Władysław Wornalkiewicz

MODELOWANIE EKONOMETRYCZNE PODSTAWĄ PROGNOZOWANIA

Monografia

Kyiv 2025

UDC 330.43:338.27
W87

Zatwierdzone na posiedzeniu Rady Naukowej
Wydziału Edukacji Fizyczno-Matematycznej, Informatycznej i Technologicznej
Państwowego Uniwersytetu Pedagogicznego w Berdiańsku (Zaporoże) protokół nr 7 (31.03.2025 r.)

Recenzenci:
dr Hanna Alieksieieva
dr Nadiya Dubrovina, PhD (Slowacja)
dr Oleksandr Nestorenko (Polska)

Rada redakcyjna:
Wojciech Duczmal (Polska), Nataliia Kravchenko, Tetyana Nestorenko, Aleksander Ostenda (Polska), Tadeusz Pokusa (Polska), Sławomir Śliwa (Polska)

Wornalkiewicz Władysław
W87 Modelowanie ekonometryczne podstawą prognozowania: monografia. Kyiv: Wydawnictwo Ludmiła, 2025. 340 s.
ISBN 978-617-555-283-4

Monografia przedstawia kompleksową analizę metod modelowania ekonometrycznego do prognozowania procesów gospodarczych i społecznych. Omówiono podstawy prognozowania, metody matematyczne i niematematyczne oraz ich zastosowanie w różnych dziedzinach. Szczególną uwagę poświęcono prognozowaniu procesów migracyjnych, analizie rynku pracy, modelowaniu regionalnego PKB. Szczegółowo opisano modele wygładzania wykładniczego, podejścia autoregresyjne, model Holta-Wintersa i model Browna. Przeanalizowano narzędzia programów WinQSB, Gretl i R do modelowania ekonometrycznego. Przedstawiono metody uwzględniania ryzyka, programowania wielokryterialnego oraz podstawy statystyki opisowej. Publikacja jest przeznaczona dla naukowców, wykładowców, doktorantów i specjalistów w dziedzinie prognozowania ekonomicznego.

Ворналкевич Владислав
W87 Економетричне моделювання як основа прогнозування: монографія. Київ: Видавництво Людмила, 2025. 340 с.
ISBN 978-617-555-283-4

Монографія представляє комплексний аналіз методів економетричного моделювання для прогнозування економічних та соціальних процесів. Розглянуто основи прогнозування, математичні та нематематичні методи, їх застосування в різних галузях. Особлива увага приділена прогнозуванню міграційних процесів, аналізу ринку праці, моделюванню регіонального ВВП. Детально висвітлено моделі експоненційного згладжування, авторегресійні підходи, модель Холта-Вінтерса та модель Брауна. Проаналізовано інструментарій програм WinQSB, Gretl та R для економетричного моделювання. Представлено методи врахування ризиків, багатокритеріального програмування та основи описової статистики. Видання призначене для науковців, викладачів, аспірантів та фахівців у сфері економічного прогнозування.

Autor ponosi pełną odpowiedzialność za tekst, cytaty i ilustracje.

© Władysław Wornalkiewicz, 2025
© Państwowy Uniwersytet Pedagogiczny w Berdiańsku (Zaporoże), 2025

Spis treści

Wstęp 7

1. Matematyczne i niematematyczne metody prognozowania 8
1.1. Wprowadzenie 8
1.2. Metody matematyczne 9
1.3. Przykładowe metody stosowane w rolnictwie 12
1.4. Publikacje dotyczące prognozowania długookresowego 13

2. Stosowane główne metody prognozowania w szerszych obszarach 16

3. Potrzeba przewidywania migracji ludności 25
3.1. Wstęp 25
3.2. Korzystanie ze współczesnej biblioteki 27
3.3. Demografia - migracja 30
3.4. Pogoń za szczęściem 35
3.5. Adaptacje środowiskowe polskich imigrantów 36
3.6. Uchodźctwo 38
3.7. Starzenie się populacji 40
3.8. Modelowanie procesów demograficznych 41
3.9. Skutki migracji 43

4. Prognozowanie migracji ludności z uwzględnieniem wag harmonicznych 46
4.1. Wprowadzenie 46
4.2. Szeregi czasowe wybranych cech statystycznych (przyczyny, skutku) 48
4.3. Założenie do cech przyczyny i skutku 57
4.4. Prognozowanie migracji (odpływ) poprzez zastosowanie wag harmonicznych 58
4.5. Obliczenie prognoz 60

5. Prognozowanie z wykorzystaniem zasady postarzania informacji 63
5.1. Wprowadzenie 63
5.2. Pracujący w handlu 63
5.3. Pracujący w transporcie, gospodarce magazynowej i łączności 70

6. Iteracyjna optymalizacja parametrów w wybranych modelach wyrównywania wykładniczego 76
6.1. Wstęp 76
6.2. Uogólniony model Holta 76
6.3. Modele wyrównywania wykładniczego w programie *WinQSB* 78
6.4. Zastosowanie dodatku *Solver Excela* 80

7. Model ekonometryczny zmiennej „*Liczba bezrobotnych*” kraju 87
7.1. Wprowadzenie 87
7.2. Dane wejściowe 87
7.3. Formułowanie modelu ekonometrycznego 90
7.4. Weryfikacja modelu 92

8. Prognozowanie liczby pasażerów lotów z zastosowaniem *WinQSB* 98
8.1. Wstęp 98
8.2. Rozpatrzenie możliwości zastosowania programu *FC* 101
8.3. Wyłonienie trendu rozwojowego 105

9. Prognozowanie zewnętrznych usług transportowych 108
9.1. Wprowadzenie 108
9.2. Logistyka w zakresie potrzeb przedsiębiorstw 112
9.3. Wstępne podejście do prognozowania krótkookresowego 114
9.4. Optymalizacja doboru parametrów w modelu multiplikatywnym 117

10. Estymowanie parametrów modelu ekonometrycznego produktu brutto programem *Gretl* dla województwa 123
10.1. Wstęp 123
10.2. Wejściowe dane statystyczne 124
10.3. Zastosowanie metody *a posteriori* 129
10.4. Testy sprawdzające model 133

11. Etapy i procedury budowy modelu produktu brutto podregionów 137
11.1. Wprowadzenie 137
11.2. Podejście do modelowania 139
11.3. Informacja o programie *Gretl* 140
11.4. Opracowanie modelu na danych empirycznych 142

12. Budowa i weryfikacja modelu ekonometrycznego w wybranych programach 150
12.1. Wstęp 150
12.2. Budowa modelu ekonometrycznego w *Gretl* 151
12.3. Przygotowanie do realizacji w programie *R* 154
12.4. Weryfikacja podstawowego modelu końcowego 160

13. Prognozowanie z zastosowaniem modelu multiplikatywnego 165
13.1. Wstęp 165
13.2. Model Holta-Wintersa bez sezonowości 167
13.3. Model Holta-Wintersa z sezonowością multiplikatywną 169
13.4. Model Holta-Wintersa z sezonowością addytywną 173

14. Modelowanie ekonometryczne z użyciem algebry liniowej macierzy 177
14.1. Wstęp 177
14.2. Estymacja parametrów modelu zmiennej X_3 w środowisku *R* 179
14.3. Zastosowanie funkcji macierzowych programem *R* 184
14.4. Wprowadzenie parametru stałego do równania liniowego modelu ekonometrycznego 187
14.5. Zastosowanie funkcji macierzowych *Excela* 190
14.6. Inne operacje *R* na macierzach 192

15. Prognozowanie z zastosowaniem modeli Holta-Wintersa 195
15.1. Wstęp 195
15.2. Wywołanie i statystyka wpisu 196
15.3. Tekst wpisu 198
15.3. Problematyka frazy *Prognozowanie z zastosowaniem modeli Holta*-Wintersa 204

16. Prognozowanie w oparciu o model autoregresyjny 207
16.1. Wywołanie i statystyka wpisu 207
16.2. Informacja na temat *Prognozowanie w oparciu o model autoregresyjny* 212

17. Podwójne wyrównywanie wykładnicze - *model Browna* 214
17.1. Wywołanie i statystyka wpisu 214
17.2. Tekst wpisu 214
17.3. Problematyka frazy *Podwójne wyrównywanie wykładnicze - model Browna* 218

18. Uwzględnienie elementu ryzyka - problem gazeciarza 221
18.1. Tekst wpisu 221
18.2. Problematyka frazy *Uwzględnienie elementu ryzyka - problem gazeciarza* 222

19. Trend pełzający 223
19.1. Tekst wpisu 223
19.2. Problematyka frazy *Trend pełzający* 226

20. Informacja o programie *WinQSB* 229
20.1. Tekst wpisu 229
20.2. Problematyka frazy *Informacja o programie WinQSB* 231

21. Programowanie wielokryterialne z priorytetem - zastosowanie *WinQSB* 234
21.1. Tekst wpisu 234
21.2. Problematyka frazy *Programowanie wielokryterialne - zastosowanie WinQSB* 236

22. Wyrównywanie wykładnicze 240
22.1. Tekst wpisu 240
22.2. Odniesienia frazy we wpisach wyszukiwarki *Google* 242

23. Aktualność modelu i proste wyrównywanie wykładnicze 244
23.1. Tekst wpisu 244
23.2. Odniesienia frazy *Aktualność modelu i proste wyrównywanie wykładnicze* we wpisach wyszukiwarki *Google* 247

24. Średnie ruchome 249
24.1. Tekst wpisu 249
24.2. Odniesienia frazy *Średnie ruchome* we wpisach wyszukiwarki *Google* 250

25. Mierzenie i prognozowanie innowacyjności 253
25.1. Tekst wpisu 253
25.2. Rozpoznanie źródeł dostępnych w *Google* według frazy *Mierzenie i prognozowanie innowacyjności* 256

26. Modelowanie ekonometryczne ruchu internetowego 259
26.1. Zastosowanie modelu wielomianowego 259
26.2. Modele Holta-Wintersa 262
26.3. Sformułowanie modelu multiplikatywnego 263
26.4. Propozycja zastosowania modelu podwójnego wyrównywania wykładniczego z trendem 267

27. Określenie modelu PKB 270
27.1. Sformatowanie pliku wejściowego 270
27.2. Charakterystyka zmiennych 272
27.3. Macierz korelacji 273
27.4. Określenie modelu ekonometrycznego 275
27.5. Dane panelowe modelu PKB województw 277

28. Podstawowe miary statystyki opisowej 281
28.1. Przedziały obserwacji 281
28.2. Miary przeciętne (położenia) 282
28.3. Przeciętne pozycyjne 286
28.4. Wykres pudełkowy oraz histogram 288

29. Miary rozproszenia oraz miary asymetrii 291
29.1. Liczność występowania oraz obszar zmienności (rozstępu) 291
29.2. Odchylenia, wariancja, odchylenie standardowe oraz współczynnik zmienności 292
29.3. Zastosowanie funkcji Excela do wyznaczania miar statystyki 294
29.4. Miary asymetrii 296
29.5. Wybrane tabele statystyczne 298

30. Słowniczek wybranych pojęć w zakresie modelowania prognoz 308

31. Moje publikacje w latach 2008-2022 314

Bibliografia 336

Wstęp

Czasy obecne charakteryzują się niespodziewanymi wydarzeniami w sferze bezpieczeństwa, zdrowia, efektywności gospodarowania, czy też zjawiskami przyrodniczymi. Potrzebę przewidywania wskaźników odczuwają przedsiębiorstwa, zakłady usługowe, a także organy administracji terytorialnej. Na jednej z konferencji lokalnej pracownik starostwa powiatowego zapytał, czy nie znam metody cyfrowej prognozowania przy zmieniających się warunkach otoczenia. Było to bodźcem do zebrania moich dotychczasowych artykułów oraz sięgnięcia po źródła internetowe.

Coraz częściej jednak klasyczne metody prognozowania zastępowane są rozwiązaniami hybrydowymi. Scalenie jednak kilku metod w jedno wyrażenie formalne napotyka na trudności doboru odpowiednich cech i parametrów.

Niniejsze opracowanie skierowane jest przede wszystkim do studentów kierunków ekonomicznych. Jednak podjęcie się trafnego prognozowania wymaga od studiujących przynajmniej podstawowej wiedzy z zakresu modelowania ekonometrycznego.

Na początku tej pracy wymieniono metody prognozowania bazujące przede wszystkim na szeregach czasowych obserwacji w wymaganym horyzoncie. Zebrane jednak dane dotyczące cech starzeją się i dlatego trzeba tą sytuacje uwzględnić stosując wzór Hellwiga.

Specjalnie w tym opracowaniu położono nacisk na zademonstrowanie przykładów prognozowania zwłaszcza krótkookresowego. Rozwinięciem tematyki prognozowania jest rekonesans po publikacjach internetowych dotyczących przewidywania cech w dłuższych okresach czasowych. Rezultaty wielu naukowców sprowadzają się do tworzenia i testowania modeli obejmujących elementy prognozowania metodami liniowymi i nieliniowymi.

Dla wygody studiujących, i nie tylko, podałem w kolejnych przykładach podejścia do podstawowego prognozowania, a w rozdziale drugim „*Stosowane główne metody prognozowania w szerszych obszarach*” zainteresowani bliżej tą tematyką mogą sięgnąć po pełne teksty oferujące metody przeważnie hybrydowe. Uzupełnieniem tej pracy są rozdziały pomocnicze dotyczące statystyki pomocne w modelowaniu zjawisk dla celów prognozowania, a są nimi:

- Podstawowe miary statystyki opisowej,
- Miary rozproszenia i miary asymetrii (z wybranymi tablicami statystycznymi),
- Słowniczek wybranych pojęć z zakresu modelowania prognoz.

W tym miejscu chciałbym serdecznie podziękować wszystkim tym studentom, których fragmenty prac, po odpowiednim przystosowaniu, były inspiracją do napisania tej książki.

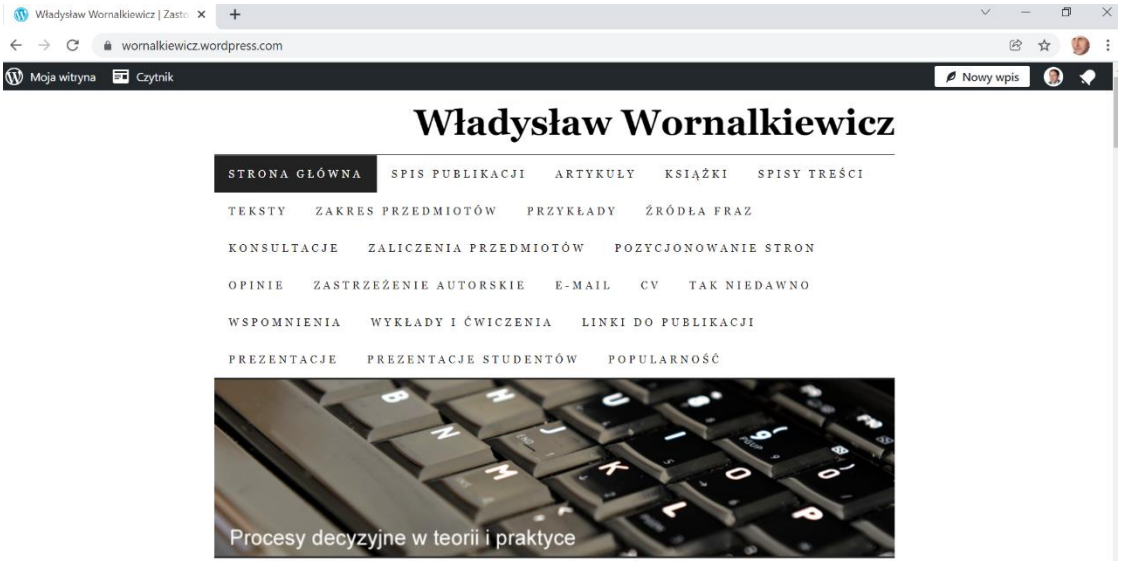
Autor

1. Matematyczne i niematematyczne metody prognozowania

1.1. Wprowadzenie

Nie na wszystkich uczelniach ekonomicznych są wykłady z ekonometrii, czyli nauce o zastosowaniach matematyki w modelowaniu zjawisk w szczególności gospodarczych. Z tego względu możliwości merytoryczne piszczących prace dyplomowe, zwłaszcza magisterskie są w znacznej mierze ograniczone. Bodźcem do napisania tego materiału było zapytanie jednej z dyplomantek co do wybrania właściwej metody ilościowej lub jakościowej, bo promotor zalecił jej wzbogacenie pracy dyplomowej o zastosowanie tych metod w obszarze sprzedaży w supermarkecie „Biedronka".

Dla popularyzacji metod ilościowych z wykorzystaniem statystyki, ekonometrii i narzędzi informatycznych prowadzę od roku 2012 blog „*Zastosowanie komputera w procesach decyzyjnych*", link: (*https://wornalkiewicz.wordpress.com/*), do którego stronę wejściową w języku polskim zaprezentowano na rysunku 1.1.



Źródło: Opracowanie własne w programie *WordPress*

Rys. 1.1. Menu główne blogu „*Zastosowanie komputera w procesach decyzyjnych*"

Ułatwieniem w podjęciu zagadnienia modelowania ekonometrycznego, i nie tylko, są też moje wcześniejsze podręczniki akademickie z zastosowaniem pakietu *WinQSB, Gretl* i *Solvera Excela*, a mianowicie:

1. Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych …*, Wydawnictwo Instytut Śląski w Opolu, Wydanie II rozszerzone, 2012, składające się z dwóch części: *Wybrane modele ekonometryczne*, *Elementy badań operacyjnych*.

2. Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, Wydawnictwo Instytut Śląski w Opolu, 2014, składające się również z dwóch części: *Środowiska programowe statystyki opisowej*, *Zagadnienia ekonometrii*.

3. Wornalkiewicz W., *Popularyzacja wybranych metod ilościowych*, Wyższa Szkoła Zarządzania i Administracji w Opolu, 2017, obejmująca części: *Ranking metod ilościowych i ich echo w zasobach wyszukiwarki o zasięgu światowym*, *Nowo wprowadzone publikacje do blogu*.

4. Wornalkiewicz W., *Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*, Skrypt Wyższej Szkoły Zarządzania i Administracji w Opolu, 2013.

Dla poczatkującego adepta sztuki modelowania, z myślą o wykorzystaniu zbudowanych modeli do prognozowania, przede wszystkim krótkookresowego konieczne jest wejście do „*świata*" różnych metod matematycznych, jak i niematematycznych. Z tego względu w tym materiale, bazując na dostępnych publikacjach, przybliżyłem wstępnie wachlarz tych metod. Oczywiście zastosowanie określonego postępowania wymaga rzetelnego poznania procedury modelowania równań, zawartej między innymi w wymienionych wcześniej publikacjach, jak też w innych źródłach bibliograficznych.

Klasyfikując rodzaje metod prognozowania sięgnijmy jeszcze do encyklopedii internetowej *Wikipedia*, aby dowiedzieć się jak definiowane jest pojęcie *prognozowanie* i jakie stosowane są techniki[1]. Metoda prognozowania to sposób przetwarzania informacji opisujących sytuację prognostyczną w prognozę, dostosowany do przyjętej zasady prognozowania. Może tu występować prognozowanie na podstawie szeregów czasowych np. bez tendencji. Metody tej grupy stosowane są, gdy w szeregu czasowym występuje składowa systematyczna w postaci stałego poziomu i wahania przypadkowe, przy czym wyróżniamy następujące metody: *naiwną*, *średniej ruchomej prostej*, *średniej ruchomej ważonej*, *prostego wyrównywania wykładniczego*. W wymienionym źródle w zakresie prognozowania ilościowego podano, że stosowane są:

- zmienna ciągła,
- modele trendu,
- analiza harmoniczna,
- analiza regresji,
- metoda najmniejszych kwadratów,
- zmienna dyskretna,
- modele probitowe,
- modele logitowe,
- analiza dyskryminacyjna.

Natomiast w zakresie prognozowania jakościowego, czyli heurystycznego wyróżnia się: *metody bezpośrednie*, wykorzystujące dane na temat dotychczasowego przebiegu procesu, *metody pośrednie*, stosujące dane odnośnie przebiegu badanego procesu oraz innych podobnych procesów, *metody analogowe*.

1.2. Metody matematyczne

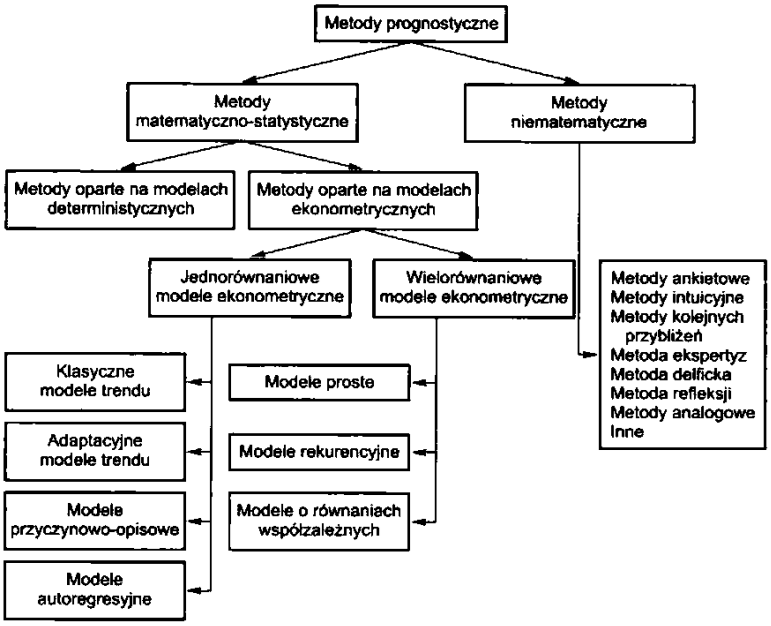
W teorii prognozowania, czyli w przewidywaniu i ocenie przyszłości wykorzystuje się[2]: *metody ilościowe o charakterze matematyczno-statystycznym*, *pojęcia i aparat analityczny rachunku prawdopodobieństwa*, *budowane modele statystyczno-ekonomiczne*. W zależności od charakteru badanego zjawiska, stosuje się metody matematyczno-statystyczne oraz metody niematematyczne, przy czym wśród metod niematematycznych wykorzystuje się w prognozowaniu następujące metody:

- *ankietowe*,
- *intuicyjne*,
- *kolejnych przybliżeń*,
- *ekspertyz*,
- *delficką*,
- *refleksji*,
- *analogowe*.

[1] https://pl.wikipedia.org/wiki/Prognozowanie.
[2] http://www.prognozowanie.info/metody-prognozowania/.

Metody niematematyczne, określone także jako *heurystyczne metody prognostyczne* wykorzystują opinie wielu ekspertów. Nie występuje tu zatem ekstrapolacja wykrytych w przeszłości prawidłowości w przyszłość, lecz jest to prognozowanie wariantów rozwoju zjawisk ze wskazywaniem tych najbardziej prawdopodobnych. Dość obszerna jest grupa metod heurystycznych do których między innymi należy *burza mózgów* i *metoda delficka*. Metody heurystyczne służą przede wszystkim do sporządzania prognoz długoterminowych. Natomiast w prognozowaniu gospodarczym stosuje się *sieci neuronowe*, bowiem dzięki zdolnościom do uczenia się *sieci neuronowe* nadają procesom cechę automatyzacji. Metody analogowe, takie jak *metoda analogii biologicznych, przestrzennych, historycznych, przestrzenno-czasowych* dostrzegają rozwój zjawisk społeczno-gospodarczych w innych obiektach np. wyżej rozwiniętych. Stosuje się tu wyrażenie prognoz ilościowo jak i jakościowo.

W metodach matematycznych dominują procedury oparte o *metody ekonometryczne* zaliczane do *technik ekstrapolacji ekonomicznych szeregów czasowych*, z doborem najlepszej metody z zastosowaniem komputerowego programu optymalizacyjnego. Schemat struktury podziału metod prognozowania pokazano na rysunku 1.2. W badaniach prognostycznych z zastosowaniem modeli ekonometrycznych wspomaganych techniką IT występują następujące modele: *klasyczne trendu, adaptacyjne, przyczynowo-opisowe, autoregresyjne, proste, rekurencyjne, o równaniach współzależnych*[3]. W wymienionych metodach ekonometrycznych wspomniano o szeregach czasowych na których bazuje procedura modelowania. Definicja tej procedury oparta jest na pojęciu zmiennej losowej x_t zależnej od parametru t (czasu). Czasami zakładamy, że odstępy pomiędzy t_i i t_{i+1} są jednakowe. Tym samym mamy do czynienia ze sparametryzowaną rodziną zmiennych losowych[4].



Źródło: http://www.prognozowanie.info/metody-prognozowania/.

Rys. 1.2. Rodzaje metod prognozowania

W obserwacji szeregu czasowego zauważyć możemy trend, sezonowość, cykliczność i losowość . Trend to czynnik kształtujący ogólną tendencję rozwojową, a czynnik trendu opisywany

[3] Zeliaś A., Pawełek B., Wanat S., *Prognozowanie ekonomiczne Teoria Przykłady Zadania*, Wydawnictwo Naukowe PWN, Warszawa 2021.

[4] http://www.kozlowski.pollub.pl/wyklady/PIS/pis_wyklad1.pdf.

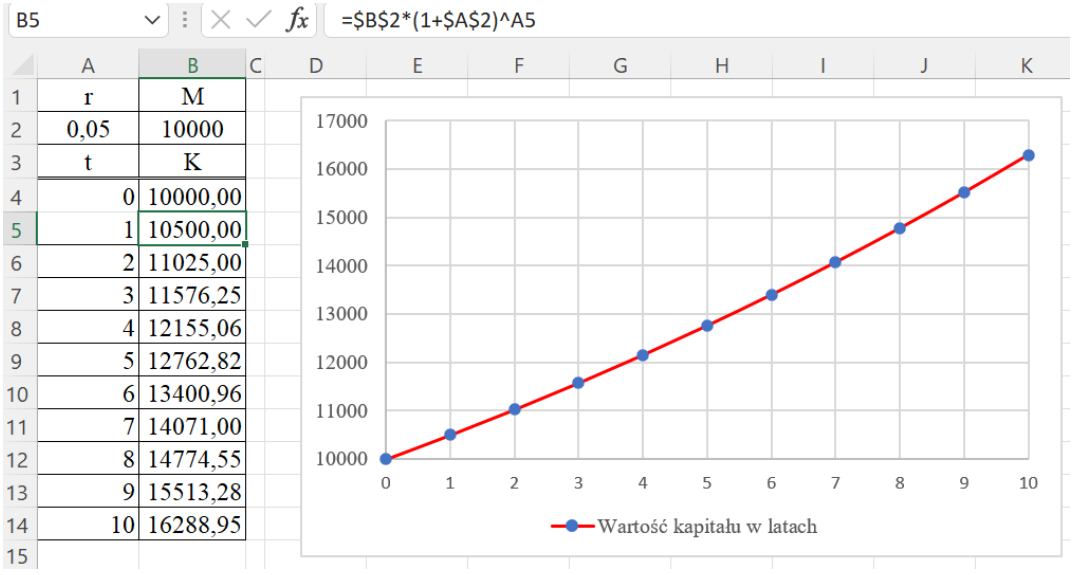
jest za pomocą nielosowej funkcji $f_{tr}(t)$. Sezonowość wyraża czynnik kształtujący okresowe wahania analizowanego zjawiska w ciągu roku. Wyniki działań czynników sezonowych modelujemy za pomocą nielosowej funkcji $\varphi(t)$. Natomiast cykliczność to czynnik kształtujący zmiany analizowanego zjawiska pod wpływem długoczasowych cykli, które opisujemy za pomocą funkcji nielosowej $\psi(t)$. Losowość wyraża czynnik kształtujący wahania (fluktuacje) o charakterze losowym. Wyniki działań losowych czynników modelujemy za pomocą zmiennych losowych ε_t, $t = 1, 2, \ldots$ przy czym szereg czasowy opisujemy następująco[5]:

$$x_t = \lambda_1 f_{tr}(t) + \lambda_2 \varphi(t) + \lambda_3 \psi(t) + \varepsilon_t$$

gdzie: $\lambda_i = 1$, jeżeli i-ty czynnik wpływa na kształtowanie zjawiska lub zero w przeciwnym razie.
Procedura budowy modeli matematycznych sprowadza się do następujących kroków postępowania:

- dobieramy odpowiednią klasę modeli,
- wybieramy model próbny,
- dokonujemy identyfikacji modelu próbnego,
- sprawdzamy za pomocą narzędzi statystyki matematycznej, czy wybrany model jest poprawny i może być wykorzystany do prognozowania,
- dokonujemy weryfikacji empirycznej.

Model, który pozwala dokładnie obliczyć wartość zmiennej zależnej w dowolnym momencie, nazywamy modelem deterministycznym. Natomiast model umożliwiający wyznaczenie przyszłych wartości z prawdopodobieństwami, nazywamy modelem probabilistycznym lub stochastycznym. Jako przykład modelu matematycznego w publikacji internetowej spotykamy model wartości kapitału, który zostanie dalej przytoczony z zastosowaniem arkusza kalkulacyjnego *Excel* (zob. rysunek 1.3)[6]. Niech x_t oznacza kapitał w chwili $t = 0, 1, 2, \ldots, N$, natomiast stopa zwrotu pozbawiona ryzyka wynosi r. Zależność wielkości kapitału od stopy zwrotu jest następująca: $x_{t+1} = (1 + r)\, x_t$. Wtedy kapitał, którym dysponuje inwestor w chwili t, jeżeli w chwili $t = 0$ posiadał $x_o = M$ możemy przedstawić jako ciąg geometryczny: $x_t = M\,(1 + r)^t$.



Źródło: Opracowanie własne w *Excelu*.

Rys. 1.3. Wartość kapitału początkowego $M = 10000$ przy stopie zwrotu $r = 5\%$ i okresie 10 lat

[5] Ibidem, wzór 1.
[6] Ibidem.

Na rysunku 1.3 pokazano również obliczone wartości kapitału w kolejnych 10. latach obliczone według wyrażenia *Excela* w postaci (=$B$2*(1+$A$2)^A5), gdzie komórki stałe, a więc kapitał początkowy *M* i stopa zwrotu kapitału *r* podane są jako adresowanie bezwzględne ze znakiem dolara.

Modele stochastyczne mają dość szerokie zastosowanie w praktyce, opisują bowiem przede wszystkim zachowania gospodarcze obiektów za pomocą zmiennych losowych, co umożliwia szacować z prawdopodobieństwem ich wartości w przyszłości.

1.3. Przykładowe metody stosowane w rolnictwie

Jak już nadmieniono, podstawowy podział metod prognozowania to dwie grupy, a mianowicie *statystyczno-matematyczne* i *niematematyczne*. Stanisław Krasowicz wskazuje, że metody niematematyczne nie są oparte na ścisłych obliczeniach i nie jest możliwe ich przedstawienie za pomocą modelu (wzoru) matematycznego[7]. Metody niematematyczne znajdują zastosowanie w tych dziedzinach, w których problemy nie zostały jeszcze wystarczająco poznane. Wymieniony wcześniej autor zaproponował więc ich zastosowanie w rolnictwie do przewidywania produkcji i uważa, że przydatne są metody: *heurystyczne* (*intuicyjne*), *kolejnych przybliżeń*, *analogowe*. *Metody intuicyjne* polegają na porządkowaniu wypowiedzi ekspertów z danej dziedziny wiedzy, bezpośrednio dotyczącej prognozowanego układu zjawisk i formułowaniu na tej podstawie wniosków prognostycznych. Znajdują szerokie zastosowanie szczególnie w przypadku prognoz jakościowych, a wśród intuicyjnych metod prognozowania występują:

- *metoda indywidualnych ekspertyz* (ocen rzeczoznawców),
- *metoda ekspertyz zespołowych* (równoległych lub kolejnych).

Metody ekspertyz mają ukazać najbardziej prawdopodobny przebieg rozwoju danej dziedziny lub (zjawiska) w określonym odcinku czasu i tu rozróżnia się:

- *metodę równoległych ekspertyz*, w której prognozę określonego zjawiska zaleca się do jednoczesnego, niezależnego od siebie do wykonania kilku osobom lub grupom rzeczoznawców;
- *metodę kolejnych ekspertyz*, przy czym pierwsza wersja prognozy podlega kolejnym ocenom, albo przez tę samą grupę ekspertów, którzy ją opracowali, albo przez nowo powołane zespoły osób. Ostatnia grupa ekspertów powinna wskazać, który z opracowanych wariantów prognozy może być uznany za najbardziej zalecany.

Według Krasowicza szeroko pojęta *metoda ekspertyz* była wykorzystana przy opracowaniach strategii rozwoju polskiego rolnictwa po akcesji z Unią Europejską. Wspomniana metoda była również stosowana przy opracowaniu prognozy zapotrzebowania rolnictwa polskiego na nawozy mineralne.

Jedną z odmian ekspertyz zespołowych jest *metoda delficka*, a prognozowanie tą metodą polega na opracowaniu szczegółowych ankiet skierowanych do specjalistów i ekspertów, a następnie na uogólnieniu opinii na podstawie statystycznej analizy uzyskanych odpowiedzi. Odmianą metody delfickiej jest metoda *SEER* (*System for Event Evaluation and Review*). Do metody ekspertyz zespołowych należą też konferencje ideowe, nazywane często *metodą panelową*. Polegają one na organizowaniu zebrań, na których eksperci przedstawiają swoje poglądy dotyczące kierunków i form rozwoju analizowanych zjawisk i procesów w przyszłości oraz propozycje osiągnięcia celów. Do tego typu konferencji zalicza się także metodę *buzz session* oraz *synektykę*. Wymieniona tu sesja *buzz* to w zasadzie niewielka nieformalna/formalna dyskusja grupowa na dany temat i podobna jest do burzy mózgów lub debaty[8]. Natomiast pojęcie *synektyka* wyraża jedną z najpopularniejszych technik

[7] Inspiracją tego materiału był artykuł: Stanisław Krasowicz, *Niematematyczne metody prognozowania w rolnictwie, Studia i raporty IUNG - PIB* 19 ZESZYT 14 2009, Opracowanie wykonano w ramach zadania 2.1. w programie wieloletnim IUNG - PIB, Instytut Uprawy Nawożenia i Gleboznawstwa - Państwowy Instytut Badawczy w Puławach.
[8] https://www.answers.com/Q/What_is_a_buzz_session.

twórczego rozwiązywania problemów. Jest to metoda heurystyczna, a więc z założenia nie gwarantuje ona osiągnięcia założonego celu, lecz jedynie zwiększa prawdopodobieństwo jego osiągnięcia. Synektyka oznacza systematyczne i intencjonalne użyciu analogii w myśleniu twórczym[9]. Skorzystajmy dalej z opracowania Stanisława Krasowicza, gdzie scharakteryzowane są dalsze metody prognozowania niematematycznego.

Metoda macierzy powiązań jest podobna do delfickiej. Macierz taką konstruuje się i przedstawia w niej oszacowane dwa typy danych ważnych dla przyszłego rozwoju. Pierwsze mówią o prawdopodobieństwie, że dane zjawisko wystąpi w określonym czasie, drugie o sile wzajemnych powiązań i ich prawdopodobieństwie. Inną formą przewidywania zjawisk w rolnictwie, którą można zaliczyć do metod heurystycznych jest *metoda refleksji*, która polega ona na sporządzaniu prognoz na podstawie badań różnic pomiędzy przyszłością a przeszłością i teraźniejszością. Przy *metodzie refleksji* koncentruje się na wyszukiwaniu tego, co może być inne od teraźniejszości i przeszłości. Ogólna idea prognozowania metodami kolejnych przybliżeń polega na korygowaniu ocen rozwoju zjawiska w przyszłości drogą coraz to nowych szacunków badanej zmiennej. Jest to klasyczny przykład weryfikacji pierwotnej wersji prognozy przez jej uściślanie lub usuwanie z niej luk i sprzeczności.

W grupie *metod macierzy powiązań* stosuje się *metody normatywne, morfologiczne, drzewa decyzji, scenariusza* i *gier ekonomicznych*. *Metody normatywne* polegają na wyborze nośników zapotrzebowania (popytu) oraz wyznaczeniu norm zapotrzebowania (zużycia) na dane produkty. Natomiast *metody morfologiczne* zmierzają do szczegółowej analizy struktury zmiennej prognozowanej, na podstawie której zostaną wyodrębnione poszczególne jej elementy czy składniki. Działanie takie jest adekwatne do prognozowania procesów, które mają charakter złożony.

Zbliżona co do sposobu jest *metoda drzewa decyzji*. Polega ona na zbudowaniu drzewa celów. Na najwyższym poziomie będzie to cel podstawowy, zasadniczy, a na poziomach niższych różne cele cząstkowe zapewniające realizację celu podstawowego. Występuje jeszcze metoda scenariusza, głównym jej celem jest jakościowy, kompletny opis systemu w założonym horyzoncie czasowym. Stosowany jest również opis w postaci *modeli agregatowych*, *submodeli*, które umożliwiają tworzenie schematu powiązań ilościowych elementów systemu w okresie objętym prognozą oraz zastosowanie wielu różnych symulacji. Dodam, że użycie *metody scenariusza* zwłaszcza do prognozowania interdyscyplinarnego wymaga zespołu badawczego.

Zwrócę jeszcze uwagę na *metody analogowe*, których istotą jest prognozowanie o przyszłości jednych obiektów na podstawie wiedzy o innych podobnych, przy czym w tej grupie metod wyróżniamy[10]: *analogie biologiczne, historyczne, techniczno-technologiczne, geograficzne, metody trendów prekursywnych, metody trendów współwiązanych*. Jednak metody analogowe zawierają w sobie pewne niebezpieczeństwa wynikające z porównywania obiektów, procesów lub zjawisk w różnych krajach, które zakładają sobie odmienne cele i zadania rozwojowe w przyszłości.

Wymienione wcześniej *metody trendów prekursywnych* i *współwiązanyc*h mogą być stosowane w określonych warunkach, bowiem w rzeczywistości społeczno-gospodarczej występują sytuacje, w których wzrastającym wartościom jednej zmiennej towarzyszą regularne opóźnienia w narastaniu innej zmiennej. Takie prawidłowości stwarzają podstawę do zastosowania *metody prognozowania zwanej trendami prekursywnymi*.

Możemy też stosować *metodę trendów współwiązanych*, gdy procesy wykazują silną zależność i można to wykorzystać do przedstawienia sytuacji w przyszłości obiektów mniej

[9] https://pl.wikipedia.org/wiki/Synektyka.
[10] Krasowicz S., *Niematematyczne metody prognozowania w rolnictwie*, op. cit.

zaawansowanych np. technologicznie. Zdaniem Stanisława Krasowicza, również *metody analogowe prognozowania* mogą być podstawą do przewidywania wielu zjawisk oraz procesów, w tak trudnym obszarze prognozowania jakim jest rolnictwo.

1.4. Publikacje dotyczące prognozowania długookresowego

Coraz częściej spotykam się z zainteresowaniem pracowników zarządzających administracji państwowej prognozowaniem długookresowym. Występuje to zwłaszcza w obszarze wyznaczania strategii rozwoju regionu, województwa czy też innej jednostki terytorialnej. Interesujące są przewidywania w zakresie możliwości napływu kapitału dla uprzemysłowienia danego obszaru, odpływu potencjalnej siły roboczej, napływu funduszy z bazy finansowej Unii Europejskiej.

Aby określić prognozy długookresowe trzeba znać metody temu służące. Nie spotkałem jednak publikacji zwartej poświęconej tylko temu tematowi. Sięgnąłem więc po źródła internetowe pozyskane poprzez wyszukiwarkę Google na frazę „*long-forecasting*", czyli prognozowanie długookresowe. Sadzę, ze taki rekonesans po publikacjach w Internecie z tego zakresu zainteresuje Czytelników i jednocześnie część z nich podejmie się implementacji wymienionych w nich metod do praktyki własnych instytucji, zakładów produkcyjnych lub usługowych. Po podaniu frazy „*long-forecasting*" uzyskano listę wielu pozycji dotyczących tego tematu. Wymienię tu jednak tylko te szersze, które moim zdaniem bardziej przybliżają problematykę prognozowania długookresowego w różnych obszarach.

W artykule „*How to Choose the Right Forecasting Technigue*" podjęto próbę odpowiedzi na pytanie „*Jak wybrać właściwą technikę prognozowania?*"[11]. Określenie miejsca Polski na mapie historii prognozowania dalekiego zasięgu spotykamy w artykule internetowym „*Poland on the map of long-range forecasting history*" pod linkiem: *https://obserwator.imgw.pl*. Przykłady użycia czterech typowych metod prognozowania przychodów występują w publikacji „*Forecasting Methods - Top 4 Types, Overview, Examples*" z dostępem: *https://corporatefinanceinstytute.com/resources/knowlege/modeling/forecasting-methods/*.

Po żądaniu: *published long-forecasting articles* uzyskujemy listing opublikowanych artykułów naukowych dotyczących prognozowania długookresowego. Wskazujemy rok 2022 ograniczenia czasowego dla sporządzenia listingu, język polski, typ artykułu - przeglądowy i w efekcie generowane jest zestawienie obejmujące 48 pozycji. Moim zdaniem warto bliżej zapoznać się z poglądami zawartymi w wersjach elektronicznych następujących artykułów:

Prognozowanie z wykorzystaniem modeli autoregresji wektorowej bayesowskiej: porównywanie bezpośrednich i iterowanych metod wieloetapowych,

Wpływ zmian klimatycznych na plony pszenicy w Turcji: heterogeniczne badania panelowe,

Prognozowanie wzrostu zatrudnienia za pomocą trendów …,

Prognozowanie w czasie rzeczywistym zmęczenia wywołanego wysiłkiem fizycznym za pomocą czujników do noszenia,

Zmiana w czasie modeli trendów do prognozowania inflacji w Australii,

Dekompozycja w trybie dynamicznym dla areo-optycznej charakterystyki czoła fali,

Przewidywanie cech wpływających na oczyszczalnie ścieków. Dobre praktyki i wyzwania,

Analiza porównawcza planowania produkcji energii wiatrowej w oparciu o sztuczną sieć neuronową i podejście teoretyczne,

Długoterminowe prognozowanie obciążenia w ramach inteligentnej sieci z wykorzystaniem szeregów czasowych i podejść ekonometrycznych (dotyczący produkcji i dystrybucji energii odnawialnej),

[11] https://hbr.org/1971/07/how-to-choose-the-right-forecasting-technique.

Poprawa dokładności prognozowania wskaźników budowy na poziomie miasta poprzez połączenie liniowych ARIMA i nieliniowych SSN (zastosowanie modelu hybrydowego),

Prognozowanie zużycia energii w budynku z wielogodzinnym wyprzedzeniem dzięki nowemu zintegrowanemu modelowi,

- Przegląd modeli przewidywania degradacji nawierzchni.

* * *

Różnorodność jest metod matematycznych i niematematycznych przydatnych do prognozowania krótko i długoterminowego. Dzielą się one na grupy, a w ramach nich można wyodrębnić rodzaje i podrodzaje, które w szczególności spotykamy w publikacjach zagranicznych internetowych. W niniejszym materiale zasygnalizowano tylko podstawowe rodzaje metod. Bliższe ich poznanie wymaga sięgnięcia do literatury przedmiotu, jak i do dość obszernych źródeł, które już opublikowano w Internecie.

Autor chciał, aby piszący pracę dyplomową z zakresu przewidywania zjawisk gospodarczych, społecznych, ekologicznych i innych, mogli poprzez ten rekonesans po metodach ilościowych jak i jakościowych, stosować je do oszacowania wartości zjawisk w przyszłości. Jest bowiem taka potrzeba, aby opracowania prac dyplomowych z zakresu zarządzania, logistyki, czy też administracji posiadały choć w części zastosowanie modeli do prognozowania cech w obiektach - miejscach pracy studentów i nie tylko. Pomocny w tym zakresie może być mój skrypt „*Badania przyczynowo skutkowe związków między cechami statystycznymi*” wymieniony na *Wstępie* tego opracowania.

2. Stosowane główne metody prognozowania w szerszych obszarach

Bieżąca i wielokrotna regresja liniowa w analizach finansowych

Analitycy finansowi w prognozowaniu dochodów budżetowych i kosztów kapitałowych wykorzystują cztery techniki, a mianowicie zastosowanie linii prostej (stała stopa wzrostu), średniej ruchomej (powtarzające się prognozy), prostej regresji liniowej (jedna zmienna zależna i jedna zmienna niezależna), wielokrotnej regresji liniowej (jedna zmienna zależna i wiele niezależnych)[12]. Przykład wykorzystania pierwszej techniki w arkuszu *Excela* pokazano na rysunku 2.1.

C4 =(C6-B6)/B6

	A	B	C	D	E	F	G	H
1	Metoda liniowa							
2		Dane→		Prognoza→				
3		2015	2016	2017	2018	2019	2020	2021
4	Procent wzrostu sprzedaży	4,0%	**4,0%**	4,0%	4,0%	4,0%	4,0%	4,0%
5								
6	Przychody	81422	84698					

Źródło: Opracowanie własne w *Excelu* na podstawie - https://corporatefinanceinstitute.com/resources/knowledge/modeling/forecasting-methods/.

Rys. 2.1. Wyznaczanie prognoz w kolejnych latach przy stałej stopie wzrostu

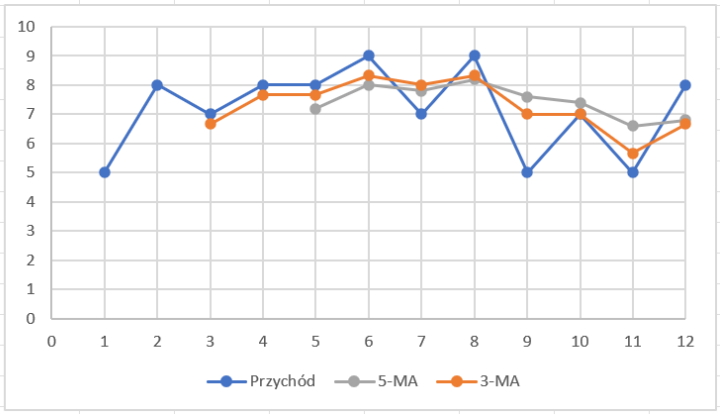
Na rysunku 2.1 dla prognozowania np. przychodów przez firmę założono, że stopa wzrostu wyznaczona według lat 2015 i 2016 jest stała i obliczona jest zależności: =(C6-B6)/B6. Wyrażona procentowo w wysokości 4% obowiązuje przez następne 5 lat tj. 2017-2021. Aby móc prognozować przychód na każdy następny rok posługujemy się formułą np.: =C6*(1+D5). Następnie kopiujemy formułę dla kolejnych lat.

Miesięczna średnia ruchoma obliczana jest dla trzech lub pięciu miesięcy. W wersji polskiej *Excela* stosowana jest formuła: =ŚREDNIA(B3:B5) obliczająca średni przychód np. od stycznia do marca w przypadku techniki średniej ruchomej (kroczącej) 3-miesięcznej (zob. rysunek 2.2). Obliczanie średniej 5-cio miesięcznej uważane jest jako starsza technika prognozowania. Ilustrację kształtowania się przychodów przy stałej stopie wzrostu oraz obliczanego jako 3 lub 5-cio okresowa średnia ruchoma pokazano na rysunku 2.2.

[12] https://corporatefinanceinstitute.com/resources/knowledge/modeling/forecasting-methods/.

C5 =ŚREDNIA(B3:B5)

	A	B	C	D	E
1	Metoda średnia ruchoma				
2		Przychód [$]	3-MA [$]	5-MA[$]	
3	Styczeń	5			1
4	Luty	8			2
5	Marzec	7	6,7		3
6	Kwiecień	8	7,7		4
7	Maj	8	7,7	7,2	5
8	Czerwiec	9	8,3	8,0	6
9	Lipiec	7	8,0	7,8	7
10	Sierpień	9	8,3	8,2	8
11	Wrzesień	5	7,0	7,6	9
12	Październik	7	7,0	7,4	10
13	Listopad	5	5,7	6,6	11
14	Grudzień	8	6,7	6,8	12



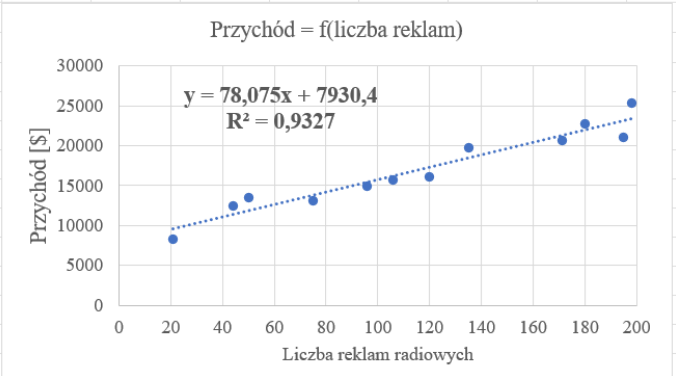
Źródło: Opracowanie własne w *Excelu* na podstawie - https://corporatefinanceinstitute.com/resources/knowledge/modeling/forecasting-methods/.

Rys. 2.2. Sposób obliczania średniej ruchomej 3 i 5-cio okresowej

Przykład zastosowania prostej regresji liniowej np. między reklamami radiowymi a przychodami widzimy na rysunku 2.3. Zastosowana formuła regresji liniowej widoczna na wykresie jest następująca: y = 78,075x +7930,4. Zwróćmy uwagę na współczynnik determinacji R^2 = 0,9327, który wskazuje na dobre dopasowanie modelu prognozy do danych - *liczby reklam radiowych*. Dysponując trendem regresji liniowe możemy swobodnie dokonywać przeliczeń symulacyjnych dla dowolnej liczby reklam radiowych, przykładowo jak na rysunku 2.3 dla 100, 150, 200. Pozostaje nam jeszcze dokończenie zredagowania wykresu zależności *Przychód = f (Liczba reklam radiowych)* poprzez wstawienie tytułu osi poziomej (*Liczba reklam radiowych*) i pionowej (*Przychody*), wpisanie tytułu wykresu *Przychód = f(liczba reklam).*

C18 =78,075*B18+7930,4

	A	B	C
1	Prosta regresja liniowa		
2	M-c	Liczba reklam	Przychód [$]
3	Styczeń	21	8350
4	Luty	180	22755
5	Marzec	50	13455
6	Kwiecień	195	21100
7	Maj	96	15000
8	Czerwiec	44	12500
9	Lipiec	171	20700
10	Sierpień	135	19722
11	Wrzesień	120	16115
12	Październik	75	13100
13	Listopad	106	15670
14	Grudzień	198	25300
15	Suma	1391	203767
16	Średnia	116	16981
17		Reklamy	Prognoza [$]
18	Przykłady:	100	15737,9
19		150	19641,65
20		200	23545,4

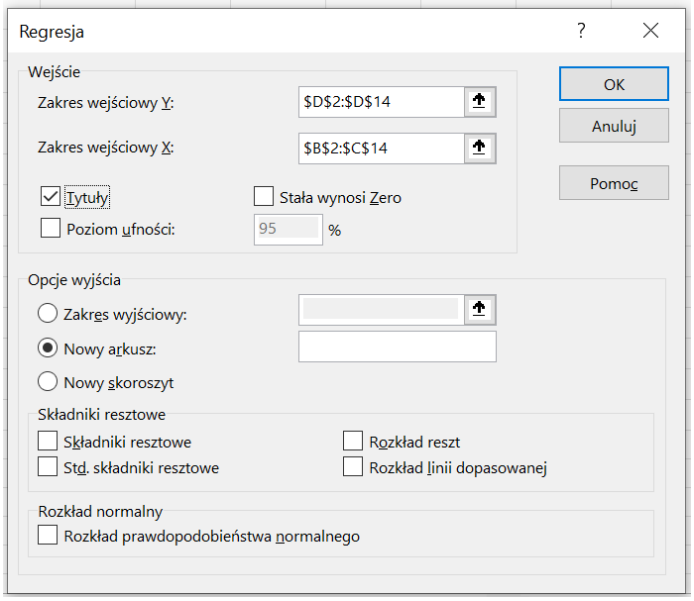


Źródło: Opracowanie własne w *Excelu* na podstawie - https://corporatefinanceinstitute.com/resources/knowledge/modeling/forecasting-methods/.

Rys. 2.3. Wyznaczenie linii regresji liniowej

Przejdźmy teraz do sytuacji, gdy trzeba zastosować zależność wielokrotną regresji liniowej, przy czym w *Excelu* skorzystamy z zakładki *Dane>Analiza danych>Regresja*. Zmienną zależną jest

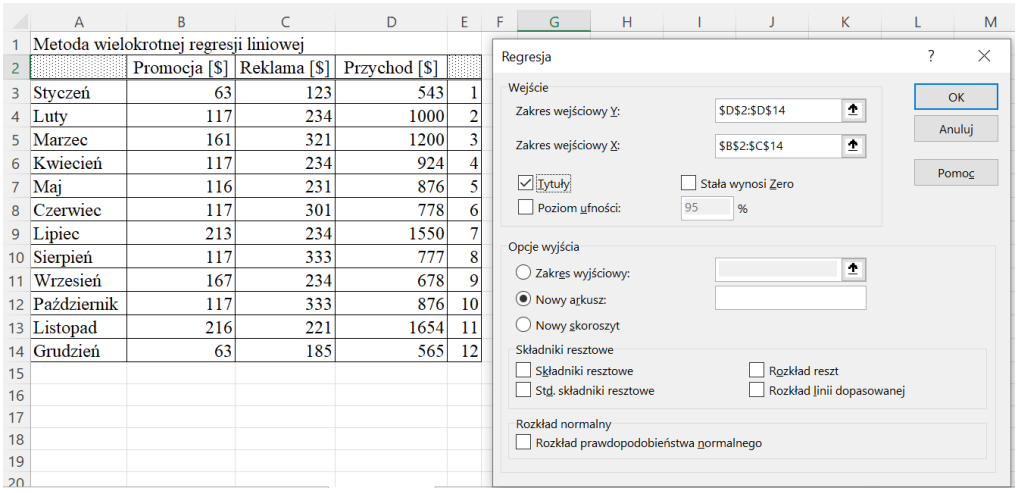
Przychód, a zmiennymi niezależnymi *Promocja sprzedaży* i *Reklama*. Reklama rozumiana jest tu raczej jako ogłoszenia, promocja sprzedaży jako wyprzedaż po cenach dogodnych dla klientów.



