+,323</td><td>,0955</td><td>11</td><td>+,088</td><td>,1026</td></tr> <tr><td>12</td><td>+,211</td><td>,0949</td><td>12</td><td>+,070</td><td>,1026</td></tr> <tr><td>13</td><td>+,283</td><td>,0943</td><td>13</td><td>+,068</td><td>,1026</td></tr> <tr><td>14</td><td>+,241</td><td>,0938</td><td>14</td><td>+,070</td><td>,1026</td></tr> <tr><td>15</td><td>+,342</td><td>,0932</td><td>15</td><td>+,017</td><td>,1026</td></tr> </tbody> </table>	Opóśn	Koz.	S.E	Opóśn	Koz.	S.E	1	+,607	,1010	1	+,607	,1026	2	+,396	,1005	2	+,043	,1026	3	+,427	,0999	3	+,270	,1026	4	+,507	,0994	4	+,248	,1026	5	+,448	,0988	5	+,051	,1026	6	+,306	,0983	6	+,078	,1026	7	+,272	,0977	7	+,004	,1026	8	+,255	,0972	8	+,070	,1026	9	+,342	,0966	9	+,191	,1026	10	+,359	,0960	10	+,110	,1026	11	+,323	,0955	11	+,088	,1026	12	+,211	,0949	12	+,070	,1026	13	+,283	,0943	13	+,068	,1026	14	+,241	,0938	14	+,070	,1026	15	+,342	,0932	15	+,017	,1026
Opóśn	Koz.	S.E	Opóśn	Koz.	S.E																																																																																												
1	+,607	,1010	1	+,607	,1026																																																																																												
2	+,396	,1005	2	+,043	,1026																																																																																												
3	+,427	,0999	3	+,270	,1026																																																																																												
4	+,507	,0994	4	+,248	,1026																																																																																												
5	+,448	,0988	5	+,051	,1026																																																																																												
6	+,306	,0983	6	+,078	,1026																																																																																												
7	+,272	,0977	7	+,004	,1026																																																																																												
8	+,255	,0972	8	+,070	,1026																																																																																												
9	+,342	,0966	9	+,191	,1026																																																																																												
10	+,359	,0960	10	+,110	,1026																																																																																												
11	+,323	,0955	11	+,088	,1026																																																																																												
12	+,211	,0949	12	+,070	,1026																																																																																												
13	+,283	,0943	13	+,068	,1026																																																																																												
14	+,241	,0938	14	+,070	,1026																																																																																												
15	+,342	,0932	15	+,017	,1026																																																																																												
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT																																																																																																	
BŁĄD MAPE [%]	29,99																																																																																																
OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY																																																																																																



Tab. 5.14. Zestawienie obliczeń modelu SVM RBF(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą wektorów nośnych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>SVM RBF(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>																																																																																																																																
ZMIENNA ZALEŻNA	v1																																																																																																																																
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16																																																																																																																																
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16																																																																																																																																
TYP JĄDRA	RBF																																																																																																																																
WZÓR REGRESYJNY	NIE JEST GENEROWANY																																																																																																																																
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH																																																																																																																																	
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	<table border="1"> <thead> <tr> <th>Opóśn</th> <th>Kor.</th> <th>S.E</th> <th>Q</th> <th>p</th> <th>Opóśn</th> <th>Kor.</th> <th>S.E</th> </tr> </thead> <tbody> <tr><td>1</td><td>+,084</td><td>,1010</td><td>,69</td><td>,4052</td><td>1</td><td>+,084</td><td>,1026</td></tr> <tr><td>2</td><td>+,134</td><td>,1005</td><td>2,48</td><td>,2890</td><td>2</td><td>+,128</td><td>,1026</td></tr> <tr><td>3</td><td>+,044</td><td>,0999</td><td>2,67</td><td>,4448</td><td>3</td><td>+,024</td><td>,1026</td></tr> <tr><td>4</td><td>+,016</td><td>,0994</td><td>2,70</td><td>,6093</td><td>4</td><td>-,006</td><td>,1026</td></tr> <tr><td>5</td><td>-,001</td><td>,0988</td><td>2,70</td><td>,7462</td><td>5</td><td>-,010</td><td>,1026</td></tr> <tr><td>6</td><td>-,334</td><td>,0983</td><td>14,27</td><td>,0268</td><td>6</td><td>-,344</td><td>,1026</td></tr> <tr><td>7</td><td>+,046</td><td>,0977</td><td>14,49</td><td>,0432</td><td>7</td><td>+,108</td><td>,1026</td></tr> <tr><td>8</td><td>-,136</td><td>,0972</td><td>16,45</td><td>,0364</td><td>8</td><td>-,073</td><td>,1026</td></tr> <tr><td>9</td><td>+,014</td><td>,0966</td><td>16,47</td><td>,0577</td><td>9</td><td>+,047</td><td>,1026</td></tr> <tr><td>10</td><td>-,005</td><td>,0960</td><td>16,47</td><td>,0869</td><td>10</td><td>+,028</td><td>,1026</td></tr> <tr><td>11</td><td>-,051</td><td>,0955</td><td>16,76</td><td>,1154</td><td>11</td><td>-,065</td><td>,1026</td></tr> <tr><td>12</td><td>+,093</td><td>,0949</td><td>17,72</td><td>,1245</td><td>12</td><td>-,010</td><td>,1026</td></tr> <tr><td>13</td><td>-,063</td><td>,0943</td><td>18,16</td><td>,1516</td><td>13</td><td>-,013</td><td>,1026</td></tr> <tr><td>14</td><td>+,059</td><td>,0938</td><td>18,56</td><td>,1827</td><td>14</td><td>-,017</td><td>,1026</td></tr> <tr><td>15</td><td>+,018</td><td>,0932</td><td>18,59</td><td>,2328</td><td>15</td><td>+,073</td><td>,1026</td></tr> </tbody> </table>	Opóśn	Kor.	S.E	Q	p	Opóśn	Kor.	S.E	1	+,084	,1010	,69	,4052	1	+,084	,1026	2	+,134	,1005	2,48	,2890	2	+,128	,1026	3	+,044	,0999	2,67	,4448	3	+,024	,1026	4	+,016	,0994	2,70	,6093	4	-,006	,1026	5	-,001	,0988	2,70	,7462	5	-,010	,1026	6	-,334	,0983	14,27	,0268	6	-,344	,1026	7	+,046	,0977	14,49	,0432	7	+,108	,1026	8	-,136	,0972	16,45	,0364	8	-,073	,1026	9	+,014	,0966	16,47	,0577	9	+,047	,1026	10	-,005	,0960	16,47	,0869	10	+,028	,1026	11	-,051	,0955	16,76	,1154	11	-,065	,1026	12	+,093	,0949	17,72	,1245	12	-,010	,1026	13	-,063	,0943	18,16	,1516	13	-,013	,1026	14	+,059	,0938	18,56	,1827	14	-,017	,1026	15	+,018	,0932	18,59	,2328	15	+,073	,1026
Opóśn	Kor.	S.E	Q	p	Opóśn	Kor.	S.E																																																																																																																										
1	+,084	,1010	,69	,4052	1	+,084	,1026																																																																																																																										
2	+,134	,1005	2,48	,2890	2	+,128	,1026																																																																																																																										
3	+,044	,0999	2,67	,4448	3	+,024	,1026																																																																																																																										
4	+,016	,0994	2,70	,6093	4	-,006	,1026																																																																																																																										
5	-,001	,0988	2,70	,7462	5	-,010	,1026																																																																																																																										
6	-,334	,0983	14,27	,0268	6	-,344	,1026																																																																																																																										
7	+,046	,0977	14,49	,0432	7	+,108	,1026																																																																																																																										
8	-,136	,0972	16,45	,0364	8	-,073	,1026																																																																																																																										
9	+,014	,0966	16,47	,0577	9	+,047	,1026																																																																																																																										
10	-,005	,0960	16,47	,0869	10	+,028	,1026																																																																																																																										
11	-,051	,0955	16,76	,1154	11	-,065	,1026																																																																																																																										
12	+,093	,0949	17,72	,1245	12	-,010	,1026																																																																																																																										
13	-,063	,0943	18,16	,1516	13	-,013	,1026																																																																																																																										
14	+,059	,0938	18,56	,1827	14	-,017	,1026																																																																																																																										
15	+,018	,0932	18,59	,2328	15	+,073	,1026																																																																																																																										
BŁĄD MAPE [%]	14,58																																																																																																																																
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY																																																																																																																																

Tab. 5.15. Zestawienie obliczeń modelu SVM Sigmoidalne(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16) metodą wektorów nośnych

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>SVM Sigmoidalne(v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)</b>
ZMIENNA ZALEŻNA	v1
ZMIENNE NIEZALEŻNE	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16
PREDYKTORY	v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16
TYP JĄDRA	Sigmoidalne
WZÓR REGRESYJNY	NIE JEST GENEROWANY
WYKRES WARTOŚCI RZECZYWISTYCH i PROGNOZOWANYCH	
WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT	
BŁĄD MAPE [%]	27,79
OCENA PRAWIDŁOŚCI MODELU REGRESYJNEGO	NIEPRAWIDŁOWY

Wykonano analizę prawidłowości i wybór najlepszego modelu prognostycznego zgodnie z punktem 4.7. w tabeli 5.16. zestawiono wszystkie obliczone modele prognostyczne. Najlepszym modelem o najmniejszym błędzie prognozy jest model GAM2(v1;v3,v4,v8,v9,v13,v10,v11,v37,v30,v22).