Źródło: Opracowanie własne.

Rys. 2.4. Okno dialogowe wyboru funkcji *Regresja* w *Excelu*

Dla wyznaczenia funkcji wielokrotnej regresji liniowej w oknie dialogowym *Regresja Excela* podajemy jako adresowanie bezwzględne zakres zmiennej zależnej ($D$2:$D$14) oraz obszaru dwóch zmiennych niezależnych ($B$2:$C$14). Zaznaczamy, że zmienne mają tytuły a wynik będzie zaprezentowany na osobnym arkuszu w ramach tego samego skoroszytu (zob. rysunek 2.5).



Metoda wielokrotnej regresji liniowej

	Promocja [$]	Reklama [$]	Przychod [$]	
Styczeń	63	123	543	1
Luty	117	234	1000	2
Marzec	161	321	1200	3
Kwiecień	117	234	924	4
Maj	116	231	876	5
Czerwiec	117	301	778	6
Lipiec	213	234	1550	7
Sierpień	117	333	777	8
Wrzesień	167	234	678	9
Październik	117	333	876	10
Listopad	216	221	1654	11
Grudzień	63	185	565	12

Źródło: Opracowanie własne w *Excelu* na podstawie danych w - https://corporatefinanceinstitute.com/resources/knowledge/modeling/forecasting-methods/.

Rys. 2.5. Wypełniony zakres parametrów wejściowych do wielokrotnej regresji liniowej

Po zaakceptowaniu OK uzyskujemy na osobnym arkuszu model oraz obszerne dane statystyczne go charakteryzujące, sprowadzone prze zemnie z dwunastu do trzech miejsc po przecinku(zob. rysunek 2.6).

B17 187,329605374008

	A	B	C	D	E	F	G	H	I
1	PODSUMOWANIE - WYJŚCIE								
2									
3	*Statystyki regresji*								
4	Wielokrotność R	0,875							
5	R kwadrat	0,765							
6	Dopasowany R kwadrat	0,713							
7	Błąd standardowy	189,608							
8	Obserwacje	12							
9									
10	ANALIZA WARIANCJI								
11		*df*	*SS*	*MS*	*F*	*Istotność F*			
12	Regresja	2	1056016,797	528008,399	14,687	0,001			
13	Resztkowy	9	323561,453	35951,273					
14	Razem	11	1379578,250						
15									
16		*Współczynniki*	*Błąd standardowy*	*t Stat*	*Wartość-p*	*Dolne 95%*	*Górne 95%*	*Dolne 95,0%*	*Górne 95,0%*
17	Przecięcie	187,330	252,282	0,743	0,477	-383,372	758,032	-383,372	758,032
18	Promocja [$]	6,392	1,201	5,323	0,000	3,676	9,109	3,676	9,109
19	Reklama [$]	-0,319	0,934	-0,342	0,740	-2,432	1,793	-2,432	1,793

Źródło: Opracowanie własne w *Excelu*.

Rys. 2.6. Raport wyników modelu wielokrotnej regresji liniowej

W raporcie występują trzy podtabele: statystyki regresji, analiza wariancji, dane modelu i oszacowania jego współczynników. *Dopasowany* R^2 będący podstawową miarą wielokrotnego modelu regresji wynosi 0,713, a sam model dla przyjętych danych szeregów czasowych przedstawia się następująco:

Przychód [$] = 187,330 + 6,392 * Promocja [$] - 0,319 * Reklama [$]

Dla przykładu dokonano obliczenia przychodu dla wskazanych danych w zakresie promocji i reklamy, z formatowaniem w dolarach amerykańskich i z dokładnością do dwóch miejsc po przecinku, wykorzystując wyrażenie naszego modelu: =$B$17+($B$18*B22)+($B$19*B23).

B24 =$B$17+($B$18*B22)+($B$19*B23)

	A	B	C	D
22	Promocja sprzedaży [$]	$125,00	$75,00	$200,00
23	Reklama (ogłoszenia) [$]	$250,00	$300,00	$300,00
24	Przychód [$]	$906,58	$570,99	$1 370,05

Źródło: Opracowanie własne w *Excelu*.

Rys. 2.7. Przykładowe obliczenia przychodu z użyciem modelu wielokrotnej regresji liniowej

Wykorzystywanie modeli statystycznych do przewidywania prognozy

Słyszałem kiedyś w radiu następującą informację meteorologa „*Dorobiliśmy się modeli ekonometrycznych do przewidywania pogody i teraz jest ona trafniejsza*". Moim zdaniem interesujący w tym zakresie jest artykuł „*Polska na mapie historii prognozowania dalekiego zasięgu*", który wystąpił w Magazynie Popularno-Naukowym „*Obserwator*" wydawanym przez IMGW[13]. Śledząc pogodę z dnia na dzień raczej zgadzamy się z tą wypowiedzią. Jednak cała trudność polega na modelowaniu przewidywań pogody w dłuższym horyzoncie czasowym i tu stale doskonali

[13] Pianko-Kluczyńska K., Konca-Kędzierska K., Jaczewski A., Mandal A., Drożdzioł R., Wdowikowski M., Bałdysz Z., Wrońska W., Kózka K., Nykiel G., Strzyżewski T., Marta Gruszczyńska M., *Polska na mapie historii prognozowania dalekiego zasięgu*, Magazyn Popularno-Naukowy „*Obserwator*", IMGW-PIB/Ośrodek Modelowania Meteorologicznego Analizy i Prognozy Dalekiego Zasięgu, lipiec 2022, wersja angielska.

się metody i techniki jej określania. Z wymienionej publikacji dowiadujemy się, że prognozowanie długookresowe w przejściowej strefie klimatycznej, a w takiej znajduje się Polska, należy do najtrudniejszych. Strefa ta bowiem cechuje się znaczną zmiennością warunków pogodowych.

Wykorzystywanie programów komputerowych rozpoczęto w roku 1956, a rok później skonstruowano analogowy analizator harmoniczny, który wykonywał obliczenia na dużych zbiorach danych z wykorzystaniem analizy harmonicznej Fouriera. Dzięki temu urządzeniu tworzono długoterminowe modele statystyczne do sporządzania prognoz średniej miesięcznej temperatury powietrza, sumy opadów dla Warszawy. Prace te prowadzone były w Zakładzie Prognoz Dalekiego Zasięgu. Opracowana metoda bazowała na założeniu „*podobne procesy atmosferyczne prowadzą do podobnych skutków*”[14]. Kluczem tej metody był dobór odpowiednich zmiennych opisujących procesy atmosferyczne w okresie poprzedzającym prognozę, tzw. predyktorów w okresie prognozowanym, przy czym często stosowanymi predyktorami były:

- średnia miesięczna i sezonowa,
- dobowe wskaźniki cyrkulacji atmosferycznej i ciśnienia atmosferycznego w wybranych punktach Europy i Północnego Atlantyku,
- temperatura powietrza i suma opadów w Warszawie.

Zdobyte doświadczenia w trakcie eksploatacji wymienionego wcześniej systemu, a także wykonanie dodatkowych aplikacji zaowocowały rozszerzeniem terytorialnym przewidywania pogody dla 18. miast w Polsce. Było to impulsem do pracy nad nowymi rozwiązaniami. Powstał spektralny model półkuli północnej na komputerze Odra. Pracowano nad klasyfikacją typów pogody dla regionów europejskich w ramach programu *European Cooperation in Science and Technology*.

W miarę rozwoju techniki cyfrowej IT zmieniały się również metody prognozowania. Powstał model statystyczny częściowo probabilistycznych prognoz 30-dniowych, oparty na zaawansowanych algorytmach logiki rozmytej. Opracowany w roku 2018 model statystyczny przewiduje średnią miesięczną temperaturę powietrza, sumę opadów oraz liczbę dni z opadami do sześciu miesięcy do przodu, przy użyciu regresji częściowej. W następnych latach 2018-2021 powstawały kolejne pakietowe wersje modeli, a w tym wykorzystujące teorię prawdopodobieństwa i twierdzenie Bayesa. Od roku 2020 w Zakładzie Analiz Meteorologicznych i Prognoz Dalekiego Zasięgu trwają prace nad wykonaniem i wdrożeniem:

- wieloletniego modelu numerycznego dla obszaru Polski (WRF-CFS),
- aktualnych modeli statystyczno-numerycznych (IMGW-TWS, IMGW-Reg),
- modeli statystycznych (IMHW-Bayes),
- modeli globalnych (ECMWF i NOAA),
- metody prezentacji prognoz.

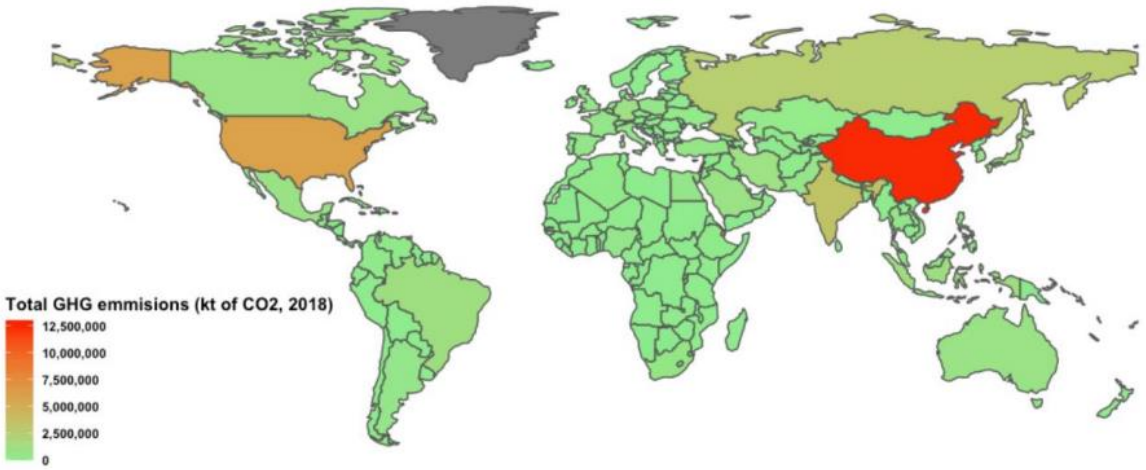
Warto w tym miejscu przybliżyć nieco choć niektóre z wymienionych modeli[15]. Długoterminowy deterministyczny model prognozy (WRF-CFS) wykorzystuje numeryczny model prognozy (*Weather Research and Forecasting Model*) oraz wyniki globalnego modelu (*Climate Forecast System*). Model IMGW-TWS opiera się na zaawansowanych metodach logiki rozmytej. Prognozuje zakresy minimalnej i maksymalnej temperatury powietrza, liczbę dni z opadami oraz liczbę dni z opadami śniegu w trzech dekadach miesiąca w odniesieniu do 87 stacji meteorologicznych w Polsce. Numeryczno-statystyczny model prognozowania IMWG-Reg określa średnią miesięczną temperaturę powietrza i miesięczne opady całkowite. Prognoza jest określona dla 20. wybranych miast w Polsce. Wyniki tej metody są wykorzystywane do przewidywania dalekosiężnej prognozy temperatury i opadów. Kolejny model statystyczny IMGW-Bayes

[14] Ibidem.
[15] Ibidem.

prognozuje prawdopodobieństwo wystąpienia średniej miesięcznej temperatury powietrza i miesięcznych opadów całkowitych w klasach: powyżej średniej, w granicach normy i poniżej średniej. Uzyskane wyniki odnoszą się też do 87. stacji meteorologicznych w Polsce.

Wspomaganie metodami ekonometrycznymi analiz związanych z ociepleniem klimatu

W celu ograniczenia ocieplenia klimatu kraje Unii Europejskiej powinny przestrzegać ograniczenia zużywania paliw kopalnych w celu redukcji emisji gazów cieplarnianych GHG[16]. Istotne jest zatem wygenerowanie wiarygodnych prognoz zagregowanych emisji tych gazów dla krajów Europy Środkowo-Wschodniej. Stosowanych jest siedem modeli statystycznych i uczenia maszynowego, aby przewidzieć GHG w 9. krajach wymienionej wcześniej części Europy w okresie 2019-2050. Opracowane szacunki wykazują, że niektórym krajom z tej dziewiątki trudno będzie osiągnąć wyznaczone progi emisji gazów cieplarnianych, obniżając ją o 55% do 2030 roku. Zwróćmy jeszcze uwagę na rysunek 2.8. pokazujący sumaryczną ilość emisji gazów cieplarnianych na świecie.



Źródło: World Development Indicators (WDI).

Rys. 2.8. Światowe emisje gazów cieplarnianych w roku 2018

Do prognozowania jakości powietrza, w tym emisji CO_2 wykorzystuje się metody statyczne. Zastosowane zostało kilka metod ekonometrycznych, a mianowicie model ARIMA, Holta-Wintersa, wygładzania wykładniczego oraz analizę widma osobliwego (SSA). Będący w użyciu również model kombinowany obejmuje SSA i prognozę dostarczaną przez *Energy Information Administration* (EIA), co umożliwia przewidywanie z 12-miesięcznym wyprzedzeniem. Badanie wysokości emisji CO_2 ma podstawowe znaczenie również dla prognoz śmiertelności i skali zmiennych zdrowotnych.

Przeprowadzone badania emisji gazów cieplarnianych we wspomnianych 9. krajach tj. Austrii, Bułgarii, Chorwacji, Czechach, Węgrzech, Polski, Rumunii, Słowacji i Słowenii umożliwiły wyznaczenie trendów emisji zanieczyszczeń w perspektywie długoterminowej. Badacze, którzy stosują modele uczenia maszynowego podkreślają ich wyższość prognozowania w stosunku do metod statystycznych. W automatycznym przewidywaniu wieloletnim stanu zanieczyszczeń powietrza stosowane są również inne modele predykcyjne a mianowicie:

- wygładzania wykładniczego w przestrzeni stanów (ETS),

[16] *Ocena realizacji polityki EU Net-Zero w wybranych państwach członkowskich za pomocą algorytmów automatycznego prognozowania*, Międzynarodowy Dziennik. Informacje geograficzne (isprs), Wydział Biznesu Międzynarodowego i Ekonomii, Uniwersytet Ekonomiczny w Bukareszcie, robert.sova@ase.ro, https://www.researchgate.net/scientific-contributions/Robert-Sova-13485445.

- Holta-Wintersa (HW),
- trygonometryczny w przestrzeni stanów z transformacją Boxa-Coxa,
- trendu i składowych sezonowych (TBATS),
- autoregresji zintegrowanej średniej ruchomej (ARIMA),
- strukturalnych szeregów czasowych (STS),
- autoregresji sieci neuronowej (NNAR).

Prognozowanie energii wiatru

Wobec globalnego ocieplenia klimatu, spowodowanego nadmierną emisją dwutlenku węgla, poszukuje się nowych technologii wytwarzania energii elektrycznej, w tym poprzez budowanie stacji turbin wiatrowych. Wiatr jest bowiem jednym ze źródeł energii odnawialnej. Szacuje się, że do roku 2024 globalna moc zainstalowania energii wiatrowej wyniesie do 1 TW (*Raport Global Wind Energy Council*)[17]. Jednak siła wiatru napędzająca turbinę w ramach określonej farmy jest zmienna i musi być przewidywalna. Poszukuje się nowatorskich metod prognozowania energii wiatru. Jedną z nich jest metoda dwukierunkowa bazująca na różnicy prognoz w przód i w tył (wygasłych). Powstał zintegrowany model sieci neuronowej i łańcucha Markowa dla krótkoterminowej prognozy prędkości wiatru, model hybrydowy składający się z ARMA i algorytmu wzmacniającego.

Występuje niska sterowalność energetyki wiatrowej, co wiąże się z niestabilnością przesyłu do ogólnego systemu sieci elektroenergetycznej. Tak więc dysponowanie adekwatnym modelem prognoz energii uzyskiwanej z wiatru daje szanse na właściwe zaplanowanie mocy dostaw energii elektrycznej. Występuje wiele technik prognozowania wiatru, które ogólnie podzielono na:

- numeryczne prognozowanie pogody (NWP),
- metody statystyczne,
- metody hybrydowe.

Uznano jednak, że najprostsza średnia godzinowa prędkość wiatru jest najbardziej odpowiednia. Jako zmienne brano też pod uwagę w modelowaniu kierunek wiatru i temperaturę otoczenia. Wymienione modele statystyczne maja na celu przewidywanie korelacji między różnymi zmiennymi objaśniającymi. Na uwagę zasługuje model prognozowania krótkookresowego energii wiatrowej opracowany przez Shao & Deng w roku 2016[18]. Wykorzystano dane empiryczne z farmy wiatrowej w prowincji Jiangsu w Chinach. W miarę użytkowania tych farm coraz istotniejsze staje się odpowiednie harmonogramowanie konserwacji turbin przy zastosowaniu m.in. sztucznej sieci neuronowej (ANN) i regresji wektora nośnego (SVR).

Systematyczne szacowanie zestawu modeli oceny ryzyka kredytowego

Wystepowanie okresowych kryzysów gospodarczych mobilizuje instytucje finansowe do poszukiwania adekwatnych modeli oceny ich odporności na niekorzystne zmiany rynkowe. Głębsze zainteresowanie budzi przełożenie danych makroekonomicznych na prawdopodobieństwo niewykonania zobowiązania kredytowego. Występuje wiele modeli regresji szacujacych ryzyko kredytowe w określonym środowisku społecznym[19]. Przeprowadzone porównanie modeli na danych z obserwacji wskazuje na zalety modeli uczenia maszynowego i kombinacji prognoz. Testy

[17] Sa'ad A., Nyoungue A., Hajej Z., *Analiza porównawcza produkcji energii wiatrowej. Planowanie oparte na sztucznej sieci neuronowej i podejście teoretyczne*, Materiały z Międzynarodowej Konferencji Inżynierii Przemysłowej i Zarządzania Operacyjnego, Stambuł, marzec 2022.
[18] Ibidem.
[19] Guth M., *Przewidywanie domyślnych prawdopodobieństw dla testów warunków skrajnych. Porównanie modeli*, artykuł internetowy, luty 2022.

warunków skrajnych sa obecnie wykorzystywane przez instytucje finansowe. Wymagają one jednak prowadzenia wieloletniego scenariusza makrofinansowego kraju i jego przełożenia na poziom określonego banku, przy czym modele ryzyka kredytowego sa następujące:

- liniowe,
- wektorowej autoregresji (VAR),
- panelowe,
- czynników latentnych,
- regresji kwantylowej,
- uczenia maszynowego.

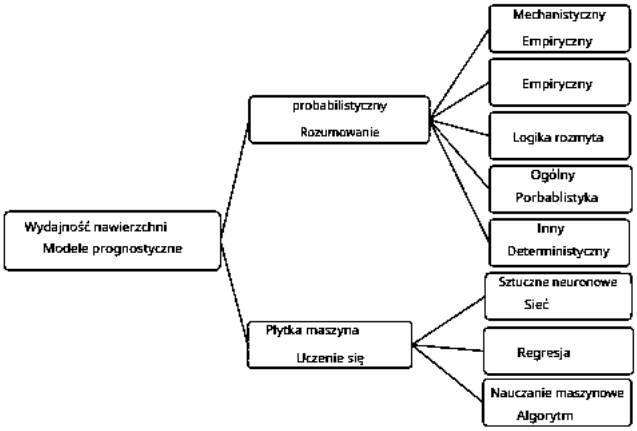
Obecnie w bankach stosowany jest zintegrowany nowoczesny model satelitarny do translacji wyładowań niezupełnych, wykorzystujacy uśrednienie modelu bayesowskiego (BMA) do generowania scenariuszy warunkowych prognoz ryzyka kredytowego. Model ten działa na puli modeli konkurencyjnych, które są ważone na podstawie ich przydatności predykcyjnej. Opracowane nowe modele ryzyka kredytowego wykorzystują języki programowania takie jak *R* lub *Python*. Jakość modelu uzależniona jest jednak od zestawu zmiennych makroekonomicznych dostarczanych instytucjom finansowym przez ERRS na potrzeby testu warunków skrajnych dla ogólnounijnego sektora bankowego. We wspomnianym w przypisie artykule dowiadujemy się, że wdrożono 43 modele z podziałem na 9 kategorii, poczawszy od tradycyjnych modeli statystycznych aż po nowoczesne modele uczenia maszynowego. W tym względzie skorzystano zarówno z modeli liniowych jak i nieliniowych dostępnych w bibliotece języka *R*.

Szacowanie kondycji nawierzchni dróg szosowych

Sięgnijmy teraz po systemy zarzadzania nawierzchniami (PMS), służące do monitorowania stanu nawierzchni i konserwacji[20]. Przeprowadzona analiza określonego odcinka drogi szosowej jest istotnym elementem tych systemów. Stosowane w tych aplikacjach modele predykcyjne służą przede wszystkim do oceny przyszłego zachowania się nawierzchni drogowej i wskazania adekwatnych technik jej konserwacji. W tym względzie ma zastosowanie posługiwanie się wskaźnikami jakości nawierzchni drogi.

W opracowaniu dokonano przeglądu modeli degradacji nawierzchni wskazując na ich mocne i słabe strony. Przeprowadzone rozpoznanie i ocena metod wskazują, że większość wcześniejszych badań preferowała podejścia do uczenia maszynowego i sztuczne sieci neuronowe, ze względu na ich zdolność radzenia sobie z dużą ilością danych. Podstawowy podział modeli szeregów czasowych stosowanych w zakresie wydajności nawierzchni drogowych zaprezentowano na rysunku 2.9.

[20] Sztajat A., Moridpour S., Best B., Rumi S., *Przegląd modeli przewidywania degradacji nawierzchni, Journal of Advanced Transportation*, tom 2022, Wiley Hindawi, Uniwersytet RMIT, Melbourne, Australia.



Źródło: Sztajat A., Moridpour S., Best B., Rumi S., *Przegląd modeli przewidywania degradacji nawierzchni*, *Journal of Advanced Transportation*, tom 2022, Wiley Hindawi, Uniwersytet RMIT, Melbourne, Australia, rys. 2.

Rys. 2.9. Modele szeregów czasowych w przewidywaniu wydajności nawierzchni drogowych

Prognozowanie wielookresowe zmiennych makroekonomicznych

W wyznaczaniu prognozy wielookresowej stosuje się metodę prognozy bezpośredniej i metodę prognozy iterowanej stosowaną, gdy model jest poprawnie określony[21]. W ocenie przyszłych zmiennych makroekonomicznych np. PKB stosowane są modele wektorowej autoregresji (VAR), lecz tu występuje problem ich parametryzacji. Dysponując techniką cyfrową możemy przeprowadzać symulacje prognoz, obserwując średni kwadratowy błąd prognozy. Możemy porównać jakość wyników uzyskanych według dwóch rodzai wcześniej wymienionych metod. W cytowanej publikacji ograniczono się do trzech zmiennych VA amerykańskiej makroekonomii, formułując algorytm postepowania w celu sporządzenia prognozy, przy czym rozwiązania dotyczą: iterowanych i bezpośrednich prognoz wielookresowych dla modelu VAR, bayesowskie VAR, symulacje Monte Carlo.

Przewidywanie kosztów budowy

Dla przewidywania kosztów budowy zastosowano odpowiednie połączenie liniowych modeli klasy ARIMA i nieliniowych typu ANN. W ten sposób powstał hybrydowy model ARIMA-ANN, który może być wykorzystany do planowania budżetu i przy licytacji kontraktów budowli zarówno na poziomie krajowym jak i miejskich[22]. W wyniku symulacji na trzech abstrakcyjnych projektach budowlanych z użyciem odrębnie modelu ARIMA, ANN oraz wspomnianego modelu hybrydowego okazało się, że to rozwiązanie ARIMA-ANN zapewnia dokładniejsze prognozy. W kreowaniu i testowaniu modelu hybrydowego uwzględniono liniową autoregresyjną średnia kroczącą oraz nieliniowe sztuczne sieci neuronowe.

[21] Sugita K., *Prognozowanie z użyciem modeli autoregresyjnych wektorów bayesowych: porównanie bezpośrednich i iterowanych metod wieloetapowych Asian Journal of Economics and Banking*, 2022, tom. 6, nr 2, strony: 142-154.

[22] Praca zbiorowa, *Poprawa dokładności w przewidywaniu wskaźników kosztów budowy na poziomie miasta dzięki połączeniu liniowej ARIMA i nieliniowej ANNS, Journal of Management in Engineering*, Uniwersytet Teksasu w Arlington, 11.2021.

3. Potrzeba przewidywania migracji ludności

3.1. Wstęp

Dokonano analizy dostępnej literatury zwartej, w formie plików *pdf*, w Internecie pod kątem modeli stosowanych w badaniach zjawisk ludnościowych. Okazało się, że publikacji podobnych do tematu „*Zjawisko i modele ilościowe badania migracji*" jest duża różnorodność i dotyczą one zarówno naszego kraju, Europy jak też odniesione są do innych obszarów naszej Ziemi. Globalizacja biznesu, łatwość przemieszczania się w poszukiwaniu pracy oraz występująca tendencja w rozluźnianiu więzi rodzinnych, zwłaszcza wielopokoleniowych spowodowała nie integrowanie się osób w wieku produkcyjnym ze swoim regionem. Obserwowane jest w wielu krajach wysoko rozwiniętych starzenie się populacji społeczeństw i występuje brak zastępowalności pokoleniowej. Powoduje to otwieranie się niektórych krajów na dopływ taniej siły roboczej, zwłaszcza z krajów niepokoju, czy też wysokiego bezrobocia.

Młode pokolenie Afryki, wzorem swoich starszych przyjaciół powracających czasowo z Europy Zachodniej, Ameryki - nie chce dalej tkwić w lokalnej nędzy. Za wszelka cenę chce się przedostać do *krajów-rajów*, aby ich dzieci mogły żyć dostatnio. Stąd obserwowany masowy *exodus* ludności z krajów Afryki przeciskających się do bogatych państw północy. Europa jednak broni się przed lawiną uchodźców, czy też migrujących w celach zarobkowych. Tworzone są specjalne strefy zaporowe straży przybrzeżnej na Morzu Śródziemnym, a niektóre kraje budują wysokie mury chroniące przed niekontrolowanym napływem cudzoziemców. Trzeba przyznać, ze część z napływającej ludności jest potrzebna do sfery produkcji i usług, ale nie nadmiar. Rodzi się potrzeba adaptacji napływających grup społecznych do nowych środowisk. Jednak wiele krajów nie rodzi sobie z tym problemem, a nawet Unia Europejska jako całość nie jest wystarczająco skuteczna w swoich działaniach.

Było to dla mnie impulsem do zebrania informacji do opisu zjawiska w świetle dostępnej literatury sygnalizowanej w Internecie zwłaszcza dotyczących modeli, przemieszczania się ludności, zmiany ich struktury wiekowej w społeczeństwach, adaptacji do współżycia w społeczeństwach wielokulturowych i różniących się religiami. W poszukiwaniu pozycji literaturowych posłużyłem się Biblioteką Internetową PWN, a procedura postepowania dotycząca dostępności do informacji o książkach z tego zakresu stanowi dalszy ciąg niniejszej pracy. Jednak poruszanie się po problematyce zjawisk ludnościowych objętych demografia wymaga przynajmniej wcześniejszego poznania obszaru tej nauki i rozróżniania jej dyscyplin.

Demografia (*grec. demos - lud, grapho - piszę*) jest dziedziną nauki zajmująca się powstawaniem, życiem i przemijaniem społeczności ludzkiej[23]. Zajmuje się ona przyrostem naturalnym, migracjami, strukturą społeczną (wiek, płeć, przynależność zawodowa, narodowość, wyznanie). Interesuje ją rozmieszczenie przestrzenne ludności oraz jej oddziaływanie społeczne i socjologiczne. Według Jerzego Z. Holzera demografia to nauka zajmująca się badaniem prawidłowości zjawisk ludnościowych w konkretnych warunkach społecznych i gospodarczych danego terytorium.

Nad kwestią „*optimum*" ludności, czyli najkorzystniejszą z punktu widzenia władzy i państwa, relacją pomiędzy liczbą ludności a wielkością danego obszaru zastanawiali się już w

[23] https://pl.wikipedia.org/wiki/Demografia.

starożytności Konfucjusz, Platon i Arystoteles. Pierwszym przykładem systematycznej analizy demograficznej jest praca J. Graunt'a *Natural and Polictical Observations* [...] *upon the Bills of Mortality*, która ukazała się w 1662 roku. Autor ten ustalił szereg prawidłowości demograficznych na podstawie których stworzył tablice trwania życia. Jednak jedną z najbardziej wpływowych teorii w myśli demograficznej aż do połowy XIX wieku była teoria T.R. Malthusa. W rozprawie *Prawo ludności* Malthus przedstawił pogląd iż "*liczba ludności wzrasta w postępie geometrycznym (co 25 lat), a produkcja środków żywności w postępie arytmetycznym*"[24].

Relacje pomiędzy zjawiskami ekonomicznymi i ludnościowymi są przedmiotem badań demografii ekonomicznej. Bezpośrednio z demografią połączona jest statystyka, która dostarcza ogólne metody pomiaru procesów masowych i jest niezbędna przy ich analizie. Ścisły związek z demografią ma socjometria, która służy do pomiaru pozycji poszczególnych jednostek w grupie i określenie stopnia jej ekspansywności, spoistości oraz zwartości. W demografii wyodrębnia się podział na następujące podstawowe dyscypliny[25]:

- *demografia formalna*, a jej istotą jest techniczny pomiar struktur i procesów ludnościowych oraz ich formalna analiza;
- *demografia opisowa*, która zajmuje się analizą struktur i procesów ludnościowych;
- *demografia potencjalna*, czyli system pomiaru i oceny procesów demograficznych, który przez nadanie wagi (wartością wyrażona długością dalszego przeciętnego trwania życia), zależnie od płci, wieku i innych cech członka, analizuje potencjał roboczy.

Teoria ludności Malthusa[26], z czasem określana jako *maltuzjanizm* to określenie koncepcji, a także ideologii i ruchów społecznych nawiązujących do teorii ludnościowej, dopracowanej przez Th.R. Malthusa. Teoria opiera się między innymi na tezie, że gatunek ludzki, mając stałą zdolność rozrodczą, wykazuje niepohamowaną skłonność do rozmnażania się (wzrostu liczebnego). Jak już wspomniałem Malthus zauważył, że liczba ludności, jeżeli nie napotyka przeszkód, powiększa się zgodnie z postępem geometrycznym; ściśle, rośnie w tempie 3% rocznie i podwaja się co 25 lat. Ponieważ podaż żywności w najlepszym razie rośnie zgodnie z postępem arytmetycznym, a zatem znacznie wolniej, to z upływem czasu (po około 50. latach od hipotetycznego momentu zapoczątkowania tej rozbieżności) nieunikniona staje się dysproporcja między przyrostem ludności a produkcją żywności; dochodzi wówczas do nasilającego się spadku produkcji rolnictwa na jednego mieszkańca. Zalecał „*ograniczenia prewencyjne*", których istotą było zmniejszenie potencjalnej płodności przez powstrzymanie się od założenia rodziny, późne wstąpienie w związek małżeński, abstynencję seksualną w przypadku już zawartych małżeństw, zwłaszcza przez osoby nie mające odpowiednich środków do życia.

Malthus był również zwolennikiem utrzymania płac na niskim poziomie, zakładał bowiem, iż ich wzrost pociąga za sobą zwiększenie rozrodczości i przez to nacisk na ograniczoną podaż żywności w okresie życia tego badacza. Podstawowym źródłem niedostatku żywności, a także przyczyną nędzy i bezrobocia jest bowiem, według Malthusa, rozrodczość tych, którzy nie dysponują odpowiednimi środkami utrzymania, czyli ubogich[27]. Trzeba dodać, że Malthus wnioski swoje wysunął na podstawie obserwacji w swoim czasie głównie proletariatu miejskiego, a ponadto nie przewidywał roli postępu technicznego we wzroście przychodów z ziemi i wydajności pracy. W wyniku późniejszych badań okazało się także, że wzrost liczby ludności nie dokonuje się w stałym tempie 3% rocznie.

[24] Ibidem.
[25] Encyklopedia zarządzania https://mfiles.pl/pl/index.php/Demografia.
[26] https://encyklopedia.pwn.pl/haslo/maltuzjanizm;3936876.html.
[27] Ibidem.

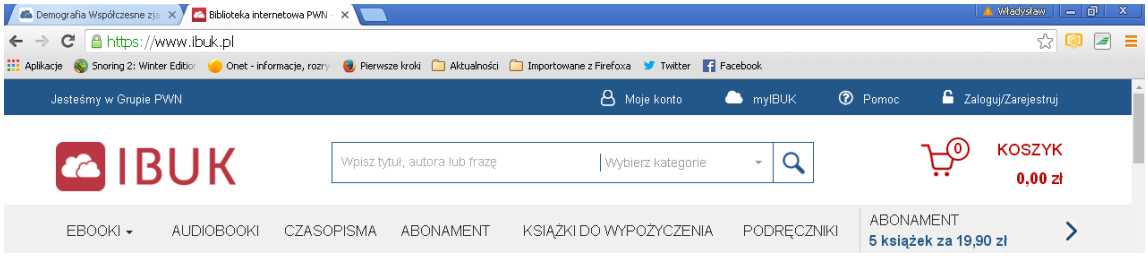
3.2. Korzystanie ze współczesnej biblioteki

Internet jako szeroka platforma dostępności do dość rozproszonej, nie w pełni udokumentowanej wiedzy z różnych obszarów nauki i praktyki jest już powszechnie uznawanym medium w zakresie zakupów, czy też zdobycia szybko krótkiej informacji. Opracowanie jednak monografii na określony temat wskazuje na dość chaotyczny i rozległy zakres dostępnych informacji, nie zweryfikowanej przez rzeczoznawców. Jednak teraz, przy prawie pustych w nowoczesne wydania książek oraz czasopism bibliotek tradycyjnych, korzystanie z jego zasobów staje się koniecznością. Coraz rzadziej występują bowiem ze względów ekonomicznych wydania papierowe. Wiele wydawnictw uczelnianych dla obniżenia kosztów publikacji drukuje tylko kilka egzemplarzy książek, skryptów, czy też monografii dla autora i wydawców. Zazwyczaj opracowania w formie plików elektronicznych formatu *pdf* umieszczają one na swoich stronach WWW lub blogach wielostronicowych.

PDF (*Portable Document Format*) jest to format plików służący do prezentacji, przenoszenia i drukowania treści tekstowo-graficznych, stworzony i promowany przez firmę Adobe Systems[28]. Dodam, że język opisu pliku PDF jest okrojoną wersją języka programowania *PostScript* wzbogaconą o elementy hipertekstowe.

Oferowane są obecnie przez księgarnie internetowe dostępy strumieniowe do publikacji w formie dzierżawy okresowej. Można też skorzystać z przesłanej w wersji elektronicznej książki do czytania na laptopach, tabletach, czytnikach, czy też popularnych obecnie smartfonach. Zamieszczona dalej przykładowa procedura dostępu do zasobów *Biblioteki internetowej PWN* wskazuje na różne, bardzo szybkie formy dostępu do wskazanej publikacji, niestety płatne w zależności od wymaganego zakresu stron przez internautę. I tak zamówienia książek w formacie *pdf* przez Internet przykładowo korzystając ze strony (*https://www.ibuk.pl*) sprowadza się zasadniczo do sześciu kroków postepowania[29]:

1. Wywołanie strony Biblioteki internetowej PWN.



Zbiór biblioteczny został podzielony na ebooki, audiobooki, czasopisma, abonament za udostępnianie pliku, książka do wypożyczenia, podręczniki i inne formy opracowań i dostępu. Możemy jednak wskazać kategorię, czy też wpisać tytuł publikacji i podać autora. Na nasze zamówienie czeka już koszyk sklepu internetowego.

Ebook to internetowa książka; publikacja elektroniczna, będąca zbiorem wiedzy na dany temat. Ebooki łączą tekst z materiałami multimedialnymi w postaci obrazów, filmów, infografik czy podcastów. Cechują się bogatą szatą graficzną, która obok tematu i wartości merytorycznej ma pozyskiwać uwagę czytelników[30]. Termin *podcast* (*podcasting*) określa

[28] https://pl.wikipedia.org/wiki/Portable_Document_Format.
[29] www.ibuk.pl, zamieszczone widoki ekranów są opracowaniem własnym autora wykonanym na podstawie wertowania tej strony.
[30] https://www.semtec.pl/slownik-seo/ebook/.

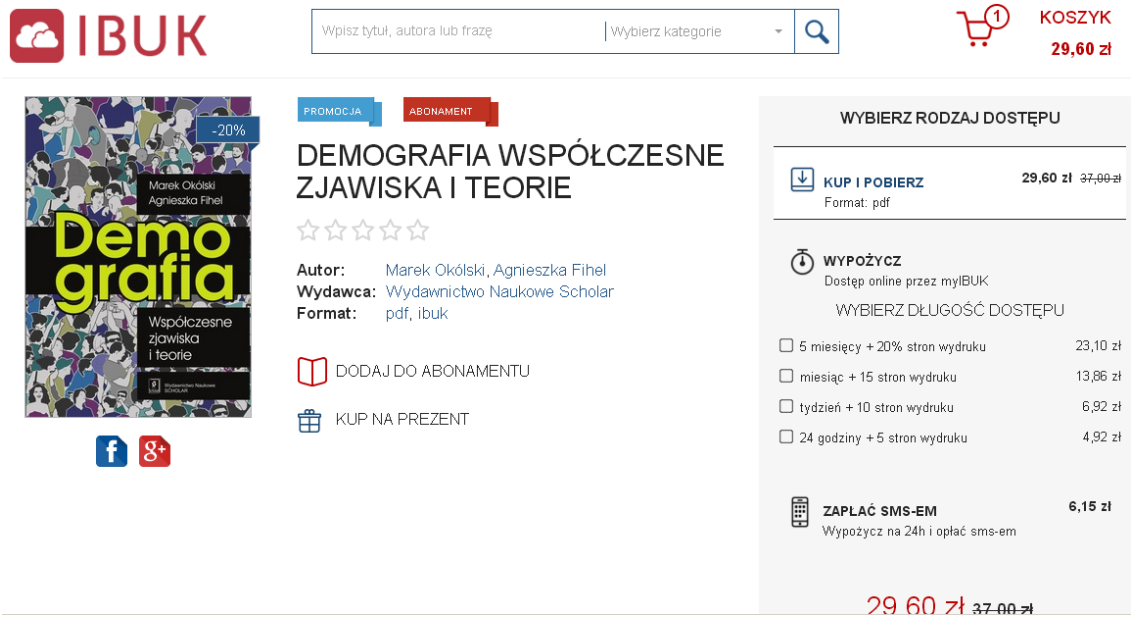
formę internetowej publikacji dźwiękowej lub filmowej, najczęściej w postaci regularnych odcinków[31].

Audiobook, czyli książka mówiona, stanowi nagranie dźwiękowe zawierające odczytany przez lektora tekst publikacji książkowej; zapisany jest na nośniku danych najczęściej w postaci cyfrowej w formatach np. MP3, M4B[32].

2. Po wpisanie tytułu książki, która może być w formie podręcznika, ebooka lub przesłana strumieniowo w formacie *pdf* klikamy na mały koszyk z ceną pozycji książkowej pod obrazem okładki książki.

Po tej operacji jedna pozycja wskazanej książki z ceną pojawi się w naszym koszyku zakupów.

3. W kolejności naszej procedury klikamy na okładkę książki i pojawi nam się dalsze uszczegółowienie do naszego zamówienia żądające wskazania formy przesłania pliku.



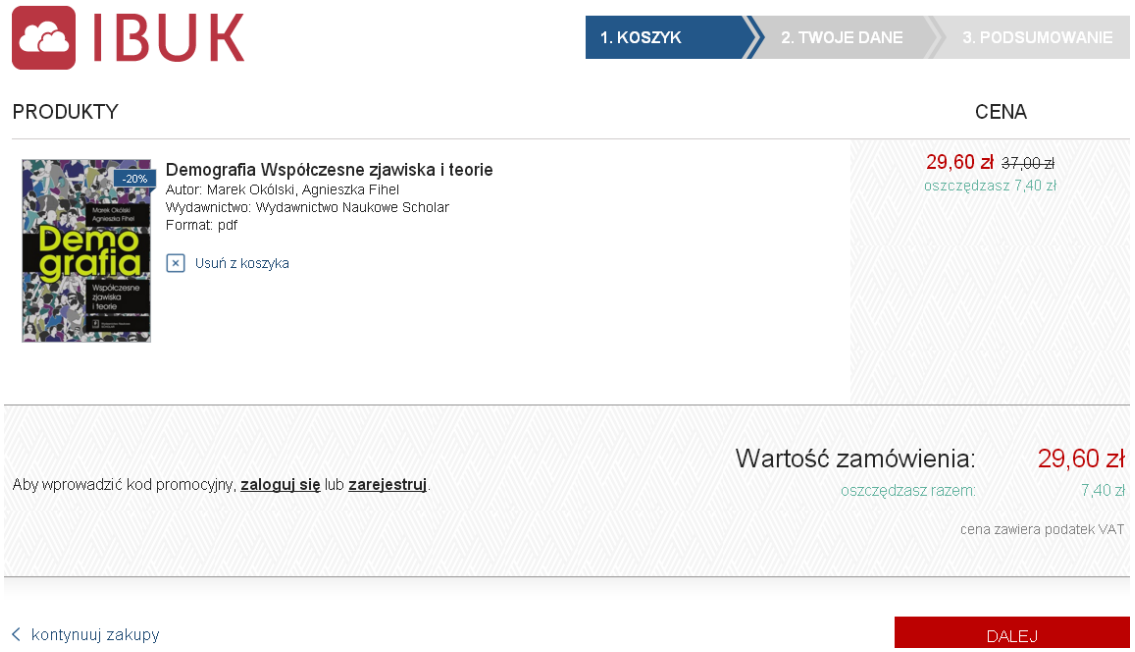
Możemy wybrać *pdf* lub *ibuk*, który umożliwia czytanie jako książka elektroniczna z dostępem online przez aplikację *myIBUK* na wcześniej już wspomnianych urządzeniach. "*myIBUK*"

[31] https://pl.wikipedia.org/wiki/Podcasting.
[32] https://pl.wikipedia.org/wiki/Ksi%C4%85%C5%BCka_m%C3%B3wiona.

to aplikacja do której potrzeba tylko przeglądarki internetowej np. *Chrome* i dostępu do Internetu. Aplikacja umożliwia czytanie wypożyczonych ebooków. Istnieje możliwość przeszukiwania treści, a ponadto powiększania lub zmniejszania tekstu. Czytnik zapamiętuje również stronę na której zakończono czytanie[33].

Istnieje możliwość wyboru czasu dostępu. W sytuacji wyboru czasowego online musimy przed emisją strumienia określonej liczby stron zapłacić SMS-em za wypożyczenie. W tym przykładzie interesuje nas jednak pobranie pełnej publikacji - pliku w formacie *pdf*, klikamy więc na KOSZYK z ceną 29,60. Warto zauważyć, że z wyborem czasu dostępu wiąże się zakres stron udostępnionej publikacji (5, 10, 15, 25% całego wydruku).

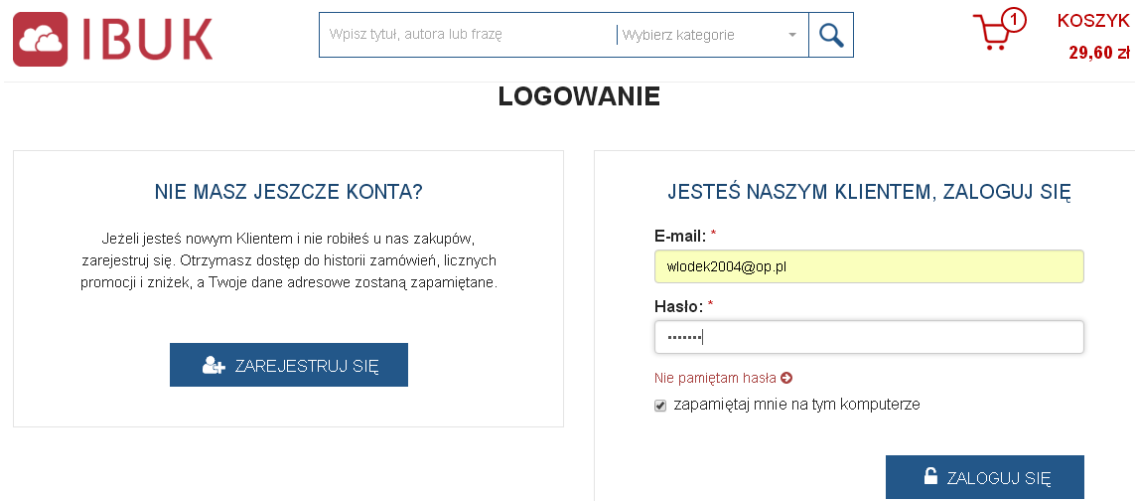
4. Możemy teraz sprawdzić swoje dane, posumowanie zamówień lub przystąpić do realizacji zakupu w koszyku. Zamówiliśmy tylko jedną książkę „*Demografia Współczesne zjawiska i teorie*". Sprawdzamy wartość zamówienia oraz naciskamy DALEJ.



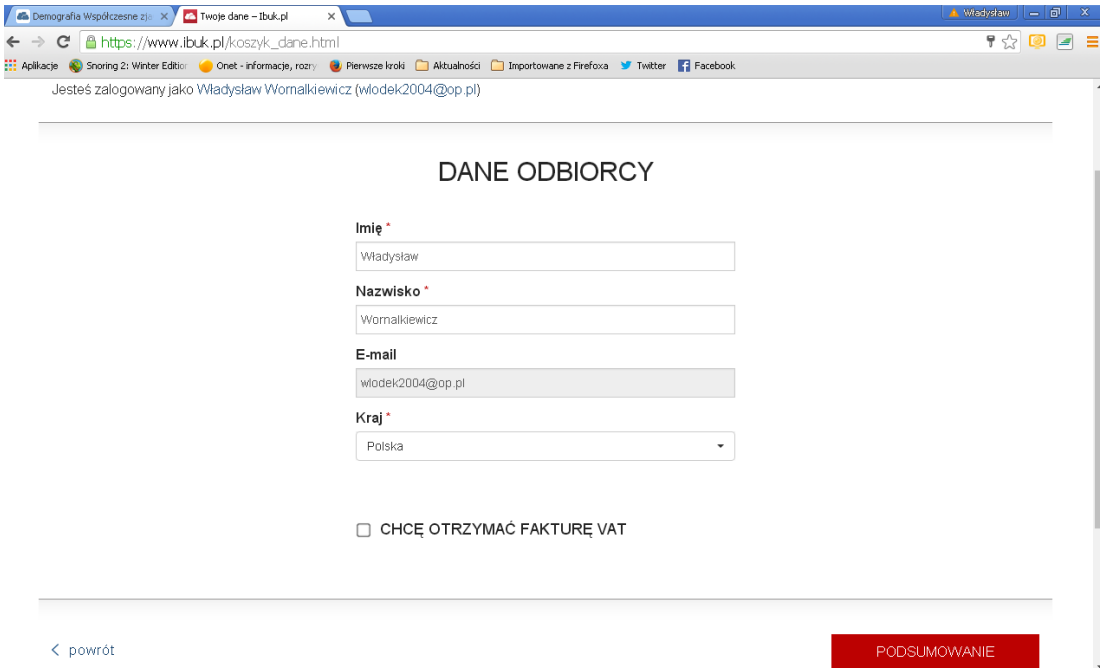
Zamówiliśmy więc pełny wydruk książki w formacie *pdf* za cenę 29,60 zł.

5. Jeśli nie mamy jeszcze konta klienta sklepu internetowego to musimy się zarejestrować i zaakceptować warunki sprzedaży. Gdy już byliśmy zarejestrowani to tylko logujemy się do sklepu internetowego podając E-mail i hasło, a następnie naciskamy ZALOGUJ SIĘ.

[33] https://www.ibuk.pl/myibuk.html.



6. Po tej operacji system informuje nas o zalogowaniu oraz wyświetla dane odbiorcy, przy czym możemy zażądać fakturę VAT. Ponieważ nasze działanie jest tylko „*na niby*" nie naciskamy PODSUMOWANIE, ale „*powrót*". Naciśnięcie bowiem PODSUMOWANIE skutkuje przesłaniem strumieniowym pliku pdf zamówionej pozycji książkowej i powoduje konieczność zapłaty za przesłaną książkę w formie elektronicznej.



W dalszych podrozdziałach niniejszego opracowania zabazowano na grafice i krótkich treściach ukazujących się po wybraniu określonej pozycji z *Biblioteki internetowej PWN*[34]. Ponadto niektóre mniej znane określenia zostały zdefiniowane przy wykorzystaniu dostępnych encyklopedii internetowych. Wybrane pozycje książkowe posortowano tematycznie następująco:

Demografia - migracje,
Pogoń za szczęściem,
Adaptacje środowiskowe polskich emigrantów,
Uchodźctwo,
Starzenie się populacji,
Modelowanie procesów demograficznych,
Skutki migracji.

[34] https://ksiegarnia.pwn.pl/.

3.3. Demografia - migracje[35]

Demografia Współczesne zjawiska i teorie. W przykładowej informacji internetowej w Onecie w dniu 28.07.2018 roku wystąpiła notatka pt. „*Do Hiszpanii przypłynęła rekordowa ilość migrantów w ciągu jednej doby*”[36], w której czytamy: „*Tylko od wczoraj do dzisiejszego poranka do Hiszpanii przypłynęło ponad 1200 nielegalnych migrantów. Łodziami do brzegów samej Andaluzji dotarło w ciągu doby 1000 osób z Maroka i był to rekordowy dzień pod względem napływu migrantów drogą wodną. Hiszpania jest obecnie głównym kierunkiem napływu migrantów z Afryki do Europy. Dotychczas największa w historii grupa migrantów dotarła morzem do Hiszpanii kontynentalnej w ciągu jednego dnia 20 lipca - było to 450 osób. Przeważali w niej obywatele Maroka i Algierii oraz państw Afryki Subsaharyjskiej. Jak poinformowała hiszpańska straż wybrzeża, wczoraj dopłynęło do południowej Hiszpanii łącznie 1000 nielegalnych migrantów z Afryki; pokonali oni Cieśninę Gibraltarską na około 50. łodziach. Dzisiaj nad ranem służby poinformowały o przybyciu kolejnych grup migrantów, w sumie ponad 200 osób*.”

Podałem tą informację wymienionego dnia dla zaprezentowania wagi problemów współczesnej migracji, której poświęcona jest książka „*Demografia Współczesne zjawiska i teorie*”. Udostępniana jest ona jako pełny wydruk w formacie *pdf* lub też w formie elektronicznego wypożyczenia wybranego zakresu stron (*ebook*)[37]. W książce przedstawiono stan współczesny demografii oraz podstawowe zagadnienia ludnościowe świata i różnych jego części. Zarysowano główne zjawiska, z jakimi mierzy się dziś demografia a mianowicie:

- wzrost zaludnienia Ziemi,
- wydłużenie się życia ludzkiego,
- słabnięcie rozrodczości,
- starzenie się populacji,
- nasilenie się mobilności terytorialnej.

Zaobserwowano liczne, przebiegające różnorako przestrzennie lub kulturowo zróżnicowanie wymienionych wcześniej zjawisk[38], przy czym w kolejnych rozdziałach książki przedstawiono:

- podstawowe kategorie dotyczące dyscypliny naukowej jaką jest demografia;
- stosowane metody analizy demograficznej, siatkę demograficzną, tablice trwania życia, techniki mikrodemografii, modele matematyczne;
- podstawowe zjawiska, takie jak reprodukcja ludności, zmiana stanu cywilnego, mobilność przestrzenna, a także migracje (ruch wędrówkowy);
- zróżnicowanie demograficzne na świecie;
- uchodźctwo;
- zaprezentowano główne nurty w demografii, przy czym trzeba tu wymienić: teorie dotyczące ludności, zaludnienia, związku partnerskiego, płodności, umieralności, długowieczności, migracji;
- wskazano na podstawowe cechy polityki ludnościowej ze wskazaniem na: rodzinę, zdrowie, migrację;
- wymieniono wybrane techniki analizy demograficznej, a w tym współczynniki i tablice eliminacji.

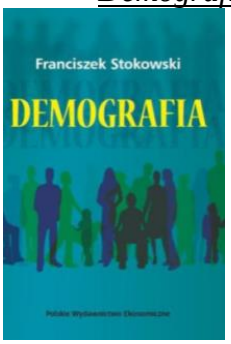
Wymieniona tu *siatka demograficzna,* czyli *diagram Lexisa* jest podstawowym graficznym instrumentem analizy zjawisk demograficznych, obserwowanych w określonej populacji/subpopulacji w danym przedziale czasu oraz zjawisk obserwowanych w określonej kohorcie[39]. Obrazem tej siatki jest zbiór jednakowo od siebie oddalonych prostych prostopadłych do osi współrzędnych (tzw. linii wieku i linii obserwacji), z których oś rzędnych (x) odzwierciedla rzeczywiste zaawansowanie (wiek) danego zjawiska demograficznego (np. lata życia,

[35] https://ksiegarnia.pwn.pl/Ksiazki/Demografia-Migracje-Statystyka,2195715,k.html?page=2.
[36] https://wiadomosci.onet.pl/swiat/hiszpania-przyplynelo-1-200-migrantow-rekord/pyhcpkr.
[37] Okólski M., Fihel A., *Demografia Współczesne zjawiska i teorie*, Wydawnictwo Scholar, 2017, *pdf*.
[38] Okólski M., Fihel A., *Demografia Współczesne zjawiska i teorie*, op. cit., streszczenie.
[39] https://encyklopedia.pwn.pl/haslo/siatka-demograficzna;3974755.html.