Tab. 5.16. Zestawienie parametrów obliczonych modeli prognostycznych

LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszt	Wartość błędu MAPE [%]	Poprawność modelu
1	MR1 (v1;v3,v4,v6, v9,v10,v12)	MR1(v1;v3,v4,v6,v9,v10,v12)= 806,4193-1,5490*v3 +13,3083*v4 +1,7623*v6 -4,2776*v9 +0,1538*v10 -2,0725*v12	NIE	21,02	TAK
2	MR2 (v1;v3,v4,v6)	MR2(v1;v3,v4,v6) = 139,5031-1,3662*v3 +12,4568*v4 +1,6958*v6	TAK	26,44	NIE
3	MR3 (v1;v4,v9,v6,v10,v29,v3, v12,v37,v22,v30,v7)	MR3(v1;v4,v9,v6,v10,v29,v3,v12,v7,v22, v30,v7) = 837,4718 +12,6576* v4 -4,0240*v9 + 2,0390*v6+0,4351*v10-0,0000*v29-2,0071*v3-2,1400*v12+ 9,9969*v37-0,1459*v22-0,0139*v30-0,4890*v7		17,72	
4	MARS1 (v1;v3,v4,v5,v8,v9,v13,v15,v11,v16)	Wydajność koparki = 1,79706819408758e+ 001 + 1,19446716463054e+001*max(0; Pojemność łyżki-1,00000000000000e+000) - 1,73912859343404e+001*max(0; 1,00000000000000e+000-Pojemność łyżki) - 4,26315025699904e+000*-max(0; Kod kategoria gruntu-1,02000000000000e+002) + 1,75923886157349e+000*max(0; Lata pracy operatora-1,00000000000000e+000) + 1,71821481315091e-001*max(0; Temperatura+5,00000000000000e+000) - 1,78708386834731e+000*max(0; Kod rodzaj łyżki-1,01000000000000e+002) + 3,33199437052750e+000*max(0; 1,02000000000000e+002-Kod wilgotność gruntu) - 5,69661471300789e-001*max(0; Godziny pracy-6,00000000000000e+000) + 4,07842963164529e+001*max(0; Awaryjność/ naprawy-8,30000000000000e-001)	NIE	17,72	TAK
5	GAM1 (v1;v4,v8,v15,v16)	GAM1(v1;v4,v6,v10,v13) = e^( -0,139001 + 0,685827*v4 + 0,065085*v8 + 0,010339*v15 + 2,243254*v16)	TAK	18,06	NIE

6	SVM Liniowy (v1;v3,v4,v5,v7,v8,v9, v11,v13,v14,v15,v16)	NIE JEST GENEROWANY	NIE	23,00	TAK
7	SVM Wielomian (v1;v3,v4,v5,v7,v8,v9, v11, v13,v14,v15,v16)		TAK	29,99	NIE
8	<b>SVM RBF</b> <b>(v1;v3,v4,v5,v7,v8,v9,</b> <b>v11,v13,v14,v15,v16)</b>		NIE	14,58	TAK
9	SVM Sigmoidalne (v1;v3,v4,v5,v7,v8,v9, v11,v13,v14,v15,v16)			27,79	
10	NN1 MLP 11-9-1 (v1;v3, v4,v5,v7,v8,v9,v11,v13,v1 4,v15,v16)			15,83	
11	NN2 MLP 11-13-1 (v1;v3,v4,v5,v7,v8,v9, v11,v13,v14,v15,v16)			22,76	
12	NN3 RBF 11-14-1 (v1;v3,v4,v5,v7,v8,v9, v11,v13,v14,v15,v16)		TAK	50,00	NIE

Zatem rozwiązaniem problemu prognostycznego jest sieć:

**SVM RBF (v1;v3,v4,v5,v7,v8,v9,v11,v13,v14,v15,v16)**

Wydajność koparek jest w tym modelu zależna od zmiennych zestawionych w tabeli 5.17.

Tab. 5.17. Zestawienie zmiennych niezależnych w modelu prognostycznym SVM RBF (v1;v3,v4,v5,v7, v8,v9,v11,v13,v14,v15,v16)

Lp.	Zmienna	Opis zmiennej
1	v3	Kodowany rodzaj łyżki koparki
2	v4	Pojemność łyżki koparki
3	v5	Ładowność środka transportowego
4	v7	Kodowany rodzaj drogi
5	v8	Lata pracy operatora
6	v9	Liczba godzin pracy w dniu roboczym
7	v11	Kodowana wilgotność gruntu
8	v13	Kodowana kategoria gruntu
9	v14	Odległość wywozu gruntu
10	v15	Temperatura
11	v16	Awaryjność/naprawy

Błąd *MAPE* prognozy wynosi 14,58% – czyli zgodnie z punktem 3.6, prognoza z błędem dopuszczalnym.

## 5.2 Roboty żelbetowe

Celem obliczeń prognostycznych jest wygenerowanie wzoru regresyjnego, który umożliwi oszacowanie czasu wykonania elementu żelbetowego. Obliczenia wykonano na podstawie danych, które zostały zebrane na budowie osiedla mieszkaniowego przy ul. Koncertowej w Lublinie w firmie budowlanej SANITEX S.A. na przełomie lipca i sierpnia 2014r. Brygady budowlane składające się z cieśli oraz zbrojarzy były zatrudnione bezpośrednio przez firmę SANITEX S.A. Dane (czas trwania zbrojenia, deskowania i betonowania) były zbierane na bieżąco podczas ośmiogodzinnej zmiany pracy robotników. Każdego dnia były zapisywane dane dotyczące czynników wpływających na czas trwania danego procesu (wysokość, długość, szerokość, rodzaj elementu, pole i stopień zbrojenia, planowane oraz rzeczywiste zużycie betonu, liczebność brygady, godzina pracy betonowania oraz sposób podawania betonu). Obliczenia przedstawiono w formie skróconej ze względu na ich zakres. w tabeli 5.18. zestawiono zmienne, arkusz kalkulacyjny do obliczeń to tabela 5.19.

Tab. 5.18. Zestawienie zmiennych do obliczeń

Lp	Zmienna	Opis zmiennej	Uwagi
1	v1	Czas wykonania elementu	min
2	v2	Czas zbrojenia	
3	v3	Czas deskowania	
4	v4	Czas betonowania	
5	v5	Wysokość elementu	cm
6	v6	Długość elementu	
7	v7	Szerokość elementu	
8	v8	Rodzaj elementu	Zmienna lingwistyczna
9	v9	Kod Rodzaj elementu	ściana – 101; słup – 102; ława – 103; stopa – 104; podciąg – 105
10	v10	Pole zbrojenia	cm <sup>2</sup>
11	v11	Stopień zbrojenia	
12	v12	Planowane zużycie betonu	m <sup>3</sup>
13	v13	Rzeczywiste zużycie betonu	
14	v14	Liczebność brygady	
15	v15	Godzina pracy betonowania	
16	v16	Podawanie betonu	Zmienna lingwistyczna
17	v17	Kod Podawanie betonu	Pompa – 101; podajnik – 102
18	v18	Czas zbrojenia wg KNR	min
19	v19	Czas deskowania wg KNR	
20	v20	Czas betonowania wg KNR	
21	v21	Czas wykonania elementu wg KNR	
22	v22	[box-cox] Stopień zbrojenia	Zastosowano przekształcenie Boxa-Coxa
23	v23	[box-cox] Planowane zużycie betonu	
24	v24	[box-cox] Rzeczywiste zużycie betonu	
25	v25	[box-cox] Liczebność brygady	
26	v26	[box-cox] wysokość elementu	
27	v27	[box-cox] długość elementu	
28	v28	[box-cox] szerokość elementu	

Tab. 5.19. Zestawienie danych – arkusz kalkulacyjny programu STATISTICA

Czas				Element					Pole zbrojenia $A_s$ [cm <sup>2</sup> ]	Stopień zbrojenia	zużycie betonu		
wykonania elementu [min]	zbrojenia [min]	deskowania [min]	betonowania [min]	wysokość [cm]	długość e [cm]	szerokość [cm]	Rodzaj	Kod Rodzaj			Planowanie	Rzeczywiste	Liczebność brygady
v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14
1468	840	555	73	395	2649	24	ściana	101	80,42	0,848312236	25,11252	23	11
187	107	65	15	300	30		słup	102	25,13	2,792222222	0,27	0,5	4
199	113	73	13		25,13	2,792222222			0,27				
1483	873	523	87	395	2966	24	ściana	101	80,42	0,848312236	28,11768	29	10
83	42	16	25	60	360	65	ława	103	9,05	0,232051282	1,404	1,5	3
179	94	78	7	285	30	30	słup	102	12,06	1,34	0,2565	0,5	4
207	107	83	17			62			16,08	0,864516129	0,5301		
456	155	248	53	390	620	24	ściana	101	80,42	0,859188034	5,8032	6	9
259	125	106	28	290	50		słup	102	58,9	2,356	0,725	1	3
166	52	94	20	350	30	35			18,85	1,795238095	0,3675	0,4	5
204	95	87	22		24	8,04			1,116666667	0,252	0,3		
230	120	91	19	285	30	8,04			0,893333333	0,2565	4		
1051	618	406	27	395	1455	24	ściana	101	80,42	0,848312236	13,7934	14	11
122	67	47	8	305	30	24	słup	102	8,04	1,116666667	0,2196	0,5	3
145	79	50	16			350			35	8,04	0,765714286		
174	128	23	23	75	736	30	ława	103	25,13	1,116888889	1,656	2	
1528	905	560	63	395	2742	24	ściana	101	80,42	0,848312236	25,99416	26	10
148	63	65	20	285	35	85	słup	102	12,06	0,405378151	0,847875	1	3
166	71	69	26			85			16,08	0,540504202	0,847875		
149	58	65				40			59	12,06	0,511016949	0,6726	
1436	826	560	50			395			2143	24	ściana	101	80,42
140	56	74	10	285	30	84	słup	102	16,08	0,638095238	0,7182	1	5
165	86	67	12			300			30	25,13	2,792222222		
157	80	68	9			285			84	16,08	0,638095238	0,7182	1
1069	597	430	42			395			1468	24	ściana	101	80,42
195	152	32	11	50	736	30	ława	103	25,13	1,675333333	1,104	1,5	3
395	330	40	25	45	1081	75			25,13	0,744592593	3,648375	4	5
95	67	23	5		160	145	stopa	104	20,36	0,312030651	1,044	1	2
1044	603	417	24	395	1570	24	ściana	101	80,42	0,848312236	14,8836	15	10
339	94	234	11	65	582		podciąg	105	9,61	0,068800115	0,90792	1	8
491	163	319	9	35	545				14,83	0,113379205	0,4578	0,5	
951	375	564	12	20	1236				4,52	0,015237325	0,59328	1	
175	77	75	23	285	30	84			słup	102	16,08	0,638095238	0,7182
1037	524	400	113	395	1355	24	ściana	101	80,42	0,848312236	12,8454	13	9
734	655	49	30	40	3400	55	ława	103	25,13	1,142272727	7,48	7,5	4
129	93	31	5		125	125	stopa	104	16,08	0,3216	0,625	1	