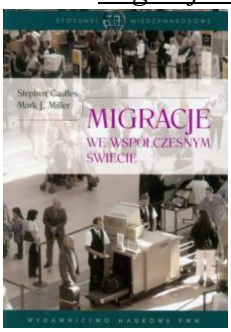
długość stażu małżeńskiego, czas pozostawania w stanie bezdzietnym itp.), a oś odciętych (*t*) - czas kalendarzowy, w jakim jest rejestrowany (obserwowany) stan owego zjawiska.

Kohorta, stosowana jest między innymi w demografii, oznacza zbiór ludzi, wyodrębniony z populacji z uwagi na zachodzące jednocześnie dla całego zbioru wydarzenie lub proces w celu przeprowadzenia analizy[40]. Kohorta powinna być wyodrębniona na podstawie istotnych statystycznie cech i jednorodna pod ich względem. Kohorta demograficzna grupuje np. osoby urodzone w danym roku, rozpoczynające w określonym roku edukację szkolną lub studia. Kohorty demograficzne często wykorzystywane są w badaniach porównawczych między przedstawicielami różnych pokoleń lub bada się też te same kohorty wiekowe po upływie określonego czasu.

Mobilność przestrzenna[41], wyraża ruchliwość społeczną i oznacza przesunięcia pozycji ludzi na skali, charakteryzującej ich cechy społeczne, takie jak przynależność do grupy społecznej, grupy zawodowej czy jednostki osiedleńczej. Mobilność przestrzenna obejmuje migracje i cyrkulację, czyli fizyczne przemieszczenia ludzi, jakim towarzyszy transfer z jednej jednostki (przestrzeni) społecznej do innej, przy czym podstawowa różnica między migracją a cyrkulacją polega na tym, że pierwsza z nich prowadzi do trwałej lub względnie trwałej zmiany miejsca zamieszkania, a druga - nie. Nadmienię, że tryb życia niektórych ludów czy grup populacji zakłada koczownictwo, czyli ciągłą mobilność przestrzenną. Próby odniesienia pojęcia „*mobilność przestrzenna*" do konkretnych zjawisk społecznych prowadzą zwykle do przyjęcia za podstawę jednostki terytorialnej, mającej określony status administracyjny, którą ludzie opuszczają celem udania się do innej[42]. Obecnie w nowoczesnym społeczeństwie obserwowane jest zwiększenie mobilności przestrzennej związanej z nauką czy też pracą.



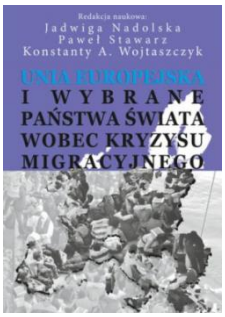
Demografia[43]. Przy podejmowaniu decyzji o charakterze ekonomicznym i społecznym niezbędna jest znajomość zjawisk demograficznych. W prezentowanej pozycji literatury spotykamy metody analizy zjawisk ludnościowych opracowane na bazie najnowszych dla autora danych liczbowych. Uzupełnienie danymi historycznymi umożliwiło pokazanie tendencji tych zmian w czasie. Dotyczy to Polski jak i ujęcia międzynarodowego. W opracowaniu oceniono prawidłowości w strukturach poszczególnych zjawisk ludnościowych. Jak już wspomniałem we *Wstępie* w szybkim tempie następuje integracja globalna nie tylko biznesu, a sprzyjają temu warunki komunikowania się poprzez Internet i media społecznościowe.



Migracje we współczesnym świecie[44]. Szczególnie lata 2016-2018 to niepokojący ruch ludności zarówno uchodźców z krajów toczących się wojen jak i porywu młodych z Azji i Afryki. Nierzadko występowało to całymi rodzinami w celu znalezienia się w świecie bogatej Zachodniej Europy, w tym w szczególności w Wielkiej Brytanii, Niemczech, Francji lub Szwecji. Publikacja „*Migracja we współczesnym świecie*" pokazuje nam typy zmian jakie występują obecnie w społeczeństwach na skutek międzynarodowej migracji ludności. Masowe migracje powodują zmiany dynamiki przepływu osób, kapitału oraz towarów. Coraz szybsze i sprawniejsze środki transportu, w tym możliwie tanie przeloty

[40] https://pl.wikipedia.org/wiki/Kohorta_(statystyka).
[41] https://encyklopedia.pwn.pl/haslo/mobilnosc-przestrzenna;3942435.html.
[42] Ibidem.
[43] Stokowski F., *Demografia*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2015.
[44] Castles S., Miller M.J., *Migracje we współczesnym świecie*, Wydawnictwo Naukowe PWN, Warszawa 2020.

samolotami, przyczyniają się także do sprawniejszej wymiany kulturalnej oraz na niwie sportowej. Obserwowana jest ucieczka kapitału do obszarów dostępnej taniej siły roboczej i bliskości pozyskiwania coraz droższych surowców. Można powiedzieć, że obecny początek XXI wieku to era migracji i to nie tylko ekonomicznej, ale także przykrej, wymuszonej przez los, jaki doświadczają uchodźcy próbujący się przedostać do Europy. Temu właśnie poświęcona jest dalej przedstawiona książka „*Uchodźcy w Europie*".



Unia Europejska i wybrane państwa świata wobec kryzysu migracyjnego[45]. Unia Europejska, chociaż coraz bardziej jednolita gospodarczo, w kwestii kryzysu migracyjnego nie jest zgodna. Kluczowym problemem jest podział uchodźców przebywających przejściowo w krajach basenu Morza Śródziemnego tj. Grecji, Włoszech, Francji. Kraje Europy Środkowej sprzeciwiły się narzucaniu liczby migrantów przez Instytucje Unii Europejskiej. Omawiana praca zbiorowa stanowi kompleksowy zbiór artykułów ukazujących zjawisko masowej migracji, w tym uchodźstwa do Europy, wywołanych tzw. arabską wiosną. Odpowiada ona na szereg pytań dotyczących strategii przyjętej przez państwa, instytucje wobec problemu emigracji. Zdaniem recenzenta Roberta Kłaczyńskiego „*…stanowi ważne źródło wiedzy, będąc zarazem rodzajem vademecum, które umożliwi badaczom problemu emigracji, a także innym czytelnikom poruszanie się po tym niezwykle aktualnym temacie*". Tyle już razy nadmieniłem słowo „*migracja*", trzeba więc przybliżyć to określenie: *migracja* [łac. *migratio* 'przesiedlenie'], *oznacza przemieszczenia terytorialne związane ze względnie trwałą zmianą miejsca zamieszkania*, przy czym migracje dzielą się na[46]:

- stałe (powodujące trwałą zmianę miejsca zamieszkania),
- czasowe (stanowiące sezonową lub okresową zmianę miejsca zamieszkania),
- wahadłowe (codzienne dojazdy z miejsca zamieszkania do miejsca pracy lub nauki).

Skorzystam dalej z tego źródła i zacytuję: „*Ze względu na odległość rozróżnia się migracje: wewnętrzne (w obrębie danego państwa - wewnątrzregionalne lub międzyregionalne), zewnętrzne (poza granice państwa - kontynentalne, międzykontynentalne). Biorąc pod uwagę organizację można wyróżnić migracje: żywiołowe, planowe (np. repatriację), legalne, nielegalne, dobrowolne, przymusowe (przesiedlenia, wysiedlenia, deportacje). W zależności od przyczyn migracje dzieli się na zarobkowe, rodzinne, narodowościowe, religijne, polityczne, rekreacyjne, turystyczne. … Napływ ludności na dane terytorium to imigracja, odpływ to emigracja. Powrót do dawnego miejsca zamieszkania (migracja powrotna), to reemigracja. Miary zjawisk migracji stosowane w demografii to: migracja brutto - suma imigracji i emigracji; saldo migracji (migracja netto)- różnica między napływem i odpływem ludności*". Do opisu migracji stosowane są modele:

- *grawitacyjne*, które przyjmują, że wielkość migracji jest funkcją liczby mieszkańców miejsc przeznaczenia i pochodzenia oraz odległości między nimi[47];
- *ekonomiczne*, które zakładają, że ludzie dostosowują się do zmian ekonomicznych, podążają więc za możliwością uzyskania pracy oraz wyższych wynagrodzeń.

Trzeba dodać, że współczesne migracje rozpoczęły się od masowego wychodźstwa ludności wiejskiej do miast. Zjawisko to zostało zainicjowane w Europy Zachodniej pod wpływem wielkich przemian społeczno-ekonomicznych, spowodowanych uprzemysłowieniem oraz urbanizacją. Stamtąd przeniosło się stopniowo na pozostałą część Europy, a także na inne kontynenty. Ocenia się,

[45] Nadolska J., Stawarz P., Wojtaszyk K.A., *Unia Europejska i wybrane państwa świata wobec kryzysu migracyjnego*, Aspra, 2017.
[46] https://encyklopedia.pwn.pl/haslo/migracje;3941121.html.
[47] http://coin.wne.uw.edu.pl/ggrotkowska/mse_msem/wyklad01.pdf.

że około roku 1800, zaledwie 3% światowej populacji mieszkało w miastach. Pomimo wyższego przyrostu naturalnego ludności wiejskiej, po 150 latach ten odsetek podniósł się do 29%, co świadczy o tym, że zmiana ta wynikała głównie z migracji. W końcu lat 70. XX w. mieszkańcy miast stanowili przeciętnie około 70% całej ich ludności [48]. Z obszernego opisu zjawiska migracji zawartego *w Internetowej Encyklopedii PWN* dowiadujemy się ponadto, ze:

a) Innym rodzajem masowych współczesnych migracji są międzynarodowe przepływy ludności o charakterze osiedleńczym. Do wybuchu II wojny światowej były one głównie udziałem mieszkańców Europy, którzy z reguły kierowali się na inne kontynenty, bowiem przesiedliło się około 60 mln osób).

b) W XX wieku udział ludności Ameryki Północnej w zaludnieniu Ziemi wzrósł z 0,7 do 5,1%. Nasiliły się migracje odbywane pod wpływem zaburzeń politycznych, szczególnie migracje uchodźców, w tym związane z wysiedleniami lub przesiedleniami na tle narodowym, etnicznym i plemiennym.

c) W roku 2000 do krajów Unii Europejskiej przybyło około 1 mln imigrantów, a do Ameryki Północnej 800 tysięcy, natomiast do Japonii ponad 250 tysięcy. W tym okresie głównymi terytoriami wychodźstwa były zaś Chiny i Meksyk.

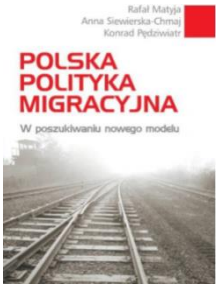
d) W okresie międzywojennym takie kraje jak Rosja i Hiszpania doświadczyły masowego odpływu ludności na tle politycznym. Szczególnie nasilone było też uchodźstwo Żydów z Niemiec. Zjawisko migracji wywołanych czynnikami politycznymi sięgnęło zenitu po wybuchu II wojny światowej.

e) Od lat sześćdziesiątych dwudziestego wieku liczba uchodźców systematycznie wzrastała i tak w II połowie lat 90. w Afryce stale rejestrowano 5,5-6,5 mln uchodźców, a w Azji około 5 mln. Czynnikiem obiektywnym tych migracji były często występujące klęski żywiołowe, wywołujące głód. Wielka skala migracji cechowała również międzynarodowe migracje pracownicze.

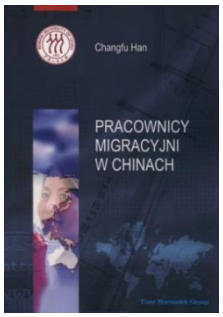
f) Przez kilka 10-leci po zakończeniu II wojny światowej napływ pracowników z zagranicy odgrywał istotną rolę w równoważeniu rynku pracy kilku krajów Europy Zachodniej, w których głęboki niedobór robotników przemysłowych wynikał z wysokiego tempa wzrostu gospodarczego. Znamiennym przypadkiem ilustrującym to zjawisko stały się Niemcy, które 1960-1971 podpisały umowy o zatrudnieniu z 9. krajami z Europy Południowej, Afryki Północnej, Turcją i Koreą. W szczytowym okresie (1971) do Niemiec napłynęło prawie 750 tysięcy nowych robotników z zagranicy. Cechami wyróżniającymi pracownicze migracje tego okresu były między innymi oprócz dużego nasilenia, znaczne geograficzne zróżnicowanie krajów pochodzenia migrantów oraz wysoki udział nieudokumentowanego ich zatrudnienia. Pięć najważniejszych krajów imigracji: USA, Rosja, Arabia Saudyjska, Indie i Kanada, uzyskało z tego tytułu łączny przyrost ludności około 31 mln, a pięć najważniejszych krajów emigracji: Meksyk, Bangladesz, Afganistan, Filipiny i Kazachstan, utraciło łącznie około 19 mln osób.

Migracje jako wyzwanie dla Unii Europejskiej i wybranych państw członkowskich[49]. Warto przeczytać tą książkę, mającą charakter monografii, gdyż ukazuje ona aktualny stan wiedzy w zakresie tematyki migracyjnej dotyczącej obszaru Unii Europejskiej (UE). Wiedza zawarta w tej monografii pozwala na usystematyzowanie wiedzy na temat mechanizmów wpływających na kształt i przebieg migracji. Pomocne to może być do badań naukowych związanych ze zdobywaniem i przekazywaniem wiedzy odnoszącej się do teoretycznych, a także praktycznych zagadnień migracji i bezpieczeństwa wewnętrznego UE.

[48] https://encyklopedia.pwn.pl/haslo/migracje;3941121.html.
[49] Cebul K., Fehler W., Podgórzanka R., *Migracje jako wyzwanie dla Unii Europejskiej i wybranych państw członkowskich*, Difin, Warszawa 2017.



Polska polityka migracyjna[50]. Wobec nasilenia się migracji zarówno wewnętrznej jak i zewnętrznej poszczególne kraje Europy starają się wprowadzić odpowiednie dla nich uregulowania prawne. Z tego względu interesująca może być książka „*Polska polityka migracyjna*". Stanowi ona podsumowanie działań związanych z procesami migracyjnymi w kontekście Polski i wybranych krajów UE. W pracy zaproponowano pewne ich modyfikacje. Warto zwrócić jeszcze uwagę na zawarty w opracowaniu materiał porównawczy, starannie zebrany na temat dotychczasowych ekspertyz, który może być pomocny dla pracowników zaangażowanych w problematykę migracyjną. W uzupełnieniu dodam, że w notatce internetowej z lipca 2018 roku na portalu Onet znajdowała się informacja dotycząca poczynań rządu Polski w zakresie walki z niedoborem pracowników[51]. Szacuje się, że nowa polityka migracyjna będzie kosztować 2,9 mld złotych w latach 2018-2025, przy czym z funduszy unijny aż 1,6 mld złotych. Przewiduje się specjalne zachęty dla studentów i wysoko wykwalifikowanych pracowników. Coraz trudniej jest bowiem ściągnąć pracowników z Ukrainy. Rozszerzono do 283 zawodów zwolnionych z tzw. testu rynku pracy.



Pracownicy migracyjni w Chinach[52]. Przenieśmy się teraz do Chin uznawanych do niedawna jako nieograniczoną podaż siły roboczej. Było to „*magnesem*" przyciągającym międzynarodowe korporacje z całego świata. Autor Chanfu Han uważa, że to one zakładając swoje bazy produkcyjne przekształciły Chiny w "*światową fabrykę*". Wywołało to masową migrację młodej ludności z obszarów wiejskich do aglomeracji miejskich, zasobnych w ośrodki produkcyjne. Jest jeszcze mimo wszystko tania siła robocza. Odbija się to jednak coraz bardziej na kondycji wiosek chińskich, w których następuje szybki proces starzenia się ludności.

3.4. Pogoń za szczęściem

Jak żyć szczęśliwie w innym kraju[53]. *Szczęście* jest emocją, spowodowaną doświadczeniami ocenianymi przez podmiot jako pozytywne[54]. Psychologia wydziela w pojęciu *szczęście* rozbawienie i zadowolenie. W rozważaniach o jego naturze szczęście najczęściej określane jest w dwu rozdzielnych aspektach: mieć szczęście, odczuwać szczęście. Migracja młodych obywateli danego kraju wiąże się z poszukiwaniem szybkiego „*dorobienia się*" i doznania szczęśliwości, w tym rodzinnej w kraju, który jest przybliżeniem tego marzenia. Zanikają stopniowo więzi z ojczyzną swoich rodziców, często i dziadków, a liczy się tylko sukces, bogactwo i to jeszcze w młodym wieku. Z ciekawości więc warto sięgnąć po książkę „*Jak żyć szczęśliwie w innym kraju*". Pojawia się tu zjawisko zwane *nuklearyzacją rodziny*[55], bowiem coraz bardziej powszechną rodzinę

[50] Pędziwiatr K., Siewierska-Chmaj A., Matyja R., *Polska polityka migracyjna W poszukiwaniu nowego modelu*, Wydawnictwo Uniwersytetu Warszawskiego, 2015.
[51] https://int.search.myway.com/search/GGmain.jhtml.
[52] Han C., *Pracownicy migracyjni w Chinach*, Adam Marszałek, Toruń 2017.
[53] Olkiewicz A., *Jak żyć szczęśliwie w innym kraju*, tłumacz: Inga Sawicka, Wydawnictwo: Czarna Owca, Warszawa 2014.
[54] https://pl.wikipedia.org/wiki/Szcz%C4%99%C5%9Bcie.
[55] http://www.wos.net.pl/rodzina.html.

nuklearną tworzą dwa pokolenia - rodzice i dzieci. Gdy dzieci usamodzielniają się pozostawiają rodziców samych. Powoduje to szczególnie dla państw Europy Zachodniej i Ameryki Północnej duże problemy socjalne, gdyż zachodzi konieczność opieki nad osobami starszymi i często schorowanymi. Dodam, że dziedziną medycyny zajmującą się osobami w wieku starczym jest geriatria[56]. Autor przeznaczył swoją książkę do ludzi planujących wyjazd lub mieszkających poza krajem ojczystym. Zastanawia się, co emigranci mogą uczynić, by zintegrować się ze społeczeństwem i nie utracić godności. Podpowiada, jak uniknąć rozczarowań w nowym środowisku społecznym. Sam przyznaje, że jako emigrant, z powodu wyniesionych z domu uprzedzeń, popełnił błędy w swoim działaniu. Chce ustrzec przybyszów do potencjalnie nowej ojczyzny przed problemami i radzi co można zrobić, by życie w nowym kraju nie było takie trudne. Nie zgadza się z opinią niektórych przybyszów, że całkowite dostosowanie się i integracja są niemożliwe.

W cieniu europejskiej twierdzy Obrazy Zachodu wśród Afrykanów Gwinea Bissau[57]. Jeśli mówimy już o szczęściu, to dla Afrykanów z Gwinei Bissau jest to dotarcie do Europy Zachodniej i zaadoptowanie się w jednym z „*rajów ziemskich*" np. w Wielkiej Brytanii. Z tego względu książka „*W cieniu europejskiej twierdzy Obrazy Zachodu wśród Afrykanów Gwinea Bissau*" jest przybliżeniem marzeń młodych Afrykanów o życiu w kraju nowoczesnym, spokojnym i łatwej egzystencji. Ze streszczenia tej książki dowiadujemy się: „…*autorka analizuje nastawienie mieszkańców Gwinei Bissau do europejskich wartości i norm kulturowych (takich jak indywidualizm, monogamia czy równouprawnienie)*…". Chcą oni przeniknąć unijną granicę wszelkimi sposobami. Zawarte w opracowaniu rozważania oparte są na badaniach etnograficznych autorki w latach 2003, 2010-2011 i 2016. Książka przybliża nam masowe zjawisko migracji Afrykanów i pozwala zrozumieć poświęcenie tych młodych ludzi, licznie przybywających dziś do Europy.

3.5. Adaptacje środowiskowe polskich imigrantów

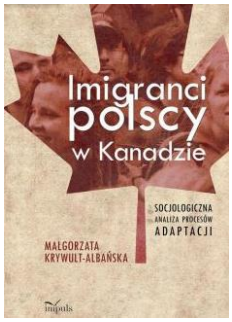
Współczesne polskie migracje[58]. Przyjrzyjmy się teraz zjawisku migracji, szczególnie do krajów zasobnych ekonomicznie, gdyż to jest bodźcem do badania tego procesu również w naszym kraju. Z tego powodu sięgnąłem po książkę „*Współczesne polskie migracje*". W prezentacji tej pracy czytamy "*Książka stanowi przegląd nowych i ważnych kierunków oraz wątków obecnych w badaniach migracyjnych w Polsce i na świecie, rozwijających się w ostatnich latach niezwykle dynamicznie. Zamieszczono w niej teksty syntetyczne, podsumowujące wieloletnie badania ogólnopolskie i regionalne, teksty podejmujące zagadnienia o szczególnym znaczeniu społecznym lub opisujące specyficzne problemy grup społecznych i społeczności polskich, w tym przebywających na obczyźnie od niedawna*". Na podstawie danych statystycznych łatwo zauważyć, że wyjazdy Polaków były szczególnie intensywne po wejściu Polski do Unii Europejskiej. Potem wystąpiły nieliczne powroty w okresie kryzysu gospodarczego, co miało wpływ na polskie społeczeństwo i gospodarkę. Migracja

[56] http://dspace.uni.lodz.pl/xmlui/bitstream/handle/11089/5353/11_Piotr_Szukalski_Demograficzne%20przemiany%20rodziny_169-186.pdf?sequence=1&isAllowed=y.
[57] Brzezińska M., *W cieniu europejskiej twierdzy Obrazy Zachodu wśród Afrykanów Gwinea Bissau*, Wydawnictwo Uniwersytetu Warszawskiego, 2017.
[58] *Współczesne polskie migracje Strategie - skutki społeczne - reakcja państwa*, Wydawnictwo Uniwersytetu Warszawskiego,

odbija się bowiem także na życiu rodzinnym wyjeżdzających za granicę. Według recenzji Jacka Schmidta omawiana książka jest zbiorem spojrzeń tematycznych różnych autorów - badaczy ilustrujących „*najnowsze kwestie skutków emigracji z Polski dla społeczności migrujących i ich rodzin, społeczności i miejsc, które opuścili, a także społeczności przyjmujących i miejsc nowego osiedlenia czy relacji wewnątrz grupy migrujących*".

Młodzi emigranci polscy pracujący w Wielkiej Brytanii[59]. Kolejną publikacją zajmującą się adaptacją zawodową emigrantów jest książka „*Młodzi emigranci polscy pracujący w Wielkiej Brytanii*". Jak wynika z podtytułu autorkę interesuje relacja między adaptacją zawodową młodych emigrantów polskich w Wielkiej Brytanii, a poczuciem ich szczęścia przebywania w kraju między innymi dobrej, zapewnionej opieki zdrowotnej.

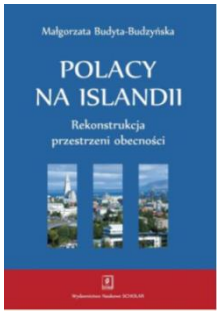
Rzeczywistość społeczna stawia człowieka wobec szeregu wyzwań życiowych, w tym zawodowych. Emigranci w stosunku do kraju swego pochodzenia - Polski spotykają się tu z potrzebą zwiększenia mobilności i elastyczności oraz otwartości na nowości. Ze streszczenia tej pracy dowiadujemy się, że: „*Książka ma na celu poszukiwanie odpowiedzi na pytanie o adaptację zawodową młodych polskich emigrantów w Wielkiej Brytanii, rozpatrywanym w kontekście poczucia szczęścia poza granicami kraju ojczystego. Poczucie szczęścia, satysfakcja i ogólne zadowolenie z życia są niezmiernie istotnym czynnikiem przetrwania oraz zaadaptowania się w warunkach migracyjnych*".



Imigranci polscy w Kanadzie[60]. W ofercie *Biblioteki Internetowej PWN* znajduje się szereg publikacji, które stanowią efekty prac badaczy, czy też dziennikarzy nad analizą pobytu i adaptacji polskich emigrantów w krajach ekonomicznie bardziej intratnych niż Polska. Spośród nich moim zdaniem na wyróżnienie zasługuje książka „*Imigranci polscy w Kanadzie*". Przedstawiono w niej polskich imigrantów z lat 80. XX wieku na tle poprzednich fal polskiej imigracji w Kanadzie. Emigracja z Polski do Kanady wymienionego okresu podobna była do migracji osiedleńczych z przełomu XIX i XX wieku. Polscy emigranci w Kanadzie stanowią zbiorowość wyodrębnioną pod względem społecznym i kulturowym, jednak dobrze zintegrowaną z rynkiem pracy. Żyjąc w nowym społeczeństwie polscy emigranci doznawali szeregu zmian dotyczących kultury, języka, wzorów zachowań, wartości norm etycznych, stosunków społecznych. Wystąpiły procesy integracji społecznej, polegające na likwidowaniu barier utrudniających lub uniemożliwiających nawiązywanie kontaktów między jednostkami lub grupami społecznymi[61]. Adaptację ułatwiała satysfakcja z warunków życia, ogólnie biorąc z decyzji o emigracji. Następowało stopniowe przyjmowanie obywatelstwa i osiedlanie się na stałe. Liczne jest polskie środowisko migracyjne w Kanadzie, a badanie jego dostarcza materiału do porównań z innymi skupiskami migrantów w świecie.

[59] Krawczyk R., *Młodzi emigranci polscy pracujący w Wielkiej Brytanii Adaptacja zawodowa a poczucie szczęścia*, Difin, Warszawa 2017.
[60] Krywult-Albańska M., *Imigranci polscy w Kanadzie Socjologiczna analiza procesów adaptacji*, Impuls, Kraków 2015.
[61] Ibidem.

Strategia - migracja - adaptacja[62]. Książka „*Strategia - migracja - adaptacja*" pokazuje strategie adaptacyjne rodzin młodych emigrantów polskich mieszkających w Londynie. Interesujące jest spojrzenie autorki z perspektywy rodziców i dzieci, które znalazły się w nowych warunkach społecznych. Nakładają się tu doświadczenia z Polski i te które wystąpiły w latach emigracji. Emigranci kierują się perspektywą przyszłościową i dostosowują do niej swoje działania w nowej „*ojczyźnie*".



Polacy na Islandii[63]. „*Polacy na Islandii*" to kolejna książka poruszająca problematykę polskich emigrantów, których coraz więcej w Islandii. Ta wyspa w XXI wieku stała się obiektem masowej turystyki, zwłaszcza ze Stanów Zjednoczonych. Ktoś jednak musi obsługiwać turystów, stąd podaż na pracowników między innymi sezonowych, również i z Polski. Autorka w swojej pracy stara się odpowiedzieć na szereg pytań, a mianowicie:

- Jak Polacy poradzili sobie w czasie kryzysu ekonomicznego na wyspie?
- Dlaczego wyspiarze do początków XX wieku realizowali agrarny model rozwoju, kompletnie nieprzystający do warunków przyrodniczych, który doprowadził do ubóstwa i wielokrotnie do klęsk głodu, a dopiero w XX wieku zdecydowali się oprzeć model gospodarczy na rybołówstwie?
- Czy pomysłem na rozwój w XXI wieku jest turystyka, która ma zastąpić marzenia z początku XXI w. o byciu centrum finansowym Europy?

W prezentowanej publikacji porównano sytuację polskiej diaspory na Islandii z sytuacja Polonii w innych krajach. Książkę można traktować również jako opowieść o oryginalności i odrębności wyspiarskiego kraju jakim jest Islandia.

3.6. Uchodźctwo

Uchodźstwo[64] - proces opuszczania stałego miejsca zamieszkania przez osoby zagrożone prześladowaniami, będącymi wynikiem konfliktów zbrojnych - zewnętrznych i wewnętrznych, polityki państwa. W myśl Konwencji Genewskiej (art. 1 A 3) za uchodźcę może być uznana osoba, która „*na skutek uzasadnionej obawy przed prześladowaniem z powodu swojej rasy, religii, narodowości, przynależności do określonej grupy społecznej lub z powodu przekonań politycznych przebywa poza granicami państwa, którego jest obywatelem, i nie może lub nie chce z powodu tych obaw korzystać z ochrony tego państwa, albo która nie ma żadnego obywatelstwa i znajdując się na skutek podobnych zdarzeń, poza państwem swojego dawnego stałego zamieszkania nie może lub nie chce z powodu tych obaw powrócić do tego państwa*". Uchodźcy przeprawiają się często z krajów Afryki czy też Azji, pokonując niebezpieczne trudy dobrnięcia np. łodziami do brzegów Grecji, Włoch czy też Hiszpanii.

[62] Winiecka K., *Strategia i migracja - adaptacja Rodziny młodych migrantów polskich w Londynie po przystąpieniu Polski do Unii Europejskiej*, Wydawnictwo Naukowe Katedra, 2018.
[63] Budyta-Budzyńska M., *Polacy na Islandii Rekonstrukcja przestrzeni obecności*, Scholar, Warszawa 2017.
[64] https://pl.wikipedia.org/wiki/Uchod%C5%BAstwo.

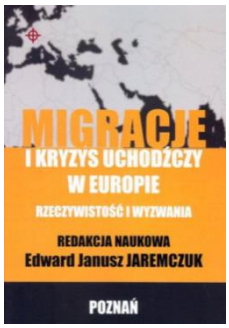
Uchodźcy w Europie[65]. Autorzy tej publikacji starają się sformułować płaszczyznę do oceny szans i zagrożeń jakie pociąga za sobą współczesna migracja, zwłaszcza w obszarze uchodźctwa. Jest to bardzo istotne zwracając uwagę na masowe zjawisko, które nie tylko niepokoi rządzących, ale także mieszkańców miast i wsi krajów Europy. Jest to zrozumiałe, gdy uświadomimy sobie, że napływa też ludność skrzywdzona przez wojny, a jednocześnie zupełnie odmienna kulturowo i religijnie. Będziemy się musieli nauczyć żyć tolerancyjnie w otoczeniu takich „*gości*". Problem migracji masowej uchodźców dotknął dość boleśnie takie kraje jak Francja i Niemcy. Również w Polsce, choć w mniejszym zakresie występuje

to zjawisko migracji i dotyczy przede wszystkim przybyszów spoza wschodniej granicy. Wróćmy jednak do omawianej książki, gdyż dostarcza ona gruntownej wiedzy o przyczynach i potencjalnych skutkach kryzysu migracyjnego, jakiego doświadcza Europa obecnego czasu. Należy docenić wyodrębnienie w niej problemów i poddanie ich wnikliwej analizie. W opracowaniu zwrócono również uwagę na ograniczenia polityki azylowej Unii Europejskiej i poszczególnych państw członkowskich.

Uchodźcy w Polsce i w Europie[66]. Kolejna książka „*Uchodźcy w Polsce i Europie*" to zbiór artykułów dotyczących niektórych problemów jakie towarzyszą uchodźcom. Wiemy, że obecnie występuje tendencja zatrzymania tej fali uchodźczej w krajach sąsiadujących z obszarami wojny, niepokoju społecznego, biedy poprzez tworzenie dofinansowanych przez Unię Europejską obozów przejściowych. Kwestie uchodźctwa to nie tylko sprawy socjalne, ale przede wszystkim wymagające jak najszybszego uregulowania prawnego i jednolitości działania przez państwa członkowskie tej organizacji gospodarczej.

Angela Merkel i kryzys migracyjny Dzień po dniu[67]. Gospodarka Niemiec odczuwała w pewnym okresie brak dostatecznej siły roboczej do swego prężnego przemysłu. Być może było to powodem otwartości tego kraju na cudzoziemców. Zachęcenie ich do przyjazdu, między innymi przez Kanclerza Niemiec, spowodowało samorzutne zorganizowanie wielkiej fali obcokrajowców, napływających tłumnie w celu znalezienia się w dobrych warunkach bytowania. Byli wśród nich uchodźcy z pogrążonej wojną Syrii, jak też przybywający w celach zarobkowych. Warto więc sięgnąć po książkę „*Angela Merkel i kryzys migracyjny Dzień po dniu*". Autor tej książki, niemiecki dziennikarz Die Welt, rekonstruuje decyzję podjętą na najwyższym szczeblu niemieckiej polityki dotyczącą wpuszczenia uchodźców. Spowodowała ona obniżenie autorytetu Kanclerza Niemiec - Angeli Merkel. Miała i nadal ma wpływ na jedność krajów Unii Europejskiej, które podzielone są w kwestii przyjmowania obcokrajowców, w tym również z obszarów Afryki. Czasopismo *The European* określiło omawianą książkę jako „*Wyjątkowe, niezwykle rzetelne dziennikarstwo. Genialnie napisana, czyta się niczym powieść detektywistyczną*." Natomiast *Der Spiegel* napisał „*Zapierający dech w piersiach raport z samego środka władzy. Polityczny kryminał strona po stronie.*"

[65] Wojtaszyk K.A., Szymańska J. (red. nauk.), *Uchodźcy w Europie Uwarunkowania, istota, następstwa*, ASPRA.
[66] Gardocka T., Sobczak J.(red. nauk.) *Uchodźcy w Polsce i Europie Stan prawny i rzeczywistość*, Adam Marszałek, Toruń 2010.
[67] Robin A., *Angela Merkel i kryzys migracyjny Dzień po dniu* (*Die Getriebenen*), tłumacz: Marcin Masny, Wydawca: Teologia Polityczna, Warszawa 2017.



Migracje i kryzys uchodźczy w Europie[68]. Spośród książek dotyczących migracji warto przeczytać: *Migracje i kryzys uchodźczy w Europie*". W opracowaniu zwraca się uwagę na to, że problem uchodźczy czy też migracyjny nie pojawił się dopiero w 2015 roku. „*Istniał już znacznie wcześniej, lecz bogate społeczeństwa Północy nie interesowały się nim. Profilaktycznie nie dociekały genezy wielu wojen, źródłem nędzy na innych lądach, wpływu, jaki wywierały na środowisko naturalne w położonych z dala od nich krainach. Nie poczuwały się do winy i nierzadko nadal tego nie robią."*

Uchodźcy polscy w latach dziewięćdziesiątych jako nowy problem społeczny[69]. Problematyka uchodźców, tak drażliwa w latach 2015-2017, nieco przygasła w środkach masowego przekazu jak i strategiach migracyjnych poszczególnych krajów Europy. Odsuwanie lawiny uchodźców poza teren Starego Kontynentu wiąże się z pilnowaniem niekontrolowanego napływu ludności z Azji i Afryki. Książka „*Uchodźcy polscy w latach dziewięćdziesiątych jako nowy problem społeczny*" omawia temat uchodźców polskich występujący już, lecz w mniejszej skali, w latach dziewięćdziesiątych XX wieku. Spotykamy tu przybliżenie aspektów prawnych, instytucjonalnych i społecznych uchodźctwa w Polsce po 1989 roku. Sięgnięcie przez autora po lata wcześniejsze od obecnych wynikało z potrzeby analizy sytuacji społeczno-prawnej uchodźców w Polsce w pierwszym okresie udzielania im pomocy.

3.7. Starzenie się populacji

Starzenie się społeczeństwa a polityka fiskalna i migracyjna[70]. Zagadnienie starzenia się społeczeństwa krajów przede wszystkim wysoko rozwiniętych stanowi kanwę wielu publikacji. Niski przyrost naturalny, często nie odtwarzanie przynajmniej stanu liczebnego rodziców powodują istotne problemy ekonomiczne. Z tego względu rządy tych krajów podejmują uregulowania prawne, w celu przyjęcia przede wszystkim emigrantów wykształconych i to w wieku produkcyjnym. Zaczyna prowadzić się też tolerancyjną politykę fiskalną wobec migrujących. Prezentowana książka jest pracą zbiorową obejmująca rozdziały:

O demograficznym starzeniu;
Pronatalistyczna polityka fiskalna;
Model podejmowania decyzji prokreacyjnych i wychowawczych;
Przegląd metod szacowania kosztów dzieci;
Koszty utrzymania dzieci w czterech krajach europejskich;
Wydatki na politykę migracyjną Austrii, Francji, Polski, Włoch - ujęcie porównawcze;
Wydatki na polską politykę imigracyjną w latach 2010-2013 - ujęcie szczegółowe;
Pronatalistyczna funkcja wydatków na politykę fiskalną i migracyjną.

Występujące tu pojęcie „*polityka pronatalistyczna*" oznacza politykę ludnościową, mającą na celu zwiększenie liczby urodzeń[71]. Wyrazem tej polityki jest tworzenie przez państwo różnego rodzaju

[68] Jaremczuk E.J., (red. nauk.), *Migracje i kryzys uchodźczy w Europie Rzeczywistość i wyzwania, FNCE,* Poznań 2017.
[69] Stawiński R., *Uchodźcy polscy w latach dziewięćdziesiątych jako nowy problem społeczny*, Difin, Warszawa 2017.
[70] Fihel A., *Starzenie się społeczeństwa a polityka fiskalna i migracyjna*, Wydawnictwo Uniwersytetu Warszawskiego, 2017.
[71] https://pl.wikipedia.org/wiki/Polityka_pronatalistyczna.

zachęt materialnych i moralnych do powiększania liczby dzieci w rodzinach i w otaczaniu szczególną opieką rodzin wielodzietnych. W definicji tego określenia czytamy: *Polityka pronatalistyczna stosowana była m.in. we Francji i w Polsce po wielkich stratach ludnościowych poniesionych w czasie pierwszej wojny światowej. Obecnie prowadzą ją kraje skandynawskie (Norwegia, Dania, Szwecja). Od 1 kwietnia 2016 obowiązuje w Polsce program Rodzina 500 plus, którego celem jest wsparcie rodzin, a którego długoterminowym skutkiem ma być wpływ na zwiększenie liczby urodzeń. Polska znajduje się aktualnie w grupie krajów o najniższym współczynniku dzietności.* Autorzy tej interesującej książki pokazują jaki wpływ na dzietność mają konkretne działania z zakresu polityki fiskalnej, jaki jest ich koszt, jakie są wydatki państwa na poszczególne działania w ramach polityki imigracyjnej i jaki jest ich koszt. Ponadto podają ile kosztuje stymulowanie imigracji kompensacyjnej, tzn. uzupełniającej braki na rynku pracy. W opracowaniu wskazano na politykę ludnościową w czterech różnych pod względem zjawisk demograficznych krajach europejskich, a mianowicie: Austrii, Francji, Polsce i Włoszech.

Demograficzne uwarunkowania i wybrane społeczno-ekonomiczne konsekwencje starzenia się ludności w krajach europejskich[72]. To kolejna wybrana prze zemnie książka poruszająca konsekwencje starzenia się ludności w odniesieniu do krajów europejskich, przy czym zakres sugerowanej pracy obejmuje rozdziały:

Ekonomiczne konsekwencje zmian rozmiarów i struktury według wieku populacji,

Proces starzenia się ludności w krajach europejskich. Ujęcie historyczne i perspektywy przemian do 2050 roku,

Zmiany oczekiwanego trwania życia a proces starzenia się populacji,

Konwergencja i dywergencja umieralności w krajach europejskich,

Starzenie się populacji a poziom życia ludności,

Zmiany demograficzne a starsi pracownicy w Polsce z perspektywy przedsiębiorstwa. Rozdział ten poparto wynikami przeprowadzonego badania ankietowego.

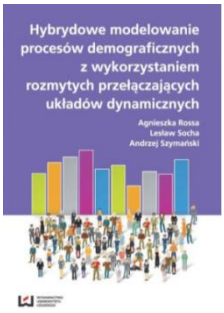
Konwergencja (łac. *convergere*, zbierać się, upodabniać się), także *teoria konwergencji, efekt doganiania* - hipoteza, że w krajach biedniejszych dochody na mieszkańca (mierzone np. PKB/osobę) rosną szybciej niż w krajach bogatych. Jak pokazują dane empiryczne, mimo znacznie niższej produktywności pracy kraje słabo rozwinięte często osiągają wyższe wskaźniki wzrostu gospodarczego niż kraje wysoko uprzemysłowione[73]. Tu odniesiono to pojęcie do umieralności, która w krajach „*zasobnych*" jest niższa niż w biednych.

W pracy podkreślono, że obecnie starzenie się populacji, wraz ze spadkiem płodności, jest zjawiskiem typowym dla demograficznego obrazu Europy. Zachodzi konieczność znalezienie rozwiązań mających na celu łagodzenie niekorzystnych skutków tego zjawiska. W streszczeniu prezentowanej książki znajdujemy akapit: „*W pracy dokonano charakterystyki zjawisk demograficznych determinujących stan i strukturę ludności Europy oraz oceny procesu starzenia się ludności w krajach europejskich do 2050 r., a także podjęto badania wpływu starzenia się populacji na zmiany poziomu życia ludności do 2050 r. Wskazano również możliwości realizacji polityki aktywnego starzenia się przez duże podmioty gospodarcze w Polsce*".

[72] Kurkiewicz J. (red.), *Demograficzne uwarunkowania i wybrane społeczno-ekonomiczne konsekwencje starzenia się ludności w krajach europejskich*, Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, format: ibuk.

[73] https://pl.wikipedia.org/wiki/Konwergencja_(ekonomia).

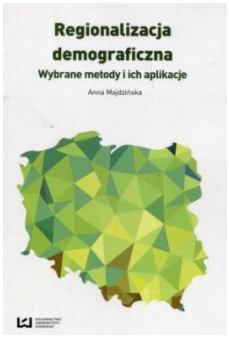
3.8. Modelowanie procesów demograficznych



Hybrydowe modelowanie procesów demograficznych z wykorzystaniem rozmytych przełączających układów dynamicznych[74]. Poszukując opisów teoretycznych i praktycznych zastosowań modeli przedstawiania oraz prognozowania wskaźników procesów ludnościowych zwróciłem uwagę na książkę: „*Hybrydowe modelowanie procesów demograficznych z wykorzystaniem rozmytych przełączających układów dynamicznych*". Zamieszczone w niej modele teoretyczne dotyczą umieralności są pomocne w szacowaniu długości życia przez firmy ubezpieczeniowe. Zakupiłem tą książkę i odniosłem wrażenie, że zawarte w niej procedury teoretyczne mogą znaleźć zastosowanie również w modelowaniu zjawisk sfer ekonomicznych, przykładowo czasu życia firm, maszyn lub instalacji.

Prawidłowości związane z umieralnością stanowią od dawna przedmiot zainteresowania wielu badaczy. Graunt - inicjator metodologii tablic umieralności opublikował w roku 1662 pracę: „*Natural and Political Observations Made upon the Bills of Mortality*". Współczesnym autorem metodologii budowy tablic wymieralności jest C. L. Chiang. W proponowanej książce zaprezentowano najnowsze modele umieralności, które umożliwiają prognozowanie procesu wymierania populacji w perspektywie średnio- i długookresowej. W streszczeniu tej książki podano: „*Autorzy omawiają kolejne modyfikacje modelu Lee-Cartera, wykorzystując teorię równań różniczkowych, algebry liczb rozmytych oraz algebry liczb zespolonych. Zastosowanie tych struktur pozwala na modelowanie umieralności, a następnie na wskazanie własności prognostycznych poszczególnych modeli*„. Zawartość omawianej książki obejmuje:

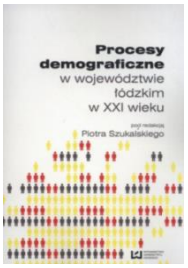
- wyszczególnienie modeli umieralności (tablicowe mierniki umieralności, modele interpolacyjne);
- opis statycznych i dynamicznych modeli hybrydowych;
- przedstawienie dynamicznych hybrydowych modeli umieralności;
- zaprezentowanie modelu Koissi-Shapiro opartego na skierowanych liczbach rozmytych;
- modele umieralności oparte na zmodyfikowanych liczbach rozmytych i funkcjach zespolonych;
- estymację i ewaluacje modeli umieralności.



Regionalizacja demograficzna Wybrane metody i ich aplikacje[75]. Procedura wyróżniania regionów nosi nazwę regionalizacji. Termin region odnosi się do okolicy, przestrzeni. Region jest to wyodrębniony, względnie jednolity fragment powierzchni Ziemi, charakteryzujący się występowaniem określonych cech, odróżniających go od obszarów sąsiednich[76]. Regionalizacja demograficzna ma na celu wyodrębnienie obszarów jednorodnych z punktu widzenia charakterystyk ludnościowych. W tym zakresie stosowanych jest obecnie szereg metod regionalizacyjnych. Na uwagę zasługują propozycje autorki dotyczące modyfikacji niektórych procedur oraz adaptacji metod wielocechowej analizy porównawczej dla celów delimitacji regionów demograficznych. Warto zainteresować się także przykładem zastosowania tych usprawnionych metod w odniesieniu do regionalizacji demograficznej obszarów Polski i Europy na początku pierwszej i drugiej dekady XXI wieku. Wymienione tu określenie

[74] Rossa A., Socha L., Szymański A., *Hybrydowe modelowanie procesów demograficznych z wykorzystaniem rozmytych przełączających układów dynamicznych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015.
[75] Majdzińska A., *Regionalizacja demograficzna Wybrane metody i ich aplikacje*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2016.
[76] http://ibrbs.pl/mediawiki/index.php/Delimitacja_granic_regionalnych.

delimitacja oznacza rozdzielenie, rozgraniczenie oraz wyznaczenie granic i jest ona przeprowadzana według określonych kryteriów[77].



Procesy demograficzne w województwie łódzkim[78]. Coraz częściej demografowie skupiają się nie tylko na całym kraju, ale także na poszczególnych obiektach struktury terytorialnej. Rządzących bowiem danym regionem, podregionem, województwem czy też powiatem interesuje bezpośrednio „*ich podwórko*". Chcą na bieżące przeciwdziałać niekorzystnym procesom ludnościowym, takim jak starzenie się populacji oraz wzmożona migracja wewnętrzna pracowników do sąsiednich regionów. Zrozumieniem chociażby w tym zakresie była zorganizowana lokalnie w maju 2018 roku konferencja: „*Sytuacja demograficzna jako wyzwanie dla polityki społecznej i gospodarczej na przykładzie miasta Nysa oraz powiatu nyskiego - stan obecny i perspektywy*". Sięgnąłem więc po książkę „*Procesy demograficzne w województwie łódzkim*". W tej pozycji literatury zajęto się określeniem zmian w liczbie i rozmieszczeniu ludności województwa łódzkiego. Zwrócono uwagę na takie czynniki jak zachowania rozrodcze i matrymonialne, stan zdrowia i umieralność oraz migracje. W porównaniach posłużono się danymi ewidencyjnymi spisu ludności, a także prognozą demograficzną GUS-u. Zaobserwowano i pokazano zmiany na poziomie powiatów województwa opolskiego. Opracowanie powinno być pomocne dla decydentów sektora publicznego w obszarach: edukacja, ochrona zdrowia, zaopatrzenie w media oraz przedsiębiorców interesujących się popytem na dobra i usługi rynkowe oraz podażą pracy.

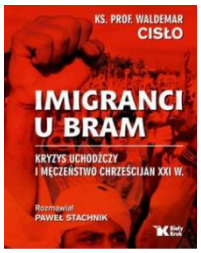
3.9. Skutki migracji

Migracje międzynarodowe i ekspatriacja[79]. Ekspatriacja (łac. *expatriare*) oznacza opuszczenie kraju ojczystego, przy czym określenie to stosowane jest zarówno do dobrowolnego opuszczenia ojczyzny, jak i przymusowego wydalenia, także w połączeniu z utratą obywatelstwa[80]. Ekspatriacja dotyczy głównie firm międzynarodowych. Współczesna ekspatriacja ekonomiczna jest używana na określenie wysokiej klasy specjalisty, który podążając za zatrudnieniem opuszcza ojczyznę. Książka „*Migracje międzynarodowe i ekspatriacja*", będąca monografią ma charakter interdyscyplinarny. Autorzy reprezentują dyscypliny nauk społecznych, jak i ekonomicznych. W prezentowanej monografii przedstawiono następujące zagadnienia:

- mobilność i migracje międzynarodowe;
- pojęcia migranta i ekspatrianta;
- motywy przemieszczeń międzynarodowych;
- cechy osobowości migranta i ekspatrianta;
- bariery i korzyści migracji i ekspatriacji;
- przedsiębiorczość migrantów;
- modele i praktyki adaptacji kulturowej,
- kulturowe, społeczne i polityczne uwarunkowania migracji międzynarodowych.

[77] http://ibrbs.pl/mediawiki/index.php/Delimitacja_granic_regionalnych.
[78] Szukalski P. (red.), *Procesy demograficzne w województwie łódzkim w XXI wieku*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015.
[79] Przytuła P., *Migracje międzynarodowe i ekspatriacja Perspektywa indywidualna, organizacyjna, społeczno-kulturalna*, Difin, Warszawa 2017.
[80] https://pl.wikipedia.org/wiki/Ekspatriacja.

Stopniowo na skutek globalizacji i znacznego obniżenia kosztów podróży międzykontynentalnych ekspatriacja została zdominowana przez migracje zarobkowe. Powstał nowy segment rynku pracy, bowiem pracownicy dużych firm coraz częściej byli wysyłani do zagranicznych oddziałów bądź podległych spółek. Obecnie ekspatrianci, zwykle zatrudniani są w międzynarodowych korporacjach i współtworzą nową, globalną klasę średnią, w której następują już większe wzajemne wpływy kulturowe.

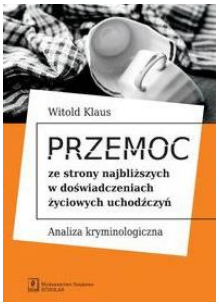


Imigranci u bram[81]. Z postawy chrześcijanina, w przystępny sposób napisana książka „*Imigranci u bram*" wprowadza nas w problematykę sytuacji wyznawców Chrystusa w obszarach wojen, czy też konfliktów społecznych. Autor Waldemar Cisło informuje nas: „*Co* o *trzy minuty gdzieś na świecie ginie jeden chrześcijanin; nie umiera naturalną śmiercią; ginie męczeńsko…*". Ich śmierć jest przeważnie w wyniku ataku muzułmanów na obiekty sakralne lub skromne domostwa wioskowe. W streszczeniu tej książki stawia się wiele pytań typu:

Dlaczego nikt z rządzących naszym kontynentem nie chce zastanowić się, co będzie z tymi krajami ogarniętymi konfliktami zbrojnymi, gdy odpłynie z nich potężna rzeka młodych ludzi - uchodźców?

Dlaczego i za czyją przyczyną świat staje dziś na głowie, kto ma w tym interes i jak to może się skończyć?

Wielkim walorem omawianej książki jest także ponad 100 dramatycznych ilustracji, które przemawiają do Czytelnika dobitniej niż słowa tej książki napisanej przez przewodniczącego polskiej sekcji światowej organizacji „*Pomoc Kościołowi w Potrzebie*". W prezentacji książki " *Imigranci u bram*" znajduje się dramatyczny akapit, który zacytuję: „*Ofiarami wojny są też muzułmanie, to prawda, ale to nie kto inny, jak bracia islamskich ofiar wciąż burzą świątynie i domostwa chrześcijan, mordują, gwałcą, palą, okaleczają, ścinają głowy, podkładają bomby. Część z nich jedzie potem, a raczej jest eksportowana, do Europy; kto z nich uchodźca, kto imigrant, kto terrorysta - nie sposób rozeznać. Faktem jest, że wojna została już zawleczona na nasz kontynent*".



Przemoc ze strony najbliższych w doświadczeniu życiowym uchodźczyń[82]. Cudzoziemcy wnoszą ze sobą nowe nawyki, zwyczaje, zachowania kulturalne. Nie zawsze to jest zrozumiałe przez mieszkańców danego kraju. Rażący może być stosunek mężczyzn do kobiet, czy też sposób wychowywania dzieci i zachowania religijne. Elementem wiedzy w tym zakresie może być książka „*Przemoc ze strony najbliższych w doświadczeniu życiowym uchodźczyń*". Jej autor Witold Klaus wykazuje jak skomplikowanym zagadnieniem jest stworzenie systemu skutecznej reakcji społecznej, gdy mamy do czynienia z problemami silnie związanymi ze zróżnicowaniem kulturowym. Spotykamy tu wnikliwą analizę zagadnień przemocy wobec migrantek na podstawie badań zagranicznych. Według recenzji Anny Kossowskiej „*Na szczególne podkreślenie zasługuje logiczna struktura pracy oraz waga wyników przeprowadzonych przez Autora innowacyjnych badań empirycznych, a także trafne wnioski, zmierzające do stworzenia skutecznego systemu ochrony i wsparcia migrantek, doświadczających przemocy ze względu na płeć*".

[81] Cisło W., *Imigranci u bram Kryzys uchodźcy i męczeństwo chrześcijan XXI w.*, Wydawca: Biały Kruk, 2017.
[82] Klaus W., *Przemoc ze strony najbliższych w doświadczeniach życiowych Analiza kryminologiczna*, Scholar, Warszawa 2017.