cd. Tab. 5.19. Zestawienie danych – arkusz kalkulacyjny programu STATISTICA

Godzina pracy betonowania	Podawanie betonu	Kod Podawanie betonu	Czas wg KNR				[cox-box]						
			zbrojenia [min]	deskowania [min]	betonowania [min]	wykonaniaelementu [min]	Stopień zbrojenia	Planowane zużycie betonu	Rzeczywiste zużycie betonu	Liczebność brygady	wysokość elementu [cm]	długość elementu [cm]	szerokość elementu [cm]
v15	v16	v17	v18	v19	v20	v21	v22	v23	v24	v25	v26	v27	v28
6	pompa	101	795	665	89	1549	-0,15877	2,01798	1,87595	1,671669	768,4621	5,224224	0,830938
7			34	52	7	93	1,29362	-1,62337	-0,78770	1,118604	562,5484	2,825504	0,835633
5			890	497	116	1503	-0,15877	2,05789	1,94754	1,626893	768,4621	5,270449	0,830938
6	podajnik pompa	102	58	18	54	131	-1,08202	0,32169	0,37716	0,925579	90,1944	4,303156	0,844981
			33	49	6	89	0,31206	-1,70173	-0,78770	1,118604	530,7573	2,825504	0,835633
			34	114	13	160	-0,14108	-0,70314	0,00000				0,844621
5	podajnik	101	184	123	72	378	-0,14688	1,34755	1,31893	1,575771	757,4446	4,574883	0,830938
			92	84	43	219	1,03796	-0,33857	0,00000	0,925579	541,3303	3,163928	0,842701
6	podajnik	101	47	66	22	134	0,66610	-1,17834	-1,08625	1,256506	669,9951	2,825504	0,838227
			32	55	15	101	0,11303	-1,72907	-1,50844	1,256506			0,830938
7	podajnik	101	33	49	15	98	-0,11008	-1,70173		1,118604	530,7573	4,969112	0,835633
			436	292	67	796	-0,15877	1,78076	1,70086	1,671669	768,4621		
4	pompa	102	28	71	5	105	0,11303	-1,94694	-0,78770	0,925579	573,1928	2,825504	0,830938
			47	99	22	167	-0,25207	-1,17834	-0,78770		669,9951	2,825504	0,838227
5	podajnik	101	69	94	39	202	0,11324	0,46610	0,61314	1,626893	768,4621	5,238397	0,830938
			823	459	92	1374	-0,15877	2,03032	1,91456	0,925579	530,7573	2,929688	0,846689
6	podajnik	101	108	99	21	227	-0,74710	-0,16942	0,00000				
			42	122	17	181	-0,58244	-0,42265					
5	podajnik	101	643	431	72	1146	-0,15877	1,93916	1,84617	1,671669	768,4621	5,135932	0,830938
6	pompa	101	91	94	18	203	-0,40820	-0,34902	0,00000	1,256506	530,7573	2,825504	0,846624
			34	77	7	118	1,29362	-1,62337	-0,78770		562,5484		0,835633
7	pompa	101	46	94	18	157	-0,40820	-0,34902	0,00000	530,7573	4,973027	0,846624	
			356	369	57	782	-0,15877	1,78462	1,70086	1,626893			768,4621
8	pompa	101	91	63	13	167	0,57832	0,09740	0,37716	0,925579	73,2062	4,657217	0,835633
			101	292	43	435	-0,27683	1,06138	1,09041	1,256506	64,8744	4,836077	0,845959
4	pompa	101	120	58	8	186	-0,91433	0,04277	0,00000	0,621356	64,8744	3,865597	0,848837
			377	394	46	817	-0,15877	1,81343	1,72711	1,626893	768,4621	5,002463	0,830938
5	podajnik	102	36	178	8	222	-1,58339	-0,09810	0,00000	1,516535	98,8344	4,544116	
			27	120	13	160	-1,40871	-0,88674	-0,78770		48,5893	4,511932	
6	pompa	101	35	189	16	240	-1,92992	-0,56784	0,00000	1,118604	25,3697	4,896624	
			27	94	28	148	-0,40820	-0,34902			530,7573	2,825504	0,846624
6	pompa	101	407	340	142	889	-0,15877	1,74945	1,67193	1,575771	768,4621	4,937635	0,830938
			206	174	58	439	0,13693	1,48710	1,43111	1,118604	56,6644	5,325517	0,843612
6	podajnik	101	36	42	7	85	-0,89599	-0,50687	0,00000		3,724265	0,848368	



Tab. 5.20. Zestawienie obliczeń – zmienna zależna v1 – Czas wykonania elementu żelbetowego

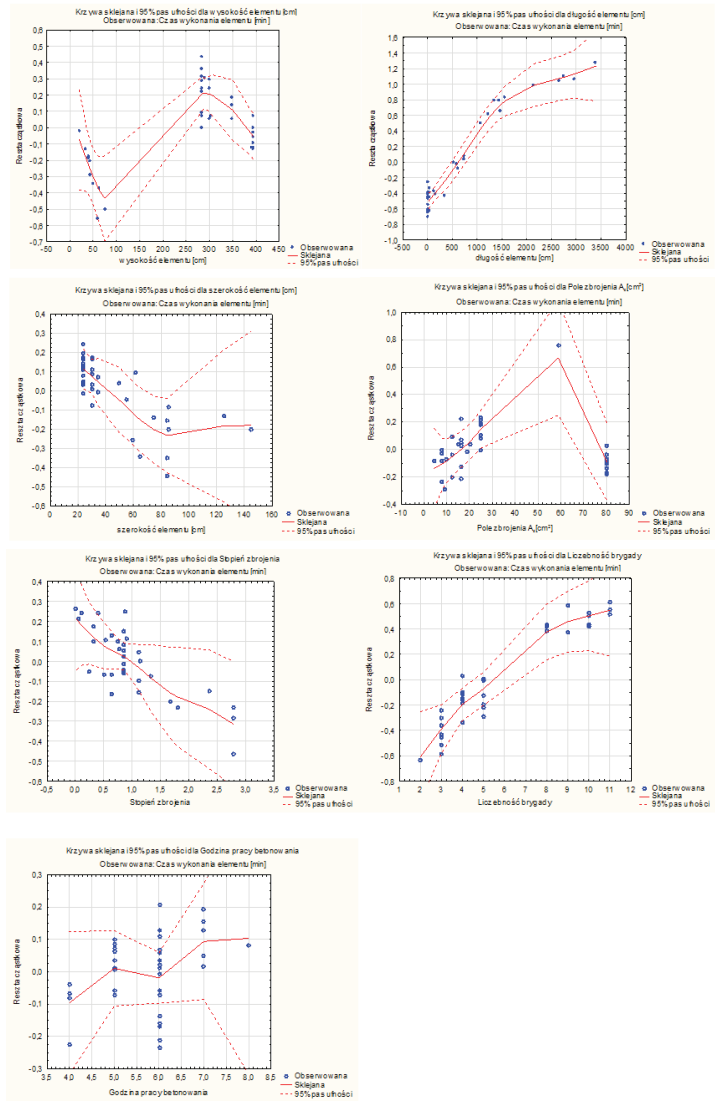
LP	Nazwa modelu	Wzór regresyjny	Występowanie autokorelacji reszt	Wartość błędu MAPE [%]	Poprawność modelu
1	MR1 (v1,v6,v13,v14)	$MR1(v1;v6,v13,v14) = -111,403\ 0,154*v6 + 20,394*v13 + 63,169*v14$	NIE	8,68871	TAK
2	MR2 (v1,v23,v26,v27,v28,v14)	$MR1(v1;v23,v26,v27,v28,v14) = -48989,0 - 502,4*v23 + 1,9*v26 + 1027,4*v27 + 52860,9*v28 + 55,8*v14$	TAK	13,34404	NIE
3	MARS1 (v1,v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)	Czas wykonania elementu [min] = 2,94778249240050e+001 + 1,78226031420766e+001*max(0; Planowanie zużycie betonu-2,19600000000000e-001) + 6,85257692783139e+001*max(0; Liczebność brygady-2,00000000000000e+000) + 1,82614203675142e-001*max(0; długość elemen- tu [cm]-3,60000000000000e+002)	NIE	8,08375	TAK
4	GAM (v1; v5,v6,v7,v10,v11,v14,v15)	$GAM(v1;v5,v6,v7,v10,v11,v14,v15) = e^{(4,472183 + 0,000948*v5 + 0,000610*v6 - 0,003976*v7 - 0,000368*v10 - 0,183670*v11 + 0,123908*v14 + 0,042312*v15)}$	NIE	2,89273	TAK
5	SVM Liniowy (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)	NIE JEST GENEROWANY	NIE	7,49031	TAK
6	SVM Wielomian (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)		NIE	10,53163	TAK
7	SVM RBF (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)		NIE	10,15406	TAK
8	SVM Sigmoidalne (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)		NIE	8,90286	TAK
9	NN1 MLP 20-5-1 (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)		NIE	3,93760	TAK
10	NN2 MLP 20-11-1 (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)		TAK	3,98501	NIE
11	NN3 MLP 20-10-1 (v1; v5,v6,v7,v9,v10,v11,v12,v13,v14,v15,v17)		TAK	3,73560	NIE

Przeprowadzono obliczenia jedenastu modeli regresyjnych, z których otrzymane wyniki (tab. 5.20.) aż w ośmiu przypadkach mogły być wykorzystane do dalszych obliczeń (błąd MAPE nie przekraczał umownej granicy 15% oraz nie wystąpiły autokorelacje i autokorelacje cząstkowe reszt). Metoda uogólnionych modeli addytywnych GAM1 okazała się najdokładniejsza, szczegółowe wyniki obliczeń modelu przedstawiono w tabeli 5.21.

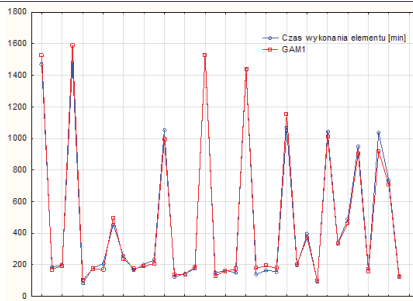
Tab. 5.21. Tabela obliczeń, uogólnione modele addytywne GAM1 (v1; v5,v6,v7,v10,v11,v14,v15)

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>GAM(v1; v5,v6,v7,v10,v11,v14,v15)</b>						
ZMIENNA ZALEŻNA	v1 – czas wykonania elementu żelbetowego						
ZMIENNE NIEZALEŻNE	v5,v6,v7,v10,v11,v14,v15						
PREDYKTORY	v5,v6,v7,v10,v11,v14,v15						
PODSUMOWANIE REGRESJI	Podsumowanie dopasowania. Obserwowana: Czas wykonania elementu [min] Rozkład: Gamma; Funkcja wiążąca: Log						
		Indeks zmiennej	Stopnie swobody	GAM współcz.	Błąd standardowy	Standaryz. wyniku	Nieliniowe p-wartość
	w.wolny	0	1,00000	4,472183	0,337476	13,25184	
	wysokość elementu v5	1	4,00225	0,00094	0,000398	2,38407	0,000197
	długość elementu v6	2	3,99512	0,00061	0,000063	9,70514	0,000820
	szerokość elementu v7	3	3,99935	-0,00397	0,001957	-2,03162	0,560217
	Pole zbrojenia v10	4	3,99853	-0,00036	0,003011	-0,12228	0,000237
	Stopień zbrojenia v11	5	4,00302	-0,18367	0,079256	-2,31743	0,977854
	Liczebność brygady v14	6	4,00066	0,12390	0,031468	3,93766	0,474997
Godzina pracy betonowania v15	7	3,99963	0,04231	0,047847	0,88432	0,865458	
WZÓR REGRESYJNY	$\text{GAM}(v1; v5,v6,v7,v10,v11,v14,v15) = e^{(4,472183 + 0,000948*v5 + 0,000610*v6 - 0,003976*v7 - 0,000368*v10 - 0,183670*v11 + 0,123908*v14 + 0,042312*v15)}$						

KRZYWE SKLEJANE  
i 95% PAS UFNOŚCI



WYKRES WARTOŚCI  
RZECZYWISTYCH  
i PROGNOZOWANYCH



<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BŁĄD MAPE [%]</p>	<p>2,89273</p>
<p>OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO</p>	<p>PRAWIDŁOWY</p>

Porównano (tab. 5.22.) rzeczywiste czasy wykonania elementów żelbetowych z czasami wg KNR. Rzeczywiste czasy wykonania elementu żelbetowego (pomierzone na budowie) różnią się od czasów wyznaczonych zgodnie z KNR. Nie występują autokorelacje i autokorelacje cząstkowe reszt (pomiarzy wykazują dużą prawidłowość). Błąd MAPE wynosi 11,35767%.