Ekonomiczno-społeczne następstwa migracji na obszarach górskich na przykładzie Karpat Polskich[83]. W okresie przedwojennym i powojennym naszego kraju wystąpiła znaczna migracja ludności z ubogich rejonów górskich do USA, Kanady i innych zamożnych państw świata. Ze tego względu interesujący się zjawiskami ludnościowymi powinien się zapoznać z książka: „*Ekonomiczno-społeczne następstwa migracji na obszarach górskich na przykładzie Karpat Polskich*". W streszczeniu tej publikacji czytamy: „*Książka stanowi kwintesencję wiedzy o współczesnych przejawach mobilności zachodzących na terenach peryferyjnych i ich konsekwencjach w wymiarze ekonomicznym, społecznym i rolno-środowiskowym. Monografia jest efektem badań Autora przeprowadzonych w gminach górskich klasyfikowanych jako obszary o niekorzystnych warunkach gospodarowania (ONW) z udziałem przedstawicieli samorządów lokalnych. W całościowej ocenie konsekwencji procesów migracyjnych wykorzystana została grupa wielokryterialnych metod analitycznych - Analityczny Proces Hierarchiczny (AHP) i Analityczny Proces Sieciowy (ANP), zastosowana w tego typu badaniach po raz pierwszy. Dokonana analiza obejmuje część diagnostyczną dotyczącą korzyści i kosztów migracyjnych, oraz prognostyczną, odnoszącą się do szans i zagrożeń wynikających z odpływów migracyjnych*".

* * *

Literatura przedmiotu dotycząca migracji, czyli wędrówki przestrzennej ludności wskazuje, że zjawisko to istniało zawsze. Objawiało się jednak w formie fal z większym lub mniejszym natężeniem. Masowe imigracje na wolniejsze tereny spowodowane były początkowo wojnami plemiennymi, szerzącymi się zarazami, możliwością dostępu do wody, łąk, żyznych dolin, czy też obszarów z łatwo dostępnymi bogactwami naturalnymi, których wydobycie przynosiło krocie zyski osiedleńcom. Przeważnie współcześnie młode pokolenie, chce jak najszybciej spełnić swoje marzenia o dostatnim życiu, zbudowaniu własnego domu, urządzeniu nowocześnie mieszkania, posiadania dziecka i wojażach po świecie. Jednak, aby temu sprostać trzeba znaleźć intratną, dobrze płatną pracę i to najlepiej zagranicą, uważa część młodego pokolenia, w tym kończący studia w Polsce.

Imigracja to nie tylko pogoń za poszukiwaniem „*szczęścia*", ale także sytuacje obiektywne, które zmuszają rzesze ludności do opuszczania swoich domostw i udawania się w nieznane, żyjąc z łaski innego zamożnego kraju. Jak trudno się w nim zadomowić, znaleźć pracę a jednocześnie zachować swoje tradycje rodzinne. Chociaż brakuje rąk do pracy, występuje również i w Polsce niechęć do przyjmowania niepewnych przybyszów, którzy traktowani są często jako terroryści Tymczasem dominująca w nich grupa to biedni uciekinierzy, chrześcijanie uciekający przed maczetami wrogich odłamów w krajach pozaeuropejskich.

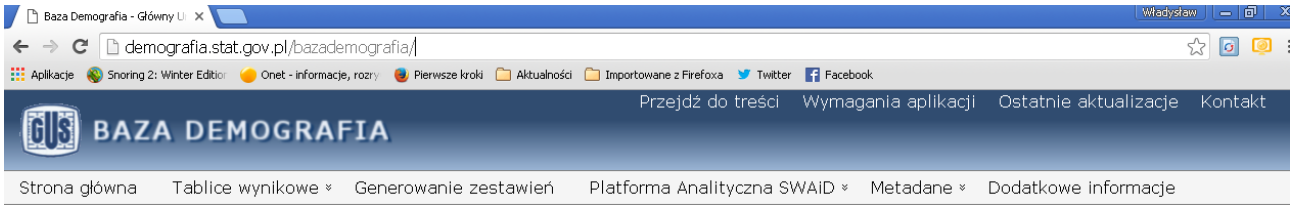
[83] Cymanow P., *Ekonomiczno-społeczne następstwa migracji na obszarach górskich na przykładzie Karpat Polskich*, Difin, Warszawa 2018.

4. Prognozowanie migracji ludności z uwzględnieniem wag harmonicznych

4.1. Wprowadzenie

Demografia (stgr. *demos* - "lud" i *grapho* - "piszę") to dziedzina nauki zajmująca się powstawaniem, życiem i przemijaniem społeczności ludzkiej[84]. Dziedzina ta zajmuje się przyrostem naturalnym, migracjami, strukturą społeczną, a więc wiekiem, płcią, przynależnością zawodową, narodowością oraz wyznaniem. Ponadto rozmieszczeniem przestrzennym ludności, oddziaływaniami społecznymi i socjologicznymi. Źródłem danych są publikacje statystyczne, reprezentatywne próby statystyczne oraz spisy ludności. Do obserwacji procesów demograficznych wykorzystuje się statystyki meldunkowe z których pozyskuje się m.in. współczynniki: urodzeń, zgonów, przyrostu naturalnego, dzietności, saldo migracji, oczekiwaną długość życia. Demografia graficznie przestawia dane w okresach czasowych w formie wykresów słupkowych, liniowych i innych.

Baza *Demografia* umożliwia stały dostęp do informacji statystycznych charakteryzujących sytuację demograficzną w Polsce[85]. Baza ta, aktualizowana przez Główny Urząd Statystyczny, stanowi źródło danych, obejmujące takie informacje jak stan i struktura ludności, ruch naturalny, migracja, dynamika zachodzących zmian w populacji ludności. Baza *Demografia* umożliwia generowanie różnych zestawień na żądanie według przekroju ustalonego przez użytkownika. Menu główne tej aplikacji programowej pokazano na rysunku 4.1. Menu obejmuje zakładki: *Strona główna*, *Generowanie zestawień*, *Platforma Analityczna* SWAiD, *Metadane*, *Dodatkowe informacje*. *Platforma Analityczna* , czyli *Dziedzinowe Bazy Wiedzy* GUS-u, dająca dostęp do aktualizowanych na bieżąco informacji w postaci gotowych raportów.

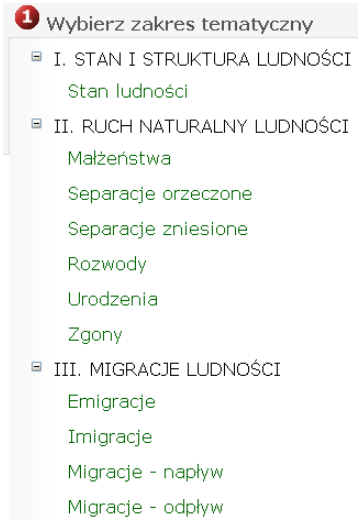


Źródło: http://demografia.stat.gov.pl/bazademografia/.

Rys. 4.1. Menu aplikacji *Baza Demografia*

W tym materiale wykorzystano możliwości jakie daje zakładka *Generowanie zestawień*, przy czym dane udostępniane są w postaci tablic w układzie wskazanym przez użytkownika. Zainteresowanie skupimy teraz na zakresie II - *Ruch naturalny ludności*, a w ramach niego na intuicyjnie wybrane przez autora cechy statystyczne: *Małżeństwa*, *Rozwody*, *Urodzenia*, *Zgony* oraz *Emigracje* i (*Migracje - odpływ*) będące składnikiem zakresu III - *Migracje ludności* (zob. rysunek 4.2).

[84] https://pl.wikipedia.org/wiki/Demografia.
[85] http://demografia.stat.gov.pl/bazademografia/.



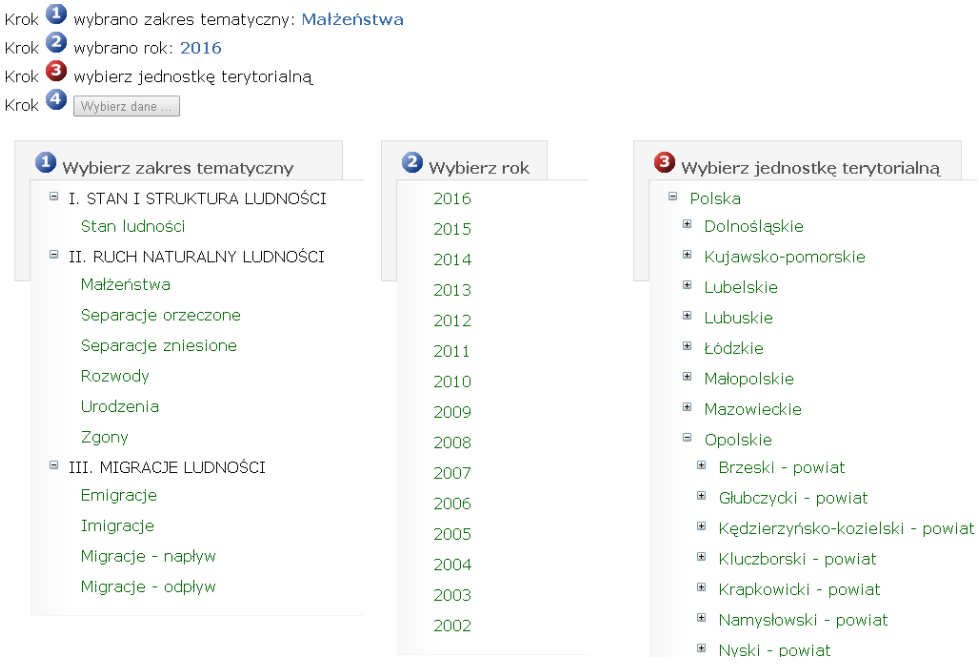
Źródło: http://demografia.stat.gov.pl/bazademografia/.

Rys. 4.2. Zakresy tematyczne zakładki *Generowanie zestawień*

Program wyszukiwania po wskazaniu zakresu tematycznego umożliwia wybór roku, jednostki terytorialnej oraz wskazanie dalszego kryterium wyszukiwania w kolejnych czterech krokach:

1- wybierz zakres tematyczny,
2 - wybierz rok,
3 - wybierz jednostkę terytorialną,
4 - wybierz dane.

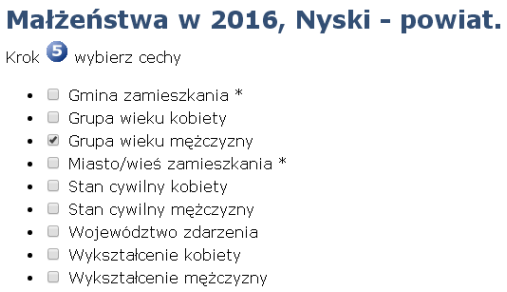
Wybierzmy przykładowo cechę *Małżeństwa*, rok 2016 i jednostkę terytorialną, kolejno zaznaczając: *Polska*, *Opolskie*, *Nyski - powiat*, co program wypisuje w kolejnych krokach (zob. rysunek 4.3).



Źródło: http://demografia.stat.gov.pl/bazademografia/.

Rys. 4.3. Wstępne zdefiniowanie kryteriów wyszukiwania

Po naciśnięciu przycisku *Wybierz dane* w kroku 4. uzyskujemy przejście do kroku 5. (zob. rysunek 4.4).



Źródło: http://demografia.stat.gov.pl/bazademografia/.

Rys. 4.4. Wybranie zakresu raportu dla cechy statystycznej *Małżeństwa*

Po wskazaniu cechy szczegółowej np. *Grupa wieku mężczyzny* przechodzimy do kroku 6. *Pokaż dane* (zob. rysunku 4.5).

Grupa wieku mężczyzny	Liczba małżeństw
do 19	3
20-24	76
25-29	272
30-34	164
35-39	74
40-44	29
45-49	19
50-54	17
55-59	11
60 i więcej	23
Razem	688

Źródło: http://demografia.stat.gov.pl/bazademografia/.

Rys. 4.5. Wybranie kryterium cechy *Małżeństwa*

Tak postępując możemy zestawić dane wskazanej cechy szczegółowej kolejnych przedziałów grupy wiekowej mężczyzn dla okresu lat (2002-2016).

4.2. Szeregi czasowe wybranych cech statystycznych (przyczyny, skutku)

Jak już nadmieniono jak cechy przyczyn skutku, czyli (*Migracji - odpływ*) przyjęto intuicyjnie: *Emigracja*, *Małżeństwa*, *Rozwody*, *Urodzenia*, *Zgony*. Powtarzając podaną wcześniej procedurę kolejnych kroków uzyskano tabele z danymi dla poszczególnych cech i przedziału lat (2002-2016) dla jednostki terytorialnej *Nyski - powiat*. Dla przejrzystości dalszego modelowania przyjmijmy następujące symbole cech statystycznych:

Em - *Emigracja* (żonaci/zamężne);

Mi - (*Migracja - odpływ*) → żonaci/zamężne;

M - *Małżeństwa* → *grupa wieku mężczyzny,* przedziały: 20-24, 25-29, 30-34;

R - *Rozwody* → według *grupa wieku mężczyzny w dniu wniesienia powództwa,* przedziały: 20-24, 25-29, 30-34;

U - *Urodzenia* → według *płeć dziecka, chłopiec;*

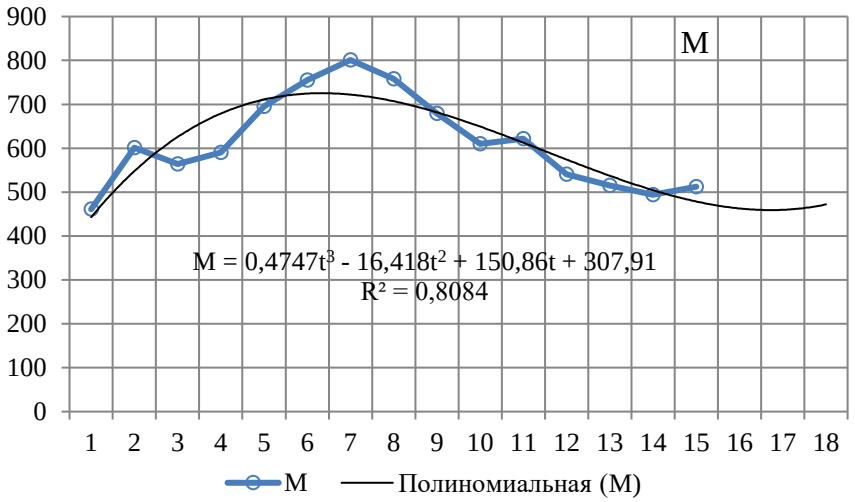
Z - *Zgony* → według *grupa wieku*, przedziały: 0 dni; 1-27 dni; 28, 29 dni, 11 miesięcy; 1-4 lata; 5-9 lat; 10-14 lat).

Małżeństwa (M), *grupa wieku mężczyzny*. Przyjęto do badań łącznie 3 przedziały: 20-24, 25-29, 30-34, co dla roku 2016 stanowi 512, czyli 74,4% *razem*[86].

[86] http://demografia.stat.gov.pl/bazademografia/CustomSelectData.aspx?s=mal&y=2016&t=00/16/07.

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Dane	461	601	564	590	695	755	801	758	679	610	622	541	515	494	**512**

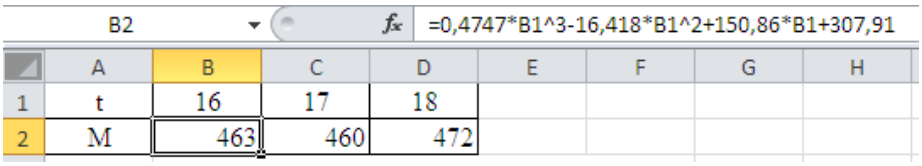
Skorzystajmy z *Excela* i zobaczmy jak te dane kształtują się na wykresie o trendzie wielomianowym trzeciego stopnia (zob. rysunek 4.6).



Źródło: Opracowanie własne w *Excelu*.

Rys. 4.6. Cecha *Małżeństwa* w latach 2002-2016

Na wykresie pokazano kształtowanie się danych rzeczywistych, trend wielomianowy trzeciego stopnia oraz oszacowanie tym trendem prognoz na trzy kolejne lata (t_{16} = 2017, t_{17} = 2018, t_{18} = 2019). Widzimy w miarę dobre dopasowanie trendu do danych empirycznych, gdyż współczynnik determinacji R^2 = 0,8084. Prognozy określono na podstawie trendu wielomianowego według przykładowej formuły podanej na rysunku 4.7.

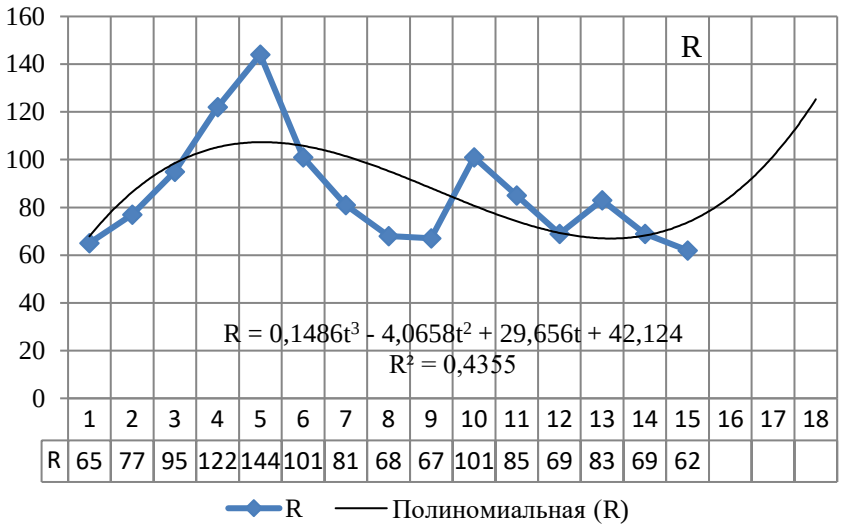


Źródło: Opracowanie własne w *Excelu*.

Rys. 4.7. Obliczenie prognoz na lata 2017-2019

Rozwody, *Nyski - powiat*, według *grupa wieku mężczyzny w dniu wniesienia powództwa*. Przyjęto do badań łącznie 3 przedziały: 20-24, 25-29, 30-34, co dla roku 2016 stanowi 62, czyli 28,2% *razem* dla roku 2016. Układ danych dla wybranej cechy statystycznej czynnika pokazano na rysunku 4.8.

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Dane	65	77	95	122	144	101	81	68	67	101	85	69	83	69	62

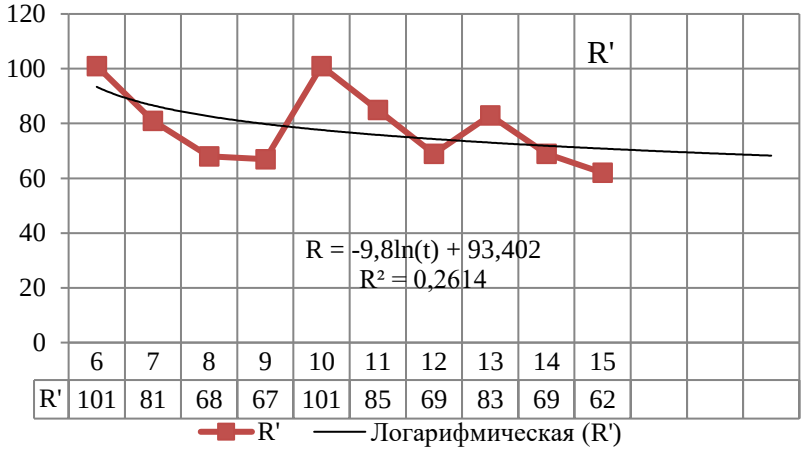


	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
R	65	77	95	122	144	101	81	68	67	101	85	69	83	69	62			

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.8. Cecha Rozw*ody* w latach 2002-2016

Trend wielomianowy 3-go stopnia ma niski współczynnik determinacji R^2 i wskazuje na wyraźny wzrost rozwodów w latach 2017-2019, dla których prognozy są następujące: $R_{16} = 84$, $R_{17} = 101$, $R_{18} = 125$. Miały na to wpływ lata (2002-2006) - gwałtownego wzrostu liczby rozwodów, pomińmy zatem ten okres i dobierzmy inny trend (zob. rysunek 4.9).



	6	7	8	9	10	11	12	13	14	15
R'	101	81	68	67	101	85	69	83	69	62

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.9. Skorygowany przedział szeregu cechy *Rozwody* w latach 2007-2016

Jest nim trend logarytmiczny, o niezadawalającym jednak R^2, o prognozach $R_{16} = 66$, $R_{17} = 66$, $R_{18} = 65$ obliczonych na podstawie podanego np. dla roku 2017 wyrażenia: =-9,8*LN(Q2)+93,402 (zob. rysunek 4.10).

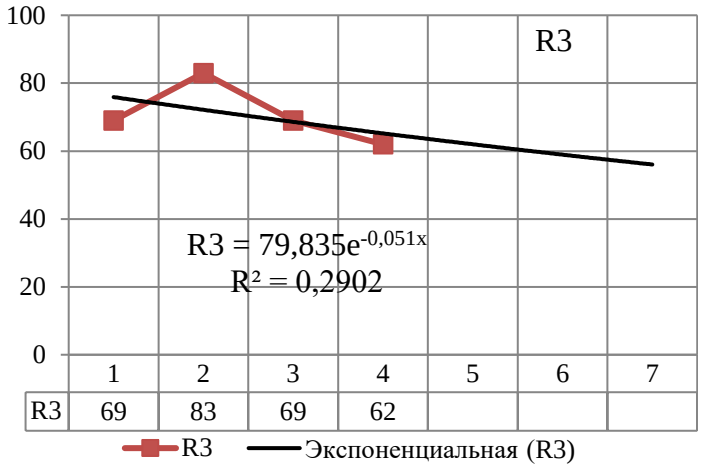
Q	R	S
16	17	18
2017	2018	2019
66	66	65

=-9,8*LN(Q2)+93,402

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.10. Prognozy na lata 2017-2019 wyznaczone trendem logarytmicznym

Zobaczmy jeszcze jak przebiega wykres według średnich ruchomych 3-y okresowych dla którego prognoza na rok 2017 jest średnią arytmetyczną: $R_{16} = (83 + 69 + 62)/3 \approx 71$. Traktując tą prognozę jako wartość to prognozy na kolejne dwa lata są następujące: $R_{17} \approx 67$, $R_{18} \approx 67$. A teraz dodatkowo określamy trend wykładniczy 4-ro okresowy wprowadzając oznaczenie okresu czasowego jako *x* (1-4), co pozwala nam na obliczenie prognoz na lata (5-7), czyli 2017-2019 (zob. rysunek 4.11).



Źródło: Opracowanie własne w *Excelu*.

Rys. 4.11. Trend wykładniczy 4-ro okresowy

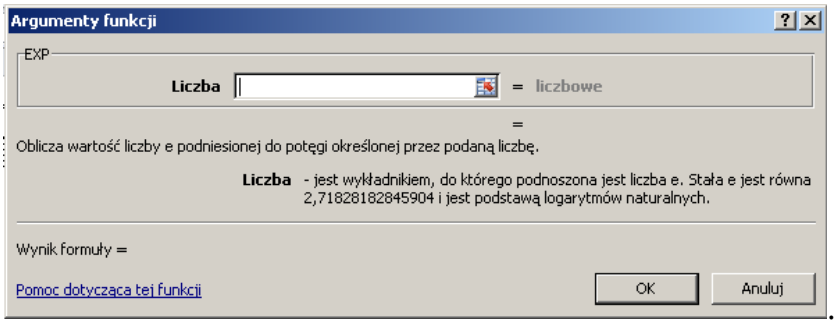
Na podstawie równania trendu wykładniczego, korzystając z formuł Excela obliczamy prognozy na kolejne lata, czyli (2017-2019) - zob. rysunek 4.12.

M	N	O	P	Q	R	S	T
12	13	14	15	16	17	18	
2013	2014	2015	2016	2017	2018	2019	Rok
69	83	69	62				Dane empiryczne
76	72	69	65	62	59	56	Dane modelowe
-0,051	-0,102	-0,153	-0,204	-0,255	-0,306	-0,357	Wykładnik
				=(-0,051*Q6)			Formuła wykładnika
1	2	3	4	5	6	7	x
=79,835*EXP(M5)							Formuła danych modelowych

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.12. Sposób obliczenia prognoz przy zastosowaniu formuły trendu wykładniczego

Podane *EXP* to funkcja *Excela* generująca wartość *e*, czyli podstawę logarytmu naturalnego (zob. rysunek 4.13).



Źródło: Opracowanie własne w *Excelu*.

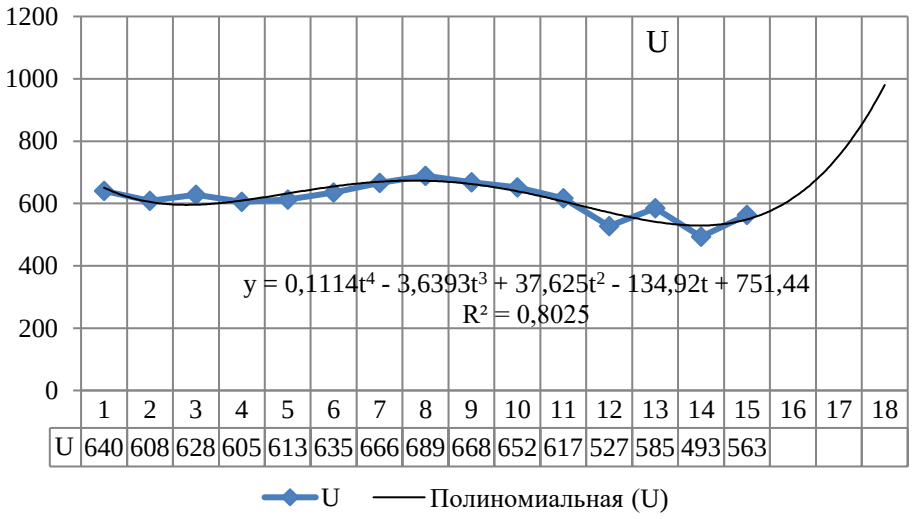
Rys. 4.13. Okno dialogowe funkcji *EXP*

Zestawmy teraz uzyskane prognozy według różnych trendów:

Rok	2017	2018	2019
Trend wielomianowy 3-go stopnia	84	101	125
Trend logarytmiczny	66	66	65
Średnie ruchome 3-okresowe	71	67	67
Trend wykładniczy 4-okresowy	62	59	56
Suma:	283	293	253
Przewidywanie średnie: Suma/4	71	73	78

Urodzenia, *Nyski - powiat*, według *płeć dziecka*. Przyjęto do badań *chłopiec*, czyli 563, co dla roku 2016 stanowi 50,6% *razem*. Zestawienie danych podaje tabela, a kształtowanie się danych rzeczywistych i modelowych zaprezentowano na rysunku 4.14.

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Dane	640	608	628	605	613	635	666	689	668	652	617	527	585	493	563



Źródło: Opracowanie własne w *Excelu*.

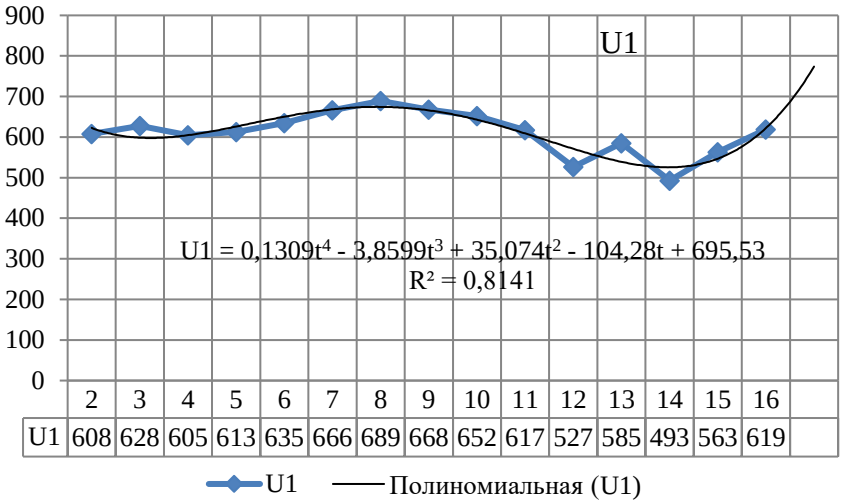
Rys. 4.14. Cecha *Urodzenia* (chłopcy) w latach 2002-2016

Według trendu wielomianowego 4-go stopnia prognoza na rok 2017, U_{16} = 619, przy zastosowaniu wyrażenia:

=0,1114*Q1^4-3,6393*Q1^3+37,625*Q1^2-134,92*Q1+751,44.

Duży wzrost trendu wielomianowego budzi obawy i z tego względu czasem celowe jest prognozowanie przy zastosowaniu „*pełzania*" szeregu czasowego corocznie o jeden okres w przód. Trzeba jednak pamiętać o przesunięciu również numeru kolejnego okresu o 1 w tył. Przyjmijmy umownie że, dane rzeczywiste na rok 2017 pokrywają się z modelowymi (619). Rozpatrzy teraz szereg czasowy od roku 2003-2017 i obliczmy prognozę na rok 2018 → U_{17-1} = 623, korzystając z wyrażenia na nowy trend wielomianowy (zob. rysunek 4.15):

=0,1309*P1^4-3,8599*P1^3+35,074*P1^2-104,28*P1+695,53.



	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
U1	608	628	605	613	635	666	689	668	652	617	527	585	493	563	619

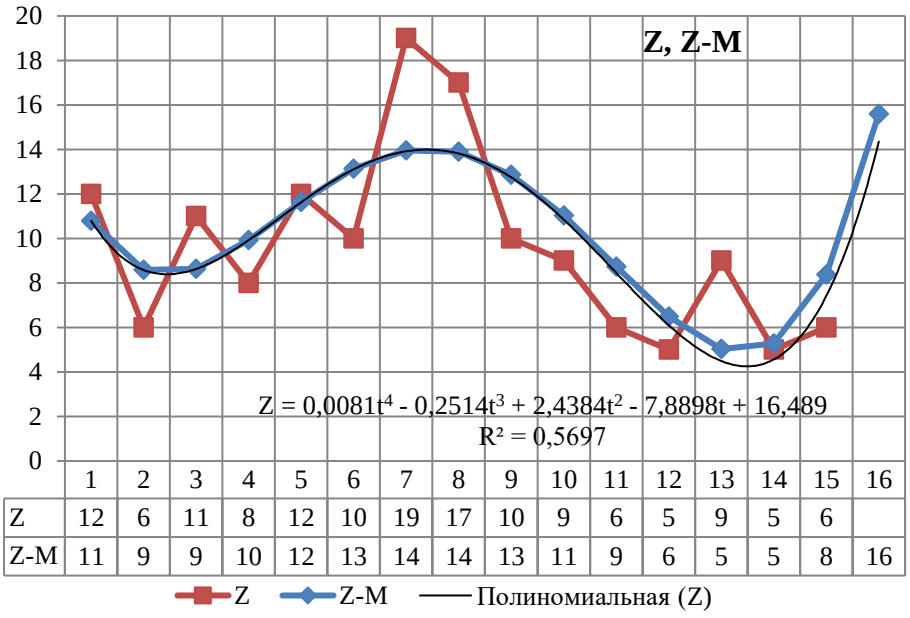
Źródło: Opracowanie własne w *Excelu*.

Rys. 4.15. Cecha *Urodzenia* (chłopcy) w latach 2003-2017

Zgony, *Nyski - powiat*, według *grupa wieku*. Przyjęto do badań łącznie 6 przedziałów: 0 dni; 1-27 dni; 28, 29 dni, 11 miesięcy; 1-4 lata; 5-9 lat; 10-14 lat, co stanowi 6 dla roku 2016, czyli 0,04% *razem*.

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Dane	12	6	11	8	12	10	19	17	10	9	6	5	9	5	6

Układ danych empirycznych oraz przebieg trendu wielomianowego 4-go stopnia oraz obliczone wartości modelowe i określenie prognozy na rok 2017 pokazano na rysunku 4.16.



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Z	12	6	11	8	12	10	19	17	10	9	6	5	9	5	6	
Z-M	11	9	9	10	12	13	14	14	13	11	9	6	5	5	8	16

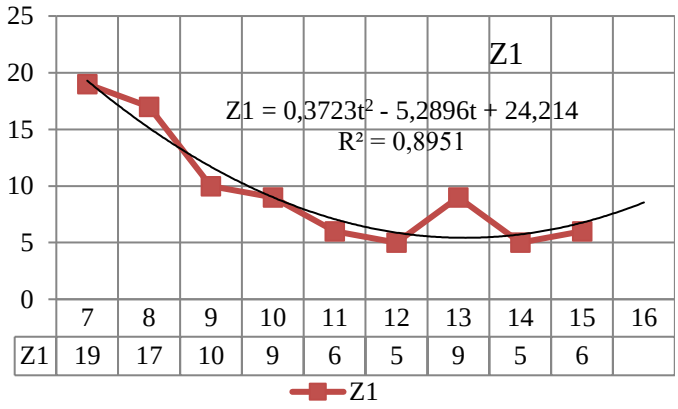
Źródło: Opracowanie własne w *Excelu*.

Rys. 4.16. Wartości rzeczywiste i teoretyczne cechy *Zgony* w latach 2002-2016

Obliczona prognoza na rok 2017, czyli okres szesnasty wynosi 16 na podstawie formuły:

=0,0081*Q1^4-0,2514*Q1^3+2,4384*Q1^2-7,8898*Q1+16,489.

Występuje duża różnica w porównaniu z okresem 15 - rok 2016, tak więc przyjmijmy krótszy szereg czasowy (7-15) - zob. rysunek 4.17.



Źródło: Opracowanie własne w *Excelu*.

Rys. 4.17. Trend wielomianowy 2-go stopnia cechy *Zgony* w latach 2008-2016

Na podstawie nowego wzoru trendu wielomianowego 2-stopnia prognoza dla roku 2017 wynosi 9. Formułę *Excela* na obliczenie tej wartości prognozy pokazano na rysunku 4.18.

Q5 =0,3723*Q4^2-5,2896*Q4+24,214

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
2	Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
3	Dane	12	6	11	8	12	10	19	17	10	9	6	5	9	5	6	
4							t	1	2	3	4	5	6	7	8	9	10
5							Z1-M	19	15	12	9	7	6	5	6	7	9

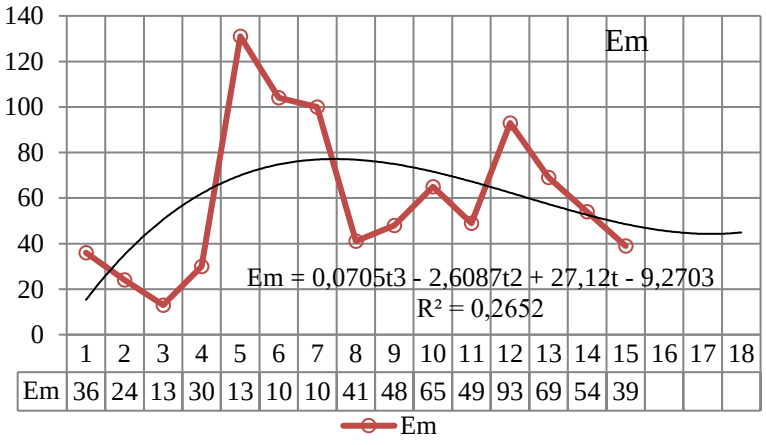
Źródło: Opracowanie własne w *Excelu*.

Rys. 4.18. Przykład danych cechy *Zgony* i formuła na obliczenie prognozy na rok 2017

Emigracja, *Nyski - powiat, według stan cywilny*. Przyjęto do badań *żonaci* (*zamężne*), co stanowi 39, czyli 37,9% *Razem* dla roku 2016.

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Dane	36	24	13	30	131	104	100	41	48	65	49	93	69	54	39

Spośród wykresów trendów dostępnych w Excelu najlepsze dopasowanie do danych rzeczywistych ma trend wielomianowy 3-go stopnia (zob. rysunek 4.19).



Źródło: Opracowanie własne w *Excelu*.

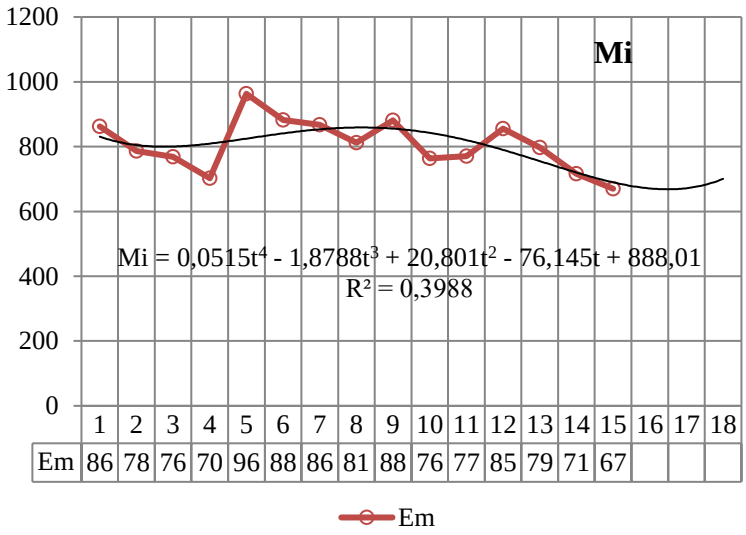
Rys. 4.19. Trend wielomianowy 3-go stopnia cechy *Emigracja* w latach 2002-2016

Obliczone na podstawie wzoru wielomianu prognozy wynoszą: $Em_{16} = 46$, $Em_{17} = 44$, $Em_{18} = 45$.

Migracje - odpływ, *Nyski - powiat*, według *stan cywilny*. Przyjęto do badań *żonaci* (*zamężne*), co stanowi 670, czyli 46,3% *Razem* dla roku 2016.

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Rok	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
Dane	862	786	768	702	963	882	867	812	881	764	770	855	797	716	670

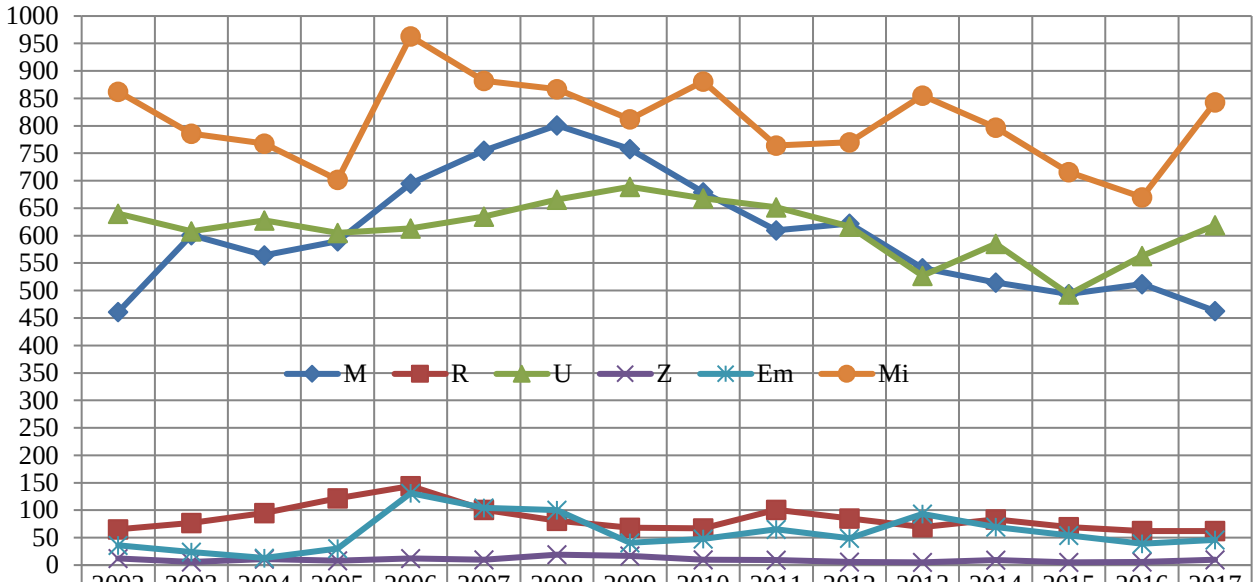
Dane z zakresu cechy *Migracja - odpływ* zaprezentowano na rysunku 8.19. Prognozy określone na podstawie wzoru wielomianu 4-go stopnia są następujące: $Mi_{16} = 674$, $Mi_{17} = 676$, $Mi_{18} = 706$.



Źródło: Opracowanie własne w *Excelu*.

Rys. 4.20. Trend wielomianowy 4-go stopnia cechy *Migracja - odpływ* w latach 2002-2016

Przedstawmy teraz wszystkie zmienne czynniki na jednym wspólnym wykresie dla w miarę stabilnego przedziału czasowego (2008-2016) i horyzontu prognozy 1, aby zobaczyć ewentualne zależności (zob. rysunek 4.21).



	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017
M	461	601	564	590	695	755	801	758	679	610	622	541	515	494	512	463
R	65	77	95	122	144	101	81	68	67	101	85	69	83	69	62	62
U	640	608	628	605	613	635	666	689	668	652	617	527	585	493	563	619
Z	12	6	11	8	12	10	19	17	10	9	6	5	9	5	6	10
Em	36	24	13	30	131	104	100	41	48	65	49	93	69	54	39	46
Mi	862	786	768	702	963	882	867	812	881	764	770	855	797	716	670	843

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.21. Dane empiryczne cech przyczyn i skutków Mi oraz ich prognozy na rok 2017

Możemy zaobserwować, że wzrostowi liczby małżeństw oraz urodzeń dzieci towarzyszy wzmożony odpływ ludności - migracja oraz emigracja. Obliczamy teraz wzajemne współczynniki korelacji między cechami dla szeregu czasowego (2002-2016) - zob. rysunek 4.22.

G2 =WSP.KORELACJI(B9:P9;B14:P14)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1		M	R	U	Z	Em	Mi									
2	M	1,00	0,29	**0,66**	**0,66**	**0,46**	**0,51**									
3	R		1,00	0,13	**0,46**	**0,42**	0,26									
4	U			1,00	**0,75**	-0,04	**0,39**									
5	Z				1,00	0,23	**0,48**									
6	Em					1,00	**0,68**									
7	Mi						1,00									
8																
9	M	461	601	564	590	695	755	801	758	679	610	622	541	515	494	512
10	R	65	77	95	122	144	101	81	68	67	101	85	69	83	69	62
11	U	640	608	628	605	613	635	666	689	668	652	617	527	585	493	563
12	Z	12	6	11	8	12	10	19	17	10	9	6	5	9	5	6
13	Em	36	24	13	30	131	104	100	41	48	65	49	93	69	54	39
14	Mi	862	786	768	702	963	882	867	812	881	764	770	855	797	716	670

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.22. Współczynniki korelacji szeregów czasowych badanych cech statystycznych

Zaobserwujmy współczynniki korelacji zaznaczone czcionką pogrubioną. Widzimy istotną zależność emigracji od liczby zawartych małżeństw(*M*) w przedziale wiekowym 20-34 lata oraz liczby rozwodów (*R*) w grupie mężczyzn w dniu wniesienia powództwa również w tym samym przedziale. Odnośnie cechy *Migracja - odpływ* obserwujemy istotną zależność małżeństw (*M*), urodzeń

chłopców (*U*) oraz zgonów (*Z*) dzieci w grupie wiekowej (0-14 lat). W miarę wysoka jest też zależność korelacyjna cech *Migracja - odpływ* od *Emigracja*.

4.3. Założenie do cech przyczyny i skutku

Biorąc powyższe pod uwagę przyjmijmy wstępnie, że migracja (odpływ) ludności z powiatu nyskiego jest funkcją liniową wielowymiarową czynników:
Migracja - odpływ (żonaci/zamężne) = f[*Emigracja* (żonaci/zamężne); *Małżeństwa* (grupa wieku mężczyzny, przedziały: 20-24, 25-29, 30-34); *Rozwody* (wg grupa wieku mężczyzny w dniu wniesienia powództwa, przedziały: 20-24, 25-29, 30-34); *Urodzenia* (*wg* płeć dziecka, chłopiec); *Zgony* (*wg* grupa wieku, przedziały: 0 dni; 1-27 dni; 28, 29 dni, 11 miesięcy; 1-4 lata; 5-9 lat; 10-14 lat)]. Tak więc ogólny model regresji liniowej wielowymiarowej jest postaci:

$$Mi = a_0 + a_1 Em + a_2 M + a_3 R + a_4 U + a_5 Z$$

Skorzystajmy teraz z funkcji REGLINP Excela dla znalezienia współczynników a_i regresji liniowej wielowymiarowej (zob. rysunek 4.23).

F9 {=REGLINP(B1:P1;B2:P6;PRAWDA;PRAWDA)}

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	Mi	862	786	768	702	963	882	867	812	881	764	770	855	797	716	670
2	Em	36	24	13	30	131	104	100	41	48	65	49	93	69	54	39
3	M	461	601	564	590	695	755	801	758	679	610	622	541	515	494	512
4	R	65	77	95	122	144	101	81	68	67	101	85	69	83	69	62
5	U	640	608	628	605	613	635	666	689	668	652	617	527	585	493	563
6	Z	12	6	11	8	12	10	19	17	10	9	6	5	9	5	6
7																
8	a_5	a_4	a_3	a_2	a_1	a_0										
9	0,035	0,943	-0,37	-0,22	2,079	269,4	parametry									
10	6,119	0,55	0,736	0,252	0,655	250,9	błąd standardowy oszacowania									
11	**0,677**	56,11	#N/D!	#N/D!	#N/D!	#N/D!	R^2, S_e									
12	3,773	9	#N/D!	#N/D!	#N/D!	#N/D!	F, iss									
13	59400	28340	#N/D!	#N/D!	#N/D!	#N/D!	ESS, RSS									

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.23. Zastosowanie funkcji REGLINP *Excela* do wyznaczenia regresji liniowej wielowymiarowej

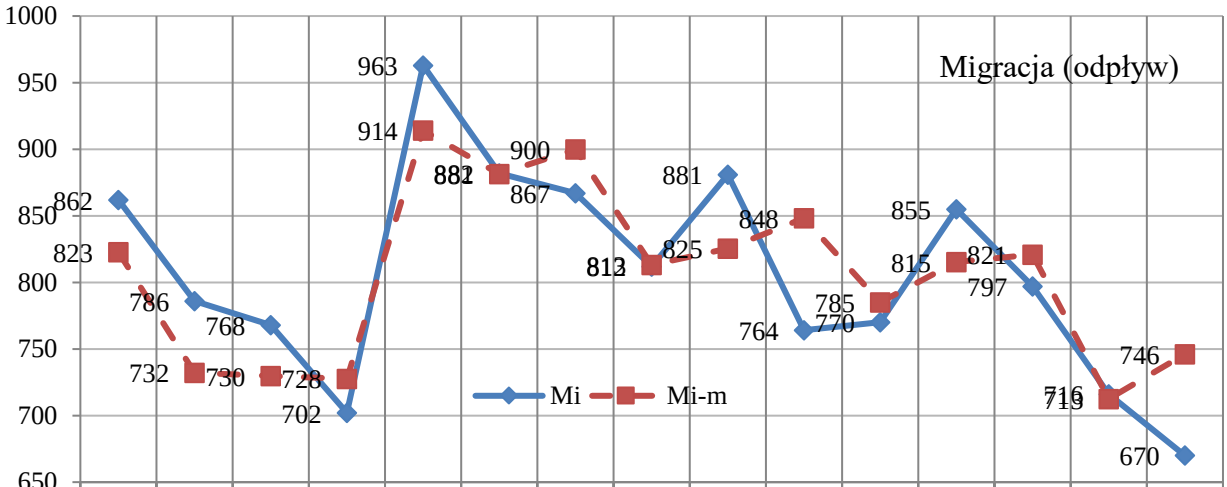
Objaśnienie występujących symboli:

R^2 - współczynnik determinacji,
S_e - odchylenie standardowe reszt,
F - statystyka Fishera,
iss - liczba stopni swobody (*n* - *m* - 1),
Ess - wyjaśniona przez model suma kwadratów,
Rss - resztowa suma kwadratów.

Tak więc model liniowy wielowymiarowy dla liczby obserwacji *n* = 15 i ilości zmiennych (cech) *m* = 5 oraz ze stałą a_0, przy zadawalającym współczynniku determinacji R^2 = 0,677 jest następujący:

$$Mi = 269{,}4 + 2{,}079\,Em - 0{,}22M - 0{,}37R + 0{,}943U + 0{,}035Z$$

Mając model cechy złożonej jaką jest (*migracja - odpływ*) oraz prognozy cech będących przyczynami tego skutku obliczmy prognozę na rok 2017, która wynosi 843. Zobaczmy teraz jak dla modelu liniowego wielowymiarowego kształtują się dane rzeczywiste i wygasłe prognozy obliczone na podstawie oszacowanego modelu (zob. rysunek 4.24).



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	20012	2013	2014	2015	2016
Mi	862	786	768	702	963	882	867	812	881	764	770	855	797	716	670
Mi-m	823	732	730	728	914	881	900	813	825	848	785	815	821	713	746

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.24. Porównanie danych rzeczywistych i modelowych cechy *migracja - odpływ*

Przyjmijmy, że prognoza na rok 2017 będzie średnią z wartości obserwacji okresu poprzedniego, wynikającą z modelu wielomianowego 4-go stopnia oraz otrzymaną z modelu ekonometrycznego:

(Mi + Mi-me + Mi-m)/3 = (670 + 674 + 843)/3 = 729.

4.4. Prognozowanie migracji (odpływ) poprzez zastosowanie wag harmonicznych

Przystępujemy teraz do określenia prognozy migracji (odpływu) ludności z powiatu nyskiego z zastosowaniem wag harmonicznych. Prognozowanie cechy złożonej typu *Mi* polega na zastosowanie trendu pełzającego, *postarzaniu* wartości danych okresów minionych, prognozowaniu np. na następny okres (n + 1), gdzie: n = 15 (kolejne lata 2002-2016). Wprowadźmy teraz szereg czasowy (1-15) obserwacji cechy *Mi*, jako liczby osób migracji (odpływ) do nowego arkusza kalkulacyjnego, w ramach wcześniej użytego skoroszytu np. *Małżeństwa - wykres2.xls*. Przyjmujemy długość segmentu $k = 3$ i za pomocą funkcji REGLINP Excela dokonujemy oszacowania parametrów trendów liniowych podszeregów 3-elementowych (zob. rysunek 8.24). Na tym rysunku pokazano estymowane wartości parametrów a_0 i a_1 poszczególnych funkcji trendów liniowych szeregów czasowych o postaci ogólnej:

$$y = a_0 + a_1 t.$$

Dalszym krokiem procedury jest obliczenie wartości teoretycznych cechy (*Migracja - odpływ*) dla poszczególnych okresów kolejnych podszeregów (1-13) - zob. kolumny L-N na rysunku 4. 25. Cecha (*Migracja - odpływ*) określana jest jako liczba osób opuszczających dany rejon (powiat nyski). Zaokrąglamy więc wartości do liczb całkowitych (zob. kolumny (O-Q) na rysunku 4.25. Przy obliczaniu parametrów korzystamy z formuły Excela np.:

=REGLINP(D2:F2;G2:I2;PRAWDA;FAŁSZ.

Oznacza to podanie zakresu zmiennej (cechy), przedziału okresów czasowych *t* oraz żądanie podania stałej a_0 i rezygnację ze statystyki opisowej danego modelu. W obliczeniach wartości teoretycznych (modelowych) tworzymy formuły *Excela* np. dla wartości w komórce L2:

=899,3-47*G2.

Pamiętać jednak musimy, że do akceptacji funkcji REGLINP np. na laptopie musimy użyć trzech klawiszy jednocześnie:

[<Shift>+ <Ctrl>] lewe klawisze+ [<Enter>] prawy klawisz.

J2 {=REGLINP(D2:F2;G2:I2;PRAWDA;FAŁSZ)}

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	i	$t_i - t_{i+2}$	y_i, y_{i+1}, y_{i+2}	M_i - podszeregi			t_1	t_2	t_3	a_1	a_0	M_{i-m}			M_{i-m} (całk.)		
2	1	1 - 3	y_1, y_2, y_3	862	786	768	1	2	3	-47,0	899,3	852,3	805,3	758,3	852	805	758
3	2	2 - 4	y_2, y_3, y_4	786	768	702	2	3	4	-42,0	878,0	794,0	752,0	710,0	794	752	710
4	3	3 - 5	y_3, y_4, y_5	768	702	963	3	4	5	97,5	421,0	713,5	811,0	908,5	714	811	909
5	4	4 - 6	y_4, y_5, y_6	702	963	882	4	5	6	90,0	399,0	759,0	849,0	939,0	759	849	939
6	5	5 - 7	y_5, y_6, y_7	963	882	867	5	6	7	-48,0	1192,0	952,0	904,0	856,0	952	904	856
7	6	6 - 8	y_6, y_7, y_8	882	867	812	6	7	8	-35,0	1098,7	888,7	853,7	818,7	889	854	819
8	7	7 - 9	y_7, y_8, y_9	867	812	881	7	8	9	7,0	797,3	846,3	853,3	860,3	846	853	860
9	8	8 - 10	y_8, y_9, y_{10}	812	881	764	8	9	10	-24,0	1035,0	843,0	819,0	795,0	843	819	795
10	9	9 - 11	y_9, y_{10}, y_{11}	881	764	770	9	10	11	-55,5	1360,0	860,5	805,0	749,5	861	805	750
11	10	10 - 12	y_{10}, y_{11}, y_{12}	764	770	855	10	11	12	45,5	295,8	750,8	796,3	841,8	751	796	842
12	11	11 - 13	y_{11}, y_{12}, y_{13}	770	855	797	11	12	13	13,5	645,3	793,8	807,3	820,8	794	807	821
13	12	12 - 14	y_{12}, y_{13}, y_{14}	855	797	716	12	13	14	-69,5	1692,8	870,8	802,3	733,8	871	802	734
14	13	13 - 15	y_{13}, y_{14}, y_{15}	797	716	670	13	14	15	-63,5	1616,7	791,2	727,7	664,2	791	728	664
15	14			716													
16	15			670													

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.25. Oszacowanie parametrów trendu liniowego segmentów oraz obliczenie wartości teoretycznych (modelowych) cechy *Migracja - odpływ*

Przystępujemy teraz do obliczenia średnich arytmetycznych z wartości teoretycznych dla danego okresu, a następnie przyrostów funkcji trendu według ogólnego wzoru[87]:

$$w_t = y_t^w - y_{t-1}^w.$$

Wymaga to jednak przedstawienia pionowo podszeregów spośród i = 1-13. Dla podszeregów 2-12 mamy po trzy wartości obliczone na podstawie wzorów trendów segmentowych. Wyznaczamy stały przelicznik dla wag jako: $1/(n - 1) = 1/(15 - 1) \approx 0{,}07$ (zob. rysunek 4.26) korzystając z formuły → =1/(15-1). W określeniu wag dla obliczonych i-tych przyrostów (zob. kolumna *Różnica*) korzystamy ze wzoru Z. Hellwiga[88]:

$$C_{t-1}^n = \frac{1}{n-1}\sum_{i=1}^{t}\frac{1}{n-i}$$

gdzie: n - ilość okresów badanego szeregu czasowego (1-15), i - numer kolejny podszeregu różnic.