Tab. 5.22. Porównanie rzeczywistych czasów wykonania elementów żelbetowych z czasami wg KNR

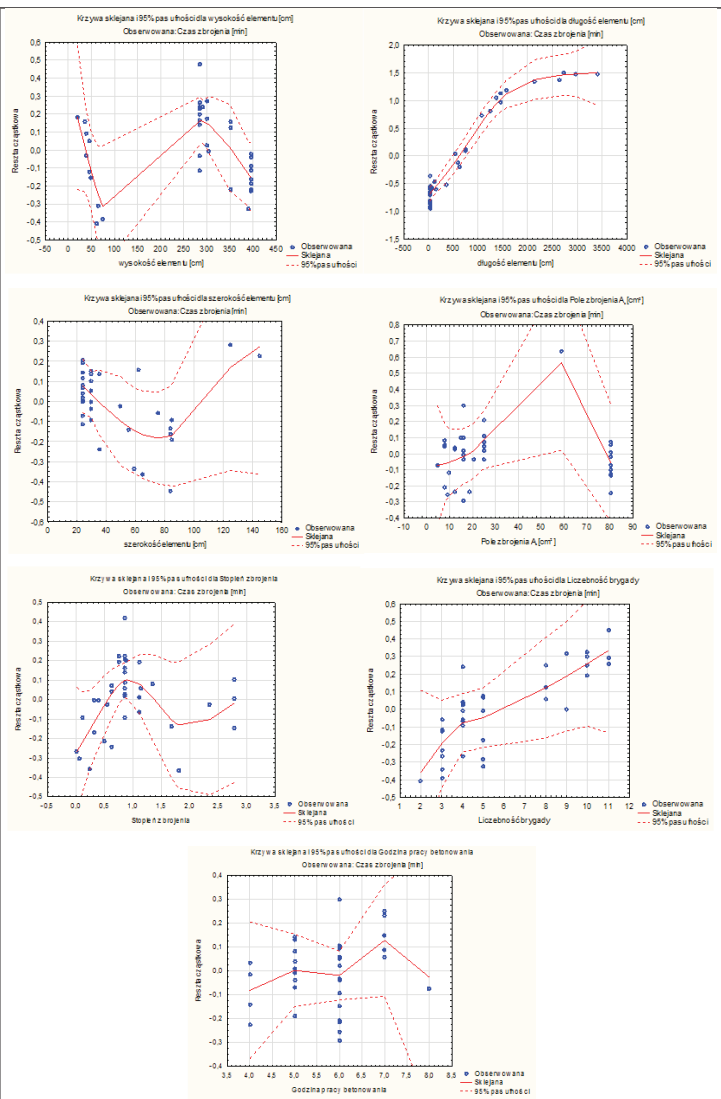
<p>WYKRES WARTOŚCI RZECZYWISTYCH i DANYCH z KNR</p>	
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BŁĄD MAPE [%]</p>	<p>11,35767</p>
<p>OCENA PRAWIDŁOWOŚCI</p>	<p>PRAWIDŁOWY</p>

Wynik obliczeń wskazuje, że na jednym lub kilku etapach prac żelbetowych występuje niezgodność z KNR. w celu wyjaśnienia wykonano obliczenia prognostyczne czasu zbrojenia elementu żelbetowego (v2) stosując metodę GAM. Wyniki obliczeń zestawiono w tabeli 5.33.

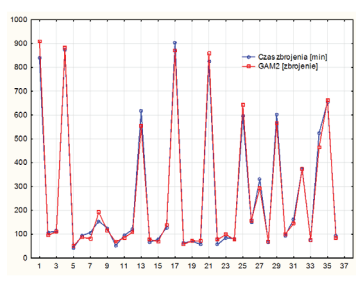
Tab. 5.33. Metoda uogólnionych modeli addytywnych GAM(v2; v5,v6,v7,v10,v11,v14,v15)

METODA OBLICZEŃ i MODEL REGRESYJNY	<b>GAM(v2; v5,v6,v7,v10,v11,v14,v15)</b>						
ZMIENNA ZALEŻNA	v2 – czas zbrojenia						
ZMIENNE NIEZALEŻNE	v5,v6,v7,v10,v11,v14, v15						
PREDYKTORY	v5,v6,v7,v10,v11,v14, v15						
PODSUMOWANIE REGRESJI	Podsumowanie dopasowania Obserwowana: Czas zbrojenia [min] Rozkład: Gamma; Funkcja wiążąca: Log						
		Indeks zmiennej	Stopnie swobody	GAM współcz.	Błąd standardowy	Standaryz. wyniku	Nieliniowe p-wartość
	w. wolny	0	1,000000	3,901474	0,438459	8,898159	
	wysokość elementu v5	1	4,002251	0,000080	0,000517	0,155451	0,017732
	długość elementu v6	2	3,995126	0,000817	0,000082	9,993069	0,000019
	szerokość elementu v7	3	3,999355	-0,000991	0,002543	-0,389742	0,202597
	Pole zbrojenia v10	4	3,998531	0,000013	0,003912	0,003200	0,111385
	Stopień zbrojenia v11	5	4,003020	0,024009	0,102972	0,233163	0,152418
	Liczebność brygady v14	6	4,000664	0,063132	0,040883	1,544183	0,905633
	Godzina pracy betonowania v15	7	3,999639	0,038451	0,062165	0,618543	0,843283
WZÓR REGRESYJNY	$GAM(v2; v5,v6,v7,v10,v11,v14,v15) = e^{(3,901474 + 0,000080*v5 + 0,000817*v6 - 0,000991*v7 - 0,000013*v10 + 0,024009*v11 + 0,063132*v14 + 0,038451*v15)}$						

KRZYWE SKLEJANE  
i 95% PAS UFNOŚCI



WYKRES WARTOŚCI  
RZECZYWISTYCH i  
PROGNOZOWA-  
NYCH



<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BŁĄD MAPE [%]</p>	<p>3,76687</p>
<p>OCENA PRAWIDŁOWOŚCI MODELU REGRESYJNEGO</p>	<p>PRAWIDŁOWY</p>

Przedstawiony model może być użyty do dalszych obliczeń. Nie stwierdzono występowania autokorelacji oraz autokorelacji cząstkowej reszt. Wartość błędu MAPE na poziomie 3,76687% oznacza, że prognoza jest dobra. Otrzymany model GAM(v2;v5,v6,v7,v10,v11, v14,v15) porównano z wartościami katalogowymi z KNR (tab. 5.34.).

Tab. 5.34. Porównanie rzeczywistych czasów zbrojenia z czasami wg KNR

<p>WYKRES WARTOŚCI RZECZYWISTYCH i DANYCH z KNR</p>	
<p>WYKRESY FUNKCJI AUTOKORELACJI RESZT i AUTOKORELACJI CZĄSTKOWYCH RESZT</p>	
<p>BŁĄD MAPE [%]</p>	<p>17,50592</p>
<p>OCENA PRAWIDŁOWOŚCI</p>	<p>NIEPRAWIDŁOWY</p>

Rzeczywiste czasy zbrojenia (pomierzone na budowie), nie pokrywają się z czasami z katalogów nakładów rzeczowych. Mimo, że autokorelacje i autokorelacje cząstkowe reszt nie występują (co świadczy o występujących prawidłowościach), błąd *MAPE* przekracza umowną wartość 15%.

Wnioski:

1. Ustalono wzory regresyjne czasu wykonania elementu żelbetowego ( $v_1$ ) oraz czasu zbrojenia elementu żelbetowego ( $v_2$ ):

$$\begin{aligned} \text{GAM}(v_1; v_5, v_6, v_7, v_{10}, v_{11}, v_{14}, v_{15}) &= e^{(4,472183 + 0,000948*v_5 + 0,000610*v_6 - 0,003976*v_7 - 0,000368*v_{10} - 0,183670*v_{11} + 0,123908*v_{14} + 0,042312*v_{15})} \\ \text{GAM}(v_2; v_5, v_6, v_7, v_{10}, v_{11}, v_{14}, v_{15}) &= e^{(3,901474 + 0,000080*v_5 + 0,000817*v_6 - 0,000991*v_7 - 0,000013*v_{10} + 0,024009*v_{11} + 0,063132*v_{14} + 0,038451*v_{15})}. \end{aligned}$$

2. Błąd *MAPE* prognoz wynosi odpowiednio:

$$\begin{aligned} \text{MAPE}(v_1) &= 2,89273\% \text{ oraz} \\ \text{MAPE}(v_2) &= 3,76687\%, \text{ prognoza dobra.} \end{aligned}$$

3. Odchylenie wartości pomierzonych od wartości katalogowych wynosi odpowiednio  $\text{MAPE}(v_1, \text{KNR}) = 11,35767\%$  oraz  $\text{MAPE}(v_2, \text{KNR}) = 17,50592\%$  i nie wykazuje jednoznacznie wartości większych lub mniejszych. Istotne są czynniki pomierzone i ich wpływ na czas wykonania tj. wysokość elementu, długość elementu, szerokość elementu, pole zbrojenia, stopień zbrojenia, liczebność brygady, godzina pracy betonowania.





## 6. Podsumowanie

Monografia dotyczy badań nad szacowaniem czasu realizacji procesów budowlanych. Praca swoją problematyką wpisuje się w nurt zagadnień prezentowanych w aktualnej literaturze krajowej i światowej, uwzględnia szczególną specyfikę budownictwa, tj. niepewność danych niezbędną w harmonogramowaniu przedsięwzięć budowlanych. Monografia stanowiąca spójną, logiczną całość, wyznacza obszar nowych rozwiązań z zastosowaniem opracowanej metodyki obliczeniowej uwzględniającej wieloczynnikowe modele w prognozowaniu czasu procesów budowlanych. Metodyka uwzględnia realizacyjne uwarunkowania techniczne, technologiczne i organizacyjne prowadzenia robót budowlanych.

Praca ta stanowi rezultat wieloletnich badań, obserwacji i doświadczeń Autorki z praktyki budowlanej oraz z wielu prac naukowych. Można stwierdzić, że czasy wykonania procesów budowlanych w rzeczywistych warunkach realizacyjnych różnią się znacznie od uśrednionych czasów trwania robót, zawartych w bazie normatywnej. Obserwuje się w praktyce budowlanej częste zjawisko nieterminowych realizacji obiektów, wynikających między innymi z przyjęcia do harmonogramowania danych normatywnych niezgodnych z aktualnymi warunkami rzeczywistymi.

W celu uaktualnienia i urealnienia podstaw określania czasów trwania procesów budowlanych, realizowanych w określonych warunkach wykonawczych, opracowano metodykę wyznaczenia czasu, wykorzystując nowe podejście polegające na zastosowaniu analizy metod i modeli statystycznych-prognostycznych.

Nowością naukową jest opracowana metodyka wyznaczenia czasu trwania procesów budowlanych z uwzględnieniem czynników (kubaturowych, organizacyjnych, technicznych, zasobowych, liczbowych, lingwistycznych) wpływających na czas realizacji zdania. Opracowana i zweryfikowana na danych rzeczywistych, Metoda Progностyczno-Regresyjna bazuje na obliczeniach statystycznych. Prognozuje się czas procesu budowlanego bazując na pomiarach i danych zebranych w warunkach rzeczywistych. Na tej podstawie wyznacza się równanie regresji lub sieć zależności, dzięki którym można obliczyć prognozowany czas trwania procesu realizowanego w rzeczywistych warunkach, uwzględniając w obliczeniach czynniki wpływu.

W wyniku przeprowadzonych obliczeń i zastosowanej metodyki wykazano, że czas wykonania danego procesu budowlanego zależy od wielu czynników wpływu i powinien być prognozowany w określonych warunkach realizacyjnych. Zastosowano metody progностyczne: regresji wielorakiej, wieloimiennej regresji adaptacyjnej z użyciem funkcji sklepanych, uogólnionych modeli addytywnych, sztucznych sieci neuronowych, wektorów nośnych i autoregresji zintegrowanej. W wyniku obliczeń różnymi metodami otrzymuje się wiele modeli progностycznych, z których wybiera się do stosowania model charakteryzujący się najmniejszym błędem progностycznym. Jako wybrany przykład skuteczności opracowanej

metody prognostyczno-regresyjnej można przedstawić wzory regresyjne czasu wykonania elementu żelbetowego( v1) oraz czasu zbrojenia elementu żelbetowego (v2), rozdz.5.2.

$$\text{GAM}(v1; v5, v6, v7, v10, v11, v14, v15) = e^{( 4,472183 + 0,000948*v5 + 0,000610*v6 - 0,003976*v7 - 0,000368*v10 - 0,183670*v11 + 0,123908*v14 + 0,042312*v15)}$$

$$\text{GAM}(v2; v5, v6, v7, v10, v11, v14, v15) = e^{( 3,901474 + 0,000080*v5 + 0,000817*v6 - 0,000991*v7 - 0,000013*v10 + 0,024009*v11 + 0,063132*v14 + 0,038451*v15)}$$

Opracowana metodyka poszerza obszar wiedzy i możliwości obliczeniowych w praktyce budowlanej. Istnieje możliwość wprowadzania do obliczeń dużej liczby czynników mających wpływ na czas realizacji procesu w określonych warunkach. Mogą to być wartości liczbowe lub lingwistyczne. Zaproponowana metoda jest uniwersalna i zapewnia jej zastosowanie do harmonogramowania procesów budowlanych, np.: robót liniowych, kubaturowych, procesów prostych i złożonych. Szczególnym obszarem zastosowania metody prognostyczno-regresyjnej jest możliwość opracowania wzorów regresyjnych i sieci zależności wyznaczających czasy procesów budowlanych z uwzględnieniem czynników wpływu i wykorzystanie ich w harmonogramowaniu złożonych przedsięwzięć budowlanych np.: kompleksów obiektów przemysłowych, infrastruktury transportowej itp.

## Bibliografia

- Adamiecki K., *O nauce organizacji*. PWE Warszawa, 1985.
- Adeli, H. and Karim, A., *Scheduling Cost Optimisation and Neural Dynamics Model for Construction*. „Journal of Construction Engineering and Management”. ASCE, vol. 123, no. 4, 1997, p. 450–458.
- Afanasjev V. A., Afanasjev A. V., *Stream scheduling of works in civil engineering*, St. Petersburg (in Russian) 2000.
- Afanasjev V.A., *Projektowanie organizacji stroitelstva letniego polia stojanych aerodromom*, LKWWIA im. Możajskiego, 1957.
- An S. H., Park U. Y., Kang K. I., Cho M. Y., Cho H. H., *Application of Support Vector Machines in Assessing Conceptual Cost Estimates*, J. Comput. Civ. Eng. 21, 259 (2007).
- Arditi, D., and Abulak, Z. M., *Line-of-balance scheduling in pavement construction*. „J. Constr. Engrg.”, ASCE, 112(3),1986, p. 411–424.
- Azaron A., Brown K. N., Tarim S. A., Modarres M., *A multi-objective stochastic programming approach for supply chain design considering risk*. „International Journal of Production Economics”, vol. 116, issue 1, November 2008, p.129–138.
- Belusic A., Herceg-Bulic I., Klaić Z. B., *Using a generalized additive model to quantify the influence of local meteorology on air quality in Zagreb*, University of Zagreb 2015.
- Bishop M. C., *Neural Networks for Pattern Recognition*, Clarendon Press, Oxford, UK 1995.
- Borne P., *Evolutionary algorithms for job-shop scheduling*, „Int. J. Appl. Math. Comput. Sci.”, 2004, vol. 14, no. 1, p. 91–103.
- Boussabaine A. H., *The use of artificial neural networks in construction management*, Construction Management and Economics, 1996.
- Box G. E. P, Muller M. E., Jenkins G. M., *Time Series Analysis Forecasting and Control*, Holden Day, San Francisco 1976.
- Box, G. E. P. and Cox, D. R., *An analysis of transformations*, „Journal of the Royal Statistical Society,” Series B, 26, 1964, p. 211–252.
- Box, G. E. P. i Pierce, D. A., *Distribution of Residual Autocorrelations in Autoregressive Moving Average Time Series Models*, „Journal of the American Statistical Association”, vol. 65, 1970, p. 1509–1526.
- Box, G. E. P., Pierce, D. A., *Distribution of residual autocorrelations in autoregressive-integrated moving average time series models*, „Journal of the American Statistical Association,” Vol. 65, 1970, p. 1509–1526.
- Bożejko W., Wodecki M., *Evolutionary Heuristics for Hard Permutational Optimization Problems*, „Internationam Journal of Computational Intelligence Research”, vol. 2, issue 2, 2005.
- Bożejko W., Wodecki M., *Parallel Evolution Heuristic Approach for the Traveling Salesman Problem*, International Conference on Numerical Analysis and Applied Mathematics 2005, WILEY-VCH Verlag 2005, p. 90–93.

- Bożejko W., Hejducki Z., Rogalska M., Wodecki M., *Scheduling of construction projects with a hybrid evolutionary algorithm's application*. W: *Evolutionary algorithms*, ed. by Eisuke Kita. Rijeka, InTech, cop. 2011. p. 295–308.
- Bożejko W., Hejducki Z., Rogalska M., Wodecki M., *Scheduling of construction projects with a hybrid evolutionary algorithm's application*. W: *Evolutionary algorithms*, ed. by Eisuke Kita. Rijeka InTech, cop. 2011. s. 295–308.
- Bożejko W., Wodecki M., *A hybrid evolutionary algorithm for some discrete optimization problems*, IEEE Computer Society P2286, 2005, p. 326–331.
- Breg D.: *Bankruptcy prediction by generalized additive models*, University of Oslo and Norwegian Computing Center 2007.
- Breiman, L., *The  $\Pi$  Method for Estimating Multivariate Functions from Noisy Data*. *Technometrics* 33, 1991a, p. 125–143.
- Buckley, Fuzzy PERT, in: G.W.Evans, W. Karwowski, M.R. Wilhelm (Eds.), *Applications of Fuzzy Set Methodologies in Industrial Engineering*, Elsevier, Amsterdam 1989, p. 103–114.
- Buda A., Jarynowski A., *Life-time of correlations and its applications*, vol.1, Wydawnictwo Niezależne: 5–21, December 2010.
- Buhmann M. D., *Multivariate interpolation using radial basis functions*, Ph. D. Thesis, University of Cambridge, England, 1989.
- Chanas S., Kamburowski J., *The Use of Fuzzy Variables In PERT*, *Fuzzy Sets and Systems* 5, 1981, p. 11–19.
- Chanas S., Zieliński P., *Critical Path Analysis in the Network with Fuzzy Activity Times*, *Fuzzy Sets and Systems* 122, 2001, p. 195–204.
- Chang C. K., Christensen M. J., Zhang T., *Genetic Algorithms for Project Management*. November 2001 *Annals of Software Engineering*, vol. 11, issue 1, Publisher: J. C. Baltzer AG, Science Publishers.
- Chang I.S, Tsujimura Y., Gen M., Tozawa T., *An Efficient Approach for Large Scale Project Planning Based on Fuzzy Delphi Method*, *Fuzzy Sets and Systems* 76, 1995, p. 277–288.
- Chang I. S, Tsujimura Y., Gen M., Tozawa T., *Project Planning Problem Using Fuzzy Activity Times and Fuzzy Delphi Method*, *Proc.5 th IFSA World Congr.*, 1993, p. 608–611.
- Chen C. T., Huang S. F., *Applying Fuzzy Method for Measuring Criticality in Project Network*, *Information Sciences*, 177, 2007, p. 2448–2458.
- Chen J. H., Lin J. Z., *Developing an SVM based risk hedging prediction model for construction material suppliers*, „Automation in Construction,” vol. 19, issue 6, October 2010, p. 702–708.
- Chen S. J., Hwang C. L., *Fuzzy Multiple Attribute Decision Making: Methods and Applications*, Springer, Berlin 1992.
- Chen S. P., *Analysis of Critical Paths in Project Network with Fuzzy Activity Times*, „European Journal of Operational Research,” 183, 2007, p. 442–459.