Musimy teraz określić składowe przeliczniki segmentów i różnic stosując wzór: $1/(n - i)$ i formułę np. =1/(15-A5), co równa się $[1/(15 – 3)] \approx 0{,}08$ w komórce Q5. Na przykład, różnicę w komórce P5 ustalono według formuły: =760-741. W korzystaniu z formuł *Excela* musimy pamiętać, że obraz widoczny w komórce wynika z naszego sformatowania i zwany jest *wartością* komórki natomiast *zawartość komórki* jest dokładniejsza do wielu miejsc po przecinku w zależności od wersji używanego arkusza kalkulacyjnego *Excel*. Podaję to dlatego, bo np. w kolumnie *R* występują czasem

[87] Nowak E., *Zaawansowana rachunkowość zarządcza*, Polskie wydawnictwo Ekonomiczne, Warszawa 2017, s. 77.
[88] Ibidem.

te same wartości mimo wykonanego kolejnego działania. Jeśli komputerowo przemnożymy wartość w komórce S2 przez wartość w komórce Q3 to otrzymamy 0,005102, co zostało zaokrąglone do drugiego miejsca jako 0,01 (patrz komórka R3).

Q21 =P20+$R$20

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S
1	i'	M_1-m	M_2-m	M_3-m	M_4-m	M_5-m	M_6-m	M_7-m	M_8-m	M_9-m	M_{10}-m	M_{11}-m	M_{12}-m	M_{13}-m	$(M_i\text{-m})'$	Różnica	Prz. i-ty	Waga	W'
2		852													852	Przelicznik	n = 15	1/(n-1)	0,07
3	1	805	794												800	-52	0,07	0,01	-0,27
4	2	758	752	714											741	-59	0,08	0,01	-0,63
5	3		710	811	759										760	19	0,08	0,02	0,31
6	4			909	849	952									903	147	0,09	0,02	3,39
7	5				939	904	889								911	4	0,10	0,03	0,12
8	6					856	854	846							852	-59	0,11	0,04	-2,25
9	7						819	853	843						838	-75	0,13	0,05	-3,53
10	8							860	819	861					847	9	0,14	0,06	0,52
11	9								795	805	751				784	-63	0,17	0,07	-4,36
12	10									750	796	794			780	-4	0,20	0,08	-0,33
13	11										842	807	871		840	60	0,25	0,10	6,08
14	12											821	802	791	805	-35	0,33	0,13	-4,38
15	13												734	728	731	-74	0,50	0,16	-11,90
16	14													664	664	-64	1,00	0,23	-14,86
17	15														W'	-32,09	Suma:	1,0000	-32,09

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.26. Obliczenie średnich teoretycznych, różnic oraz wag harmonicznych i postarzałych przyrostów dla poszczególnych segmentów cechy (*Migracja - odpływ*)

Jak już nadmieniłem, waga harmoniczna (zob. kolumna R) odnosząca się do danego przyrostu funkcji trendu stanowi iloczyn stałego przelicznika w komórce S2 oraz sumy przeliczników (z uwzględnieniem wcześniejszych) z kolumny Q - zob. rysunek 8.26, np. w komórce R5 jest wyrażenie:

=$S$2*SUMA(Q3:Q5).

Pozostaje nam teraz obliczenie wartości odchyleń ważonych wagami harmonicznymi W_i' (zob. kolumna S na rysunku 4.26), które stanowią iloczyny pól *Różnica* i *Waga*. Przykładowo dla komórki S5 jest wyrażenie: =P5*R5.

Suma składników średnich odchyleń jest średnią ważoną naszego szeregu czasowego obserwacji liczby cechy (*Migracja – odpływ*), ta więc np. w komórce S17 jest wyrażenie: =SUMA(S3:S16), które równa się (-32,09). Tą średnią ważoną możemy też obliczyć jako sumę iloczynów odpowiednich wartości w kolumnach P oraz R. Warto zwrócić uwagę, że suma wszystkich wag równa się jeden (komórka R17):

=P3*R3+P4*R4+P5*R5+P6*R6+P7*R7+P8*R8+P9*R9+P10*R10+P11*R11+P12*R12+P13*R13+P14*R14+P15*R15+P16*R16

4.5. Obliczenie prognoz

Mamy teraz wszystkie składniki na obliczenie prognozy na kolejne okresy T po $n = 15$ stosując wzór ogólny[89]:

$$Y_T^* = y_n^w + (T - n)\overline{w}.$$

W celu obliczenia przykładowo prognozy na okres 16, czyli rok 2017 o wartości 632 (komórka Q21) zastosujemy wyrażenie: =P20+$R$20. Kopiujemy nasze wyrażenie i otrzymujemy wygasłe prognozy dla okresów (1-15) - zob. rysunek 4.27.

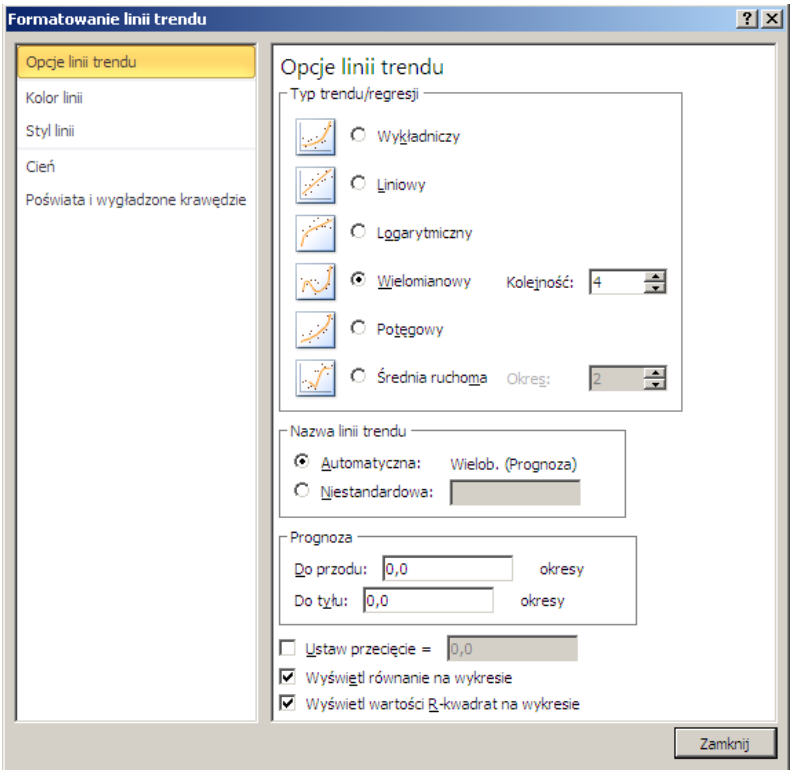
[89] Ibidem.

19	t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	W
20	(Mi-m)-	852	800	741	760	903	911	852	838	847	784	780	840	805	731	664		-32,09
21	Prognoza		820	767	709	728	871	879	820	806	815	752	748	808	773	699	632	

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.27. Wygasłe prognozy i przewidywana prognoza cechy (*Migracja - odpływ*) na rok 2017[90]

Z ciekawości zobaczmy teraz układanie się wartości rzeczywistych, średnich z trendów oraz uzyskanych prognoz. Na załączonym dalej wykresie zbiorczym dodatkowo wprowadzono wykres wielomianowy 4–stopnia dla szeregu wartości prognoz cechy (*Migracja - odpływ*). W tym względzie korzystamy z okna dialogowego pokazanego na rysunku 4.28.



Źródło: Opracowanie własne w *Excelu*.

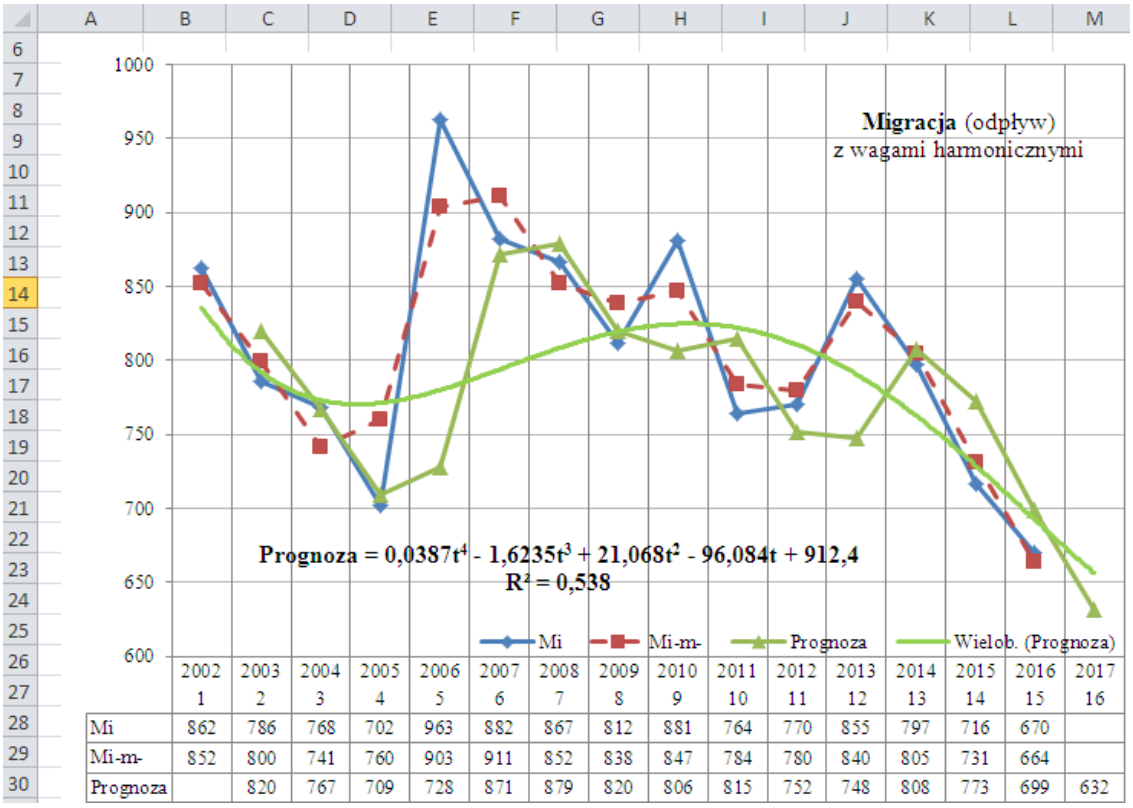
Rys. 4.28. Okno dialogowe określenia wykresu wielomianu

Możemy przyjąć np. horyzont prognozy o jeden okres do przodu, wskazano na potrzebę zapisania na wykresie równania wielomianu, a także podanie współczynnika determinacji R^2, który jest w miarę zadawalający, bo równy 0,538. Równanie wielomianu 4-stopnia jest następujące: Prognoza = $0{,}0387t^4 - 1{,}6235t^3 + 21{,}068t^2 - 96{,}084t + 912{,}4$.

Na kolejnym rysunku 4.29 zaprezentowano dla cechy (*Migracja - odpływ*) zbiorcze zestawienie wykresów:

- dane rzeczywiste,
- średnie teoretyczne wartości cechy uzyskane z trendów określonych podszeregów,
- prognozy z uwzględnieniem postarzania minionych wartości cechy,
- wykres wielomianowy 4-stopnia.

[90] Zamieszczone dane dotyczą listopada 2017 roku.



	2002 1	2003 2	2004 3	2005 4	2006 5	2007 6	2008 7	2009 8	2010 9	2011 10	2012 11	2013 12	2014 13	2015 14	2016 15	2017 16
Mi	862	786	768	702	963	882	867	812	881	764	770	855	797	716	670	
Mi-m-	852	800	741	760	903	911	852	838	847	784	780	840	805	731	664	
Prognoza		820	767	709	728	871	879	820	806	815	752	748	808	773	699	632

Źródło: Opracowanie własne w *Excelu*.

Rys. 4.29. Zbiorcze zestawienie wykresów dotyczących cechy (*Migracja - odpływ*) z uwzględnieniem wag harmonicznych

* * *

Zaprezentowano kilka sposobów modelowania cechy złożonej na przykładzie (*Migracji - odpływ*) w powiecie Nyskim. Dane statystyczne pochodzą z bazy danych D*emografia* prowadzonej przez Główny Urząd Statystyczny. Intuicyjnie założono, że liczba mieszkańców opuszczających wspomniany powiat zależna jest od sytuacji rodzinnej jaka wytwarza się w wyniku zawarcia małżeństwa, rozwodów, urodzeń dzieci, czy też ich zgonów, a ponadto od warunków zewnętrznych jakie stwarza emigracja do krajów wyżej rozwiniętych. Ukazano dobór trendów dla czynników oraz skutku, jakim jest cecha (*Migracja - odpływ*) kierując się współczynnikiem determinacji, jako miarą dopasowania danych empirycznych do proponowanego modelu. Kolejnym krokiem była próba sformułowania modelu ekonometrycznego liniowego wielowymiarowego.

Efektem końcowym niniejszego opracowania jest propozycja postępowania w zakresie możliwości prognozowania krótkookresowego, z uwzględnieniem oceny różnic wartości teoretycznych średnich kolejnych okresów wagami zwanymi harmonicznymi. Zamieszczone propozycje modelowania wspomagane techniką komputerową należy traktować jako sugestię postępowania także z innymi cechami statystycznymi demografii, które jako wynikowe zależne są od szeregu czynników dla których prowadzona jest statystyka GUS-u.

5. Prognozowanie z wykorzystaniem zasady postarzania informacji

5.1. Wprowadzenie

W ustalaniu cech statystycznych w przyszłych okresach zachodzi potrzeba uwzględnienia wartości informacji z minionego czasu. W tym celu dokonujemy wyrównywania szeregu czasowego obserwacji trendem pełzającym, a następnie przeprowadzamy tzw. postarzanie wyników z zastosowaniem wag harmonicznych. Powstaje w ten sposób model adaptacyjny obejmujący model trendu pełzającego z wagami harmonicznymi. Postępowanie w tym zakresie spotykamy w książce Edwarda Nowaka oraz na forum Internetu[91] [92] i sprowadza się ono do realizacji następujących etapów:

- określenie stałej wygładzania dla trendów segmentowych np. $k = 3$, przy $n = 12$ obserwacji;
- oszacowanie parametrów a_0 i a_1 dla liniowych funkcji trendów poszczególnych segmentów;
- obliczenie wartości teoretycznych według funkcji trendów segmentowych dla kolejnych okresów danego segmentu;
- obliczenie średnich arytmetycznych z wartości teoretycznych dla danego okresu;
- obliczenie przyrostów funkcji trendu według:

$$w_t = y_t^w - y_{t-1}^w$$

- określenie wag dla obliczonych *i*-tych przyrostów według zależności:

$$C_{t-1}^n = \frac{1}{n-1}\sum_{i=1}^{t}\frac{1}{n-i}$$

- wyznaczenie prognozy na kolejny okres *T* korzystając ze wzoru:

$$Y_T^* = y_n^w + (T-n)\overline{w}$$

W obliczenia według podanej procedury warto korzystać z funkcji elementarnych *Excela* oraz funkcji regresji liniowej REGLINP o formacie: =REGLINP(zakres y; zakres x, PRAWDA; PRAWDA).

gdzie:

zakres *y* - szereg czasowy zmiennej objaśnianej określonego segmentu,
zakres *x* - szereg zmiennej czasowej *t*,
PRAWDA - żądanie stałej we funkcji regresji liniowej segmentu,
PRAWDA - zgoda na obliczenie podstawowych parametrów statystyki opisowej.

Dla bliższego przedstawienia metody adaptacyjnej prognozowania z uwzględnieniem wag harmonicznych zaprezentowane zostaną dwa przykłady. Dane do nich zaczerpnąłem z mojej wcześniejszej książki „*Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*”[93].

5.2. Pracujący w handlu

Dla zademonstrowania metody prognozowania z postarzaniem się informacji skorzystałem z tabeli 4.14 w zakresie zmiennej *H* „ *Liczba pracujących w handlu* (*przeciętnie w roku*) *w tysiącach*”. Dane pochodzą z roczników statystycznych GUS-u, a w ramach nich z tabeli: *Ważniejsze dane o*

[91] Nowak E., *Zaawansowana rachunkowość zarządcza*, Polskie wydawnictwo Ekonomiczne, Warszawa 2003, s. 77.
[92] http://dydaktyka.polsl.pl/roz6/mwolny/Shared%20Documents/Prognozowanie%20i%20symulacje/Repozytorium/Instrukcja_trend_pelzajacy_PIS.pdf.
[93] Wornalkiewicz W., *Metoda badania przyczynowo-skutkowego związków miedzy cechami statystycznymi*, Skrypt Wyższej Szkoły Zarządzania i Administracji, Opole 2013.

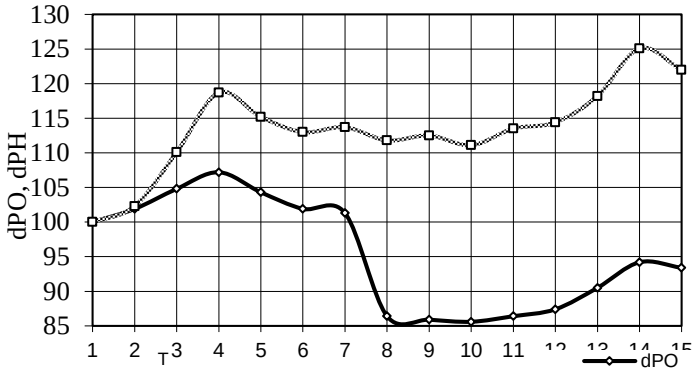
sytuacji społeczno-gospodarczej kraju, handel, str. 54-55, poz. 9, pracujący - przeciętne w roku w tys. Dynamikę pracujących ogółem dPO oraz w handlu dPH określono przyjmując dane roku 1995 = 100.

Tab. 5.1. Dynamika pracujących w handlu (przeciętne w roku) w tys.

T	Rok	dPO	H	dPH
1	1995	100,0	1858	100,0
2	1996	101,9	1901	102,3
3	1997	104,8	2046	110,1
4	1998	107,2	2205	118,7
5	1999	104,3	2140	115,2
6	2000	101,9	2100	113,0
7	2001	101,3	2113	113,7
8	2002	86,4	2078	111,8
9	2003	85,9	2090	112,5
10	2004	85,6	2065	111,1
11	2005	86,4	2108	113,5
12	2006	87,4	2126	114,4
13	2007	90,5	2196	118,2
14	2008	94,2	2324	125,1
15	2009	93,4	2267	122,0

Źródło: Opracowanie własne w *Excelu* na podstawie roczników GUS-u.

Kształtowanie się dynamiki zmiennych dPO oraz dPH pokazano na rysunku 5.1, która stanowi wykres 4.14 w wymienionej wcześniej publikacji[94].



Źródło: Opracowanie własne w *Excelu*.

Rys. 5.1. Dynamika pracujących ogółem i w handlu (przeciętne w roku) w tys.

Na podstawie dynamik dPO i dPH możemy w *Excelu* dobrać trendy wielomianowe takie, aby współczynnik determinacji R^2 był możliwie jak najlepszy:

dPO	$y = 0{,}0682t^3 - 1{,}5224t^2 + 7{,}7586t + 93{,}74$	$R^2 = 0{,}8374$
dPH	$y = 0{,}0522t^3 - 1{,}2732t^2 + 9{,}6836t + 91{,}079$	$R^2 = 0{,}8287$

Widzimy, że zarówno dla dynamiki dPO oraz dPH są to trendy wielomianowe trzeciego stopnia z dobrym dopasowanie modeli do danych empirycznych. Przystąpmy teraz przykładowo do prognozowania liczby pracujących w handlu na kolejne trzy okresy, tj. 13, 14, 15 z uwzględnieniem metody wag harmonicznych autorstwa Zdzisława Hellwiga[95]. Autor ten był profesorem Akademii

[94] Ibidem, s. 34.
[95] https://pl.wikipedia.org/wiki/Zdzis%C5%82aw_Hellwig.

Ekonomicznej we Wrocławiu. Przez wiele lat pełnił funkcję dyrektora Instytutu Cybernetyki Ekonomicznej i był inicjatorem utworzenia Wydziału Zarządzania i Informatyki. Jest twórcą szkoły naukowej statystyki, ekonometrii i cybernetyki. Powróćmy jednak do naszego przykładu i wprowadzamy wybrany szereg 12 liczb (lata 1998-2009) pracujących w handlu w tysiącach do kolumny np. E arkusza kalkulacyjnego Excel (zob. rysunek 5.2).

E
Handel-pracujący
2 205
2 140
2 100
2 113
2 078
2 090
2 065
2 108
2 126
2 196
2 324
2 267

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.2. Dane wejściowe pracujących w handlu [tys.]

Dla przyjętej długości segmentu k = 3 korzystamy z funkcji REGLINP i dokonujemy estymacji parametrów a_0 i a_1 poszczególnych funkcji trendów liniowych szeregów czasowych o postaci ogólnej: $y = a_0 + a_1 t$. Zrealizowane prace w ramach tego etapu pokazano na rysunku 5.3. Wymaga to jednak przedstawienia pionowo podszeregów spośród i = 1-10 i skorzystania z funkcji o przykładowej formule: =REGLINP(I26:I28;A26:A28;PRAWDA;PRAWDA).

	A	B	C	D	E	F	G
16	i	t_i - t_{i+2}	y_i, y_{i+1}, y_{i+2}	Y - (Han-prac)			$f_i(t)$
17	1	1 - 3	y_1, y_{1+1}, y_{i+2}	2 205	2 140	2 100	2253,3-52,5t
18	2	2 - 4	y_2, y_{2+1}, y_{2+2}	2 140	2 100	2 113	2158,2-13,5t
19	3	3 - 5	y_3, y_{3+1}, y_{3+2}	2 100	2 113	2 078	2141,0-11,0t
20	4	4 - 6	y_4, y_{4+1}, y_{4+2}	2 113	2 078	2 090	2151,2-11,5t
21	5	5 - 7	y_5, y_{5+1}, y_{5+2}	2 078	2 090	2 065	2116,7-6,5t
22	6	6 - 8	y_6, y_{6+1}, y_{6+2}	2 090	2 065	2 108	2024,7+9,0t
23	7	7 - 9	y_7, y_{7+1}, y_{7+2}	2 065	2 108	2 126	1855,7+30,5t
24	8	8 - 10	y_8, y_{8+1}, y_{8+2}	2 108	2 126	2 196	1747,3+44,0t
25	9	9 - 11	y_9, y_{9+1}, y_{9+2}	2 126	2 196	2 324	1225,3+99,0t
26	10	10 - 12	y_{10}, y_{10+1}, y_{10+2}	2 196	2 324	2 267	1871,8+35,5t
27	11			2 324			
28	12			2 267			

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.3. Określone funkcje trendów liniowych 10-ciu segmentów

Efekt skorzystania z funkcji REGLINP dla segmentu dziesiątego pokazano na rysunku 5.4. Trzeba jednak pamiętać, że akceptacja funkcji wymaga skorzystania z trzech klawiszy jednocześnie:

<Ctrl>+ <Shift>+<Enter>.

J	K	L
a_1	a_0	
35,5	1871,8	parametry
53,4	589,1	błąd standardowy
0,3	75,5	R^2; S_e
0,4	1,0	F; iss
2520,5	5704,2	ESS, RSS

gdzie:

a_0, a_1 - parametry modelu trendu liniowego,
błąd standardowy - oszacowania parametrów.
R^2 - współczynnik determinacji,
Se - odchylenie standardowe reszt,
F - statystyka Fishera,
iss - ilość stopni swobody,
ESS - estymowana suma kwadratów,
RSS - resztowa suma kwadratów.

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.4. Rezultat działania funkcji *REGLINP* dla segmentu 10-tego

Mając modele trendów liniowych poszczególnych 10-ciu segmentów obliczamy wartości teoretyczne liczby pracujących w tys. w poszczególnych okresach *t*. Następnie obliczamy średnie arytmetyczne $\bar{y}_t$ (zob. rysunek 5.5).

	A	B	C	D	E	F	G	H	I	J	K	L
14	t	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	$\bar{y}_t$
15	1	2201										2 201
16	2	2148	2131									2 140
17	3	2096	2118	2108								2 107
18	4		2104	2097	2105							2 102
19	5			2086	2094	2084						2 088
20	6				2082	2078	2079					2 080
21	7					2071	2088	2069				2 076
22	8						2097	2100	2099			2 099
23	9							2130	2143	2116		2 130
24	10								2187	2215	2227	2 210
25	11									2314	2262	2 288
26	**12**										2298	**2 298**

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.5. Obliczone wartości teoretyczne oraz średnie dla okresów

Dla pierwszego okresu i pierwszej funkcji jest tylko jedna liczba, stąd średnia równa się 2201.

L15 f_x =B15

	L	M	N
14	$\bar{y}_t$	y_t (Han-prac)	
15	2 201	2 205	Różnica

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.6. Przykład wyrażenia (=B15) na określenie średniej

Dla okresu trzeciego mamy już trzy wyniki teoretyczne, stąd wyrażenie komputerowe jest następujące: =(B17+C17+D17)/3.

L17 =(B17+C17+D17)/3

	L	M	N	O
17	2 107	2 100	-33	0,1000

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.7. Przykład wyrażenia na obliczenie średniej z 3-ch wartości teoretycznych

W określeniu wag harmonicznych dla naszego przykładu występuje stały przelicznik: $1/(n - 1) = 1/(12 - 1)$.

P12 =1/($A$26-1)

	N	O	P
12	Przelicznik	1/(n-1)	0,0909

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.8. Wyznaczenie stałego przelicznika wag

Obliczamy teraz przyrosty funkcji trendu nazwane „*Różnica*" (zob. rysunek 5.9 oraz rysunek 5.10).

	L	M	N	O	P	Q	R
14	$\bar{y}_t$	y_t (Han-prac)					
15	2 201	2 205	Różnica	Przelicznik i-ty	Waga	i	$\bar{W}_i$
16	2 140	2 140	-61	0,0909	0,0083	1	-0,50
17	2 107	2 100	-33	0,1000	0,0174	2	-0,57
18	2 102	2 113	-5	0,1111	0,0275	3	-0,14
19	2 088	2 078	-14	0,1250	0,0388	4	-0,55
20	2 080	2 090	-8	0,1429	0,0518	5	-0,44
21	2 076	2 065	-4	0,1667	0,0670	6	-0,23
22	2 099	2 108	23	0,2000	0,0851	7	1,92
23	2 130	2 126	31	0,2500	0,1079	8	3,38
24	2 210	2 196	80	0,3333	0,1382	9	11,04
25	2 288	2 324	79	0,5000	0,1836	10	14,41
26	**2 298**	2 267	10	1,0000	0,2745	11	2,61
27	y^*_T	$\bar{W}$	**30,93**	Suma wag:	1,0000	Suma:	**30,93**
28	2 329						
29	2 360						
30	2 391						

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.9. Wyznaczenie średniej ważonej wagami harmonicznymi

N17 =L17-L16

	L	M	N
14	$\bar{y}_t$	y_t (Han-prac)	
15	2 201	2 205	Różnica
16	2 140	2 140	-61
17	2 107	2 100	-33

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.10. Przykład wyznaczenia przyrostu funkcji trendu

W kolejności obliczamy dalsze przeliczniki „*Przelicznik i-ty*" z uwzględnieniem wartości ***i*** w kolumnie Q (zob. rysunek 5.11).

O17 =1/($A$26-Q17)

	L	M	N	O	P	Q	R
14	$\bar{y}_t$	y_t (Han-prac)					
15	2 201	2 205	Różnica	Przelicznik i-ty	Waga	i	$\bar{W}_i$
16	2 140	2 140	-61	0,0909	0,0083	1	-0,50
17	2 107	2 100	-33	0,1000	0,0174	2	-0,57

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.11. Przykład wyznaczenia przelicznika dla $i = 2$

Jak już nadmieniłem, waga harmoniczna odnosząca się do danego przyrostu funkcji trendu stanowi iloczyn stałego przelicznika w komórce P12 oraz sumy przeliczników (z uwzględnieniem wcześniejszych) - zob. rysunek 5.12, wiersz 17 o wyrażeniu: =P12*SUMA(Q16:Q17).

P17 =P12*SUMA(O16:O17)

	L	M	N	O	P
12			Przelicznik	1/(n-1)	0,0909
13					
14	$\bar{y}_t$	y_t (Han-prac)			
15	2 201	2 205	Różnica	Przelicznik i-ty	Waga
16	2 140	2 140	-61	0,0909	0,0083
17	2 107	2 100	-33	0,1000	0,0174

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.12. Przykład wyznaczenia wagi dla $i = 2$

Pozostaje nam teraz obliczenie wartości odchyleń ważonych wagami harmonicznymi w_i' (zob. kolumna R na rysunku 5.13. Stanowią one iloczyny pól „Różnica" i „Waga".

R17 =N17*P17

	L	M	N	O	P	Q	R
14	$\bar{y}_t$	y_t (Han-prac)					
15	2 201	2 205	Różnica	Przelicznik i-ty	Waga	i	$\bar{W}_i$
16	2 140	2 140	-61	0,0909	0,0083	1	-0,50
17	2 107	2 100	-33	0,1000	0,0174	2	-0,57

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.13. Przykład wyznaczenia składnika średniego odchylenia dla $i = 2$

Suma składników średnich odchyleń jest średnią ważoną naszego szeregu czasowego obserwacji liczby pracowników pracujących w handlu tj. po zaokrągleniu wartość 31.

R27 =SUMA(R16:R26)

	R	S	T	U
27	**30,93**			

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.14. Wyrażenie na sumowanie składników średnich

Tą średnią ważoną możemy też obliczyć jako sumę iloczynów odpowiednich wartości w kolumnach N oraz P. Warto zwrócić uwagę, że suma wszystkich wag równa się jeden (zob. rysunek 5.15).

N27 =N16*P16+N17*P17+N18*P18+N19*P19+N20*P20+N21*P21+N22*P22+N23*P23+N24*P24+N25*P25+N26*P26

	L	M	N	O	P	Q	R	S	T	U
27	y^*_T	$\bar{W}$	**30,93**	Suma wag:	1,0000	Suma:	**30,93**			

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.15. Wyrażenie obliczenie średniej ważonej wagami harmonicznymi (komórka N27)

Przystępujemy teraz do obliczenia prognoz dla przykładowych okresów T: $n + 1$, $n + 2$, $n + 3$, przy czym $n = 12$. Dla okresu 13. zastosujemy wyrażenie: =$L$26+(K28-$A$26)*$N$27. Analogicznie obliczamy dalsze prognozy dla okresu 14. i okresu 15.

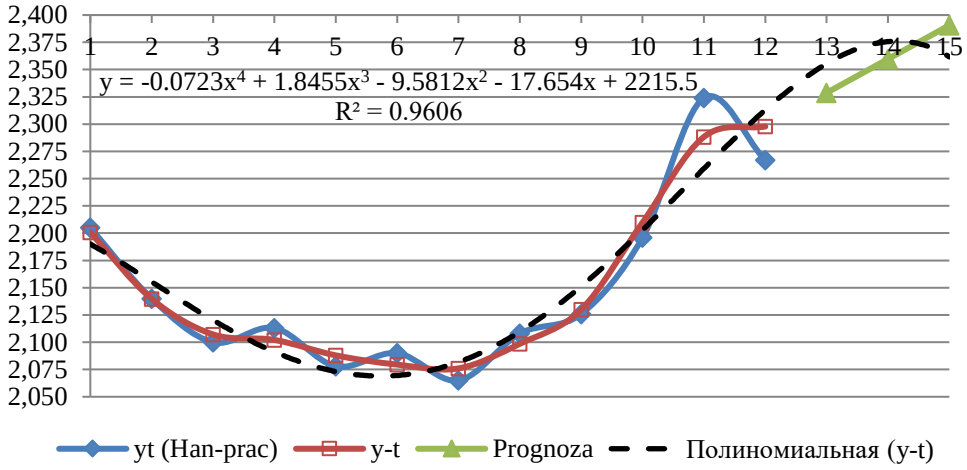
L28 =$L$26+(K28-$A$26)*$N$27

	K	L	M	N
27	*t*	y^*_T	$\bar{W}$	**30,93**
28	**13**	2 329		

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.16. Wyrażenie obliczenie prognozy na okres 13

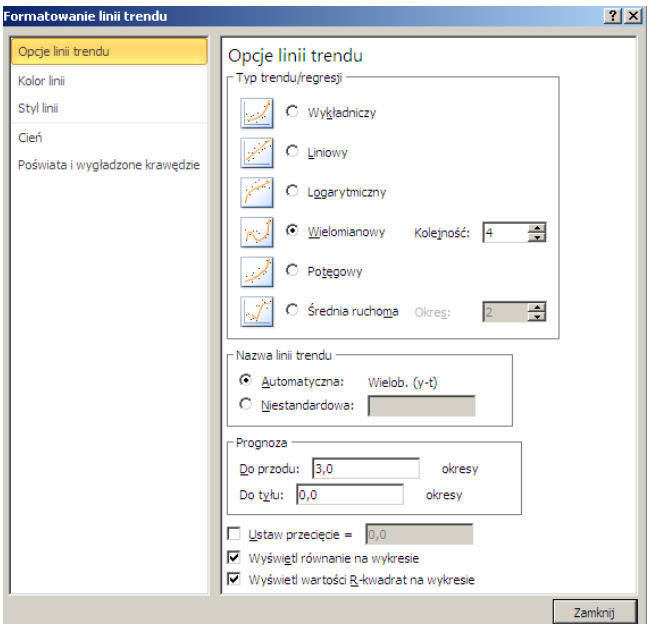
Z ciekawości zobaczmy na wykresie jak układają się graficznie wartości empiryczne szeregu czasowego liczby pracujących w handlu, wartości teoretyczne - modelowe oraz prognozy (zob. rysunek 5.17).



Źródło: Opracowanie własne w *Excelu*.

Rys. 5.17. Graficzne przedstawienie kształtowania się liczby pracujących w handlu w tys.

Program Excel umożliwia dobór stopnia wielomianu i w tym względzie korzystamy z okna dialogowego (zob. rysunek 5.18).



Źródło: Opracowanie własne w *Excelu*.

Rys. 5.18. Okno dialogowe określenia stopnia wielomianu i horyzontu prognozowania

5.3. Pracujący w transporcie, gospodarce magazynowej i łączności

Podejmujemy teraz kolejny przykład dotyczący liczby pracujących w transporcie, gospodarce magazynowej i łączności. Podobnie jak poprzednio dane zaczerpnięto z wcześniejszego skryptu autora[96]. Dane pochodzą z roczników statystycznych GUS-u (lata 1995-2009). Występują w tabelach o nazwie: *Ważniejsze dane o sytuacji społeczno-gospodarczej kraju, transport, gospodarka magazynowa i łączność*, str. 52, *pracujący - przeciętne w roku w tys.* Dodam, że dynamikę pracujących ogółem dPO, pracujących w transporcie, gospodarce magazynowej i łączności dPT określono przyjmując dane z roku 1995 = 100.

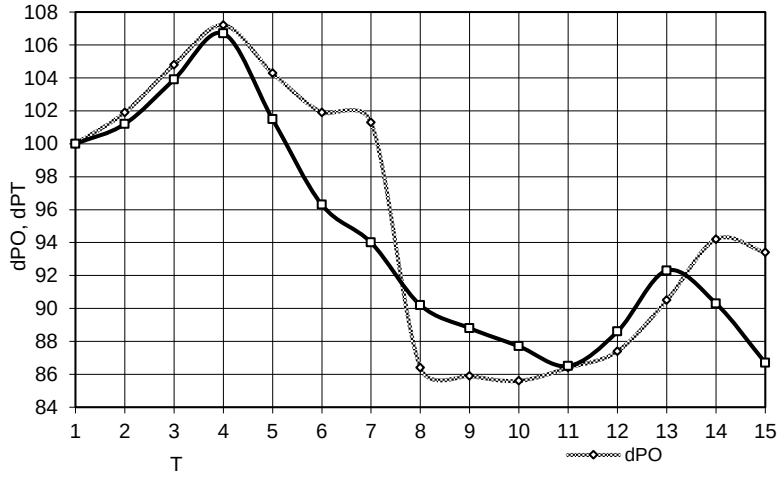
Tab. 5.2. Dynamika pracujących w transporcie, gospodarce magazynowej i łączności (przeciętne w roku) w tys.

T	Rok	dPO	T	dPT
1	1995	100,0	845	100,0
2	1996	101,9	855	101,2
3	1997	104,8	878	103,9
4	1998	107,2	902	106,7
5	1999	104,3	858	101,5
6	2000	101,9	814	96,3
7	2001	101,3	794	94,0
8	2002	86,4	762	90,2
9	2003	85,9	750	88,8
10	2004	85,6	741	87,7
11	2005	86,4	731	86,5
12	2006	87,4	750	88,6
13	2007	90,5	780	92,3
14	2008	94,2	763	90,3
15	2009	93,4	733	86,7

Źródło: Opracowanie własne w *Excelu* na podstawie roczników GUS-u.

[96] Wornalkiewicz W., *Metoda badania przyczynowo-skutkowego związków miedzy cechami statystycznymi*, op. cit., s. 37.

Dynamiki dPO i DPT przedstawiono na rysunku 5.19.



Źródło: Opracowanie własne w *Excelu*.

Rys. 5.19. Dynamika pracujących ogółem i w transporcie, gospodarce magazynowej i łączności [tys.]

Oszacowanie równań wielomianowych oraz współczynnika determinacji R^2 w *Excelu* jest następujące:

dPO	$y = 0{,}0682t^3 - 1{,}5224t^2 + 7{,}7586t + 93{,}74$	$R^2 = 0{,}8374$
dPT	$y = 0{,}0331t^3 - 0{,}728t^2 + 2{,}9117t + 99{,}412$	$R^2 = 0{,}8305$

Podobnie jak w przykładzie wcześniejszym określamy funkcje dziesięciu trendów segmentowych dla zmiennej „*Pracujący w transporcie, gospodarce magazynowej i łączności* (przeciętne w roku [*tys.*]" lecz dla czasookresu 1998-2009 (zob. rysunek 5.20).

	A	B	C	D	E	F	G
16	i	$t_i - t_{i+2}$	y_i, y_{i+1}, y_{i+2}	Y - (Tran-prac)			$f_i(t)$
17	1	1 - 3	y_1, y_{1+1}, y_{i+2}	902	858	814	946,0-44,0t
18	2	2 - 4	y_2, y_{2+1}, y_{2+2}	858	814	794	918,0-32,0t
19	3	3 - 5	y_3, y_{3+1}, y_{3+2}	814	794	762	894,0-26,0t
20	4	4 - 6	y_4, y_{4+1}, y_{4+2}	794	762	750	878,7-22t
21	5	5 - 7	y_5, y_{5+1}, y_{5+2}	762	750	741	814,0-10,5t
22	6	6 - 8	y_6, y_{6+1}, y_{6+2}	750	741	731	807,2-9,5t
23	7	7 - 9	y_7, y_{7+1}, y_{7+2}	741	731	750	704,7+4,5t
24	8	8 - 10	y_8, y_{8+1}, y_{8+2}	731	750	780	533,2+24t
25	9	9 - 11	y_9, y_{9+1}, y_{9+2}	750	780	763	699,3+6,5t
26	10	10 - 12	$y_{10}, y_{10+1}, y_{10+2}$	780	763	733	1017,2-23,5t
27	11			763			
28	12			733			

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.20. Określenie funkcji segmentowych

Przykład określenia funkcji trendu liniowego segmentu dziesiątego korzystając z *REGLINP* zaprezentowano na rysunku 5.21.

J17 | {=REGLINP(I26:I28;A26:A28;PRAWDA;PRAWDA)}

	J	K	L	M	N	O	P
15	$f_{10(3)}$						
16	a_1	a_0					
17	-23,500	1017,167	parametry				
18	3,753	41,394	błąd standardowy				
19	0,975	5,307	R^2; S_ε				
20	39,213	1,000	F; iss				
21	1104,500	28,167	ESS; RSS				

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.21. Parametry i statystyka opisowa funkcji trendu segmentu dziesiątego

Mając określone funkcje segmentowe obliczamy wartości teoretyczne dla trzech okresów każdego ze segmentów. Następnie w *Excelu* wyznaczane są średnie arytmetyczne z uzyskanych wartości teoretycznych dla każdego z dziesięciu segmentów trendu pełzającego (zob. rysunek 5.22).

B15 | =946-44*A15

	A	B	C	D	E	F	G	H	I	J	K	L
14	t	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	$\bar{y}_t$
15	1	902										902
16	2	858	854									856
17	3	814	822	816								817
18	4		790	790	791							790
19	5			764	769	762						765
20	6				747	751	750					749
21	7					741	741	736				739
22	8						731	741	725			732
23	9							745	749	758		751
24	10								773	764	782	773
25	11									771	759	765
26	**12**										735	**735**

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.22. Wartości teoretyczne i średnie segmentów trendu pełzającego

Na kolejnym rysunku 5.23 pokazano efekt obliczania wag harmonicznych.

P17 =P12*SUMA(O16:O17)

	M	N	O	P	Q	R
14	y_t (Tran-prac)					
15	902	Różnica	Suma przeliczników	Waga	i	W_i
16	858	-46	0,0909	0,0083	1	-0,38
17	814	-39	0,1000	0,0174	2	-0,67
18	794	-27	0,1111	0,0275	3	-0,74
19	762	-26	0,1250	0,0388	4	-0,99
20	750	-15	0,1429	0,0518	5	-0,80
21	741	-10	0,1667	0,0670	6	-0,68
22	731	-7	0,2000	0,0851	7	-0,58
23	750	18	0,2500	0,1079	8	1,98
24	780	23	0,3333	0,1382	9	3,11
25	763	-8	0,5000	0,1836	10	-1,56
26	733	-30	1,0000	0,2745	11	-8,11
27	W	**-9,42**	Suma wag:	1,0000	Suma:	**-9,42**

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.23. Obliczenia pomocnicze (różnice, przeliczniki, wagi, wartości średnie dla segmentów)

Analogicznie jak w przykładzie wcześniejszym określamy teraz prognozy dla okresów: 13, 14, 15.

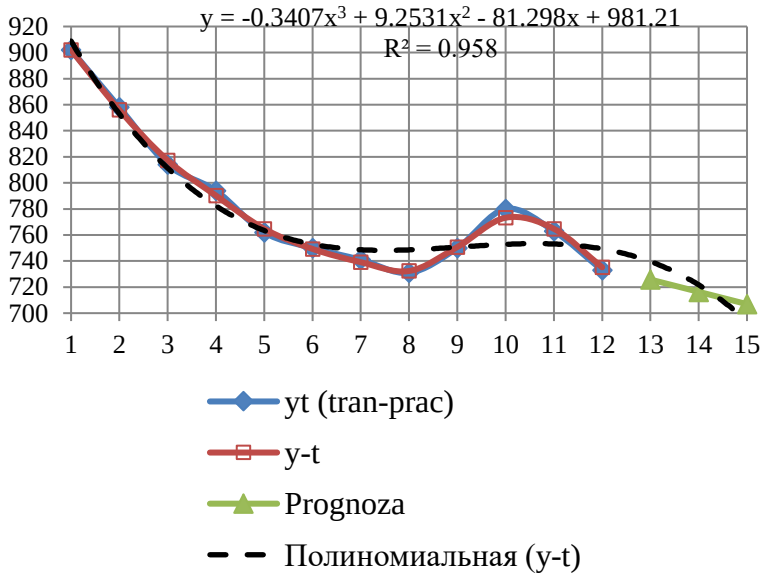
L28 =$L$26+(K28-$A$26)*$N$27

	K	L	M	N
27	*t*	y^*_T	W	**-9,42**
28	**13**	726		
29	14	716		
30	15	707		

Źródło: Opracowanie własne w *Excelu*.

Rys. 5.24. Prognozy z uwzględnieniem wag harmonicznych

Przedstawmy jeszcze wyniki naszej pracy na wykresie obejmującym: dane rzeczywiste, dane modelowe, prognozy na lata (2010-2012), trend wielomianowy.



Źródło: Opracowanie własne w *Excelu*.

Rys. 5.25. Porównanie danych rzeczywistych z modelowymi

* * *

Zaprezentowano dwa przykłady zastosowania wag harmonicznych do cech statystycznych. W skrypcie „*Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*" zestawiono podstawowe cechy statystyczne z okresu (1995-2009) prowadzone w ramach roczników statystycznych GUS-u w obszarach:

- podstawowe dynamiki statystyki okresu 1995-2009 (nakłady inwestycyjne, wartość brutto środków trwałych, produkt krajowy brutto);
- dynamiki popytu według sektorów (rachunki narodowe, rolnictwo, przemysł, budownictwo, handel, transport i gospodarka magazynowa oraz łączność);
- dynamika na mieszkańca kraju (rachunki narodowe, handel, inwestycje i środki trwałe, rynek pracy);
- cechy z różnych działów (stopa bezrobocia, wynagrodzenia, budżety gospodarstw domowych, działalność badawcza i rozwojowa, wskaźniki cen, handel zagraniczny);
- finanse (finanse przedsiębiorstw, sytuacja pieniężna, finanse publiczne);
- inwestycje i środki trwałe w sektorach.

Przez analogię do zaprezentowanych dwóch przykładów, gdzie prognoza jest wyznaczana przez liniową ekstrapolację ostatniego wygładzania szeregu czasowego, można zastosować wagi harmoniczne do prognozowania również innych cech statystycznych w sektorach gospodarki narodowej.

Na zakończenie porównajmy jeszcze dane rzeczywiste w latach 2010-2012[97] z otrzymanymi wcześniej metodą trendu pełzającego po uwzględnieniu starzenia się informacji:

Liczba zatrudnionych w handlu - przeciętne w roku [tys.]	2010	2011	2012
Prognoza	2329	2360	2391
Dane rzeczywiste	2218	2208	2176
Liczba zatrudnionych w transporcie, gospodarce magazynowej i łączności – przeciętne w roku [tys.]			
Prognoza	726	716	707
Dane rzeczywiste	730	745	748

W stosunku do pierwszej zmiennej - *Liczba pracujących w handlu*, prognoza na lata 2010-2012 wykazuje trend rosnący, natomiast dane empiryczne wskazują na tendencje odwrotną. Sugeruje to nam przyjęcie dłuższego segmentu trendu pełzającego. Na podstawie rysunku 3.17 wydaje się za celowe dla uchwycenia potencjalnej zmiany trendu przyjęcie segmentu 5-cio okresowego. Natomiast dla drugiej zmiennej - *Liczba pracujących w transporcie, gospodarce magazynowej i łączności* prognoza ma trent malejący, a dane rzeczywiste wykazują trend odwrotny. Według rysunku 5.25 występują trzy zmiany trendów szeregu czasowego. Z ciekawości spójrzmy jednak na dłuższy horyzont obserwacji, tj. lata 1995-2015 (zob. rysunek 5.26).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V
1	t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
2	Lata	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015
3	y_t (tran-prac)	845	855	878	902	858	814	794	762	750	741	731	750	780	763	733	730	745	748	737	744	769

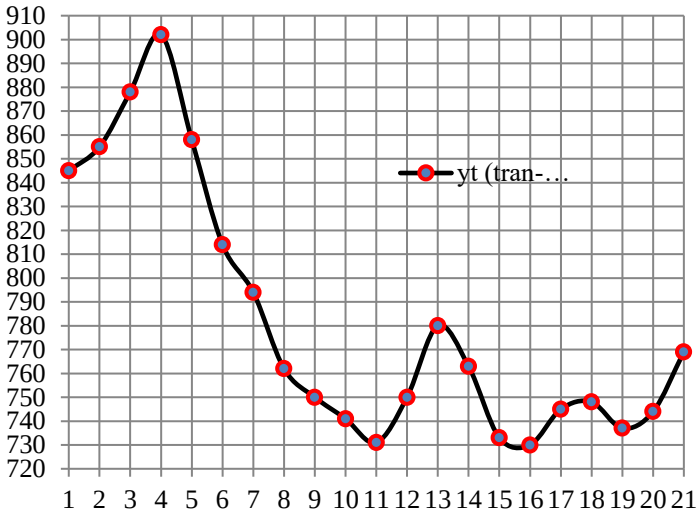
Źródło: Opracowanie własne w Excelu na podstawie roczników statystycznych GUS-u.

Rys. 5.26. Dane rzeczywiste zmiennej *Liczba pracujących w transporcie, gospodarce magazynowej i łączności* (lata 1995-2015)

Kształtowanie się zmiennej w czasookresie 21 lat podpowiada nam, że wskazane było by dla przykładu zmiennej *Liczby pracujących w transporcie, gospodarce magazynowej i łączności*

[97] Roczniki statystyczne GUS, lata 2011-2016.

zastosowanie segmentów np. 4-ro okresowych - dla uchwycenia zmieniających się w czasie tendencji trendów.



Źródło: Opracowanie własne w *Excelu*.

Rys. 5.27. Obserwowana zmiana trendu zmiennej *Liczba pracujących w transporcie, gospodarce magazynowej i łączności* (lata 1995-2015)

6. Iteracyjna optymalizacja parametrów w wybranych modelach wyrównywania wykładniczego

6.1. Wstęp

Zagadnienie doboru najlepszych parametrów równań modeli wyrównywania wykładniczego przewija się w różnych publikacjach. Jest przedmiotem zarówno ustaleń przez specjalistę, obliczenia według określonego wzoru, czy też postępowań iteracyjnych zarówno manualnych jak i komputerowych. W niniejszym materiale zwrócono uwagę na modele Holta, Wintersa, Holta-Wintersa oraz modele stanowiące uogólnienia rozwiązań „*klasycznych*". Tematykę tę rozpatrzono w układzie przeglądu literaturowego włączając w to źródła internetowe. Realizacje komputerowe bazujące na module prognozowania programu *WinQSB* oparto o dane statystyczne dotyczące konsumpcji piwa przez jednego mieszkańca w latach 1989-2010.

6.2. Uogólniony model Holta

W modelu Holta do wygładzania szeregu czasowego stosuje się średnią ruchomą z odpowiednim parametrem. Szeregi takie cechuje występowanie wahań przypadkowych oraz trend, czyli tendencja rozwojowa badanego zjawiska. Model Holta jest jedną z metod wygładzania (wyrównywania) wykładniczego. Oprócz omawianego modelu m.in. stosowany jest także prosty model wyrównywania wykładniczego, model Wintersa, model Holta-Wintersa, a także do prognoz długoterminowych model trendu tłumionego z uwzględnieniem sezonowości addytywnej i multiplikatywnej. Zagadnienie optymalizacji iteracyjnej parametrów uogólnionego modelu Holta, każdorazowo w odniesieniu do kolejnych okresów prognozowania spotykamy w publikacji Romana Szostka[98]. Proponuje on modyfikację metody Holta polegającą na:

- nie ograniczaniu przedziału parametrów α oraz β do [0, 1];
- ustalaniu najlepszych wartości tych parametrów kolejno dla okresu prognozowanego [n + 1], [n + 2], ..., [n = T], przy czym n to ilość obserwacji, a T ostatni z okresów prognozowania.

Umożliwia to stosowanie omawianego modelu Holta również do bardziej odległych okresów czasowych. Model ten dla okresu t bazuje na równaniu:

$$F_t = \alpha\, x_t + (1 - \alpha)(F_{t-1} + S_{t-1})$$

oraz

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

gdzie:

t = 2, ..., n - 1,

F_t - wygładzona wartość szeregu czasowego,

S_t - wygładzona wartość przyrostu trendu na okres t,

α, β - parametry modelu,

przy czym:

x_o, x_1, ..., x_{n-1} - wartości empiryczne prognozowanego szeregu czasowego; $F_1 = x_1$; $S_1 = x_1 - x_0$.

[98] Szostek R., *Uogólniony model Holta na przykładzie prognozowania liczby pasażerów w transporcie lotniczym w Polsce*, Ekonometria, 2(36) 2012.

Prognozy na kolejne okresy k = 1,2, 3, przy założeniu okresu początkowego t = 0 obliczamy rekurencyjnie według wzoru[99]:

$$x^*_{n+k-1} = F_{n-1} + k\, S_{n-1}$$

Wygasłe prognozy dla okresów t = 2, 3, ..., n-1 obliczamy według wzoru:

$$x^*_t = F_{t-1} + S_{t-1}$$

Pozwala to nam na obliczenie błędów prognoz, a do oceny jakości stosować możemy średnią liniową lub średnią kwadratową J_2 o postaci:

$$J_2 = \sqrt{\frac{1}{n-2}\sum_{t=2}^{n-1}(F_{t-1} + S_{t-1} - x_t)^2}$$

W kolejnych krokach dobieramy parametry α oraz β tak, aby otrzymać najmniejszą - optymalna wartość błędu przykładowo według miary J_2. Roman Szostek proponuje nie przestrzeganie założenia że[100]: $\propto \in [0,1]$ oraz $\beta \in [0,1]$. Dla przyjętych dwóch miar jakości (liniowej, kwadratowej) oraz obliczeń ze skokiem 0,0001 dla parametrów α oraz β przy wykorzystaniu specjalnego programu w języku (C++) uzyskano optymalne wartości tych parametrów. Dla miary kwadratowej α = 1,5676 a β = 0,1720, przy czym zwróćmy uwagę, że $\alpha > 1$. Umożliwia to sporządzenie wykresu punktowego na podstawie wcześniej wymienionych wzorów modelu Holta obejmującego :

X_t - kolejne obserwacje zmiennej,
F_t, - wygładzone wartości szeregu czasowego,
S_t - wygładzone wartości przyrostu trendu na okres t,
X^*_t - dokładnie wyznaczone wygasłe prognozy .

Natomiast przy zachowaniu ograniczenia przedziału [0,1] i przyjętej przykładowo miary kwadratowej parametrów α oraz β wynoszą: α = 1, β = 0,3057.

Przy ustalaniu prognoz wygasłych krok prognozy (k = 1). Jeśli chcemy jednak, aby przy pomocy modelu Holta ustalać prognozy na k okresów do przodu, to również musimy tak samo obliczać prognozy wygasłe oraz odpowiednio dobrać (przykładowo na podstawie miary kwadratowej) optymalne wartości parametrów α oraz β. W takim przypadku minimalizujemy miarę jakości o następującej postaci:

$$J_2(k) = \sqrt{\frac{1}{n-k-1}\sum_{t=k=1}^{n-1}(F_{t-k} + kS_{t-k} - x_t)^2}$$

gdzie: k = 1, 2, 3, ...

Badania Szostka wykazały, że dzięki modyfikacji postępowania ze swobodą wartości parametrów α oraz β jak również ich odniesienia do ilości okresów prognozowania przy ustalaniu prognoz wygasłych doprowadzono do większej wiarygodności prognoz. Przez analogię można opracować kod programu dla przykładu danych spożycia piwa w przeliczeniu na 1. mieszkańca Polski zamieszczonych w tabeli 2.1[101].

[99] Ibidem, wzory zaczerpnięto z podanego w przypisie 1. artykułu.
[100] Ibidem, s. 18.
[101] Ibidem, zob. s. 22-23.

6.3. Modele wyrównywania wykładniczego w programie *WinQSB*

Dostępne oprogramowanie *WinQSB* w Internecie na system operacyjny *Windows XP* umożliwia dla danego modelu danych zastosowanie kolejno kilku modeli wyrównywania wykładniczego, a mianowicie[102]:

- prostego wyrównywania wykładniczego (SES),
- prostego wyrównywania wykładniczego z trendem (SEST),
- podwójnego wyrównywania wykładniczego (DES),
- podwójnego wyrównywania wykładniczego z trendem (DEST),
- adaptacyjnego wyrównywania wykładniczego (AES),
- modelu Holta-Wintersaz sezonowościa addytywną (HWA),
- modelu Holta-Wintersa z sezonowościa multiplikatywną (HWM).

W zakresie SES istnieje w programie *WinQSB* opcja określenia najlepszych wartości parametrów. Po zainstalowaniu W*inQSB* postępowanie sprowadza się do:

a) wywołania programu z dysku;

b) wyboru modułu FC, a następnie wywołania z menu głównego kolejno: *File* (zbiór), *New Problem* (nowy problem), *Problem type* (typ problemu) i wybór w ramach niego: *Time Series Forecasting* (szereg czasowy);

c) wprowadzamy informacje: tytuł problemu, jednostka czasu, liczba okresów obserwacji;

d) wprowadzamy dane rzeczywiste - przykładowo spożycie piwa w kwartałach roku gospodarczego[103],

e) zapisanie pliku na dysku i wywołanie funkcji: *Solve and Analyze Perform Forecasting,*

f) zaznaczamy interesującą nas metodę prognozowania, np. SES,

g) wybieramy żądanie wyszukiwania iteracyjnego przez program wartości najlepszych (*Search the best*) dla naszego modelu wyrównywania wykładniczego według kryterium błędu:

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{n}\sum\left(\hat{Y}_n - Y_n\right)}$$

gdzie:

n - liczba sumowanych reszt,

$\hat{Y}_n$ - wartość określona z modelu,

Y_n - wartość z obserwacji w okresie *t.*

Zatem MSE stanowi wartość średnią sumy kwadratów reszt. Ponadto po wyborze *Search the best* podajemy programowi tylko zakres danego parametru z krokiem iteracyjnym przykładowo 0,01. W przypadku SES wpisujemy także liczbę okresów do prognozowania. Jeśli nie podamy prognozy inicjującej to program przyjmuje $F(0) = F_2 = Y_1$ i oblicza prognozę według wzoru[104]:

$$F_n = aY_{n-1} + (1-\alpha)F_{n-1}$$

Tu możemy zauważyć, że symbolizacja i znaczenia poszczególnych elementów wzorów są różne w zależności od autora publikacji zarówno zwartej jaki internetowej. Otrzymany wynik błędu jest mniejszy po wyborze przez komputer optymalnego parametru α.

Przetestujmy teraz metodę SEST posiadającą analogię do przedstawionego już modelu Holta z zastosowaniem opcji *Search the best* na przykładzie danych zamieszczonych w tabeli 6.1. Podajemy ilość okresów prognozowania - 1, nie określamy parametrów α oraz β, a także wartości początkowej $F(0)$. Jako kryterium błędu przyjmujemy MSE oraz krok iteracyjny 0.01 dla parametrów α i β. Trzeba

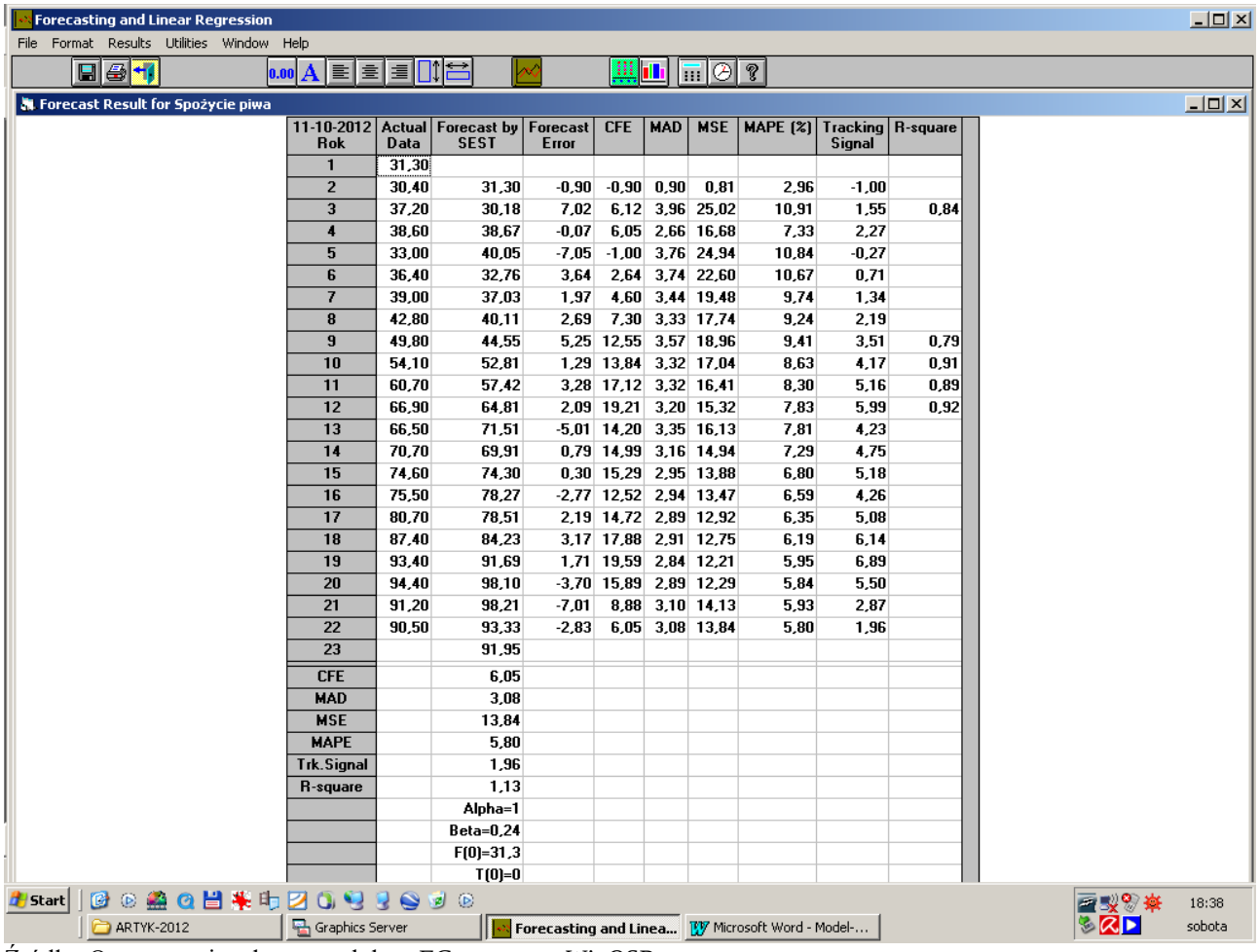
[102] Wornalkiewicz W., *Dobór metod i parametrów w budowaniu modeli ekonometrycznych*, *Ekonomiczno-społeczne problemy współczesnego zarządzania i komunikacji*, (red.) Duczmal. M, Pokusa T, Wyższa Szkoła Zarządzania i Administracji , Opole 2008.
[103] Ibidem, tab. 2, s. 203.
[104] Ibidem, s. 196.

tu jednak zaznaczyć, że zarówno krok iteracyjny oraz dane empiryczne podajemy z kropką dziesiętną zamiast przecinka.

Tab. 6.1. Ilość piwa wypitego w Polsce w litrach przez jednego mieszkańca w latach 1989-2010

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
Rok	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010
litry	**31,3**	30,4	37,2	38,6	33	36,4	39	42,8	49,8	54,1	60,7	66,9	66,5	70,7	74,6	75,5	80,7	87,4	93,4	94,4	91,2	**90,5**

Źródło: „Rocznik Statystyczny" za lata: 1988-2006, GUS, Warszawa, GUS; strona internetowa www gov.pl - lata 2006-2010.



11-10-2012 Rok	Actual Data	Forecast by SEST	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-square
1	31,30								
2	30,40	31,30	-0,90	-0,90	0,90	0,81	2,96	-1,00	
3	37,20	30,18	7,02	6,12	3,96	25,02	10,91	1,55	0,84
4	38,60	38,67	-0,07	6,05	2,66	16,68	7,33	2,27	
5	33,00	40,05	-7,05	-1,00	3,76	24,94	10,84	-0,27	
6	36,40	32,76	3,64	2,64	3,74	22,60	10,67	0,71	
7	39,00	37,03	1,97	4,60	3,44	19,48	9,74	1,34	
8	42,80	40,11	2,69	7,30	3,33	17,74	9,24	2,19	
9	49,80	44,55	5,25	12,55	3,57	18,96	9,41	3,51	0,79
10	54,10	52,81	1,29	13,84	3,32	17,04	8,63	4,17	0,91
11	60,70	57,42	3,28	17,12	3,32	16,41	8,30	5,16	0,89
12	66,90	64,81	2,09	19,21	3,20	15,32	7,83	5,99	0,92
13	66,50	71,51	-5,01	14,20	3,35	16,13	7,81	4,23	
14	70,70	69,91	0,79	14,99	3,16	14,94	7,29	4,75	
15	74,60	74,30	0,30	15,29	2,95	13,88	6,80	5,18	
16	75,50	78,27	-2,77	12,52	2,94	13,47	6,59	4,26	
17	80,70	78,51	2,19	14,72	2,89	12,92	6,35	5,08	
18	87,40	84,23	3,17	17,88	2,91	12,75	6,19	6,14	
19	93,40	91,69	1,71	19,59	2,84	12,21	5,95	6,89	
20	94,40	98,10	-3,70	15,89	2,89	12,29	5,84	5,50	
21	91,20	98,21	-7,01	8,88	3,10	14,13	5,93	2,87	
22	90,50	93,33	-2,83	6,05	3,08	13,84	5,80	1,96	
23		91,95							
CFE		6,05							
MAD		3,08							
MSE		13,84							
MAPE		5,80							
Trk.Signal		1,96							
R-square		1,13							
		Alpha=1							
		Beta=0,24							
		F(0)=31,3							
		T(0)=0							

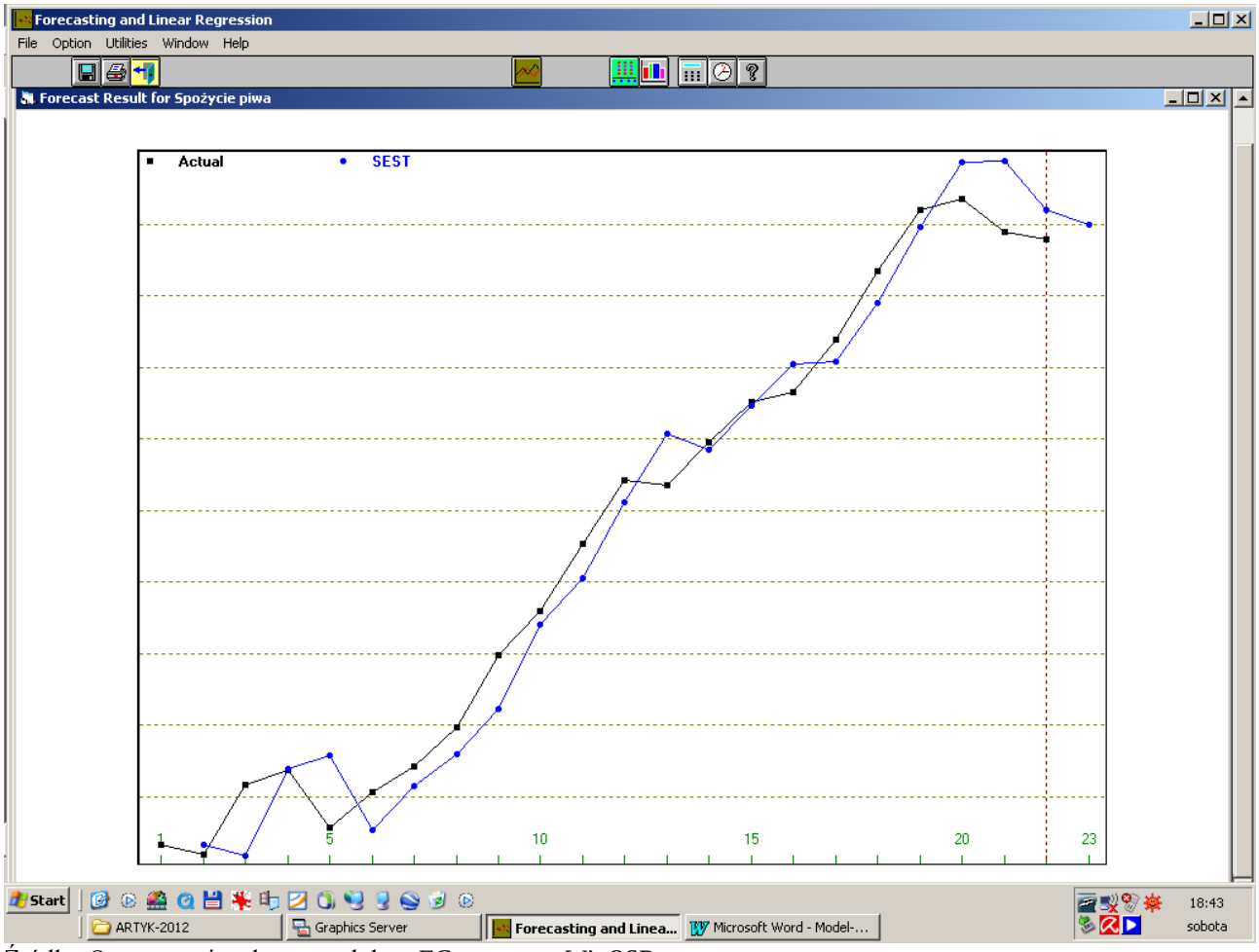
Źródło: Opracowanie własne modułem *FC* programu *WinQSB*.

Rys. 6.1. Raport z określenia prognozy na okres 23.

Program ustalił parametry optymalne $\alpha = 1$ oraz $\beta = 0{,}24$. Zainicjował prognozę $F(0) = F_2 = Y_1 = 31{,}3$. Kryterium błędu MSE = 13,84. Zatem RMSE jako pierwiastek z MSE wynosi: 3,72. Oprócz miary MSE w raporcie są także inne miary błędu, tj. CFE, MAD oraz występujący często w literaturze średni procentowy absolutny błąd:

$$MAPE = \frac{1}{n}\sum \left|\frac{\hat{y}_n - y_n}{y_n}\right| 100$$

Wartość prognozy na okres 23 wynosi 91,95.



Źródło: Opracowanie własne modułem *FC* programu *WinQSB*.

Rys. 6.2. Graficzne przedstawienie empirycznych (*Actual*) i modelowych (SEST) wartości zmiennej prognozowanej

Jak łatwo zauważyć spożycie piwa w litrach w przeliczeniu na jednego mieszkańca wykazuje zmienność, a także trend rosnący, przy czym w kolejnych latach kształtowało się w przedziale (od 31,3-90,5), czyli nastąpił wzrost prawie trzykrotny.

Dobór metody wygładzania wykładniczego zależy od cech badanego szeregu czasowego do których zaliczamy: wahania przypadkowe, trend, sezonowość. Cechy takie posiada spożycie piwa w kwartałach roku gospodarczego (III, IV oraz I i II następnego roku). Procedury zastosowanie modułu *FC* do dalszych metod wyrównywania wykładniczego, a w tym określania wartości parametrów znajdzie Czytelnik w artykule „*Dobór metod i parametrów w budowaniu modeli ekonometrycznych*”[105]. Temu tematowi poświęcony jest także m.in. podrozdział „*Przykłady doboru najlepszych parametrów modelu...*” w książce „*Wstęp do ekonometrii i badań operacyjnych ...*[106]. Podano w nim także postępowanie w zakresie korzystania programu *WinQSB* obejmujące modele:

- podwójne wyrównywanie wykładnicze z trendem (DEST),
- model Holta-Wintersa z sezonowością multiplikatywną.

6.4. Zastosowanie dodatku *Solver Excela*

Dodatek *Solver* arkusza kalkulacyjnego *Excel* służy do znalezienia rozwiązania optymalnego przy zadanej funkcji celu i ograniczeniach. Może wiec być zastosowany do znalezienia wartości

[105] Ibidem, s. 196-202.

[106] Duczmal M., Wornalkiewicz W.: *Wstęp do ekonometrii i badań operacyjnych ...”,* Wydawnictwo Instytut Śląski, Opole 2010, podrozdział 1.13.2.

optymalnych parametrów równań modelu wygładzania wykładniczego. W publikacji Radzikowskiej użyto go do znalezienia wartości minimalnego średniego względnego błędu *ex post* (Ψ) wygasłych już prognoz[107], przy czym zaproponowano stosowanie *Solvera* do: prostego modelu wygładzania wykładniczego, modelu liniowego Holta. W cytowanej publikacji prosty model wygładzania wykładniczego określony jest równaniem:

$$y_1^* = \alpha\, y_{t-1} + (1-\alpha)\, y_{t-1}^*$$

gdzie:

α - parametr wygładzania,
y_{t-1} - wartość obserwacji zmiennej szeregu czasowego dla okresu (t-1),
y_{t-1}^*- wartość prognozy zmiennej dla okresu (t-1).

Jako już wspomniano jako miarę błędu stosuje się (Ψ) w %, który według notacji we wspomnianej książce określony jest wzorem[108]:

$$\psi = \frac{1}{n-k} \sum_{t=k+1}^{n} \frac{|y_t - y_t^*|}{y_t}\ 100$$

gdzie: k - stała wygładzania (ilość okresów).

Zmieniła się zatem symbolizacja we wzorze stanowiącym analogię do poznanej wcześniej miary błędów MAPE. Wykorzystajmy nadal szereg czasowy podany w tabeli 3.1. Jako wartość początkową prognozy przyjmuje się: $y_1^* = y_1 = 31{,}3$ lub średnią arytmetyczną z trzech pierwszych obserwacji:

$$y_1^* = (y_1 + y_2 + y_3)/3 = (31{,}3 + 30{,}4 + 37{,}2)/3 = 32{,}97$$

Następnie zastosowano *Solvera* do określenia optymalnej wartości parametru α. Wymagało to wcześniejszego wprowadzenia wartości obserwacji do arkusza kalkulacyjnego *Excel* (zob. tabela 6.2).

Tab. 6.2. Dane rzeczywiste, prognoza, błędy jej oszacowania modelem prostego wygładzania wykładniczego

A	**B**	**C**	**D**	**E**	**F**
2	*t*	Spożycie	Prognoza	Błąd	
3	1	**31,3**	**32,97**	5,32	
4	2	**30,4**	29,71	2,28	
5	3	**37,2**	31,06	16,50	
6	4	38,6	43,06	11,57	
7	5	33,0	34,33	4,04	
8	6	36,4	31,73	12,84	
9	7	39,0	40,87	4,79	
10	8	42,8	37,22	13,04	
11	9	49,8	48,13	3,35	
12	10	54,1	51,39	5,01	
13	11	60,7	56,69	6,61	
14	12	66,9	64,53	3,54	
15	13	66,5	69,16	4,00	
16	14	70,7	63,96	9,54	
17	15	74,6	77,14	3,41	

[107] Radzikowska B. (red.), *Metody prognozowania Zbiór zadań,* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004, rozdz. 1.1.
[108] Ibidem, s. 12.

18	16	75,5	72,17	4,41
19	17	80,7	78,68	2,50
20	18	87,4	82,63	5,46
21	19	93,4	91,96	1,54
22	20	94,4	94,78	0,40
23	21	91,2	94,04	3,11
24	22	90,5	88,49	2,22
25	23		92,42	
26				125,48
27	Ψ			**5,70**

gdzie: ψ - średni względny błąd prognoz *ex post*, przy czym $n = 22$, a $k = 0$.

Źródło: Opracowanie własne w *Excelu*.

W tabeli zaznaczono wiersze i kolumny arkusza kalkulacyjnego programu *Excel*. Obliczenia wykonano pod systemem operacyjnym *Windows XP*. Obliczono prognozy spożycia piwa w litrach na 1. mieszkańca Polski dla poszczególnych okresów. Błędy dla okresów (1-22) określono na podstawie wzoru:

$$\psi_t = \frac{|y_t - y_t^*|}{y_t} 100$$

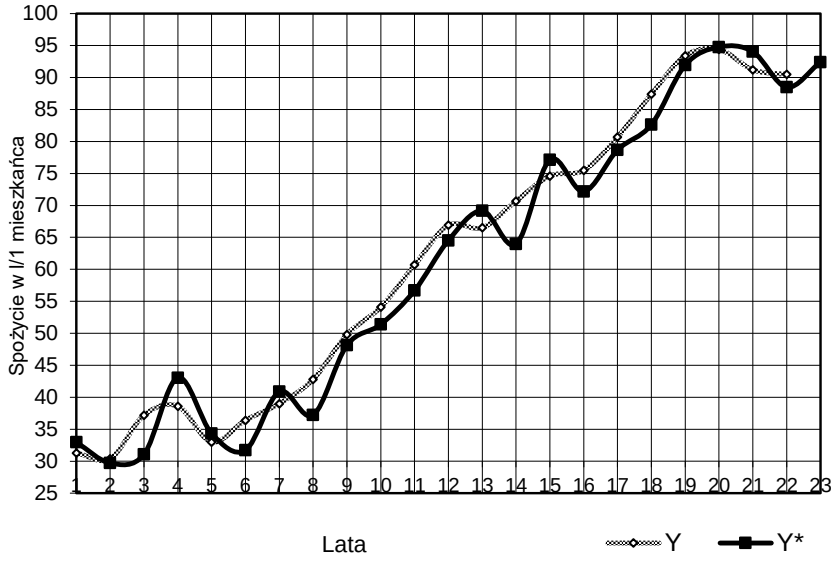
po wcześniejszym wyznaczeniu wartości optymalnej parametru $\alpha = 1{,}9554$ w *Solverze* (tab. 6.3).

Tab. 6.3. Raport z określenia parametru α przy minimum wartości średniego względnego błędu prognoz *ex post*

F			
α			
1,9554			
Microsoft Excel 8.0 Raport wyników			
Arkusz: [Optymalizacja.xls]Arkusz1			
Raport utworzony: 2012-11-11 15:12:34			
Komórka celu (Min)			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
$E$27	Psi Błąd	38,71	**5,70**
Komórki decyzyjne			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
$G$3	***α***	0	**1,9553954**
Warunki ograniczające			
BRAK			

Źródło: Opracowanie własne w dodatku *Excel-Solver*.

Wynik dla α otrzymano w *Solverze* bez podawania typowego ograniczenia $\alpha = (0, 1)$. Przy $\alpha = 1{,}9553954 \approx 1{,}9554$, $\psi = 5{,}70$ stanowi wartość minimalną. Zobaczmy jeszcze jak układają się na wykresie punktowym wartości rzeczywiste i modelowe zmiennej „*Spożycie piwa*”.



Źródło: Opracowanie własne w *Excelu*.

Rys. 6.3. Spożycie piwa w litrach w przeliczeniu na 1. mieszkańca Polski

Mimo użycia prostego modelu wygładzania wykładniczego, lecz nie ograniczania parametru α, widzimy dobre dopasowanie danych teoretycznych do empirycznych.

Rozwiążmy jeszcze nasze zadanie w *Solverze* stosując model liniowy Hola. Jak już wiemy jest to model dwurównaniowy. Ze względu na oznaczenie zmiennej jako *y* oraz w nawiązaniu do wcześniejszego prostego modelu wygładzania wykładniczego skorzystajmy ze wzorów[109]:

a) Wygładzone wartości szeregu czasowego w okresie *t*-1:

$$F_{t-1} = \alpha\, y_{t-1} + (1-\alpha)\,(F_{t-2} + S_{t-2})$$

b) Wygładzone wartości przyrostu trendu w okresie *t*-1:

$$S_{t-1} = \beta\,(F_{t-1} - F_{t-2}) + (1-\beta)\,S_{t-2}$$

gdzie: α, β - parametry wygładzania modelu o typowym przedziale [0,1], dobierane również według średniego względnego błędu prognoz. Wartość prognozy na okres *T*, który jest większy od *n* , czyli liczby wyrazów zmiennej prognozowanej w szeregu czasowym określamy z zależności[110]:

$$y_T^* = F_n + (T-n)\,S_n$$

Istotną rolę odgrywa wyznaczenie wartości początkowych. Przyjmijmy dla naszego przykładu, że $F_1 = y_1 = 31{,}3$; $S_1 = y_2 - y_1 = 30{,}4 - 31{,}3 = -0{,}9$. I tu również posłużono się dodatkiem *Solver* dla określenia najlepszych parametrów α oraz β. Obliczenia pomocnicze elementów oraz wygasłych prognoz wykonano w arkuszu kalkulacyjnym Excel (zob. tabela 6.4).

Dla początkowych wartości $\alpha = 0{,}35$ i $\beta = 0{,}55$ suma kwadratów błędów wygasłych prognoz dla obserwacji (3-22) wynosi 491,1. Średnia wartość kwadratu błędów przy 20. obserwacjach równa się 24,555. Tak więc miara RMSE jest pierwiastkiem średniego kwadratu błędów i wynosi 4,96.

[109] Radzikowska B. (red.), *Metody prognozowania. Zbiór zadań*, op. cit., s. 33.
[110] Ibidem, s. 34.

Tab. 6.4. Rozplanowanie arkusza kalkulacyjnego do optymalizacji parametrów α i β modelu Holta

1	A	B			C	D
2	t	y_t	F_t	S_t	y^*_t	
3	1	**31,3**	31,30	-0,90		
4	2	30,4	30,40	-0,90	30,40	
5	3	37,2	44,65	0,70	29,50	
6	4	38,6	32,08	-0,70	45,34	
7	5	33,0	34,57	-0,36	31,38	
8	6	36,4	38,52	0,09	34,21	
9	7	39,0	39,37	0,17	38,61	
10	8	42,8	45,95	0,85	39,55	
11	9	49,8	52,71	1,47	46,79	
12	10	54,1	54,03	1,45	54,18	
13	11	60,7	65,75	2,54	55,48	
14	12	66,9	65,56	2,25	68,29	
15	13	66,5	65,23	1,98	67,81	
16	14	70,7	74,07	2,70	67,21	
17	15	74,6	72,50	2,25	76,78	
18	16	75,5	76,23	2,41	74,75	
19	17	80,7	82,70	2,84	78,64	
20	18	87,4	89,21	3,22	85,53	
21	19	93,4	94,34	3,42	92,43	
22	20	94,4	91,15	2,73	97,76	
23	21	91,2	88,61	2,17	93,87	
24	22	90,5	90,22	2,11	90,79	
25	23			Prognoza	92,34	207,6
26					**Średni**	10,38
27					RMSE	**3,22**

gdzie: Średni (=G25/20).

Źródło: Opracowanie własne w *Excelu*.

W obliczeniu wygasłych prognoz oraz prognozy na okres 23. zastosowano wyznaczone przez *Solvera* wartości optymalne parametrów α = 1,9671 i β = 0,1055. Ponadto posłużono się funkcją *Excela* (=SUMA.XMY.2(C5:C24;F5:F24) dla szybkiego obliczenia (207,6) tj. sumy kwadratów różnic wartości empirycznych od prognozowanych dla obserwacji 3-22. Miara RMSE jest pierwiastkiem z wartości w komórce (*Średni*).

Raport z procedury wyznaczenia wartości optymalnych parametrów α i β przy nieograniczaniu typowym ich przedziałów [0, 1] wykonany dodatkiem *Solver Excela* prezentuje tabela 6.5. Jako funkcję celu przyjęto minimum RMSE. Komórkami zmienianymi były F3 oraz G5, których początkowe wartości były odpowiednio: 0,3500 oraz 0,5500.

Tab. 6.5. Raport *Solvera* przy nieograniczaniu parametrów

F	**G**
α	***β***
1,9671	0,1055
Microsoft Excel 8.0 Raport wyników dla RMSE	
Arkusz: [Optymalizacja.xls]Holta	
Raport utworzony: 2012-11-20 21:49:15	
Parametry bez ograniczeń	

Komórka celu (Min)			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
$G$27	**RMSE**	4,96	3,22
Komórki decyzyjne			
Komórka	Nazwa	Wartość początkowa	Wartość końcowa
$I$3	α	**0,3500**	1,9671
$J$3	β	**0,5500**	0,1055
Warunki ograniczające:			Brak

Źródło: Opracowanie własne w *Solverze*.

Dla przykładu zobaczmy jak kształtować się będą wartości parametrów α i β przy nadaniu im przedziału [0,1].

Tab. 6.6. Dane i obliczenia pomocnicze przy ograniczeniu parametrów α i β do [0, 1]

1	**A**	**B**			**C**	**D**
2	t	y_t	F_t	S_t	y^*_t	
3	1	31,3	31,30	-0,90		
4	2	30,4	30,40	-0,90	30,40	
5	3	37,2	37,20	1,38	29,50	
6	4	38,6	38,60	1,39	38,58	
7	5	33,0	33,00	-0,68	39,99	
8	6	36,4	36,40	0,53	32,32	
9	7	39,0	39,00	1,14	36,93	
10	8	42,8	42,80	1,93	40,14	
11	9	49,8	49,80	3,43	44,73	
12	10	54,1	54,10	3,69	53,23	
13	11	60,7	60,70	4,55	57,79	
14	12	66,9	66,90	5,04	65,25	
15	13	66,5	66,50	3,43	71,94	
16	14	70,7	70,70	3,66	69,93	
17	15	74,6	74,60	3,73	74,36	
18	16	75,5	75,50	2,89	78,33	
19	17	80,7	80,70	3,57	78,39	
20	18	87,4	87,40	4,50	84,27	
21	19	93,4	93,40	4,94	91,90	
22	20	94,4	94,40	3,78	98,34	
23	21	91,2	91,20	1,71	98,18	
24	22	90,5	90,50	1,00	92,91	
25	23			Prognoza	91,50	**299,4**
26					**Średni**	14,97
27					RMSE	**3,87**

Źródło: Opracowanie własne w *Excelu*.

Widzimy wyraźny wzrost miary błędu RMSE wygasłych już prognoz dla okresów (3-22). Obliczenia wykonano po wcześniejszym wyznaczeniu najlepszych wartości parametrów α = 1 i β = 0,2963 przyjmując dla nich ograniczenia przedziału [0, 1] (zob. tabela 6.7).

Tab. 6.7. Raport *Solvera* przy zadanych ograniczeniach parametrów α i β

F	G
α	***β***
1,0000	0,2963
0	0

1	1			
Komórka celu (Min)				
	Komórka	Nazwa	Wartość początkowa	Wartość końcowa
	$G$27	RMSE	4,96	3,87
Komórki decyzyjne				
	Komórka	Nazwa	Wartość początkowa	Wartość końcowa
	$I$3		0,3500	1,0000
	$J$3		0,5500	0,2963

* * *

Zamieszczone rozważania potwierdzają celowość pozostawienia otwartości wartości parametrów modeli w procesie ich optymalizacji. Zaprezentowano trzy możliwości doboru najlepszych wartości parametrów w równaniach modeli jednorównaniowych oraz dwurównaniowych stosując:

- specjalny program napisany w języku PL/I,
- moduł programowania liniowego i całkowitoliczbowego programu *WinQSB,*
- dodatek *Solver* arkusza kalkulacyjnego *Excel*.

Na uwagę zasługuje też propozycja zachowania zgodności „*skoku*" aktywnej prognozy np. ($T = n + 2$) ze skokiem doboru obserwacji przy obliczaniu miary błędu wygasłych prognoz. Iteracyjna optymalizacja parametrów z zastosowaniem programu *WinQSB* jest też pomocna w doborze parametrów w modelach trzyrównaniowych Holta-Wintersa o parametrach α, β, γ z sezonowością multiplikatywną lub addytywną[111].

[111] Przykład doboru najlepszych parametrów z punktu widzenia miary błędów RMSE występuje w podrozdziale 1.13.2.2 *Model Holta-Wintersa z sezonowościa multiplikatywną* książki: Duczmal. M., Wornalkiewicz W.: *Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski, Opole 2010.

7. Model ekonometryczny zmiennej „*Liczba bezrobotnych*” kraju

7.1. Wprowadzenie

Zmienną, która powinna wejść do modelu budżetu powiatów Województwa Dolnośląskiego jest zmienna *Bezrobotni zarejestrowani ogółem w tys.* Liczba bezrobotnych jest bowiem istotną cechą w ocenie gospodarki danego obiektu struktury terytorialnej. Określa bowiem jej niewyczerpane dodatkowe możliwości. Z tego względu w niniejszym rozdziale uwaga skoncentrowana będzie na zbudowaniu modelu ekonometrycznego regresji tej zmiennej z innymi zmiennymi objaśniającymi. W rozdziale 5. *Podejścia do określania modeli ekonometrycznych PKB* główny nacisk położono na sformułowaniu modelu produktu krajowego brutto PKB w zależności od wybranych zmiennych z poszczególnych rynków[112]. Jedną ze zmiennych rynku pracy jest bowiem liczba bezrobotnych w tys. Interesuje nas zależność tej zmiennej od innych zmiennych objaśniających. W tym względzie posłużymy się programem *Gretl* oraz zastosujemy eliminację *a posteriori* dla selekcji zmiennych objaśniających. Wyjściem do opracowania modelu są zebrane dane statystyczne dla celu projektowania modelu ekonometrycznego PKB Polski. Obejmują one okres szesnastu lat (1992-2007). Wybrano z nich niżej wymienione zmienne pochodzące z określonych rynków, zachowując jednak ich symbole dla celów przetwarzania komputerowego:

- rynek pracy/aktywność ekonomiczna ludności w wieku 15 i więcej - przeciętne w roku w tys./bezrobotni (*X*3);
- przeciętne zatrudnienie w tys.: przemysł (*X*9), budownictwo (*X*10), handel i naprawy (*X*11), transport, gospodarka magazynowa i łączność (*X*12), obsługa nieruchomości i firm (*X*13), działalność usługowa komunalna, społeczna i indywidualna, pozostałe (*X*14);
- koszty pracy w zł/ogółem na jednego zatrudnionego przeciętnie miesięcznie (*X*25);
- rynek płacy/wynagrodzenia w mln/wskaźnik - przeciętne miesięczne wynagrodzenie nominalne brutto - rok poprzedni = 100 (*X*38);
- budżety gospodarstw domowych/przeciętne miesięczne wydatki na 1. gospodarstwo domowe w zł (*X*48);
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych): zużycie wody z wodociągów w ciągu roku w hm^3 (*X*49), zużycie energii elektrycznej w ciągu roku w GWh (*X*50), zużycie gazu w sieci w ciągu roku w hm^3 (*X*51);
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci - policealne w tys. (*X*59);
- rynek mieszkaniowy/wskaźnik zawartych małżeństw na 10000 ludności (*X*78).

Zmienne dobrano intuicyjnie, a ich ilość wynikała z warunku *Gretl*-a:

ilość obserwacji > ilość parametrów (zmiennych + 1); 16 obserwacji > 14 zmiennych objaśniających + 1.

7.2. Dane wejściowe

Wartości obserwacji poszczególnych zmiennych podano w tabeli 7.1.

Tab. 7.1. Wartości obserwacji zmiennych wyjściowych do modelu ekonometrycznego

Rok	*t*	*X*3	*X*9	*X*10	*X*11	*X*12	*X*13	*X*14	*X*25	*X*38	*X*48	*X*49	*X*50	*X*51	*X*59	*X*78
1992	1	2394	3467,6	823,0	1042,1	794,7	476,7	287,7	295,16	138,9	404,70	1921,9	18430	4626,8	41,1	5,70
1993	2	2595	3391,8	721,9	999,4	746,5	381,1	239,3	384,25	134,8	550,29	1856,5	18206	4504,2	40,1	5,40
1994	3	2375	3361,4	678,8	996,1	728,1	366,4	235,4	515,49	134,5	708,30	1750,0	18206	4472,6	45,8	5,40
1995	4	2233	3461,1	689,2	1078,6	723,4	414,4	243,5	704,99	131,6	878,29	1648,3	18075	4775,1	59,2	5,40
1996	5	1961	3436,0	684,2	1119,7	721,9	425,3	250,5	900,58	126,7	1112,95	1565,0	19224	4258,9	70,5	5,30
1997	6	1737	3433,4	723,5	1244,8	730,7	487,4	273,3	1088,03	121,9	1300,00	1514,7	19771	4437,4	77,4	5,30

[112] Wornalkiewicz W., *Modele ekonometryczne PKP obiektów struktury* terytorialnej, Wydawnictwo Instytut Śląski, Opole 2013.

1998	7	1827	3378,7	736,9	1321,1	725,2	552,5	253,0	1277,06	115,7	1527,52	1452,6	20288	4243,5	83,3	5,40
1999	8	2641	3138,4	710,4	1318,4	686,3	583,0	286,9	1428,01	112,5	1700,54	1406,5	20800	4138,6	86,4	5,70
2000	9	2760	2955,0	661,9	1325,0	654,9	614,2	286,7	2622,76	111,1	1904,57	1360,3	21037	3750,8	78,4	5,50
2001	10	3186	2820,6	627,8	1295,6	530,4	637,3	252,1	2821,30	108,0	1897,60	1310,4	21376	3905,8	75,1	5,00
2002	11	3431	2670,5	545,5	1291,1	601,9	671,3	260,0	2930,70	102,6	1886,10	1284,3	21659	3708,1	85,7	5,00
2003	12	3329	2639,1	496,4	1308,0	586,3	699,5	261,8	3017,64	104,2	1918,71	1268,6	22052	3877,2	78,4	5,10
2004	13	3230	2663,1	453,1	1295,4	583,0	700,6	264,6	3111,04	104,0	1966,72	1229,1	22804	3779,1	98,0	5,00
2005	14	3045	2665,4	483,6	1360,7	577,7	695,8	265,2	3252,02	103,8	1954,20	1219,4	26565	3855,3	104,3	5,40
2006	15	2344	2714,3	511,5	1394,2	596,7	723,3	267,6	3421,05	104,9	2106,28	1221,5	27547	3967,0	88,8	5,90
2007	16	1619	2842,4	576,8	1482,8	621,0	780,3	277,0	3720,1	107,9	2290,78	1280,1	27713	3807,0	88,8	6,50

Źródło: Opracowanie własne na podstawie roczników statystycznych *GUS*-u (lata 1992-2007).

Teraz posłużymy się *Gretlem*-em dla określenia statystyki opisowej zarówno zmiennej objaśnianej *X*3 jak i zmiennych objaśniających *X*9-*X*14, *X*25, *X*38, *X*48-*X*51, *X*59, *X*78.

Tab. 7.2. Statystyka opisowa zmiennych

Zmienna	Średnia	Mediana	Min	Max	*s*	*V*	*S*	*K*
*X*3	2544,19	2494,50	1619,00	3431,00	585,608	0,230175	-0,00926730	-1,18380
*X*9	3064,93	3046,70	2639,10	3467,60	345,773	0,112816	-0,0272987	-1,77077
*X*10	632,781	670,350	453,100	823,000	108,517	0,171492	-0,210728	-1,05959
*X*11	1242,06	1295,50	996,100	1482,80	147,797	0,118993	-0,469248	-0,925767
*X*12	663,044	670,600	530,400	794,700	78,2167	0,117966	-0,0562422	-1,28409
*X*13	575,569	598,600	366,400	780,300	134,639	0,233924	-0,202479	-1,33192
*X*14	262,788	263,200	235,400	287,700	16,6576	0,0633880	0,0388373	-1,01035
*X*25	1968,13	2025,38	295,163	3720,05	1238,71	0,629383	-0,0405670	-1,65828
*X*38	116,444	111,800	102,600	138,900	12,9368	0,111099	0,531208	-1,28171
*X*48	1506,72	1793,32	404,700	2290,78	600,455	0,398517	-0,604703	-1,03547
*X*49	1455,58	1383,40	1219,40	1921,90	231,951	0,159353	0,780318	-0,678815
*X*50	21484,6	20918,5	18075,0	27713,0	3225,56	0,150134	0,904225	-0,362171
*X*51	4131,71	4052,80	3708,10	4775,10	347,356	0,0840706	0,407472	-1,20955
*X*59	75,0812	78,4000	40,1000	104,300	19,3223	0,257351	-0,628182	-0,582073
*X*78	5,43750	5,40000	5,00000	6,50000	0,384491	0,0707110	1,25941	1,72878

gdzie:

Min - wartość minimalna, *Max* - wartość maksymalna, *s* - odchylenie standardowe,
V - współczynnik zmienności, *S* - współczynnik skośności, *K* - współczynnik kurtozy.

Źródło: Opracowanie własne programem *Gretl*.

Wyznaczenie procentowe współczynnika zmienności V_j na podstawie powyższej statystyki opisowej stanowi wstępny krok zbadania wystarczającego zróżnicowania danej *j*-tej zmiennej objaśniającej. Współczynnik ten jak wiemy określany jest w procentach i stanowi moduł z ilorazu odchylenia standardowego i średniego poziomu wartości rozpatrywanej zmiennej.

Tab. 7.3. Współczynniki zmienności

Zmienna	*X*3	*X*9	*X*10	*X*11	*X*12	*X*13	***X*14**	*X*25	*X*38	*X*48	*X*49	*X*50	***X*51**	*X*59	***X*78**
V_j	23,02	11,28	17,15	11,90	11,80	23,39	**6,34**	62,94	11,11	39,85	15,93	15,01	**8,41**	25,73	**7,07**

Źródło: Opracowanie własne w *Excel*-u według danych z tabeli 1.2.

Z tabeli 7.3 wynika, że dla trzech zmiennych (patrz czcionka pogrubiona) współczynnik zmienności nie przekracza wartości krytycznej $V^* = 10\%$, czyli zmienne te z tego punktu widzenia można już wstępnie wyeliminować z budowy modelu ekonometrycznego. Są to następujące zmienne: *X*14, *X*51, *X*78.

Zależność między zmiennymi możemy przedstawić przy pomocy macierzy korelacji korzystając w *Gretl* z menu: *Widok/Macierz korelacji* w wyniku czego uzyskujemy poniższy raport.

Tab. 7.4 Współczynniki korelacji liniowej przy wykorzystaniu obserwacji 1-16

***X*3**	***X*9**	***X*10**	***X*11**	***X*12**	
1,0000	-0,6628	-0,5486	0,0927	-0,6162	***X*3**
	1,0000	0,9060	-0,7265	0,9378	***X*9**
		1,0000	-0,5920	0,8645	***X*10**
			1,0000	-0,7106	***X*11**
				1,0000	***X*12**
***X*13**	***X*14**	***X*25**	***X*38**	***X*48**	
0,3582	-0,0196	0,4105	-0,4713	0,2917	***X*3**
-0,9017	-0,2365	-0,9460	0,9094	-0,8529	***X*9**
-0,7603	0,0197	-0,8578	0,8072	-0,7515	***X*10**
0,9170	0,5132	0,8694	-0,9003	0,9562	***X*11**
-0,8305	-0,0760	-0,9152	0,8936	-0,8571	***X*12**
1,0000	0,4839	0,9600	-0,9241	0,9309	***X*13**
	1,0000	0,3139	-0,3192	0,3729	***X*14**
		1,0000	-0,9327	0,9440	***X*25**
			1,0000	-0,9691	***X*38**
				1,0000	***X*48**
***X*49**	***X*50**	***X*51**	***X*59**	***X*78**	
-0,3683	0,0815	-0,5123	0,2173	-0,6201	***X*3**
0,8644	-0,7667	0,9032	-0,7209	0,0912	***X*9**
0,8165	-0,7333	0,7610	-0,7093	0,1542	***X*10**
-0,9063	0,8381	-0,8049	0,8931	0,2838	***X*11**
0,8849	-0,7065	0,8526	-0,7210	0,1910	***X*12**
-0,8925	0,8747	-0,8751	0,8199	0,1921	***X*13**
-0,2782	0,3514	-0,3389	0,3678	0,4193	***X*14**
-0,9243	0,8780	-0,9134	0,8063	0,1052	***X*25**
0,9800	-0,7645	0,9229	-0,9056	0,0783	***X*38**
-0,9779	0,8202	-0,9054	0,9037	0,1080	***X*48**
1,0000	-0,7807	0,8819	-0,9366	0,0463	***X*49**
	1,0000	-0,7086	0,7562	0,4510	***X*50**
		1,0000	-0,7789	0,0534	***X*51**
			1,0000	0,0116	***X*59**
				1,0000	***X*78**

Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,4973 dla $n = 16$

Źródło: Opracowanie własne w *Gretl*.

Dobór zmiennych objaśniających do modelu zmiennej *X*3 *Liczba bezrobotnych* możemy przeprowadzić stosując analizę współczynników korelacji. W tym celu możemy skorzystać z podanej powyżej wartości krytycznej współczynnika korelacji $r^* = 0{,}4973$. Postępowanie w tej analizie jest następujące[113]:

a) Sprawdzamy, czy $|r_j| \leq r^*$ i eliminujemy zmienne X_j nieistotnie skorelowane z *X*3:

$$\left|r_j\right| \leq r^* = 0{,}4973$$

[113] Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania,* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003.

Kierując się tym kryterium i przeglądając w poniższej macierzy korelacji pary relacji zmiennej objaśnianej *X*3 względem odpowiadających zmiennych objaśniających *X*9, *X*10-*X*14, *X*25, *X*38, *X*48, *X*49-*X*51, *X*59, *X*78 możemy wyeliminować 9 zmiennych:

*X*11, *X*13, ***X*14**, *X*25, *X*38, *X*48, *X*49, *X*50, *X*59.

Zwróćmy uwagę, że spośród powyższych zmiennych według współczynnika zmienności eliminacji podlegała zmienna *X*14.

b) Z pozostałych potencjalnych zmiennych wybieramy zmienną X_h wg kryterium:

$$r_h = \max_j \left(r_j\right)$$

c) Badamy relacje tak wybranej zmiennej z pozostałymi zmiennymi i eliminujemy te dla których $|r_{hi}| > r^*$, gdyż są one zbyt silnie skorelowane ze zmienną X_h.

Pełny dobór zmiennych objaśniających analizą współczynnika korelacji jest pracochłonny. Z tego względu łatwiej jest skorzystać z metody zwanej eliminacją *a posteriori* zawartej w *Gretl*.

7.3. Formułowanie modelu ekonometrycznego

Bazując na danych przekrojowych zapisanych (z uwzględnieniem zmiennych podanych w tabeli 7.1) w własnej bazie danych *Gretla*-a pod nazwą *Korelacja.gdt* możemy skorzystać z menu:

Model/klasyczna metoda najmniejszych kwadratów

definiując zmienną zależną *X*3 oraz zmienne niezależne (objaśniające): *X*9, *X*10-*X*14, *X*25, *X*38, *X*48-*X*51, *X*59, *X*78. Rezultatem pracy programu jest raport *modelu*1 zamieszczony w tabeli 7.5.

Tab. 7.5. Oszacowanie parametrów modelu zmiennej *X3*

Zmienna	Współczynnik	Błąd stand.	*t*-Studenta	Wartość *p*
const	21921,7	17844,1	1,229	0,4349
*X*9	-3,33581	4,46911	-0,7464	0,5918
*X*10	1,60023	6,38934	0,2505	0,8438
*X*11	1,12999	8,71286	0,1297	0,9179
*X*12	-4,62118	9,97778	-0,4631	0,7239
*X*13	-2,88898	4,35578	-0,6633	0,6272
*X*14	2,72167	13,2955	0,2047	0,8715
*X*25	0,0943435	0,763999	0,1235	0,9218
*X*38	-42,7163	152,104	-0,2808	0,8257
*X*48	-2,10554	4,81048	-0,4377	0,7373
*X*49	-0,0850932	5,72343	-0,01487	0,9905
*X*50	-0,194893	0,208678	-0,9339	0,5217
*X*51	0,0184532	1,42619	0,01294	0,9918
*X*59	16,4829	23,3847	0,7049	0,6091
*X*78	625,299	1577,31	0,3964	0,7597

Średnia arytmetyczna zmiennej zależnej = 2544,19
Odchylenie standardowe zmiennej zależnej = 585,608
Suma kwadratów reszt = 51525,8
Błąd standardowy reszt = 226,993
Współczynnik determinacji (R^2) = 0,98998
Skorygowany współczynnik determinacji = 0,84975
Statystyka F (14, 1) = 7,05961 (wartość p = 0,288)
Logarytm wiarygodności = -87,321
Kryterium informacyjne Akaike'a (*AIC*) = 204,642
Wyłączając stałą, największa wartość p jest dla zmiennej 53 (X51).

Źródło: Opracowanie własne w *Gretl*.

W powyższym raporcie szereg pojęć jest nam znanych ze statystyki. Na uwagę zasługuje jednak k*ryterium informacyjne Akaike'a*:

$$AK = N \ln\left(\sum Z_n^2\right) + 2k$$

W *Gretl* obliczany jest wg wzoru:

$$AIC = -2l\left(\hat{\Theta}\right) + 2\mathrm{k}$$

przy czym logarytm wiarygodności oznaczany jest symbolem:

$$l\left(\hat{\Theta}\right)$$

Kierując się podpowiedzią na końcu kolejnego raportu, budowano w *Gretl* modele eliminując kolejno zmienne:

*X*51, *X*49, *X*25, *X*11, *X*14, *X*38, *X*78, *X*12

uzyskując w efekcie *model10* uznany przez program *Gretl* jako końcowy (bez sugestii dalszej eliminacji zmiennych).

Tab. 7.6. *Model10* - estymacja parametrów zmiennej *X*3

Zmienna	Współczynnik	Błąd stand	*t*-Studenta	Wartość *p*	
const	17753,3	964,731	18,40	1,89E-08	***
*X*9	-4,33449	0,380482	-11,39	1,20E-06	***
*X*10	3,19675	0,826869	3,866	0,0038	***
*X*13	*-2,29688*	*0,950987*	*-2,415*	*0,0389*	**
*X*48	-0,887720	0,175447	-5,060	0,0007	***
*X*50	-0,116366	0,0203389	-5,721	0,0003	***
*X*59	16,1492	3,88416	4,158	0,0025	***

Średnia arytmetyczna zmiennej zależnej = 2544,19
Odchylenie standardowe zmiennej zależnej = 585,608
Suma kwadratów reszt = 101629
Błąd standardowy reszt = 106,264
Współczynnik determinacji = 0,98024
Skorygowany współczynnik determinacji = 0,96707
Statystyka *F* (6, 9) = 74,4243 (wartość $p < 0{,}00001$)
Logarytm wiarygodności = -92,7549
Kryterium informacyjne Akaike'a (*AIC*) = 199,51.

Źródło: Opracowanie własne w *Gretl*.