- Cheng M. Y., Peng H. S., Wu Y. W., Chen T. L., *Estimate at Completion for construction projects using Evolutionary Support Vector Machine Inference Model*, „Automation in Construction”, vol. 19, issue 5, August 2010, p. 619–629.
- Cheng M. Y., Roy A. F. V., Chen K. L., *Evolutionary risk preference inference model using fuzzy support vector machine for road slope collapse prediction*, Expert Systems with Applications, In Press, Corrected Proof, Available online 11 August 2011.
- Cheng M. Y., Roy A. F. V., *Evolutionary fuzzy decision model for construction management using support vector machine*, Expert Systems with Applications, vol. 37, issue 8, August 2010, p. 6061–6069.
- Cheng M. Y., Wu Y. W., Chen K. L., *Risk Preference Based Support Vector Machine Inference Model for Slope Collapse Prediction*, Automation in Construction, In Press, Corrected Proof, Available online 4 August 2011.
- Cheng M. Y., Wu Y. W., *Evolutionary support vector machine inference system for construction management*, „Automation in Construction”, vol. 18, issue 5, August 2009, p. 597–604.
- Christodoulos Ch., Michalakelis Ch., Varoutas D., *Forecasting with limited data: Combining ARIMA and diffusion models*, Original Research Article, „Technological Forecasting and Social Change”, vol. 77, issue 4, May 2010, p. 558–565.
- Cieślak M., *Prognozowanie gospodarcze. Metody i zastosowania*, PWN, Warszawa 2004.
- Cohen, J., *Statistical power analysis for the behavioral sciences*. (2<sup>nd</sup> ed.). Hillsdale 1988.
- Coussement K., Benoit D. F., Van den Poel D., *Improved marketing decision making in a customer churn prediction context using generalized additive models*, Expert Systems with Applications 2010.
- Czarnigowska A., Sobotka A., *Przewidywalność czasu trwania budowy. Publiczne przedsięwzięcia drogowe*, „Budownictwo – czasopismo techniczne” 1-B/2010, p. 23–25, Wydawnictwo Politechniki Krakowskiej.
- Czerwiński Z., *Dylematy ekonomiczne*, PWE, Warszawa 1992.
- Czyżycki R., Klóska R., *Ekonometria i prognozowanie zjawisk ekonomicznych w przykładach i zadaniach*, Economicus, Szczecin, 2011.
- Dawson R. J., Dawson C. W., *Practical proposals for managing uncertainty and risk in project planning*, „International Journal of Project Management”, vol. 16, Issue 5, October 1998, p. 299–310.
- Dickey, D. A., Fuller W. A., *Likelihood ratio statistics for autoregressive time series with a unit root*, „Econometrica”, vol. 49, no. 4, 1981, p. 1957–1972.
- Dittmann P., *Prognozowanie w przedsiębiorstwie. Metody i przykłady zastosowań*. Oficyna Ekonomiczna, Kraków 2004.
- Dominici F., McDermott A., Zeger S. L., Samet J. M., *On the use of Generalized Additive Models in Time-Series Studies of Air Pollution*, Baltimore, Johns Hopkins University 2002.
- Ediger V. E., Akar S., *ARIMA forecasting of primary energy demand by fuel in Turkey*, Original Research Article, „Energy Policy”, vol. 35, issue 3, March 2007, p. 1701–1708.

- El-Rayes, K. and Moselhi, O. (2001), *Optimal Resource Utilization for Repetitive Construction Projects*, „Journal of Construction Engineering and Management”. ASCE, vol. 127, no. 1, p. 18–27.
- Etienne Ch., Lechmann A., Goyette S., *Spatial Predictions of Extreme Wind Speeds over Switzerland Using Generalized Additive Models*, „Journal of applied meteorology and climatology” 2009.
- Faruk D. O., *A hybrid neural network and ARIMA model for water quality time series prediction*, „Engineering Applications of Artificial Intelligence”, vol. 23, issue 4, June 2010, p. 586–594.
- Fijorek K., Mróz K., Niedziela K., Fijorek D., *Prognozowanie cen energii elektrycznej na Rynku Dnia Następnego metodami data mining*. „Rynek Energii” 2010, 6 (91), s. 46–50.
- Friedman, J., *Multivariate Adaptive Regression Splines*, „Annals of Statistics” vol. 19, 1991b, s. 1–67.
- Friedman J., *Rejoinder*, „Annals of Statistics”, vol. 19, 1991c, p. 123–141.
- Gab-Allah A. A., Ibrahim A. H., Hagrais O. A., *Predicting the construction duration of building projects using artificial neural networks*, „International Journal of Applied Management Science”, 2015.
- Gajda J., *Prognozowanie i symulacje a decyzje gospodarcze*, wyd. C. H. Beck, Warszawa 2001.
- Gilbert K. C., Chatpattananan V., *An ARIMA supply chain model with a generalized ordering policy*, „Journal of Modelling in Management”, vol. 1, issue: 1 2006.
- Gilbert, K. C., *An ARIMA supply chain model*, „Management Science”, vol. 51 no. 2, 2005, p. 305–310.
- Goldratt E. M., *Critical Chain*, Great Barrington: The North River Press 1997.
- Goldratt E. M., *The goal, 2nd revised edition*, Great Barrington, The North River Press 1992.
- Gong D., Rowings Jr J. E., *Calculation of safe float use in risk-analysis-oriented network scheduling*. „International Journal of Project Management”, vol. 13, issue 3, June 1995, p. 187–194.
- Gong D., *Optimization of float use in risk analysis-based network scheduling*. „International Journal of Project Management”, vol. 15, issue 3, June 1997, p. 187–192.
- Greń J., *Modele i zadania statystyki matematycznej*. PWN, Warszawa 1984.
- Guan B. T., Hsu H., Wey T., Tsao L., *Modeling monthly mean temperatures for the mountain regions of Taiwan by generalized additive models*, Architectural and forest meteorology 2009.
- Hamerlink D. J., Rowings J. E., *Linear scheduling model development of controlling activity path*, „Journal of Construction Engineering and Management”, vol. 124, no. 4, p. 266–268, 1998.
- Hamerlink D. J., *Linear scheduling model: the development of a linear scheduling model with micro computer application for highway construction control*. PhD thesis, Iowa State University, USA 1995.
- Hammersley J. M., Handscomb D. C., *Monte Carlo Methods*, Methuen, London 1964.



- Han P., Wang P. X., Zhang S. Y., Zhu D. H., *Drought forecasting based on the remote sensing data using ARIMA models*, Original Research Article. „Mathematical and Computer Modelling”, vol. 51, issues 11-12, June 2010, p. 1398–1403.
- Han S., Guikema S. D., Quiring S. M., *Improving the Predictive Accuracy of Hurricane Power Outage Forecasts Using Generalized Additive Models*, Risk Analysis 2009.
- Harris, R. B. and Ioannou, P. G., *Scheduling Projects with Repeating Activities*, „Journal of Construction Engineering and Management”. ASCE, vol. 124, no. 4, 1998, p. 269–278.
- Hastii T. J., Tibshirani R. J., *Generalized Additive Models* Chapman & Hall/CRC Monographs on Statistics & Applied Probability 1990.
- Hejducki Z., *Scheduling Model of Construction Activity with Time Couplings*, „Journal of Civil Engineering and Management”, vol. 9, no. 4, 2003, p. 284–291.
- Hejducki Z., *Sequencing problems in methods of organizing construction processes*, „Engineering, Construction and Architecture Management”, vol. 11, no. 1, 2004, p. 20–32.
- Hejducki Z., Rogalska M., *Time coupling methods. Construction scheduling and time, cost optimization*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław 2011.
- Hejducki, Z., *Time couplings in scheduling methods of complex construction processes (Sprzężenia czasowe w metodach organizacji złożonych procesów budowlanych)*, Wrocław University of Technology (in Polish), 2000.
- Herroelen W., Leus R., *On the merits and pitfalls of critical chain scheduling*, „Journal of Operations Management” vol. 19, 2001, p. 559–577.
- Hisaki T., Kaneko M. A., Yamaguchi M., Sasa H., Kouzuki H., *Development of QSAR models using artificial neural network analysis for risk assessment of repeated-dose, reproductive, and developmental toxicities of cosmetic ingredients*, „Journal of Toxicological Sciences”, 2015.
- Ho S. L., Xie M., *The use of ARIMA models for reliability forecasting and analysis*, Original Research Article. „Computers & Industrial Engineering”, vol. 35, issues 1–2, October 1998, p. 213–216.
- Hodge B. M., Zeiler A., Brooks D., Blau G., Pekny J., Reklatis G., *Improved Wind Power Forecasting with ARIMA Models*. „Computer Aided Chemical Engineering”, vol. 29, 2011, p. 1789–1793.
- Hoła B., Mrozowicz J., *Modelowanie procesów budowlanych o charakterze losowym*, DWE, Wrocław 2003.
- Huang J., Wang X., *Risk Analysis of Construction Schedule Based on PERT and MC Simulation*. December 2009 ICIII '09: Proceedings of the 2009 International Conference on Information Management, Innovation Management and Industrial Engineering, vol. 02, Publisher: IEEE Computer Society.
- Jang D.H., *Fuzzy Resource Leveling and Allocation in PERT*, Proc. Korea-Japan Joint Conf. Fuzzy Systems and Engineering (1992), p. 123–126.
- Jaskowski P., Czarnigowska A., Sobotka A. (2014), *Decision model for selecting supply sources of road construction aggregates*. Inżynierne Ekonomika-Engineering Economics, vol. 15, nr 1, s. 13–20.