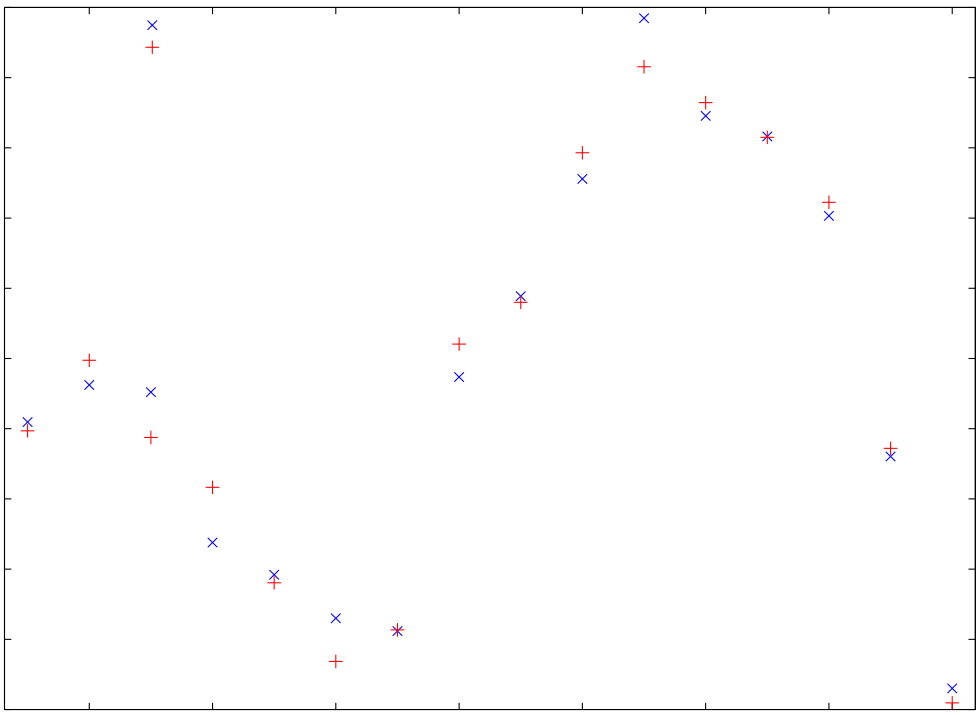
Program *Gretl* oprócz obliczania statystyki empirycznej *t*-Studenta podaje również graficznie informacje o poziomie istotności dla obliczonego parametru (zwanego w raporcie współczynnikiem), a mianowicie: *** - $\alpha = 0{,}01$; ** - $\alpha = 0{,}05$; *- $\alpha = 0{,}1$; puste - $\alpha > 0{,}1$. Występuje zatem niski poziom istotności - popełnienia błędu w oszacowaniu parametrów oraz wysoki jest współczynnik determinacji $R^2 = 0{,}98024$. Nasz model jeszcze przed etapem weryfikacji jest następujący:

$$\hat{X}_3 = 17753{,}3 - 4{,}33449X_9 + 3{,}19675X_{10} - 2{,}29688X_{13} - 0{,}88772X_{48} - 0{,}116366X_{50} + 16{,}1492X_{59}$$

Przypomnijmy sobie zmienne występujące w modelu końcowym:

- bezrobotni (*X*3);
- przeciętne zatrudnienie w tys*:* przemysł (*X*9), budownictwo (*X*10), obsługa nieruchomości i firm (*X*13);
- budżety gospodarstw domowych/przeciętne miesięczne wydatki na 1. gospodarstwo domowe w zł (*X*48);
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych)/zużycie energii elektrycznej w ciągu roku w GWh (*X*50);
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci – policealne w tys. (*X*59).

Posługując się *Gretl*-em pokazano wartości z obserwacji i obliczonych - modelowych (wyrównanych) na podstawie modelu na poniższym wykresie punktowym.



Źródło: Opracowanie własne w *Gretl*.

Rys. 7.1. Wartości empiryczne i modelowe zmiennej zależnej *X*3 *Liczba bezrobotnych*

7.4. Weryfikacja modelu

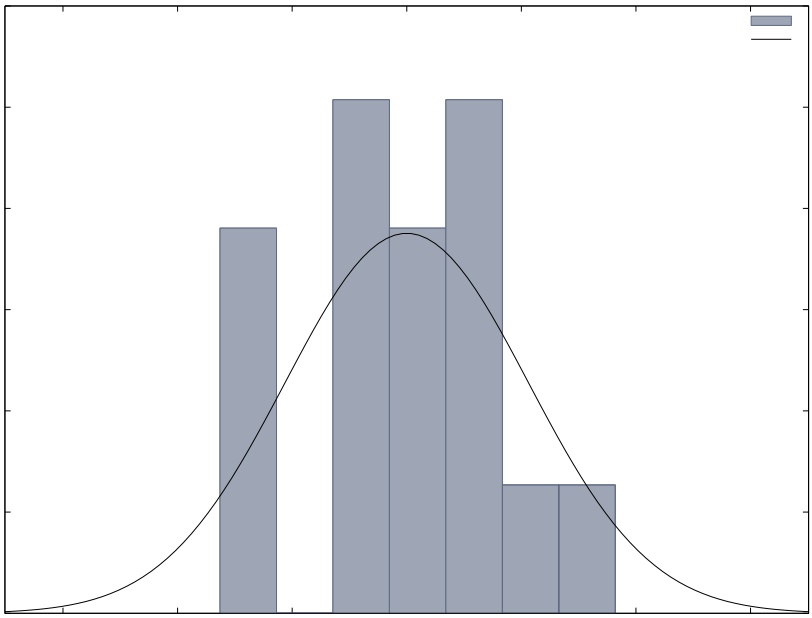
Etap weryfikacji modelu ekonometrycznego estymowanego klasyczną metodą najmniejszych kwadratów może obejmować wiele testów, którymi dysponuje program *Gretl*. W niniejszym temacie ograniczymy się jednak do sprawdzenia: normalności reszt, heteroskedastyczności (*zmienność wariancji resztowej*), współliniowości zmiennych objaśniających. Po ponownym uzyskaniu na ekranie *modelu10* wywołujemy menu: *Testy/test na normalność rozkładu reszt* i w efekcie otrzymujemy rozkład częstości reszt (tabela 4.7, rysunek 7.2).

Tab. 7.7. Szereg rozdzielczy reszt dla 16 obserwacji

Przedziały	Średnia	Liczba	Częstość [%]	Skumulowana [%]
< -113,68	-138,32	3	18,75	18,75
-113,68 - -64,410	-89,047	0	0,00	18,75
-64,410 - -15,135	-39,772	4	25,00	43,75
-15,135 - 34,140	9,5025	3	18,75	62,50
34,140 - 83,415	58,777	4	25,00	87,50
83,415 - 132,69	108,05	1	6,25	93,75
>= 132,69	157,33	1	6,25	100,00

Liczba przedziałów = 7, średnia = -1,08002e-012, odch. std. = 106,264.

Źródło: Opracowanie własne w *Gretl*.



Źródło: Opracowanie własne w *Gretl*.

Rys. 7.3. Przedstawienie częstości występowania zakresów reszt

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny. W *Gretl* zastosowano *test Doornika-Hansena* bazujący na współczynniku skośności i kurtozie. Z wykresu odczytamy statystykę testu: $\chi^2_{(2)} = 0{,}346$ z wartością $p = 0{,}84120$. Wartość krytyczna $\chi^2_{(\alpha,r)}$ przy poziomie istotności $\alpha = 0{,}05$ o dwóch stopniach swobody wynosi 5,991. Zatem spełnione jest hipoteza H_0, gdyż wartość statystyki jest mniejsza niż wartość krytyczna.

Kolejnym krokiem jest sprawdzenie heterskedastyczności, czyli zmienności wariancji resztowej. W tym celu wywołujemy poniższe menu stosując *test White'a*:

Testy/test heteroskedastyczności/test White'a (zmienność wariancji resztowej)

Tab. 7.8. Raport testu heteroskedastyczności

Zmienna	Współczynnik	Błąd stand.	*t*-Studenta	Wartość *p*
const	7,01813E+06	4,36139E+06	1,609	0,2060
*X*9	-2059,16	1544,73	-1,333	0,2747
*X*10	-2002,50	1883,47	-1,063	0,3657
*X*13	-2670,96	1840,64	-1,451	0,2427
*X*48	276,069	189,069	1,460	0,2404
*X*50	-219,474	99,1065	-2,215	0,1136
*X*59	1366,18	3093,86	0,4416	0,6887
sq_X9	0,288788	0,228593	1,263	0,2957
sq_X10	2,24569	1,89877	1,183	0,3221
sq_X13	1,91037	1,22645	1,558	0,2172
sq_X48	-0,0443028	0,0399304	-1,110	0,3482
sq_X50	0,00442355	0,00200754	2,203	0,1148
sq_X59	-0,551215	17,7976	-0,03097	0,9772

Współczynnik determinacji = 0,86696

Statystyka testu: TR^2 = 13,871395.

Źródło: Opracowanie własne w *Gretl*, gdzie: TR^2 $\rightarrow$ TR^2

Hipoteza zerowa H_0: heteroskedastyczność reszt nie występuje. Raport podaje statystykę testu, która można obliczyć również według wzoru:

$$LM = TR^2 = 16 * 0{,}86686 = 13{,}87136.$$

Test White'a zakłada wprowadzenie dodatkowo do modelu kwadratów zmiennych. Wartość krytyczna $\chi^2_{(\alpha,r)}$ wg rozkładu χ^2 przy poziomie istotności $\alpha = 0{,}05$ o dwóch stopniach swobody wynosi 5,991, co potwierdza hipotezę zerową.

Sprawdzamy teraz współliniowość zmiennych objaśniających**.** W sytuacji występowania wysokich współczynników korelacji między zmiennymi objaśniającymi modelu T. Kufel zaleca zbadanie dodatkowo współliniowości zmiennych objaśniających[114]. Sytuacja taka powoduje zaniżanie wartości statystyki *t*-Studenta w ocenie istotności parametru. Ocenę stopnia współliniowości zmiennych objaśniających w programie *Gretl* można dokonać posługując się miarą VIF_j określaną jako czynnik inflacji wariancji i korzystając z menu *Testy/ Ocena współliniowości VIF* (czynnika powiększania wariancji).

Tab. 7.9. Raport z oceny współliniowości

Minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości-rozdęcia wariancji
*X*9 22,992
*X*10 10,695
*X*13 21,778
*X*48 14,742
*X*50 5,717
*X*59 7,482

Źródło: Opracowanie własne w *Gretl*.

Widzimy, że tylko dla zmiennych *X*50 oraz *X*59 miara $VIF_j < 10$. Miarę tę obliczyć możemy również według wzoru:

$$VIF_j = \frac{1}{1 - R_j^2}$$

gdzie: R_j^2 - współczynnik korelacji wielorakiej między zmienną x_j a pozostałymi zmiennymi modelu[115].

Dla naszego modelu współczynniki korelacji między zmiennymi objaśniającymi podawane przez *Gretl* są wysokie.

Tab. 7.10. Współczynniki korelacji między zmiennymi objaśniającymi

*X*9	*X*10	*X*13	*X*48	*X*50	*X*59	
1,0000	0,9060	-0,9017	-0,8529	-0,7667	-0,7209	*X*9
	1,0000	-0,7603	-0,7515	-0,7333	-0,7093	*X*10
		1,0000	0,9309	0,8747	0,8199	*X*13
			1,0000	0,8202	0,9037	*X*48
				1,0000	0,7562	*X*50
					1,0000	*X*59

Źródło: Opracowanie własne w *Gretl*.

W zaistniałej sytuacji dokonamy korekty w zestawie zmiennych objaśniających dodatkowo dobierając intuicyjnie zmienne, które mogą mieć wpływ na zmienną zależna *Liczba bezrobotnych*:

- przeciętne zatrudnienie w tys./gospodarka magazynowa i łączność (*X*12);

[114] Kufel T.: *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu Gretl*, Wydawnictwo Naukowe PWN, Warszawa 2007, rozdz. 4.3.6.

[115] Nowak E., *Zarys metod ekonometrii. Zbiór zadań,* Wydawnictwo Naukowe PWN, Warszawa 2002.

- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach): absolwenci - wyższe w tys. (*X*60), absolwenci - dla dorosłych w tys. (*X*61);

- rynek mieszkaniowy/wskaźnik rozwodów na 1000 ludności (*X*79).

Teraz rozpoczynamy procedurę eliminacji *a posteriori* od nowa realizując również każdorazowo test współliniowości zmiennych objaśniających. Zmienną zależną jest nadal *X*3. Raport *Gretl*-a bez podawania statystyki opisowej podano w tabeli 7.11.

Tab. 7.11. Model po ponownej estymacji parametrów

Zmienna	Współczynnik	łąd stand.	*t*-Studenta	Wartość *p*	
const	20051,9	3633,49	5,519	0,0027	***
***X*9**	-4,77960	1,07768	-4,435	0,0068	***
***X*10**	4,25611	1,71349	2,484	0,0556	*
***X*13**	-2,12944	1,33769	-1,592	0,1723	
***X*48**	-0,882948	0,273869	-3,224	0,0234	**
***X*50**	-0,141497	0,0638607	-2,216	0,0775	*
***X*59**	20,4479	13,4971	1,515	0,1902	
*X*12	-1,70175	1,79905	-0,9459	0,3876	
X60	-1,84050	1,98024	-0,9294	0,3953	
*X*61	-3,49460	9,19190	-0,3802	0,7194	
*X*79	303,742	667,829	0,4548	0,6683	

Wyłączając stałą, największa wartość p jest dla zmiennej 63 (*X*61)

Źródło: Opracowanie własne w *Gretl*.

Postępując dalej przy korzystaniu z podpowiedzi na końcu raportu oraz śledząc VIF_j w teście współliniowości wyeliminowano kolejno zmienne: *X*61, *X*79, *X*60, *X*13 otrzymując model wykazujący poprawę tej miary przy poszczególnych parametrach zmiennych:

Tab. 7.12. Parametry modelu uznanego za końcowy wg kryterium *p*

Zmienna	Współczynnik	Błąd stand.	*t*-Studenta	Wartość *p*	
const	16170,6	862,311	18,75	4,02E-09	***
*X*9	-3,69031	0,330486	-11,17	5,73E-07	***
*X*10	2,07071	0,831721	2,490	0,0320	**
*X*48	-1,01061	0,204500	-4,942	0,0006	***
*X*50	-0,147759	0,0190528	-7,755	1,54E-05	***
*X*59	14,2650	4,63422	3,078	0,0117	**

Źródło: Opracowanie własne w *Gretl*.

Tab. 7.13. Ponowna ocena współliniowości powyższego modelu

Ocena współliniowości VIF - czynnika powiększania wariancji
Minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia wariancji

*X*9	11,694
*X*10	7,295
*X*48	13,503
*X*50	3,382
*X*59	7,180

Źródło: Opracowanie własne w *Gretl*.

Największa wartość VIF_j występuje przy zmiennej *X*48. Decydujemy się więc na usunięcie tej zmiennej z modelu i w rezultacie otrzymujemy podany niżej model oraz wynik testu współliniowości.

Tab. 7.14. Skorygowany model wg testu współliniowości

Zmienna	Współczynnik	Błąd stand.	*t*-Studenta	Wartość *p*	
const	14435,8	1393,28	10,36	5,18E-07	***
*X*9	-2,59783	0,434569	-5,978	9,21E-05	***
*X*10	0,218794	1,31349	0,1666	0,8707	
*X*50	-0,175889	0,0321647	-5,468	0,0002	***
*X*59	-3,84933	5,01594	-0,7674	0,4590	

Średnia arytmetyczna zmiennej zależnej = 2544,19
Odchylenie standardowe zmiennej zależnej = 585,608
Suma kwadratów reszt = 576569
Błąd standardowy reszt = *228,944*
Współczynnik determinacji (R^2) = 0,88792
Skorygowany współczynnik determinacji $\bar{R}^2$ = 0,84716
Statystyka *F* (4, 11) = 21,785 (wartość *p* = 3,48e-005)
Logarytm wiarygodności = -106,641
Kryterium informacyjne Akaike'a (*AIC*) = 223,282.

Źródło: Opracowanie własne w *Gretl*.

Tab. 7.15. Ocena współliniowości modelu podanego w tabeli 7.14

Wartości > 10.0 mogą wskazywać na problem współliniowości - rozdęcia wariancji	
*X*9	6,461
*X*10	5,814
*X*50	3,080
*X*59	2,688

Źródło: Opracowanie własne w *Gretl*.

Z powyższej tabeli wynika, że dla zmiennych spełniony jest warunek wartości współczynnika $VIF_j < 10$, co wskazuje na nie wystąpienie problemu rozdęcia.

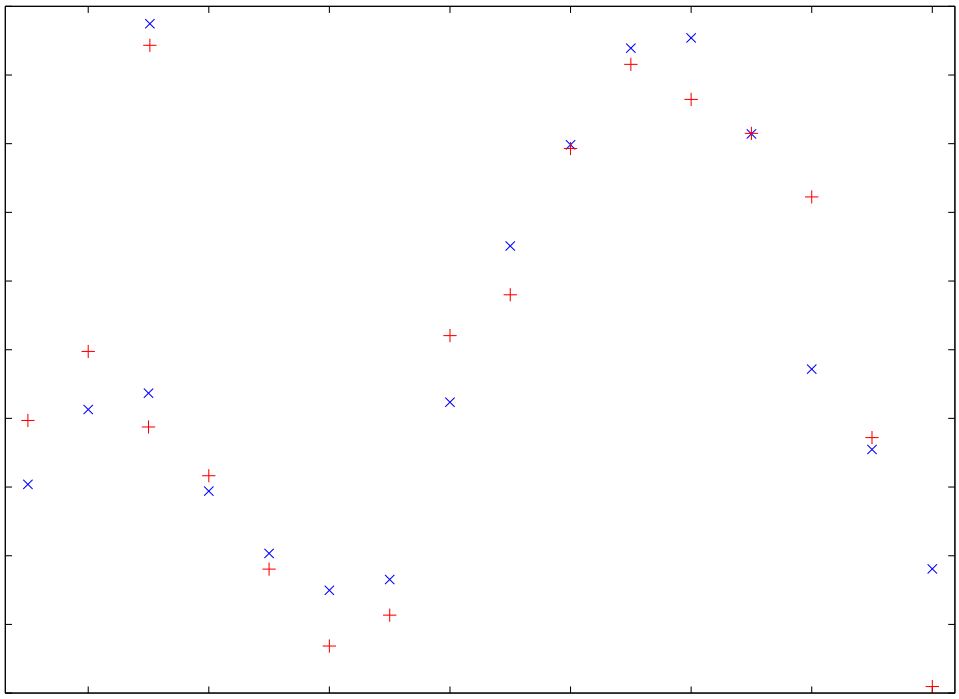
* * *

Analiza współczynników korelacji eliminuje zmienne pozostające poza modelem zbyt silnie skorelowane ze zmienną ostatnio dobraną do modelu. Jednakże procedura eliminacji *a posteriori* *Gretl*-a wymaga dodatkowo śledzenia miary VIF_j po każdym kolejnym kroku budowy modelu ekonometrycznego. Rezultatem końcowym pracy jest model stanowiący consensus w budowie modelu ekonometrycznego według dwóch miar jednocześnie: statystyki *t*-Studenta, miary VIF_j współliniowości danej zmiennej. Tak więc w rezultacie końcowym nasza zmienna zależna *X*3 *Liczba bezrobotnych* zależy od zmiennych objaśniających:

- przeciętne zatrudnienie w tys.: przemysł (*X*9), budownictwo (*X*10);
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych): zużycie wody z wodociągów w ciągu roku w hm^3 (*X*49), zużycie energii elektrycznej w ciągu roku w GWh (*X*50);
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci - policealne w tys. (*X*59).

Zaobserwujmy teraz obraz danych empirycznych oraz modelowych na wykresie punktowym. Widzimy jak możliwie dokładnie dla poszczególnych 16. obserwacji, wartości określone przez poniższy model są przybliżone do wartości empirycznych (statystycznych) zmiennej *X*3 *Liczba bezrobotnych* w tys.

$$\hat{x}_3 = 14435{,}8 - 2{,}59783\, x_9 + 0{,}218794\, x_{10} - 0{,}175889\, x_{50} - 3{,}84933\, x_{59}$$



Źródło: Opracowanie własne w *Gretl*.

Rys. 7.16. Wartości empiryczne i modelowe zmiennej *X*3

8. Prognozowanie liczby pasażerów lotów z zastosowaniem *WinQSB*

8.1. Wstęp

Trafne prognozowanie zjawisk, czy też wskaźników na podstawie szeregów czasowych jest niełatwą sztuką natury merytorycznej. Wiele temu poświęcono publikacji, szczególnie jest to istotne w dłuższych okresach czasowych. Jednak coraz bardziej odczuwamy potrzebę wiedzy co nas czeka w przyszłości, zwłaszcza w odniesieniu do większych grup ludności. Niepokoi nas ocieplenie klimatu, pandemie oraz potencjalne zagrożenia wojenne w coraz to innych obszarach naszego Globu. Wyrazem szerokiego zainteresowania problematyką prognozowania, w szczególności długookresowego są również liczne artykuły, które spotykamy w domenie Google. Na frazę „*Prognozowanie długookresowe*" w tej wyszukiwarce Google otrzymano 344 odpowiedzi w 0,57 sekundy, zróbmy zatem rekonesans choć po części tych materiałów.

Paweł Dittmann w publikacji „*Prognozowanie w zarządzaniu przedsiębiorstwem*" podejmuje temat prognozowania długookresowego sprzedaży[116]. Jego zdaniem prowadzenie tego działania jest konieczne dla zaprogramowania przyszłych zmian organizacyjnych . Sprowadzają się one do wykreowania innowacyjnych form sprzedaży wyrobów, sposobów wejścia na nowe rynki zbytu, czy też poprzez odpowiednią reklamę uformowania dodatkowych kanałów dystrybucji. Branżowe podejście do przewidywania sytuacji spotykamy w opracowaniu „*Prognozowanie długoterminowe - metody i uwarunkowania na przykładzie rynku wieprzowiny*"[117]. Autor zaznacza jednak, że wskazane przewidywania należy traktować z dużą dozą ostrożności. Jak już wspomniano w rozdziale 1. Stanisław Krasowicz w artykule „*Niematematyczne metody prognozowania w rolnictwie*" sugeruje potrzebę zastosowania różnych kompozycji tych metod, aby estymować parametry w obszarze dość trudnym jakim jest produkcja i zbyt wytworów rolnictwa[118].

Interesujący i specyficzny temat, lecz z zakresu prognozowania krótkoterminowego zużycia energii elektrycznej z wykorzystaniem sieci neuronowej, podejmują autorzy: Tomasz Jeżyk i Andrzej Tomczewski[119]. W opracowaniu „*Ważny głos o prognozach makroekonomicznych*" wskazuje się na istotność prognozowania długookresowego, zwłaszcza w tak trudnym obszarze jakim jest handel zagraniczny[120]. Długookresowe prognozowanie popytu/sprzedaży/cen/kursów walut jest tematem publikacji „*Zarządzanie zapasami - metody uzupełniania, magazynowanie i dystrybucja w łańcuchach dostaw*"[121]. Sięgnijmy teraz po artykuł Elżbiety Mączyńskiej „*Globalna niepewność a społeczny futuryzm*"[122]. Występuje tu zwrócenie uwagi na zależność prognozowania długookresowego i tzw. ekonomii niepewności, w sytuacji bezprecedensowego dynamizmu współczesnych przemian w gospodarce globalnej. Kolejny autor Jarosław Hermaszewski artykułem „*Problemy wyodrębniania procesów zarządzania finansami samorządu terytorialnego*" uważa za celowe prowadzenie prognoz długoterminowych wydatków finansowych, pomimo ryzyka jakimi są

[116] https://media.statsoft.pl/_old_dnn/downloads/prprognozowanie1.pdf.
[117] https://www.ierigz.waw.pl.
[118] https://iung.pl/sir/zeszyt14_2.pdf.
[119] http://www.iee.put.poznan.pl/wydawnictwa/wydawnictwaFiles/%5C2014%5C5_Academic%20Journals%20Poznan%20University%20of%20Technology_79%5C!121_15_Tomczewski.pdf.
[120] http://www.pte.pl/pliki/0/251/k18nrk.pdf.
[121] https://www.projektgamma.pl/uploads/pdf/Zarzadzanie_zapasami.pdf.
[122] https://cor.sgh.waw.pl/bitstream/handle/20.500.12182/822/Biuletyn_2_2009-strony-4-7.pdf?sequence=2&isAllowed=y.

one obarczone[123]. Obserwowane są coraz wyższe koszty pracy, a ich prognozowanie długookresowe jest złożonym przedsięwzięciem merytorycznym, gdyż trzeba rozpatrywać szanse i zagrożenia z tym związane. Jest to przedmiotem artykułu „*Rosnące koszty pracy to główne zmartwienie dyrektorów finansowych*"[124].

Zwróćmy jeszcze uwagę na jedną z publikacji pozyskanych hasłem „*long-term forecasting*", czyli długoookresowe prognozowanie. W "*Short & Long Cash Forecasting*" wymieniono różne techniki prognozowania i autor zadaje sobie pytanie „*Which method is best for term forecasting*?", tj. *Która z metod jest najlepsza do prognozowania długookresowego*?[125]. W kolejności warto jeszcze bliżej zapoznać się z opracowaniem Gawona Yoona pt. „*Porównanie długoterminowych metod prognozowania: komentarz"*[126]. W tej pracy poleca się stosowanie różnych metod regresji z autokorelacją dla długoterminowej prognozy koreańskiego indeksu cen towarów i usług konsumpcyjnych. Nadmienia się w niej, że w roku 1998 Seo i Choi uznali tą metodę jako dającą dobre wynik ze średnim błędem kwadratowym (*MSE*) równym 0,1896. Nadmienię tu, że *MSE* jest często używaną miarą poziomu błędu prognozy i wyrażone jest wzorem:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$$

gdzie: n - liczba obserwacji, y_i - obserwacja w okresie t_i, a $\hat{y}_i$ jest oszacowaniem prognozy wynikającej z zastosowanej metody.

Po tych słowach wstępnych podejmijmy próbę prognozowania natężenia ruchu pasażerów na lotniskach, co wymaga odpowiedniego doboru i zaimplementowania parametrów modeli szeregów czasowych. Ruch ten zmienia się w ciągu doby, przy czym największe obciążenie notowane jest w godzinach przedpołudniowych i wczesnych popołudniowych. Obserwacje natężenia ruchu można też prowadzić w tygodniach miesiąca, miesiącach roku, czy też w kwartałach lub latach horyzontu prognozowania.

Inspiracją do tego opracowania była statystyka liczby pasażerów prowadzona na lotnisku Bolonia (Włochy) w latach 2017-2015 wykorzystana w pracy magisterskiej opublikowanej w Internecie[127]. Dość złożone jest postępowanie związane z rozpatrzeniem trafności doboru modelu szeregu czasowego liczby pasażerów dla prognozowania obciążenia lotniska w kolejnych przykładowo kwartałach lat rozpatrywanego okresu. W ciągu roku występuje sezonowość odpowiadająca czterem kwartałom. Szczególnie istotne jest to dla badania występowania natężenia ruchu pasażerów w okresie lata i zimy. W okresie wiosny i jesieni obserwowane jest zmniejszenie liczby pasażerów. Zatem nie tylko występuje zmienność liczby pasażerów w poszczególnych godzinach doby, ale także sezonowość w okresie roku oraz wyraźna tendencja rosnąca w okresie lat 2007-2015. Zanim przejdziemy do przykładu empirycznego przypomnijmy sobie z ekonometrii wielorównaniowy Holta-Wintersa z trendem i sezonowością. Rozpocznijmy od modelu prostego wyrównywania wykładniczego o skrócie SES, który jak już nadmieniono określony jest zależnością[128]:

[123] http://hermaszewski.glogow.pl/wp-content/uploads/2011/09/Karpacz-zarz%C4%85dzanie-procesami.pdf.
[124] https://picm.pl/wp-content/uploads/2019/09/CMM-11-2018.pdf.
[125] https://www.highradius.com.
[126] Gawon Yoon, *A Comparison of Long-term Forecasting Methods: A Comment*, http://www.kostat.go.kr/attach/journal/3-2-6.PDF.
[127] Wargentin Robin, *Long-term and Short-term Forecasting Techniques for Regional Aiport Planning*, Degree Project in Mathematics second cycle, 30 credits Stockholm, Sweden 2016.
[128] Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.

$$F_n = \alpha Y_{n-1} + (1 - \alpha) \cdot F_{n-1}$$

gdzie: α - parametr wyrównania (0;1); gdy celem jest silne wygładzanie szeregu, to α powinno mieć małą wartość, np. 0,1 lub 0,2.

Postać tego modelu doprowadziła do uformowania addytywnego modelu Holta-Wintersa. Model *SES* wymaga znajomości wartości początkowej, którą przyjmuje się $F_2 = Y_1$. Prognoza w tym modelu na kolejny okres jest ważoną średnią ruchomą wartości rzeczywistej z okresu bezpośrednio poprzedzającego oraz wartości prognozowanej. Do podanego modelu Holta-Wintersa (bez sezonowości) o parametrze wygładzania wykładniczego α wprowadza się drugi parametr β w celu wygładzania trendu[129]. Wygładzona jest wartość trendu z poprzedniego okresu, by usunąć błędy przypadkowe. Następujące równania służą do zapisu tego modelu[130]:

$$S_n = \alpha Y_n + (1 - \alpha) \cdot (S_{n-1} + b_{n-1}),$$
$$b_n = \beta \cdot (S_n - S_{n-1}) + (1 - \beta) \cdot b_{n-1},$$
$$F_{n+m} = S_n + b_n \cdot m.$$

przy czym: b_n oznacza wyrównywany trend w okresie n.

W obliczeniach z użyciem modelu Holta-Wintersa przyjmujemy następujące wartości początkowe:

$$S_1 = Y_1 \qquad b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2.$$

Staramy się dobrać parametry α, β tak, aby doprowadzić do minimalnej wartości pierwiastka kwadratu średniego błędu *RMSE* prognozy *ex post*. Wymieniony model Holta-Wintersa stosuje się wtedy, gdy występuje trend i wahania okresowe. Warto w tym miejscu przytoczyć wzór na pierwiastek średniego kwadratu błędu[131]:

$$RMSE = \sqrt{\frac{1}{1+h}\sum_{n=S}^{S+h}(\hat{Y}_n - Y_n)^2}$$

Zatem jest on pierwiastkiem ze stosowanej w programie *FC* (*Forecasting and Linear Regression*) pakietu *WinQSB* miary *MSE*. Wymieniony symbol Y_n oznacza wartość prognozowaną w okresie *n*, a *h* - horyzont prognoz wygasłych.

Ogólnie biorąc zmiany w czasie wywołane są takimi przyczynami, jak trend, wahania sezonowe/okresowe, wahania cykliczne oraz wahania przypadkowe[132]. Trendem nazywamy długookresowe zmiany powodujące wzrost lub spadek zmiennej objaśniającej. Wahania sezonowe to zmiany poziomu badanego zjawiska w tym samym czasie każdego roku. Jeśli występują nieregularne zmiany określonego zjawiska (badanej zmiennej) i mają one charakter falowy wokół trendu, to mówimy o wahaniach cyklicznych. Wahania przypadkowe (losowe) to zakłócenia w regularnych zmianach obserwowanej zmiennej objaśniającej. W rzeczywistych szeregach czasowych zmiany nie występują tylko w wymienionych formach, lecz są z nich złożone. Jeśli np. dla kolejnych okresów amplituda zmian jest stała przy istniejącym trendzie i wahaniach cyklicznych, to taki model nazywamy addytywnym, a jeśli nie - to modelem multiplikatywnym.

W modelu Holta-Wintersa bez sezonowości występowały dwa parametry wyrównywania, przy czym α odnosiło się do wartości szeregu zmiennej objaśnianej, a parametr β dotyczył trendu. Wyrównywanie sezonowości w modelu Holta-Wintersa przeprowadza się według określonej

[129] Ibidem, rozdz. 2.14.
[130] Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski, Opole 2010, rozdz. 1.13.
[131] Ibidem, s. 574.
[132] Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, t. 2: *Ekonometria*, op. cit., rozdz. 2.2.

procedury poprzez zastosowanie parametru γ (gamma), stosowanego do obliczania indeksów sezonowości, wówczas sezonowość multiplikatywna modelu Holta-Wintersa ma postać:

$$Y_{n+1} = (S_n + b_n) \cdot I_{n-L+1} + z_{n+1},$$

gdzie: I_{n-L+1} - wyrównana wartość indeksu sezonowości na okres $n + 1$; z_{n+1} - błąd w okresie $n + 1$, L - długość cyklu sezonowości (12 dla danych miesięcznych, 4 dla kwartalnych).

Wyrównana wykładniczo wartość Y_n po usunięciu sezonowości przyjmuje postać:

$$S_n = \alpha \cdot (Y_n/I_{n-L}) + (1 - \alpha) \cdot (S_{n-1} + b_{n-1}),$$

Natomiast wyrównana wykładniczo wartość trendu w okresie n:

$$b_n = \beta (S_n - S_{n-1}) + (1 - \beta) \cdot b_{n-1}.$$

Wyrównany wykładniczo indeks sezonowości w okresie n:

$$I_n = \gamma \cdot (Y_n/S_n) + (1 - \gamma) I_{n-L}.$$

Prognoza na okres $(n + m)$, przy czym m to horyzont prognozy, obliczana jest według:

$$F_{n+m} = (S_n + b_n \cdot m) \cdot I_{n-L+1}.$$

Dla naszego przykładu trzeba jeszcze podać równania stanowiące model Holta-Wintersa z sezonowością addytywną[133]:

$$S_n = \alpha (Y_n - I_{n-L}) + (1 - \alpha) (S_{n-1} + b_{n-1}),$$
$$b_n = \beta (S_n - S_{n-1}) + (1 - \beta) b_{n-1},$$
$$I_n = \gamma (Y_n - S_n) + (1 - \gamma) I_{n-L},$$
$$F_{n+m} = (S_n + b_n m) + I_{n-L+1}.$$

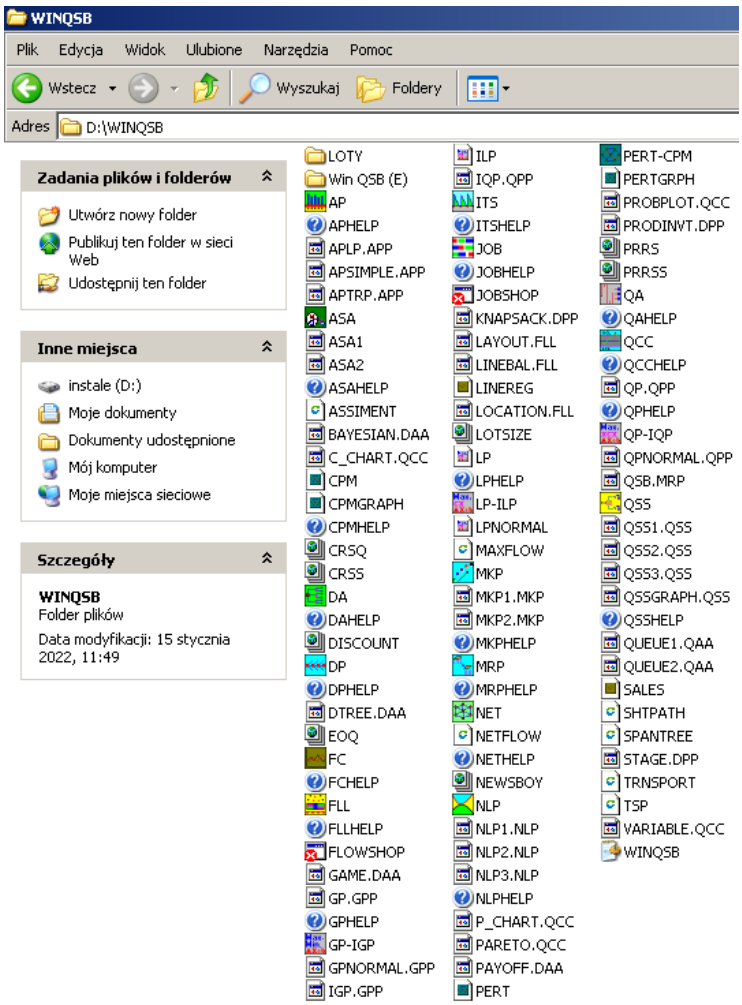
Już z tego pobieżnego przytoczenia postępowania widzimy, że manualne budowanie modeli adaptacyjnych jest pracochłonne. Z tego względu trzeba zastosować odpowiedni program komputerowy do automatyzacji obliczeń. Istotne to jest zwłaszcza w doborze najlepszych parametrów α, β, γ z punktu widzenia miary prognozy *MSE*. Automatyczny dobór parametrów dokonuje program *FC* poprzez wskazanie opcji (*Search the best*) z punktu widzenia minimum *MSE*. W programie *FC* stosowane są również inne miary błędu oszacowania wygasłych prognoz, a niektóre z nich znajdujemy w Aneksie[134].

8.2. Rozpatrzenie możliwości zastosowania programu *FC*

Po tych wstępnych rozważaniach przejdźmy do wykorzystania programu *FC* pracującego w środowisku systemu operacyjnego *Windows XP*. Program ten wchodzi w skład pakietu *WinQSB* składającego się z wielu modułów z zakresu statystyki, ekonometrii i optymalizacji (zob. rysunek 8.1).

[133] Ibidem, s. 94.

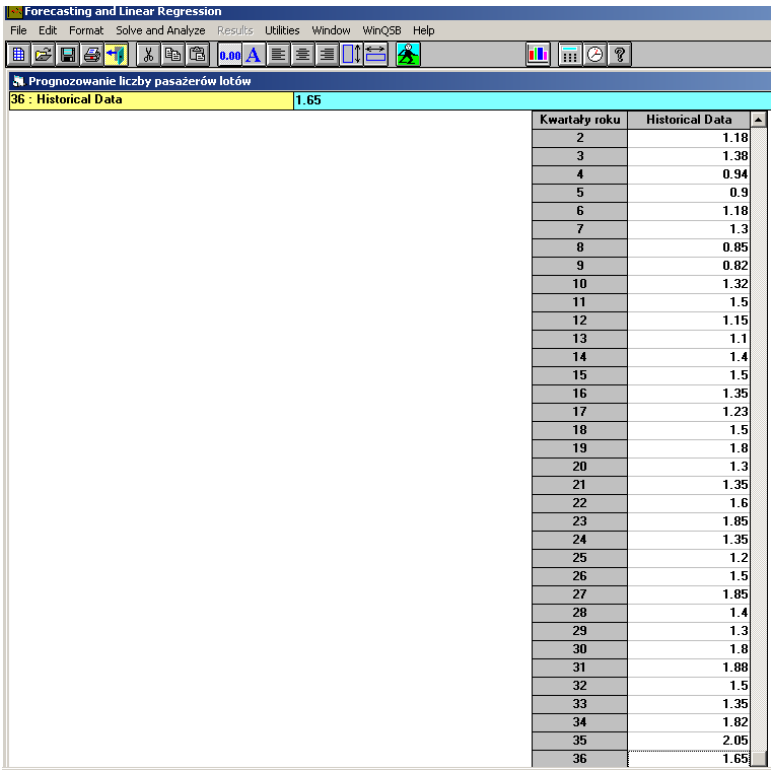
[134] Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*, op. cit., s. 555.



Źródło: Opracowanie własne na podstawie pakietu *WinQSB*.

Rys. 8.1. Moduły programowe pakietu *WinQSB* zawarte w katalogu *D:\WINQSB*

Po wywołaniu programu *FC* ukazuje się nam okno dialogowe do podania specyfikacji nowego problemu. Wybieramy typ problemu „*Time Series Forecasting*" (*szereg czasowy prognozowania*) i wprowadzamy: *Problem Title - Prognozowanie lotów pasażerów*, *Time Unit - Kwartały roku*, *Number of Time Units (Periods) - Liczba okresów obserwacji* - 36. Po naciśnięciu OK udostępniona jest na ekranie tabela do wpisywania wartości 36 obserwacji (zob. rysunek 8.2).

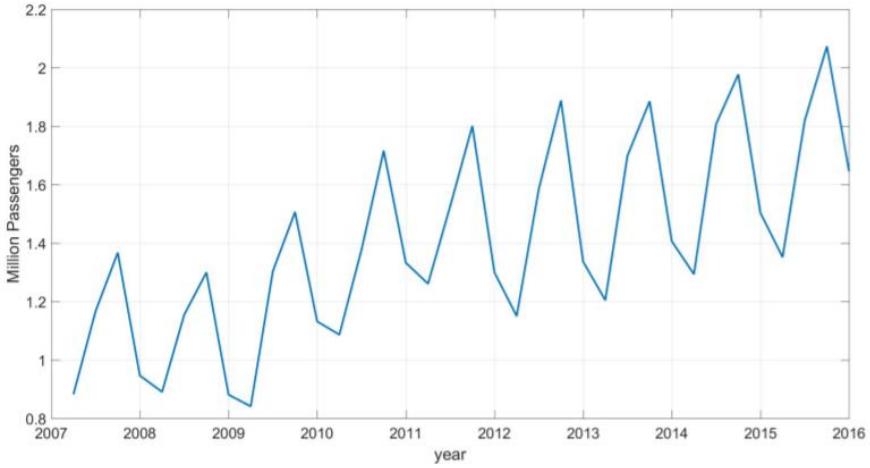


Kwartały roku	Historical Data
2	1.18
3	1.38
4	0.94
5	0.9
6	1.18
7	1.3
8	0.85
9	0.82
10	1.32
11	1.5
12	1.15
13	1.1
14	1.4
15	1.5
16	1.35
17	1.23
18	1.5
19	1.8
20	1.3
21	1.35
22	1.6
23	1.85
24	1.35
25	1.2
26	1.5
27	1.85
28	1.4
29	1.3
30	1.8
31	1.88
32	1.5
33	1.35
34	1.82
35	2.05
36	1.65

Źródło: Opracowanie własne na podstawie pakietu *WinQSB*.

Rys. 8.2. Dane empiryczne wprowadzone do prognozowania, przy czym kwartał pierwszy ma 0,9 mln pasażerów

W naszym przykładzie w horyzoncie lat 2007-2015 mamy (9 x 4 = 36) kwartałów obserwacji liczby pasażerów w milionach odlatujących lub przylatujących na przykładowe lotnisko w Bolonii we Włoszech. Możemy dokonać symulacji doboru adekwatnego modelu prognozowania na podstawie przyjętego szeregu czasowego, biorąc za kryterium jedną z miar oceny błędu prognozowania np. *MSE*. Jednak jak już nadmieniono, bliższe przyjrzenie się wykresowi zamieszczonemu w opracowaniu Wargentina Robina (zob. rysunek 8.3) podpowiada nam, że należy wybrać od razu model Holta-Wintersa z sezonowością addytywną (*HWA*).



Źródło: https://www.diva-portal.org/smash/get/diva2:954470/FULLTEXT02[135], Dodatek rys. A.1.

Rys. 8.3. Obserwowane liczby pasażerów w ujęciu kwartalnym na lotnisku w Bolonii w latach 2007-2015

[135] Wargentin Robin, *Long-term and Short-term Forecasting Techniques1,35 for Regional Aiport Planning*, op. cit.

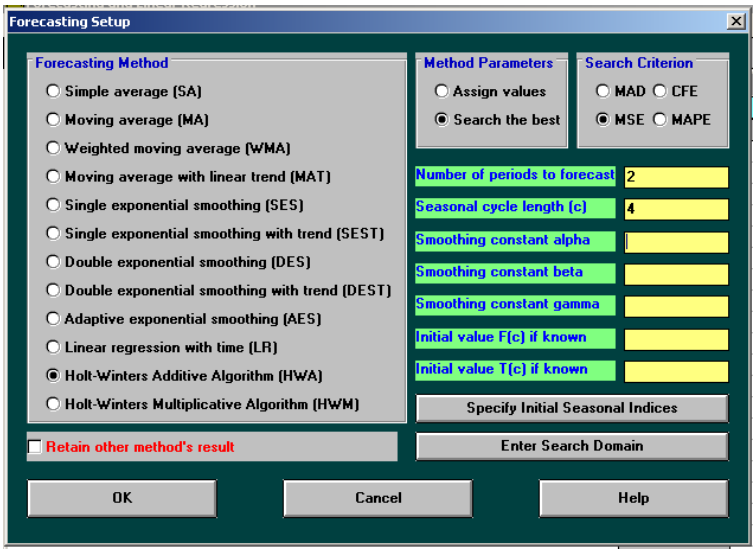
Skorzystajmy więc z tej możliwości, lecz w programie *FC* musimy najpierw określić elementy wejściowe (zob. rysunek 8.4). Po tej czynności ustawiamy skok przeprowadzenia komputerowych obliczeń na 0.01 (z kropką dziesiętną) dla następujących parametrów wielorównaniowego modelu Holta-Wintersa z sezonowością addytywną:

$\propto$ - wyrównywanie zmienności obserwacji,

β - wyrównywanie zmienności trendu,

γ - wyrównywanie zmienności sezonowości.

Trzeba tu podkreślić, że program *FC* wymaga wpisywania kropki dziesiętnej zamiast przecinka, przy podawaniu danych liczbowych. Określenie wstępne wyżej wymienionych trzech parametrów jest kłopotliwe, stąd powierzamy to programowi dając opcję „*Search the best*" w oknie dialogowym „*Forecasting Setup*".



Źródło: Opracowanie własne w *WinQSB*.

Rys. 8.4. Okno dialogowe po wywołaniu funkcji *Solve and Analyze/Perform Forecasting*

Program rozpatruje kolejne modele HWA biorąc skok parametru $\propto$, potem parametru β, a następnie γ zeskokiem co 0,01. Ten komputerowy proces optymalizacji trwa dość długo, aż w efekcie uzyskamy wynik optymalny dla prognoz (zob. rysunki 8.5 i 8.6). Widzimy tu wyszczególnienie wartości obserwacji (1-36) i odpowiadające im wygasłe prognozy. Program *FC* prognozowania według modelu HWA podaje nam również oszacowanie liczby pasażerów na żądane dwa dalsze kwartały, a ponadto optymalne wartości parametrów $\propto$, β, γ wielorównaniowego modelu addytywnego Holta-Wintersa. Obliczone są także dla kolejnej obserwacji błędy oszacowania według innych różnych miar, jednak jako podstawową miarę przyjęto MSE (średni kwadratowy błąd prognoz).

01-23-2022 Kwartały roku	Actual Data	Forecast by HWA	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-square
1	0,9000								
2	1,1800								
3	1,3800								
4	0,9400								
5	0,9000	0,9000	0	0	0	0	0	0	0
6	1,1800	1,1800	0	0	0	0	0	0	1,0000
7	1,3000	1,3800	-0,0800	-0,0800	0,0267	0,0021	2,0513	-3,0000	
8	0,8500	0,8942	-0,0442	-0,1242	0,0311	0,0021	2,8387	-4,0000	
9	0,8200	0,8263	-0,0063	-0,1305	0,0261	0,0017	2,4249	-5,0000	
10	1,3200	1,0987	0,2213	0,0908	0,0586	0,0096	4,8153	1,5485	0,8479
11	1,5000	1,4167	0,0833	0,1741	0,0622	0,0092	4,9205	2,8006	0,8033
12	1,1500	1,0293	0,1207	0,2948	0,0695	0,0099	5,6174	4,2429	0,8273
13	1,1000	1,0661	0,0339	0,3287	0,0655	0,0089	5,3354	5,0160	0,8299
14	1,4000	1,3876	0,0124	0,3410	0,0602	0,0080	4,8902	5,6643	0,8773
15	1,5000	1,5933	-0,0933	0,2477	0,0632	0,0081	5,0112	3,9183	
16	1,3500	1,1150	0,2350	0,4827	0,0775	0,0120	6,0443	6,2259	
17	1,2300	1,2148	0,0152	0,4979	0,0727	0,0111	5,6742	6,8450	
18	1,5000	1,5305	-0,0305	0,4673	0,0697	0,0104	5,4143	6,7030	
19	1,8000	1,7113	0,0887	0,5560	0,0710	0,0102	5,3818	7,8329	
20	1,3000	1,3594	-0,0594	0,4966	0,0703	0,0098	5,3309	7,0685	
21	1,3500	1,2886	0,0614	0,5581	0,0697	0,0094	5,2850	8,0020	
22	1,6000	1,6287	-0,0287	0,5294	0,0675	0,0090	5,0910	7,8474	
23	1,8500	1,8191	0,0309	0,5603	0,0655	0,0085	4,9111	8,5497	
24	1,3500	1,4279	-0,0779	0,4824	0,0662	0,0084	4,9541	7,2921	
25	1,2000	1,3533	-0,1533	0,3291	0,0703	0,0091	5,3265	4,6811	
26	1,5000	1,5651	-0,0651	0,2640	0,0701	0,0089	5,2816	3,7685	

Źródło: Opracowanie własne w *WinQSB*.

Rys. 8.5. Efekt prognozowania według modelu HWA po wywołaniu funkcji „*Forecast Results*”

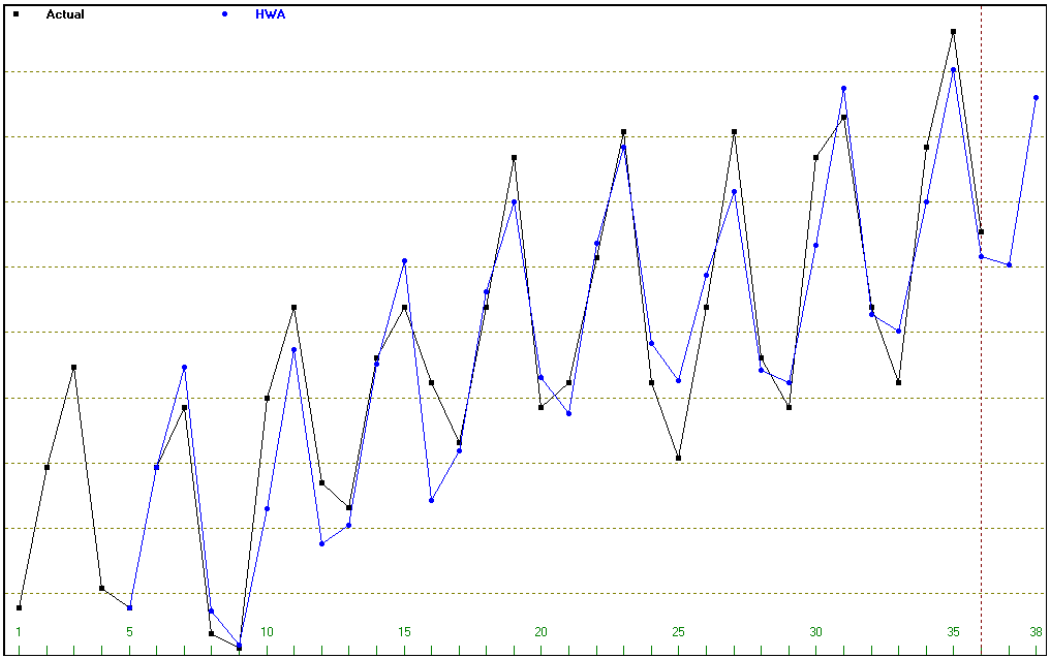
27	1,8500	1,7305	0,1195	0,3836	0,0722	0,0091	5,3328	5,3112	0,9950
28	1,4000	1,3754	0,0246	0,4081	0,0702	0,0088	5,1837	5,8111	0,9948
29	1,3000	1,3496	-0,0496	0,3585	0,0694	0,0085	5,1290	5,1657	0,9929
30	1,8000	1,6232	0,1768	0,5354	0,0735	0,0094	5,3095	7,2799	0,9372
31	1,8800	1,9382	-0,0582	0,4772	0,0730	0,0092	5,2275	6,5390	0,9813
32	1,5000	1,4850	0,0150	0,4922	0,0709	0,0089	5,0766	6,9418	0,9818
33	1,3500	1,4526	-0,1026	0,3896	0,0720	0,0089	5,1636	5,4111	0,9829
34	1,8200	1,7111	0,1089	0,4985	0,0732	0,0090	5,1910	6,8077	0,9559
35	2,0500	1,9753	0,0747	0,5732	0,0733	0,0089	5,1412	7,8231	0,9384
36	1,6500	1,6010	0,0490	0,6223	0,0725	0,0087	5,0734	8,5809	0,9356
37		1,5847							
38		1,9178							
CFE		0,6223							
MAD		0,0725							
MSE		0,0087							
MAPE		5,0734							
Trk.Signal		8,5809							
R-square		0,9356							
		c=4							
		Alpha=0,54							
		Beta=0,06							
		Gamma=0,12							
		F(0)=1,1							
		T(0)=0							
		S(1)=-0,2000							
		S(2)=0,0800							
		S(3)=0,2800							
		S(4)=-0,1600							

Źródło: Opracowanie własne w W*inQSB*.

Rys. 8.6. Efekt prognozowania według modelu HWA po wywołaniu funkcji „*Forecast Results*” cd.

8.3. Wyłonienie trendu rozwojowego

Mając oszacowane prognozy wygasłe oraz dwie wartości prognoz przyszłych możemy w *Excelu* sporządzić wykres z uwzględnieniem doboru trendu dla wskazania dynamiki natężenia ruchu pasażerów na lotnisku w Bolonii. Nadmienię jeszcze, że w latach 2007-2015 dominowały tam tanie linie lotnicze (LCC), choć też działały wcześniejsze profesjonalne linie lotnicze. A teraz skorzystajmy jeszcze z menu programu *FC* o *ścieżce: Results/Show forecasting in Graph* (*Pokaż prognozy wygasłe i przyszłe*). Okno dialogowe „*Chart Vertical Range*” umożliwia nam podanie zakresu osi pionowej wykresu. Dokonujemy podania zakresu liczby pasażerów w milionach (0.8-2.1), przy czym pamiętamy o kropce dziesiętnej i w rezultacie uzyskujemy wykres pokazany na rysunku 8.7.