- Jaworski K. M., *Metodologia projektowania realizacji budowy*, PWN, Warszawa 1999.
- Jaworski K. M., *Projektowanie realizacji budowy według kryterium niezawodności. Teoria i metoda.*, Budownictwo z. 66, Wydawnictwa Politechniki Warszawskiej, Warszawa 1980.
- Jia J., Zhao J., Deng H., Duan J., *Ecological footprint simulation and prediction by ARIMA model—A case study in Henan Province of China.* „Ecological Indicators”, vol. 10, issue 2, March 2010, p. 538–544.
- Jiang A. N., Wang S. Y., Tang S. L., *Feedback analysis of tunnel construction using a hybrid arithmetic based on Support Vector Machine and Particle Swarm Optimisation*, Original Research Article, „Automation in Construction”, vol. 20, issue 4, July 2011, p. 482–489.
- Kapliński O., *Development and usefulness of planning techniques and decision-making foundations on the example of construction enterprises in Poland*, „Technological and Economic Development of Economy”, vol. 14, no. 4, 2008, s. 492–502.
- Khashei M., Bijari M., *A novel hybridization of artificial neural networks and ARIMA models for time series forecasting*, Original Research Article. „Applied Soft Computing”, vol. 11, issue 2, March 2011, p. 2664–2675.
- Kim H. W., Lee J. H., Choi Y. H., Chung Y. U., Lee H., *Dynamic bandwidth provisioning using ARIMA-based traffic forecasting for Mobile WiMAX*, Original Research Article. „Computer Communications”, vol. 34, issue 1, 15 January 2011, p. 99–106.
- King, *Minium. Statystyka dla psychologów i pedagogów*, PWN, 2009.
- Kister T. C., Hawkins B., *Special Case. Maintenance Planning and Scheduling for Maintenance Outages – The Plant Shutdown*, „Maintenance Planning and Scheduling”, 2006, p. 211–229.
- Kornacki J., Ćwik J., *Statystyczne systemy uczące się*, Warszawa, WNT, 2005.
- Kot S., Jakubowski J., Sokołowski A., *Statystyka*, Difin 2011.
- Kowalczyk Z., Zabielski J., *Kosztorysowanie i normowanie w budownictwie*. WSiP, Warszawa 2012.
- Kryszpin D., *Internetowa Encyklopedia Zarządzania*, 2008.
- Kulejewski J., *Metody harmonogramowania budowy z uwzględnieniem rozmytego charakteru danych*, Oficyna Wydawnicza Politechniki Warszawskiej, Warszawa 2010.
- Kusiak A., Li M., Tang F., *Modeling and optimization of HVAC energy consumption*, Applied Energy, vol. 87, issue 10, October 2010, p. 3092–3102.
- Kusiak A., Li M., Zhang Z., *A data-driven approach for steam load prediction in buildings*, „Applied Energy”, vol. 87, issue 3, March 2010, p. 925–933.
- Lam K. Ch., Palaneeswaran E., Yu Ch., *A support vector machine model for contractor pre-qualification*, „Automation in Construction”, vol. 18, issue 3, May 2009, p. 321–329.
- Lee Ch. M., Ko Ch. N., *Short-term load forecasting using lifting scheme and ARIMA models*, Original Research Article. „Expert Systems with Applications”, vol. 38, issue 5, May 2011, p. 5902–5911.

- Leu S. S., Yang C. H., Huang J. C., *Resource leveling in construction by genetic algorithm-based optimization and its decision support system application*. „Automation in Construction”, vol. 10, 2000, p. 27–41.
- Lewis, P. A. W., Stevens J. G., *Nonlinear Modeling of Time Series using Multivariate Adaptive Regression Splines (MARS)*, „Journal of the American Statistical Association” vol. 86, 1991, p. 864–877.
- Li C., Wang K., *The risk element transmission theory research of multi-objective risk-time-cost trade-off*, June 2009 Computers & Mathematics with Applications, vol. 57, issue 11–12, Publisher: Pergamon Press, Inc.
- Li Q., Meng Q, Cai J., Yoshino H., Mochida A., *Predicting hourly cooling load in the building: a comparison of support vector machine and different artificial neural networks*, „Energy Conversion and Management”, vol. 50, issue 1, January 2009, p. 90–96.
- Li X., Chen Q., Li Y., *Impact on Genetic Algorithm of Different Parameters*. December 2008 ISICA '08. *Proceedings of the 3rd International Symposium on Advances in Computation and Intelligence*, Publisher: Springer-Verlag.
- Ligas A., *Modeling the growth of recruits of European hake (Merlucciusmerluccius) in the northwestern Mediterranean Sea with generalized additive models*, Fishery Bulletin 2015.
- Lin S., Goodman E., Punch W., *A genetic algorithm approach to dynamic job-shop scheduling problems*, Proc. Of the Seventh Internat. Conf. On Genetic Algorithms, San Francisco 1997, p. 481–488.
- Ljung G. M. i Box, G. E. P., *On a Measure of Lack of Fit in Time Series Models*, „Biometrika”, vol. 65, 1978, p. 297–303.
- Lute V., Upadhyay A., Singh K. K., *Support vector machine based aerodynamic analysis of cable stayed bridges*, „Advances in Engineering Software”, vol. 40, issue 9, September 2009, p. 830–835.
- Maddala G.S., *Introduction to Econometrics*, Wiley 2002.
- Mahfouz T., Kandil A., *Construction Legal Decision Support Using Support Vector Machine (SVM)* ASCE Conf. Proc., Construction Research Congress 2010: Innovation for Reshaping Construction Practice, Proceedings of the 2010 Construction Research Congress.
- Marcinkowski R., *Metody rozdziału zasobów realizatora w działalności inżyniersko – budowlanej*, Wydawnictwo WAT, Warszawa 2002,.
- Marcinkowski R., *Harmonogramowanie produkcji przedsiębiorstwa budowlanego*. „Przegląd budowlany” nr 2, 2007.
- Marcinkowski R., Koper A., *Projektowanie zespołu maszyn zapewniających ciągłość betonowania konstrukcji monolitycznej*. Budownictwo i Inżynieria Środowiska, Wydaw. Politechniki Białostockiej, nr 2, 2011, s. 583–587.
- Marcinkowski R., *Metody rozdziału zasobów realizatora w działalności inżyniersko – budowlanej*, Wydawnictwo WAT, Warszawa 2002.
- Marcinkowski R., *Planowanie organizacji robót budowlanych na podstawie analizy nakładów pracy zasobów czynnych*. Budownictwo i Architektura, nr 12, z.1, 2013, p. 39–46.

- Mccullagh P., Nelder J. A., *Generalized Linear Models*, 2nd ed. London, Chapman and Hall, 1989.
- McGough T., Tsolacos S., *Forecasting commercial rental values using ARIMA models*. „Journal of Property Valuation and Investment”, vol. 13, issue: 5, 1995.
- McGough, T., Tsolacos, S., *Forecasting office rental values using vector autoregressive models*, Cutting Edge 1994, p. 303–320.
- Mitchell T. M., Ben-Zion Y., Shimamoto T., *Pulverized fault rocks and damage asymmetry along the Arima-Takatsuki Tectonic Line, Japan*. „Earth and Planetary Science Letters”, vol. 308, issues 3–4, 15 August 2011, p. 284–297.
- Mohammadi H., Su L., *International evidence on crude oil price dynamics: Applications of ARIMA-GARCH models*, Original Research Article. „Energy Economics”, vol. 32, issue 5, September 2010, p. 1001–1008.
- Momeni E., Nazir R., Armaghani D.J., Maizir H., *Application of Artificial Neural Network for Predicting Shaft and Tip Resistances of Concrete Piles*, „Earth Science Research Journal, 2015.
- Moon J. W., Jung S. K., Lee Y. O., Choi S., *Prediction Performance of an Artificial Neural Network Model for the Amount of Cooling Energy Consumption in Hotel Rooms*, Energies, 2015.
- Nassar K., *Evolutionary optimization of resource allocation in repetitive construction schedules ITcon*, vol. 10, 2005, p. 265.
- Ndongo M., Diongue A.K., Diop A., Dossou-Gbété S., *Estimation of long-memory parameters for seasonal fractional ARIMA with stable innovations*, Original Research Article. „Statistical Methodology”, vol. 7, issue 2, March 2010, p. 141–151.
- Nisbet R., Elder J., Miner G., *Advanced Algorithms for Data Mining, Handbook of Statistical Analysis and Data Mining Applications*, 2009, p. 151–172.
- Nisbet R., Elder J., Miner G., *Handbook of Statistical Analysis and Data Mining Applications*, 2009, p. 77–97.
- Nowak E., *Prognozowanie gospodarcze*, PLACET 1998.
- Östermark, R., *Efficient Time Series Models of Finnish Stock Returns Obtained with a Cartesian ARIMA Search Algorithm*, Abo Academy, Henriksgatan 1988.
- Östermark, R., Höglund, R., *The Cartesian ARIMA Search Algorithm*, The Eighth International Symposium on Forecasting, Amsterdam 1988.
- Pal M., Deswal S., *Modeling Pile Capacity Using Support Vector Machines and Generalized Regression Neural Network*, J. Geotech. Geoenviron. Eng. 134, 1021 (2008).
- Pawłowski Z., *Zasady predykcji ekonometrycznej*. PWN, Warszawa 1982.
- Pearson K., *On the General Theory of Skew Correlation and Non-linear Regression*. London: Dulau & Co. 1905.
- Platon R., Dehkordi V. R., Martel J., *Hourly prediction of a building's electricity consumption using case-based reasoning, artificial neural networks and principal component analysis*, Energy & Buildings, 2014.

- Podolski M., *Analiza nowych zastosowań teorii szeregowania zadań w organizacji robót budowlanych*. Praca doktorska. Raporty Inst. Bud. PWroc. 2008, Ser. PRE nr 5/08, 2008.
- Prakash A., Felix T. S. Chan F. T. S., Deshmukh S. G., *FMS scheduling with knowledge based genetic algorithm approach*. April 2011 Expert Systems with Applications: An International Journal, vol. 38 Issue 4. Publisher: Pergamon Press, Inc.
- Premachandra I. M., *An approximation of the activity duration distribution in PERT*. Computers & Operations Research, Volume 28, Issue 5, April 2001, p. 443–452.
- Radovilsky Z. D., *A quantitative approach to estimate the size of the time buffer in the theory of constraints*. „International Journal of Production Economics”, vol. 55, 1998, p. 113–119.
- Raymond Y. C. Tse R. Y. C., *An application of the ARIMA model to real-estate prices in Hong Kong*. „Journal of Property Finance”, vol. 8 Issue: 2 1997.
- Reeves C. R. Yamada T. Genetic Algorithms Path Relinking on the Flowshop Sequencing Problem, „Evolutionary Computation Journal MIT press”, vol. 6. No 1, p. 230–234, 1998.
- René van Dorp J., *Statistical dependence through common risk factors: With applications in uncertainty analysis*. „European Journal of Operational Research”, vol. 161, issue 1, 16 February 2005, p. 240–255.
- Rodger J. A., *A fuzzy nearest neighbor neural network statistical model for predicting demand for natural gas and energy cost savings in public buildings*, Expert Systems With Applications, 2014.
- Rogalska M., Bożejko W., Hejducki Z., *Sterowanie poziomem zatrudnienia z zastosowaniem algorytmów genetycznych*, Pięćdziesiąta Pierwsza Konferencja Naukowa KIL i w PAN i KN PZITB Gdańsk-Krynica 2005, s. 185–192.
- Rogalska M., Bożejko W., Hejducki Z., *Time/cost optimization using hybrid evolutionary algorithm in construction project scheduling*. „Automation in Construction”, vol. 18, 2008, p. 24–31.
- Rogalska M., Hejducki Z., Łodożyński Ł., *Harmonogramowanie liniowych procesów budowlanych z uwzględnieniem metody Branch & Bound*. W: *Organizacja przedsięwzięć budownictwa drogowego*, red. nauk. Zbigniew Tokarski. Bydgoszcz: Zarząd Oddziału Stowarzyszenia Inżynierów i Techników Komunikacji RP: Rełpol, 2011, s. 107–118.
- Rogalska M., Hejducki Z., *Shortening the realisation time of building project with application of theory of constraints and critical chain scheduling*. „Journal of Civil Engineering and Management”, 2004, vol. 10, Suppl 2, p. 99–105.
- Rogalska M., Hejducki Z., *Time buffers in construction process scheduling*. „Journal of Civil Engineering and Management”, vol. 13, no. 2, 2007, p. 143–148.
- Rumelhart, D. E., McClelland, J. L., & the PDP research group. (1986). *Parallel distributed processing. Explorations in the microstructure of cognition*. vol. 1. Cambridge, MA: MIT Press.
- Russell, A. D. and Wong, W., *New Generation of Planning Structures*. „Journal of Construction Engineering and Management”. ASCE, vol. 119, no. 2, 1993, p. 196–214.
- Scavino S. J., *Effect of multiple calendars on total float and critical path*. „Cost Engineering”, vol. 45, no. 6, June 2003.

- Schimek M., *Smoothing and regression. Approaches, computation, and application*, Wiley 2000.
- Serinaldi F., *Distributional modeling and short-term forecasting of electricity prices by Generalized Additive Models for Location, Scale and Shape*, Energy Economics 2011.
- Shapiro S. S., Wilk M. B., *An Analysis of Variance Test for Normality (Complete Samples)*. „*Biometrika*”, vol. 52, no. 3/4 (Dec., 1965), p. 591–611.
- Shin Y., Kim D. W., Kim J. Y., Kang K. I., Cho M. Y., Cho H. H., *Application of AdaBoost to the Retaining Wall Method Selection in Construction* „*J. Comput. Civ. Eng.*” vol. 23, no. 188, 2009.
- Shipley M. F., Korvin A., Omer K., *BIFPET Methodology Versus PERT in Project Management: Fuzzy Probability Instead of the Beta Distribution*, „*Journal of Engineering and Technology Management JET-M 14*”, 1997, p. 49–65.
- Skorupka D., *Kompleksowa ocena ryzyka realizacji inwestycji budowlanych*. *Zeszyty Naukowe WSOWL 2006*, nr 1, s. 86–92.
- Skorupka D., *Kwantyfikacja i alokacja czynników ryzyka w sieciowych harmonogramach budowlanych*. *Zeszyty Naukowe WSOWL 2008*, nr 1, s. 145–159.
- Skorupka D., *Metoda oceny ryzyka realizacji przedsięwzięć inżynierjno-budowlanych*. *Zeszyty Naukowe WSOWL 2007*, nr 3, s. 79–88.
- Skorupka D., *Zarządzanie projektem inwestycyjnym w warunkach ryzyka*. *Zeszyty Naukowe WSOWL 2004*, nr 1, s. 88–94.
- Skorupka D., *Zarządzanie ryzykiem w przedsięwzięciach budowlanych*. *Zeszyty Naukowe WSOWL 2008*, nr 3, s. 120–129.
- Skorupka D., *Identification and initial risk assessment of construction projects in Poland*. „*Journal of Management in Engineering, American Society of Civil Engineers*” 2007.
- Skorupka D., *Metoda identyfikacji i oceny ryzyka realizacji przedsięwzięć budowlanych*. WAT, Warszawa 2007.
- Sobotka A., Jaśkowski P., *Zarządzanie łańcuchem dostaw w realizacji przedsięwzięć budowlanych*. „*Gospodarka Materiałowa i Logistyka*” nr 11/2009, p. 11–20.
- Sobotka A., Pawluś D., Wałach D., Czarnigowska A., *Badanie procesów logistycznych przedsięwzięć drogowych*. *Zeszyty Naukowe Politechniki Rzeszowskiej, Budownictwo i Inżynieria Środowiska*, z. 59 (3/2012/III), s. 113–120.
- Sokołowski A., *Prognozowanie i analiza szeregów czasowych. Course materials*. Statsoft, Kraków 2010.
- Son H., Kim Ch., Kim Ch., *Automated Color-Model-Based Concrete Detection in Construction-Site Images Using Machine Learning Algorithms*, „*Journal of Computing in Civil Engineering*”, June 2011.
- Song C., Huang C., Liu G., *Simulating potential distribution of *Tamarix chinensis* in Yellow River Delta by generalized additive model*, Wetland Science 2010.
- Soroush H., *Risk taking in stochastic PERT networks*. „*European Journal of Operational Research*”, vol. 67, issue 2, 11 June 1993, p. 221–241.
- Stevenson S., *A comparison of the forecasting ability of ARIMA models*, „*Journal of Property Investment & Finance*”, vol. 25, issue: 3, 2007.



- Steyn H., *Project management applications of the theory of constraints beyond critical chain scheduling*. „International Journal of Project Management” vol. 20, 2002, p. 75–80.
- Stokes H.H., *Multivariate Adaptive Regression Spline Modeling Using SCAB34S SPLINES and SCA WorkBench*, Department of Economics University of Illinois at Chicago, William J. Lattyak, Scientific Computing Associates Corp. 2008. p. 1–95.
- Suryanita R., Adnan A., *Intelligent Monitoring System on Prediction of Building Damage Index using Neural-Network*, Telkmnika, 2012.
- Tadeusiewicz R., Lula P., *Neural Network Analysis of Time Series Data*, „Informatica – An International Journal of Computing and Informatics”, vol. 25, nr 1, 2001, p. 3–10.
- Tadeusiewicz R., *Computer simulation of the alive brain*. Warszawska Wyższa Szkoła Informatyki, Warszawa 2010.
- Tadeusiewicz R., *How computer science can help to look into the human body*, Warszawska Wyższa Szkoła Informatyki, Warszawa 2010.
- Tadeusiewicz R., *Neural Networks*, Akademicka Oficyna Wydawnicza, Warszawa 1993.
- Tan Z., Zhang J., Wang J., Xu J., *Day-ahead electricity price forecasting using wavelet transform combined with ARIMA and GARCH models*, Original Research Article. „Applied Energy”, vol. 87, issue 11, November 2010, p. 3606–3610.
- Theil H., *Zasady ekonometrii*, PWN, Warszawa, 1979.
- Toklu Y. C., *Application of genetic algorithms to construction scheduling with or without resource constraints*, „Can. J. Civ. Eng.” vol. 29, 2012, p. 421–429.
- Trocki M., Grucza B., Ogonek K., *Zarządzanie Projektami*, PWE, Warszawa 2003.
- Tse, R. Y. C., *An application of the ARIMA model to real estate prices in Hong Kong*, „Journal of Property Finance”, vol. 8, 1997, p. 152–163.
- Tserng H. P., Lin G. F., Tsai L. K., Chen P. Ch., *An enforced support vector machine model for construction contractor default prediction*, Automation in Construction, In Press, Corrected Proof, Available online 22 June 2011.
- Tsujitani M., Takana Y., *Analysis of Heart Transplant Survival Data Using Generalized Additive Models*, Hindawi 2013.
- Ugwu O. O., Tah J. H. M., *Development and application of a hybrid genetic algorithm for resource optimization and management*, „ECAM” 2002 vol. 9/4, p. 304–317.
- Underwood F. M., *Describing long-term trends in precipitation using generalized additive models*, Journal of Hydrology 2009.
- Venezian E. C., Leng Ch. Ch., *Application of spectral and ARIMA analysis to combined-ratio patterns*. „Journal of Risk Finance”, vol. 7, issue: 2, 2006.
- Vidoli F., *Evaluating the water sector in Italy through a two stage method using the conditional robust nonparametric frontier and multivariate adaptive regression splines*, „European Journal of Operational Research”, vol. 212, issue 3, 1 August 2011, p. 583–595.
- Wang Ch. Ch., *Acomparison study between fuzzy time series model and ARIMA model for forecasting Taiwan export*, Original Research Article, „Expert Systems with Applications”, vol. 38, issue 8, August 2011, p. 9296–9304.

- Wang Y., Hou S., Masson-Delmotte V., Jouzel J., *A generalized additive model for the spatial distribution of stable isotopic composition in Antarctic surface snow*, Chemical Geology 2010.
- Yalcin N., Tezel G., Karakuzu C., *Epilepsy diagnosis using artificial neural network learned by PSO*, „Turkish Journal of Electrical Engineering & Computer Sciences”, 2015.
- Yamada T., *Solving the Csum Permutation Flow Shop Scheduling Problem by Genetic Local Search ICEC98* (International Conference on Evolutionary Computation) p. 230–234.
- Yang I., Yeo M., Kwang W., *Application of artificial neural networks to predict the optimal start time for heating system in building*, „Energy Conversion and Management”, 2003.
- Yang J. B., *Applying the theory of constraints to construction scheduling. System-based Vision for Strategic and Creative Design*, Bontempi (ed.) 2003 Swets & Zeitlinger, Lisse.
- Young C. C., Liu W-C., Hsieh W, *Predicting the water level fluctuation in an Alpine Lake Using Physically Based, Artificial Neural Network, and Time Series Forecasting Models*, „Mathematical Problems in Engineering”, 2015.
- Yunning Z., Xixi S.: *Research on Improved PERT Model in Analysis of Schedule Risk of Project*. May 2010 ICEE '10: Proceedings of the 2010 International Conference on E-Business and E-Government Publisher: IEEE Computer Society.
- Zeliaś A., *Prognozowanie ekonomiczne. Teoria. Przykłady. Zadania*”, PWN, Warszawa 2003
- Zhang H., Gao S. Seiler M. J., Zhang Y., *Identification of real Estate cycles in China based on artificial neural networks*, „Journal of Real Estate Literature”, 2015.
- Zhang Y., Huang J. J., Chen L., Q i L., *Eutrophication forecasting and management by artificial neural network: a case study at Yuqiao Reservoir in North China*, „Journal of Hydroinformatics”, 2015.