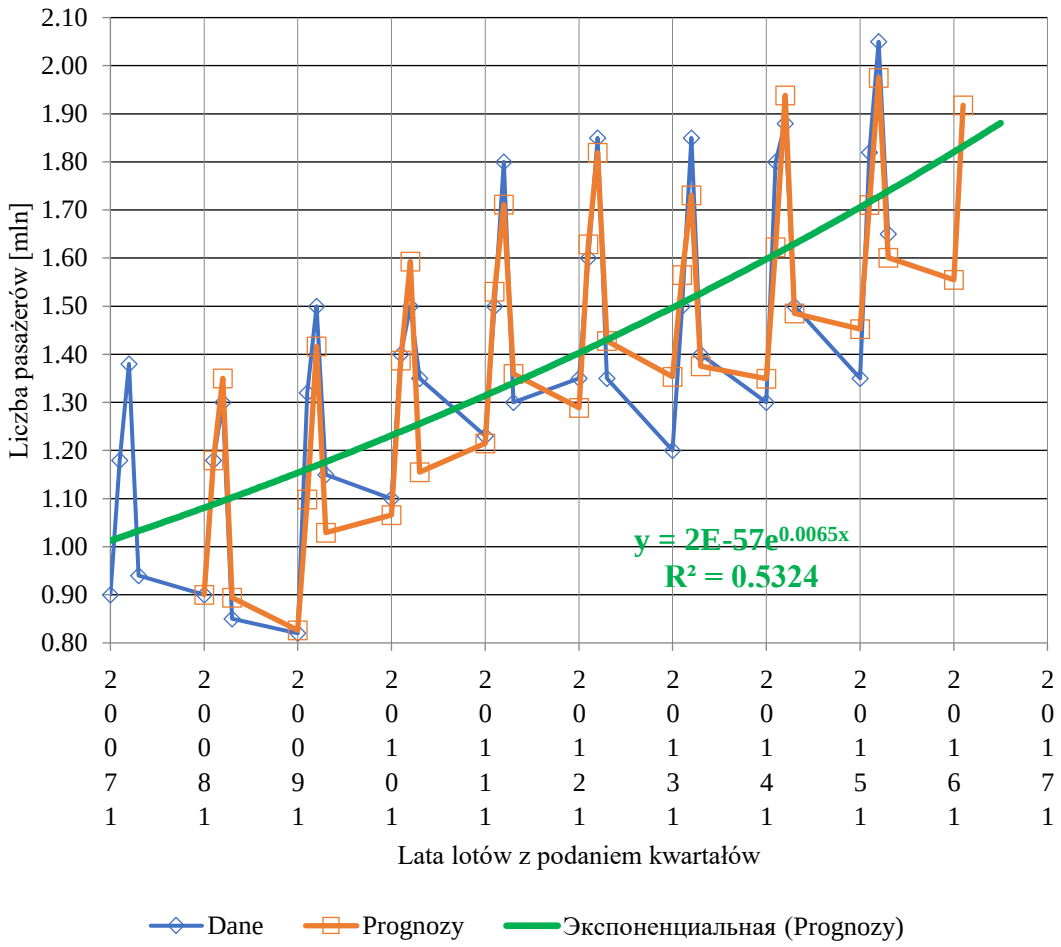


Źródło: Opracowanie własne w *WinQSB*.

Rys. 8.7. Dane rzeczywiste i prognozy kwartalne liczby pasażerów na lotnisku Bolonia

Program wykonał stopniowanie skali pionowej co 0,13 (0,8; 0,93; 1,06; 1.19; 1,32; 1,45; 1,58; 1,71, 1,84; 1,97; 2,1). Stopniowanie skali poziomej co 1 kwartał (1 -36) dane z obserwacji, (37-38) prognozy. Nad wykresem program podaje nam: *Mean of Actual Date* = 1,3903, średnia z danych rzeczywistych; S.D. [Sigma] = 0,3125; suma odchyleń standardowych; *#Forecasting* =2 (liczbę okresów prognozowania).

Powróćmy do *Excela*, gdzie bazując na wynikach z prognozowania według modelu HWA założono plik „*Dane - prognozy*", a następnie wyznaczono trend rozwojowy obciążenia pasażerami lotniska w Bolonii (Włochy) na przykładzie pomyślnej tendencji w latach 2007-2015 (zob. rysunek 8.8). Na wykresie numery okresów kwartalnych (1-36) danych z obserwacji i (37-38) z prognoz zastąpiono oznaczeniem na początku danego roku a potem kwartału. Do uzyskanych prognoz programem *FC* dopasowano trend wykładniczy, którego równanie zamieszczono na wykresie. Widzimy, że ten trend, w miarę najlepszy z proponowanych w *Excelu*, wykazuje dopasowanie do danych rzeczywistych, mierzone współczynnikiem determinacji R^2 = 0,5424. Pozwala jednak w przybliżeniu oszacować tendencję wzrostu natężenia ruchu pasażerów na rozpatrywanym lotnisku w Bolonii, oczywiście w sytuacjach normalnych. Czasy obecnej pandemii zakłócają jednak, miejmy nadzieję przejściowo, skorzystanie z zaproponowanego zastosowania programu *FC* do prognozowania na kolejne kwartały.



Źródło: Opracowanie własne w *Excelu*.

Rys. 8.8. Wyznaczenie trendu wykładniczego jako tendencji rozwojowej liczby pasażerów na przykładowym lotnisku w warunkach normalnych

* * *

Doszukiwanie się adekwatnego modelu prognozowania dla szeregu czasowego o różnorodnej zmienności liczby pasażerów w horyzoncie np. 36 okresów kwartalnych stanowiło wiele stron wspomnianej pracy magisterskiej[136]. Zastosowanie natomiast postępowania symulacyjnego programem *FC*, np. poprzez przejścia przez poszczególne modele podstawowe szeregów czasowych, z punktu widzenia założonego minimum kryterium miary błędu np. *MSA*, daje pewien „*automatyzm*" w procesie wskazania wiarygodnych prognoz na przyszłe kwartały według modelu Holta-Wintersa z sezonowością addytywną. Można to, przy zbieraniu danych statystycznych w krótszych okresach, zastosować do dni tygodnia lub miesięcy okresu stanowiącego podstawę modelowania. Dodam jeszcze, że wykorzystanie programu „*Forecasting and Linear Regression*" (*FC*) to duża wygoda dla badacza, tym bardziej, że umożliwia on optymalne dobranie parametrów wyrównywania wykładniczego wartości obserwacji, trendu i sezonowości.

[136] Wargentin Robin, *Long-term and Short-term Forecasting Techniques for Regional Aiport Planning*, op. cit.

9. Prognozowanie zewnętrznych usług transportowych

9.1. Wprowadzenie[137]

W średnich przedsiębiorstwach produkujących elementy instalacyjne dla budownictwa sytuacja gospodarcza zależy od koniunktury - w szczególności na rynku developerów budujących mieszkania. Występują wahania comiesięczne, pojawia się zmienność trendu oraz cykliczność okresowa. Przedstawione w tej pracy przedsiębiorstwo wytwarza armaturę budowlaną. Zebrane przez licencjata, pod moim kierunkiem, dane świadczą o takiej tendencji. Zatem pojawia się trudność w zakresie niezbędnego prognozowania przewozów przez firmy zewnętrzne np. na najbliższe trzy miesiące.

Przeprowadzone badania na szeregach czasowych dotyczących transportu zaopatrzenia podstawowych grup materiałowych, wyłonionych analizą ABC, jak również dystrybucji wyrobów gotowych ilościowo i wartościowo, wykazały małą przydatność trendów liniowego, potęgowego oraz wykładniczego. Podjęto się więc zamodelowania według trendu wielomianowego wyższego stopnia, co stanowi w miarę dobre dopasowanie do danych empirycznych.

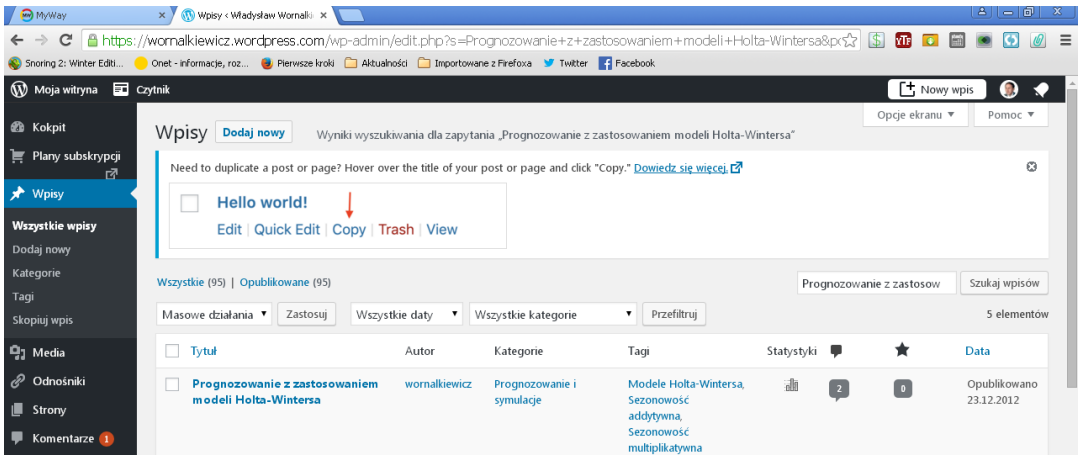
W ramach tej pracy zaobserwowano układanie się zarówno ilościowo jak i wartościowo obserwacji według sezonowości zbliżonej do multiplikatywnej, którą spotykamy w modelach Holta-Wintersa[138]. Skorzystano więc z modułu *Forecasting and Linear Regression* (*Prognozowanie i liniowa regresja*) będącego częścią pakietu *WinQSB* i komputerowo dobrano najlepsze parametry wyrównywania α odnoszące się do wartości obserwacji, β charakteryzujący zmienność trendu oraz γ dotyczący sezonowości. W wyniku przeprowadzonej symulacji zaproponowano testowanie według kilku różnych cykli sezonowości. Ze względu na ograniczone ramy niniejszej monografii, pokazano procedury tylko dla usługi zewnętrznej dotyczącej dystrybucji armatury ilościowo w tonach oraz wartościowo w zł. Przez analogię można bowiem przeprowadzić testowanie dla podstawowych grup materiałów zaopatrzeniowych, a więc odlewów, odkuwek, materiałów hutniczych lub innych.

Prognozowanie krótkookresowe z zastosowaniem modeli Holta-Wintersa stało się ostatnio bardzo popularne. Jak już nadmieniono, zasadniczo możemy wyróżnić trzy rodzaje tych modeli, a więc bez sezonowości, z sezonowością multiplikatywną i z sezonowością addytywną. Świadczy o tym zainteresowanie internautów wpisem zawierającym formuły postępowania przy stosowaniu tej klasy modeli z sezonowością multiplikatywną (zob. rysunek 9.1). Bliższe informacje dotyczące formułowania równań tych modeli oraz przykłady zastosowań można znaleźć na blogu *wornalkiewicz.wordpress.com* (*Zastosowanie komputera w procesach decyzyjnych*) - obecnie po tytułem „*Quantitative methods (statistics, econometrics, optimization, computer systems)*" po kliknięciu jako *Podgląd* wpisu *Prognozowanie z zastosowaniem modeli Holta-Wintersa*[139].

[137] W niniejszym opracowaniu częściowo zabazowano na badaniach przeprowadzonych pod moim kierunkiem, a zamieszczonych w pracy licencjackiej: Migacz J., *Prognozowanie zewnętrznych usług transportowych* (*na podstawie Fabryki Armatur „Głuchołazy"*), Wyższa Szkoła Zarządzania i Administracji, Opole 2017.

[138] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, Opole 2015, rozdział.2.12. 2. *Przykłady doboru najlepszych parametrów modelu (zastosowanie modułu FC programu WinQSB*).

[139] https://wornalkiewicz.wordpress.com/2012/12/23/prognozowanie-z-zastosowaniem-modeli-holta-wintersa/./



Źródło: Opracowanie własne w *Wordpress*.

Rys. 9.1. Wpis *Prognozowanie z zastosowaniem modeli Holta-Wintersa*

Wyszczególniony na rysunku 9.1 wpis jest na moim blogu *Zastosowanie komputera w procesach decyzyjnych*[140]. Gdy klikniemy na ikonę *Statystyki*, to pojawia się nam zestawienie liczbowe wizyt gości-internautów w tekście tego wpisu, począwszy od miesiąca jego wprowadzenia aż do miesiąca bieżącego roku. Przedstawia to rysunek 9.2, przy czym możemy zaobserwować tendencję rosnącą w liczbach wizytujących.

miesiące i lata

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012												2	2
2013	18	12	5	20	48	46	15	12	22	40	57	66	361
2014	165	42	35	65	77	63	11	26	22	29	147	104	786
2015	150	50	80	33	76	71	22	19	35	39	85	93	753
2016	183	77	44	81	110	118	15	13	33	25	46	61	806
2017	107	60	58	81	84	50	16	34	29	52	76	84	731
2018	186	91	35	57	189	146	13	15	26	31	44	66	899
2019	97	35	35	31	132	150	4						484

Źródło: Opracowanie własne w *Wordpress*.

Rys. 9.2. Statystyki wpisu *Prognozowanie z zastosowaniem modeli Holta-Wintersa*

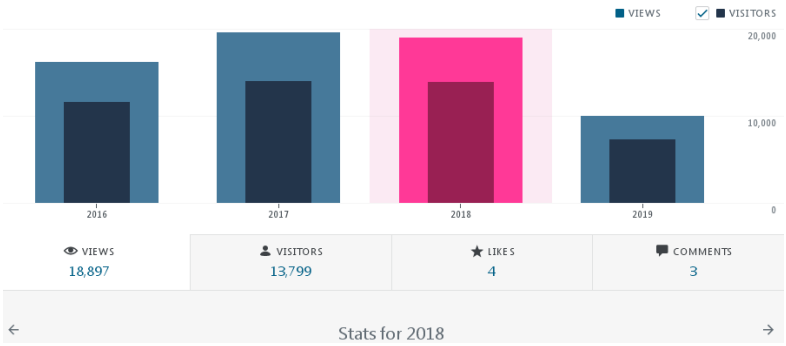
Zwróćmy uwagę, że najwięcej zainteresowanych jest w miesiącach przed sesją zimową lub letnią, a szczyt wystąpił w maju 2018 roku.

Model bez sezonowości ma zastosowanie, gdy występuje trend i wahania przypadkowe. W tym modelu oprócz parametru α występuje drugi parametr β odpowiednio do wygładzania wartości poziomu i trendu. Kryterium rozstrzygającym o trafności doboru tych parametrów w modelu bez sezonowości jest miara błędu prognozy o skrócie RMSE. W modelu Holta-Wintersa z sezonowością multiplikatywną występuje dodatkowo zmienna sezonowość charakteryzowana w równaniu parametrem γ. Zamiast dość uciążliwego obliczania manualnego prognoz możemy zastosować, jak już wspomniałem, moduł *Forecasting and Linear Regression* wchodzący w skład pakietu *WinQSB* dla doboru najlepszej wartości parametrów α, β, γ przy minimum RMSE, co będzie jeszcze bliżej rozpatrywane w niniejszym materiale. Jak już wspomniano model Holta-Wintersa z sezonowością

140 https://wornalkiewicz.wordpress.com/wp-admin/edit.php?s=Prognozowanie+z+zastosowaniem+modeli+Holta-Wintersa&post_status=all&post_type=post&action=-1&m=0&cat=0&paged=1&action2=-1.

addytywną stanowi analogię do modelu multiplikatywnego, przy czym występuje w nim odejmowanie (zamiast dzielenia) indeksu sezonowości.

Zwróćmy jeszcze uwagę na sumaryczną statystykę w latach 2016-2019 z graficznym pokazaniem liczby wejść do blogu (*views*) i liczby gości wpisów (*visitors*) dla roku 2018 na tle innych lat[141].

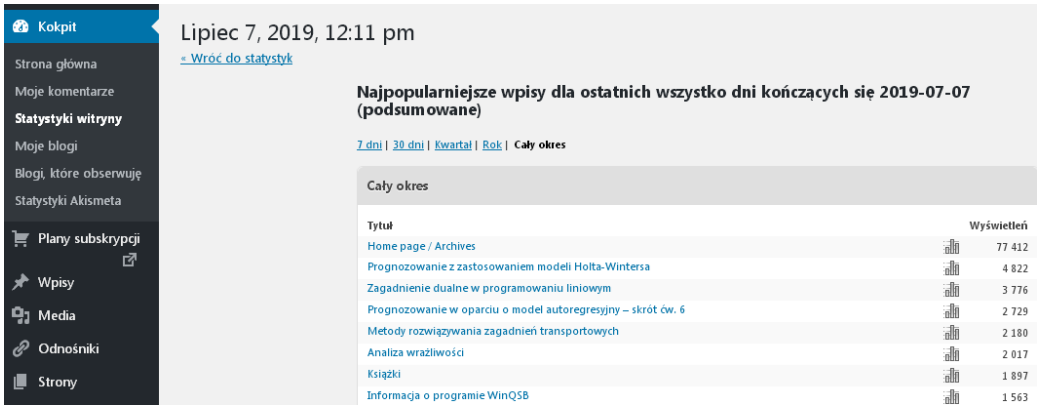


Źródło: Opracowanie własne w *Wordpress*.

Rys. 9.3. Statystyki zainteresowania internautów stroną główną blogu i wpisami

Oprócz wymienionego wcześniej zagadnienia, przez cały czas trwania blogu, do najpopularniejszych należały wpisy wyszczególnione na rysunku 9.4, a mianowicie:

- zagadnienie dualne w programowaniu liniowym,
- prognozowanie w oparciu model autoregresyjny,
- metody rozwiazywania zagadnień transportowych,
- analiza wrażliwości,
- książki (autora niniejszej publikacji),
- informacja o programie *WinQSB*.



Źródło: Opracowanie własne w Wordpress.

Rys. 9.4. Początkowy fragment zestawienia wpisów według popularności

Ciekawe jest też obserwowanie na bieżąco krajów pochodzenia internatów - poza Polską, dominują goście ze Stanów zjednoczonych. W roku 2018 (zob. rysunek 9.5) zainteresowanie blogiem pod względem liczby wejść internautów (początek zestawienia) było następujące: Polska - 16442, Stany Zjednoczone - 1866, SRA Honkong (Chiny) - 109, Wielka Brytania - 71, Niemcy - 66, Unia Europejska - 58.

141 https://wornalkiewicz.wordpress.com/wp-admin/index.php?page=stats&blog=31774799&view=postviews&summarize&numdays=-1.

Country	Views
Polska	16,442
Stany Zjednoczone	1,866
SRA Hongkong (Chiny)	109
Wielka Brytania	71
Niemcy	66
Unia Europejska	58

Źródło: Opracowanie własne w *Wordpress*.

Rys. 9.5. Zainteresowanie blogiem w roku 2018 (początkowy fragment zestawienia)

Szerszą panoramę metod ilościowych, w tym dotyczących prognozowania, w oparciu o szeregi czasowe można znaleźć w publikacji *Popularyzacja wybranych metod ilościowych w Internecie*[142], a dostęp do tej książki w wersji elektronicznej jest możliwy ze strony WSZiA w Opolu[143] (zob. rys. 9.6).

Źródło: Opracowanie własne.

Rys. 9.6. Strona tytułowa książki *Popularyzacja wybranych metod ilościowych w Internecie*

[142] Wornalkiewicz W., *Popularyzacja wybranych metod ilościowych w Internecie, Wyższa Szkoła Zarzadzania i Administracji*, Opole 2017.

[143] https://www.wszia.opole.pl/wp-content/uploads/2018/03/Popularyzacja_wybranych_metod_ilosciowych_w_internecie_Wornalkiewicz.pdf.

9.2. Logistyka w zakresie potrzeb przedsiębiorstw[144]

Logistyka obejmuje sferę zaopatrzenia, produkcji i dystrybucji w przedsiębiorstwie. Procesy logistyczne to fizyczny przepływ towarów, informacji i środków finansowych. Przedsiębiorstwo stanowi składnik łańcucha logistycznego[145]. Jednym z celów logistyki jest oferowanie klientom odpowiedniego poziomu i jakości obsługi logistycznej przy minimalnym poziomie kosztów, a wzroście efektywności w przedsiębiorstwie[146]. W zagwarantowaniu przepływów towarowych i realizacji przepływów informacyjnych bezpośrednio uczestniczą zewnętrzne przedsiębiorstwa transportowe[147]. Zadaniem logistyki zaopatrzenia jest koordynacja przepływu dóbr i uzyskanie pewności i rytmiczności dostaw poprzez zorganizowanie trwałej współpracy dostawców z przedsiębiorstwem. Transport zaopatrzeniowy związany jest z bieżącymi kosztami transportu oraz sprzedaży na określonych warunkach, które definiują właściciela towarów w czasie transportu, wyznaczają zasięg odpowiedzialności dostawcy i odbiorcy oraz wskazują wymagania dotyczące regulowania należności[148]. Efektywność korzystania z usług firm transportowych zależy od sprawnego zorganizowania procesu transportowego zapewniającego właściwą jakość usługi i niezawodność jej wykonania oraz zastosowanie przy jej wykonaniu najniższych nakładów kosztowych[149].

Dystrybucja to jedno z najważniejszych ogniw w łańcuchu logistycznym, związana z dostarczaniem klientom oczekiwanych przez nich wyrobów, od miejsc ich wytworzenia do finalnych odbiorców. Wyroby gotowe powinny dotrzeć do właściwych nabywców, w określonym czasie, miejscu, właściwej ilości i jakości. Istotne jest również odpowiednie dobranie środka transportu, właściwego opakowania i załadunku[150]. W kanałach dystrybucji występują takie ogniwa przepływu jak: sprzedaż bezpośrednia, magazyn centralny wyrobów gotowych producenta, magazyny regionalne, hurtownie, agenci i brokerzy[151].

Przy organizowaniu transportu należy brać pod uwagę najefektywniejsze rozwiązania transportowe np. w przypadku małych przesyłek zaleca się korzystanie z usług firm kurierskich, natomiast do przewozu dużych ładunków wskazane jest wykorzystanie wyspecjalizowanych firm transportowych lub spedycyjnych. Przedsiębiorstwa o charakterze spedycyjnym są w posiadaniu dużej bazy transportowej i magazynowej. Dysponują własnymi sortowniami, przeładowniami oraz składami niezbędnymi do obsługi logistycznej towarów. Organizacja procesu dystrybucji wyrobów gotowych wymaga odpowiedniego opakowania towaru. Stosowane do przewozów opakowania powinny posiadać znormalizowane wymiary i kształty.

Metoda prognozowania to sposób postępowania zastosowany z pełną świadomością do wyjaśnienia prognostycznego zadania. To sposób przetworzenia informacji z przeszłości wraz ze sposobem przejścia z informacji przetworzonych do prognozy[152]. W metodach prognozowania

---

[144] Ta część niniejszego opracowania bazuje na rozdziale *Usługi transportowe w zakresie zaopatrzenia materiałowego* w pracy licencjackiej: Migacz J., *Prognozowanie zewnętrznych usług transportowych (na podstawie Fabryki Armatur „Głuchołazy" w Opolu*, op. cit.

[145] Blaik P., *Logistyka*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2001, s. 52.

[146] Coyle J.J., Bardi E.J., Langley C.J. Jr , *Zarządzanie Logistyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2002, s. 85.

[147] Kisperska-Moroń D., Krzyżaniak S., *Logistyka*, Biblioteka Logistyka, Poznań 2009, s. 140.

[148] http://www.czasopismologistyka.pl/artykuly-naukowe/send/252-artykuly-na-plycie-cd-2/3145 -artykul.

[149] Mendyk E., *Ekonomika i organizacja transportu*, Wydawnictwo Wyższa Szkoła Logistyki, Poznań 2002, s. 122.

[150] http://www.czasopismologistyka.pl/artykuly-naukowe/send/-/4106, pobrano: 7.06.2016.

[151] Ibidem, s. 167.

[152] Czerwiński Z., Guzik B., *Prognozowanie ekonometryczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa 1980, s. 24.

można wyróżnić podział na metody ilościowe i jakościowe[153]. Jak już nadmieniono, ogólnie pojęte metody ilościowe dzielą się na metody: szeregów czasowych, ekonometryczne, analogowe, inne. Natomiast w ramach metod jakościowych wyróżniamy metody: ankietowe, ekspertyz, delficką, inne. W ramach metody prognozowania według szeregów czasowych występuje kilka modeli, a mianowicie bazujące na średniej, modele wygładzania wykładniczego, modele Holta-Wintersa.

Komputerową realizację określenia prognoz według różnych modeli umożliwia pakiet programowy *WinQSB,* a zaliczamy do nich[154] modele prostych średnich ruchomych (SMA) i modele ważonych średnich ruchomych (WMA). Ponadto są modele bazujące na wygładzaniu wykładniczym np. *model prostego wyrównywania wykładniczego* (*SES*), *model Browna* (*podwójnego wyrównywania wykładniczego*).

W przykładowym przedsiębiorstwie tj. Fabryce Armatur „Głuchołazy" proces transportowy realizowany jest poprzez odpłatne świadczenie usług transportowych. W niniejszym omawianiu prognozowania potrzeb tych usług skupiono się tylko na przewozie produkcyjnych materiałów zaopatrzeniowych, a w szczególności na dystrybucji wyrobów gotowych. Zaopatrzenie materiałowe oraz dystrybucja wyrobów odbywa się transportem zamawianym przez sprzedającego, zamawianym przez kupującego oraz transportem własnym kupującego. Jak już wspomniałem usługi dotyczące przywozu materiałów oraz dystrybucji wyrobów świadczą zewnętrzne firmy przewozowe, spedycyjne oraz kurierskie. Fabryka współpracuje z 30. stałymi dostawcami, przy czym dominująca liczba jej dostawców pochodzi z południowej części Polski (zob. rysunek 9.7).

Źródło: Migacz J., *Prognozowanie zewnętrznych usług transportowych (na podstawie Fabryki Armatur „Głuchołazy" w Opolu*, op. cit. rys. 8.

Rys. 9.7. Lokalizacja podstawowych dostawców Fabryki Armatur „Głuchołazy"

Koszty transportu w zakresie zaopatrzenia materiałowego stanowią w omawianym przedsiębiorstwie około 4% kosztów poniesionych na zakup materiałów. Odlewy i odkuwki

[153] Dittman P., *Metody prognozowania sprzedaży w przedsiębiorstwie,* Akademia Ekonomiczna we Wrocławiu, Wrocław 2000.

[154] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania,* część II, *Zagadnienia ekonometrii,* Wydawnictwo Instytut Śląski, Opole 2014.

transportowane są w metalowych pojemnikach o wymiarach 1200x800 oraz w skrzyniach drewnianych. Transport kolejnej grupy materiałów, którymi są wyroby hutnicze czyli pręty, blachy i rury odbywa się luzem. „Głuchołazy" sprzedają swoje wyroby zarówno na rynek krajowy jak i na rynki zagraniczne - Europy, Azji i Ameryki Północnej. Sprzedaż bezpośrednia i sprzedaż poprzez dystrybutorów wynosi po 50%, przy czym dąży się do zwiększenia udziału w sprzedaży odbiorców bezpośrednich. Organizacja dystrybucji wyrobów realizowana jest w 85% poprzez zewnętrzne firmy transportowe. Dystrybucja wyrobów gotowych na terenie kraju i Europy odbywa się drogą lądową za pośrednictwem transportu samochodowego. Natomiast na rynek azjatycki i Ameryki Północnej wykorzystuje się transport multimodalny, prowadzony przez firmy spedycyjne.

Różnorodność komponentów pod względem ilościowym i wartościowym potrzebnych w procesie produkcyjnym, była powodem wcześniejszego przed prognozowaniem przeprowadzenia klasyfikacji ABC stosowanych materiałów. Umożliwiło to wyłonienie grup materiałowych mających znaczenie strategiczne dla rozpatrywanej Fabryki. Rozszerzeniem klasyfikacji ABC jest analiza Pareto a z niej wynika, że wyznaczona w klasyfikacji grupa A, klasycznie obejmuje około 80% całkowitej wartości, przy czym stanowi ona 20% rozpatrywanych pozycji pod względem ilościowym. Jest to tzw. reguła 80/20), przy czym grupa B (15/30), a grupa C (5/50). W prezentowanym przedsiębiorstwie jednym z podstawowych składników kosztowych w produkcji armatury przemysłowej są materiały. W procesie produkcji zaworu konieczne jest zastosowanie wielu różnych rodzajów materiałów, części i podzespołów. Większość z nich to materiały, które po zakupie podlegają dalszej obróbce, a należą do nich odlewy, odkuwki, materiały hutnicze, a także elementy złączne i uszczelnienia. Tabela 9.1 prezentuje wynik klasyfikacji ABC na podstawie dostępnych danych z roku 2015. Zgodnie z analizą A, B, C w grupie A znalazły się materiały, których skumulowany udział nie przekracza 80% ogólnej wartości zakupu materiałów. Grupa B to materiały z przedziału powyżej 80% do 95%, natomiast pozostałe materiały to grupa C.

Tab. 9.1. Klasyfikacja ABC materiałów zaopatrzeniowych ze względu na wartość

Rodzaj	Wartość zakupu	Wartość skumulowana	Udział [%]	Udział skumulowany [%]	Grupa
Odlewy	1992028,85	1992028,85	50,26	50,26	A
Odkuwki	711545,89	2703574,74	17,95	68,21	A
Materiały hutnicze	514183,27	3217758,01	12,97	81,18	B
Elementy złączne	237877,66	3584026,66	6,00	87,18	B
Materiały różne	155542,40	3868605,20	3,92	91,10	B
Materiały kolorowe	76548,87	3660575,02	1,93	93,03	B
Druty spawalnicze	68532,19	3286290,20	1,73	94,76	B
Uszczelki	59858,29	3346148,49	1,51	96,27	C
Farby	59125,17	3,927730,37	1,49	97,77	C
„Mieszki"	52487,78	3713062,78	1,32	99,09	C
Kooperacja	36078,50	3963808,87	0,91	100,00	C

Źródło: Opracowanie własne wg - Migacz J., *Prognozowanie zewnętrznych usług transportowych* (na podstawie Fabryki Armatur „Głuchołazy", s. 45.

9.3. Wstępne podejście do prognozowania krótkookresowego

Przy użyciu funkcji regresji liniowej *REGLINP Excela* określono prognozę przewozów w tonach na okres (n + 1), przy czym szeregi czasowe 24. elementowe dotyczą miesięcy roku 2015 i 2016 dystrybucji wyrobów (zob. tabela 9.2). Tabela ta może być poszerzona o prognozy wygasłe $\hat{y}$

obliczone po oszacowaniu parametrów trendu liniowego. Okazało się, że prosty trend liniowy nie jest adekwatnym modelem do ustalania prognoz na najbliższe miesiące, które są niezbędne w zakresie:

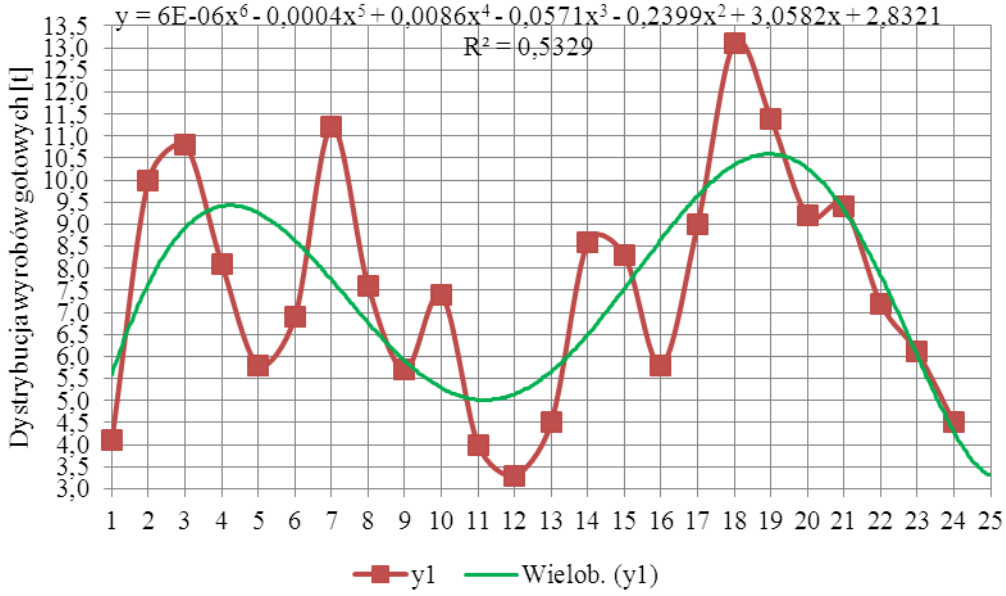
- dystrybucji wyrobów gotowych [*t*],
- dystrybucji wyrobów gotowych pod względem kosztów transportu [zł],
- zaopatrzenia materiałowego [*t*],
- zaopatrzenia materiałowego pod względem kosztów transportu [zł],
- podstawowych kluczowych materiałów dla przedsiębiorstwa [*t*].

Tab. 9.2. Szereg czasowy dystrybucji wyrobów w tonach (y_1)

t	1	2	3	4	5	6	7	8	9	10	11	12
y_1	4,1	10,0	10,8	8,1	5,8	6,9	11,2	7,6	5,7	7,4	4,0	3,3
t	13	14	15	16	17	18	19	20	21	22	23	24
y_1	4,5	8,6	8,3	5,8	9,0	13,1	11,4	9,2	9,4	7,2	6,1	4,5

Źródło: Opracowanie własne wg - Migacz J., *Prognozowanie zewnętrznych usług transportowych* (na podstawie Fabryki Armatur „Głuchołazy", s. 58.

Zauważono także, że dalsze trendy (potęgowy, wykładniczy, logarytmiczny), jak również trendy wielomianowe (stopni 1-6) możliwe do uzyskania w *Excelu*, również nie dają dobrego dopasowania do danych empirycznych. Współczynniki determinacji R^2 są bowiem niskie i wynoszą od 0,0029 do 0,5329. Tak więc te modele dla naszych danych nie mogą być użyte do wyznaczania prognoz. Na rysunku 9.8. przedstawiono układanie się danych empirycznych i modelu wielomianowego zmiennej y_1.



Źródło: Opracowanie własne na podstawie tabeli 10.2.

Rys. 9.8. Porównanie danych empirycznych i modelu wielomianu 6. stopnia dla y_1

Występują duże wahania danych rzeczywistych i poszukiwano innego modelu, po czym zdecydowano się na model trendu pełzającego. Trend pełzający bowiem to model adaptacyjny, służący do budowy prognoz krótkoterminowych. W tym względzie skorzystano z procedury przeprowadzonej w ramach pracy licencjackiej obejmującej kroki[155]:

1. Ustalenie długości segmentu - stałej wygładzania, przy czym: $k \leqslant n$, w naszym przypadku $k = 3$, $n = 24$.

[155] Migacz J., *Prognozowanie zewnętrznych usług transportowych* (*na podstawie Fabryki Armatur „Głuchołazy"*), op. cit., strony: 61-64.

2. Określenie podszeregów *i*-tych czasowych trzyelementowych o liczbie $(n - k + 1) = 24 - 3 + 1 = 22$ (tab. 9.3).

Tab. 9.3. Podszeregi *i*-te czasowe zmiennej y_1

t	y_1				*i*	*t*	y_1				*i*
1	4,1	4,1	10,0	10,8	1	13	4,5	4,5	8,6	8,3	13
2	10,0	10,0	10,8	8,1	2	14	8,6	8,6	8,3	5,8	14
3	10,8	10,8	8,1	5,8	3	15	8,3	8,3	5,8	9,0	15
4	8,1	8,1	5,8	6,9	4	16	5,8	5,8	9,0	13,1	16
5	5,8	5,8	6,9	11,2	5	17	9,0	9,0	13,1	11,4	17
6	6,9	6,9	11,2	7,6	6	18	13,1	13,1	11,4	9,2	18
7	11,2	11,2	7,6	5,7	7	19	11,4	11,4	9,2	9,4	19
8	7,6	7,6	5,7	7,4	8	20	9,2	9,2	9,4	7,2	20
9	5,7	5,7	7,4	4,0	9	21	9,4	9,4	7,2	6,1	21
10	7,4	7,4	4,0	3,3	10	22	7,2	7,2	6,1	4,5	22
11	4,0	4,0	3,3	4,5	11	23	6,1				
12	3,3	3,3	4,5	8,6	12	24	4,5				

Źródło: Opracowanie własne wg - Migacz J., *Prognozowanie zewnętrznych usług transportowych* (*na podstawie Fabryki Armatur „Głuchołazy”*), s. 61.

3. Wyznaczenie parametrów a_0 i a_1 dla 22. modeli segmentowych, z użyciem funkcji *REGLINP Excela* (zob. tabela 9.4).

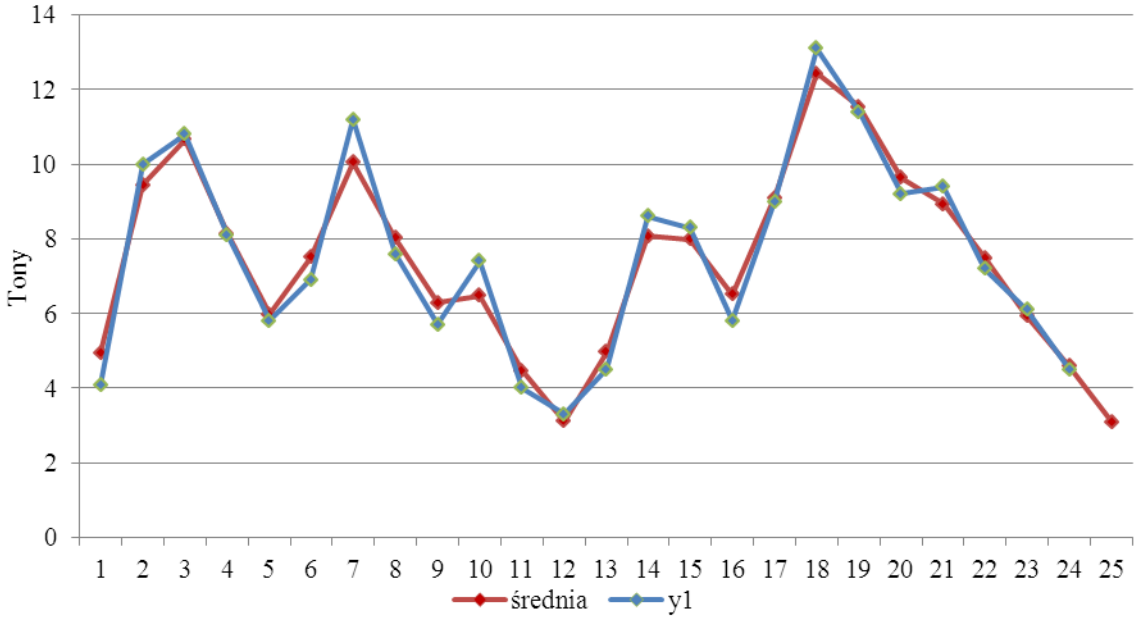
4. Obliczenie wartości modelowych dla trzech okresów każdego z 22. segmentów według wzoru: $\hat{y} = a_0 + a_1 \cdot t$.

5. Obliczenie średnich dla każdego z 24. okresów (zob. tabela 9.4).

6. Obliczenie prognozy na pierwszy miesiąc nowego roku tj. 25, przyjmując ostatnie trzy średnie jako elementy funkcji REGLINP Excela z okresów: 22 → 7,48; 23 → 5,93; 24 → 4,58 i uzyskano:

$$\hat{y}_{25} = 39{,}35 - 1{,}45 * 25 = 3{,}1\ t.$$

7. Porównanie danych empirycznych z teoretycznymi (średnimi) - zob. rysunek 9.9.



Źródło: Migacz J., *Prognozowanie zewnętrznych usług transportowych* (na podstawie Fabryki Armatur „Głuchołazy”), s. 63.

Rys. 9.9. Porównanie danych empirycznych i średnich z segmentów oraz prognoza na okres 25.

8. Przesunięcie (przepełznięcie) szeregu o kolejny miesiąc i obliczenie prognozy. W ten sposób model stale nadąża za zmieniającą się rzeczywistością w zbieranych miesięcznych danych statystycznych w zakresie dystrybucji wyrobów gotowych, a także zaopatrzenia materiałowego.

Tab. 9.4. Współczynniki trendów liniowych segmentów, wartości modelowe okresów oraz średnie

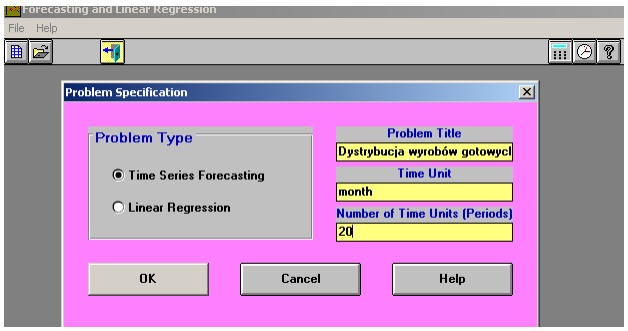
i	a_0	a_1	1	2	3	4	5	6	7	8	9	10	11	12		
			11	12	13	14	15	16	17	18	19	20	21	22	23	24
1	1,60	3,35	4,95	8,30	11,65											
2	12,48	-0,95		10,58	9,63	8,68										
3	18,23	-2,50			10,73	8,23	5,73									
4	9,93	-0,60				7,53	6,93	6,33								
5	-8,23	2,70					5,27	7,97	10,67							
6	6,12	0,35						8,22	8,57	8,92						
7	30,17	-2,75							10,92	8,17	5,42					
8	7,80	-0,10								7,00	6,90	6,80				
9	14,20	-0,85									6,55	5,70	4,85			
10	27,45	-2,05										6,95	4,90	2,85		
11	0,93	0,25	3,68	3,68	4,18											
12	-28,98	2,65		2,82	5,47	8,12										
13	-19,47	1,90			5,23	7,13	9,03									
14	28,57	-1,40				8,97	7,57	6,17								
15	2,10	0,35					7,35	7,70	8,05							
16	-52,75	3,65						5,65	9,30	12,95						
17	-10,43	1,20							9,97	11,17	12,37					
18	48,28	-1,95								13,18	11,23	9,28				
19	30,00	-1,00									11,00	10,00	9,00			
20	29,60	-1,00										9,60	8,60	7,60		
21	43,87	-1,65											9,22	7,57	5,92	
22	36,98	-1,35												7,28	5,93	4,58
Średnia			4,95	9,44	10,67	8,15	5,98	7,51	10,05	8,03	6,29	6,48				
			4,48	3,12	4,96	8,07	7,98	6,51	9,11	12,43	11,53	9,63	8,94	7,48	5,93	4,58

Źródło: Opracowanie własne wg - Migacz J., *Prognozowanie zewnętrznych usług transportowych* (*na podstawie Fabryki Armatur „Głuchołazy”*), s. 62.

9.4. Optymalizacja doboru parametrów w modelu multiplikatywnym

Zamiast uciążliwego zastosowania modelu pełzającego, proponuje teraz wykorzystanie komputerowego modułu *Forecasting and Linear Regression* (FC), będącego składnikiem pakietu *WinQSB*. Zastosujemy go do modelu multiplikatywnego i to od razu w zakresie optymalnego doboru trzech parametrów wyrównywania tj. α, β, γ. Jako dane posłużą nam obserwacje w tonach oraz wartościowo w zakresie zewnętrznych usług transportowych dotyczących dystrybucji wyrobów gotowych, produkowanych przez wcześniej wymienioną Fabrykę. Rozpoczynamy od wprowadzenia informacji do okna dialogowego *Problem Specification* (zob. rysunek 9.10), a mianowicie:

Problem Type → *Time Series Forecasting* - prognozowanie w oparciu o szeregi czasowe,
Problem Title - Dystrybucja wyrobów gotowych,
Time Unit → *month* (*miesiąc*),
Number of Time Units (*Periods*) - 20.



Źródło: Opracowanie własne w module *FC*.

Rys. 9.10. Okno wstępne definiujące problem decyzyjny

Rozpoczynamy teraz od wprowadzania danych 20. obserwacji zmiennej y_1, gdyż dla nich po sporządzeniu wykresu zauważamy układanie się wartości według modelu zbliżonego do multiplikatywnego Holta-Wintersa. Musimy jednak pamiętać o użyciu kropki dziesiętnej zamiast przecinka.

Month	Historical Data
1	4.1
2	10.0
3	10.8
4	8.1
5	5.8
6	6.9
7	11.2
8	7.6
9	5.7
10	7.4
11	4.0
12	3.3
13	4.5
14	8.6
15	8.3
16	5.8
17	9.0
18	13.1
19	11.4
20	9.2

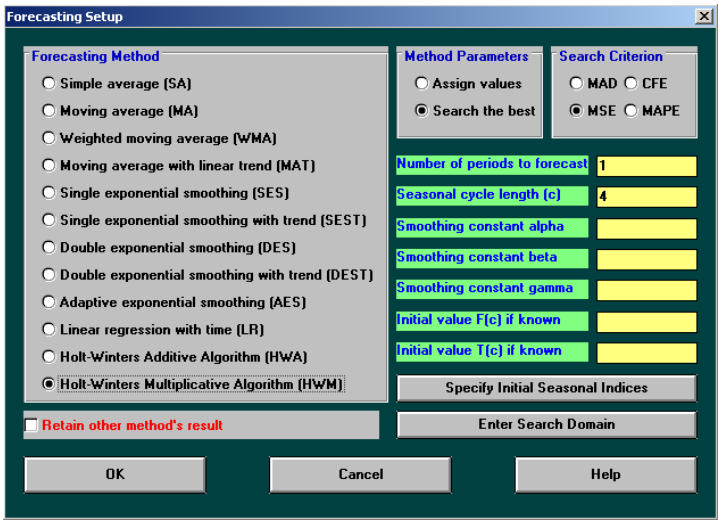
Źródło: Opracowanie własne w module *FC* na podstawie tabeli 11[156].

Rys. 9.11. Wprowadzone dane szeregu czasowego zmiennej y_1

Po wybraniu zakładki *Solve and Analyze* w menu głównym *WinQSB* pojawia się nam okno dialogowe *Forecasting Setup* w którym musimy wskazać model prognozowania zwany tu *Forecasting Method*. Wybieramy *Holt-Winter Multiplicative Algorithm (HWM)* oraz podajemy kryteria obliczeń (zob. rysunek 9.12) , a mianowicie:

- *Method Parameters* → *Search the best* (wybór najlepszego zestawu parametrów wyrównywania wykładniczego α, β, γ);
- *Search Criterion* → MSE (miara błędu prognozy stanowiąca kwadrat innej miary tj. RMSE);
- *Number of periods to forecast* - 1 (liczba okresów prognozowania, czyli okres 25.),
- *Seasonal cycle length* (*c*) - 4 (długość cyklu sezonowości),
- *Enter Search Domain) - ustawienie kroku obliczeń iteracyjnych dla parametrów*, ustawiamy stopniowanie obliczeń co 0,01dla każdego z trzech parametrów, z tym że zamiast przecinka dajemy kropkę dziesiętną.

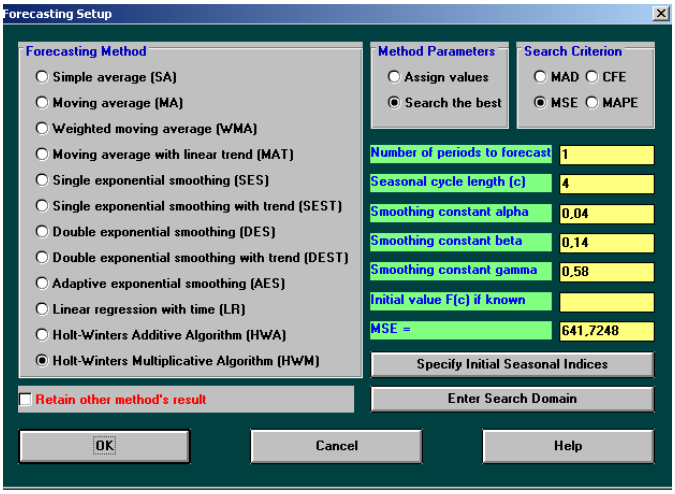
[156] Migacz J., *Prognozowanie zewnętrznych usług transportowych (na podstawie Fabryki Armatur „Głuchołazy" w Opolu*, op. cit, tab. 11.



Źródło: Opracowanie własne w module *FC*.

Rys. 9.12. Okno definiujące problem optymalnego doboru parametrów modelu zmiennej y_1

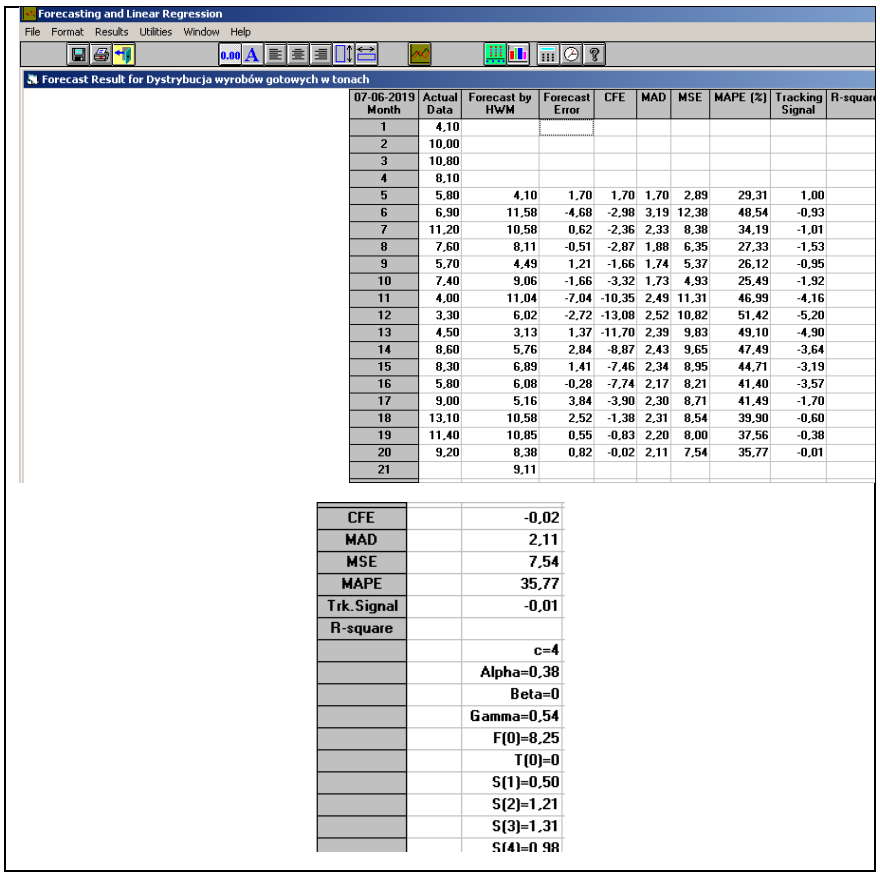
Po naciśnięciu OK mamy możliwość obserwowania pracy modułu, gdyż w kolejnych krokach iteracyjnych następuje zmiana parametrów i co jakiś czas korygowana jest też miara MSE, która jest coraz mniejsza (zob. rysunek 9.13).



Źródło: Opracowanie własne w module *FC*.

Rys. 9.13. Jeden ze stanów procesu określenia prognozy zmiennej y_1

Po osiągnięciu najlepszego wyniku z punktu widzenia funkcji celu - minimum MSE generowany jest raport *Forecast and Result for Dystrybucja wyrobów gotowych w tonach* (zob. rysunek 9.14). Występują w nim wygasłe już prognozy *Forecast by* HWM oraz prognoza na okres 21. Są też różne miary błędów prognoz (CFE, MAD, MAPE (%), *Tracing Signal*) liczone po poszczególnych obserwacjach. Porównajmy teraz prognozę na okres 21 z daną rzeczywistą zmiennej y_1 w tym okresie: $F_{21} = 9{,}11$ a $y_{21} = 9{,}40$, a więc wartości są zbliżone. Zobaczmy jeszcze do wyniku osiągniętego w trendzie pełzającym dla okresu 21 (zob. tabela 9.4) który wynosi 8,94.



07-06-2019 Month	Actual Data	Forecast by HWM	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-squar
1	4,10								
2	10,00								
3	10,80								
4	8,10								
5	5,80	4,10	1,70	1,70	1,70	2,89	29,31	1,00	
6	6,90	11,58	-4,68	-2,98	3,19	12,38	48,54	-0,93	
7	11,20	10,58	0,62	-2,36	2,33	8,38	34,19	-1,01	
8	7,60	8,11	-0,51	-2,87	1,88	6,35	27,33	-1,53	
9	5,70	4,49	1,21	-1,66	1,74	5,37	26,12	-0,95	
10	7,40	9,06	-1,66	-3,32	1,73	4,93	25,49	-1,92	
11	4,00	11,04	-7,04	-10,35	2,49	11,31	46,99	-4,16	
12	3,30	6,02	-2,72	-13,08	2,52	10,82	51,42	-5,20	
13	4,50	3,13	1,37	-11,70	2,39	9,83	49,10	-4,90	
14	8,60	5,76	2,84	-8,87	2,43	9,65	47,49	-3,64	
15	8,30	6,89	1,41	-7,46	2,34	8,95	44,71	-3,19	
16	5,80	6,08	-0,28	-7,74	2,17	8,21	41,40	-3,57	
17	9,00	5,16	3,84	-3,90	2,30	8,71	41,49	-1,70	
18	13,10	10,58	2,52	-1,38	2,31	8,54	39,90	-0,60	
19	11,40	10,85	0,55	-0,83	2,20	8,00	37,56	-0,38	
20	9,20	8,38	0,82	-0,02	2,11	7,54	35,77	-0,01	
21		9,11							

CFE		-0,02
MAD		2,11
MSE		7,54
MAPE		35,77
Trk.Signal		-0,01
R-square		
		c=4
		Alpha=0,38
		Beta=0
		Gamma=0,54
		F(0)=8,25
		T(0)=0
		S(1)=0,50
		S(2)=1,21
		S(3)=1,31
		S(4)=0,98

Źródło: Opracowanie własne w module *FC*.

Rys. 9.14. Raport wyników (w dwóch fragmentach) z prognozowania y_1

Jak widzimy na rysunku 9.14 miara błędu RMSE, jako pierwiastek z MSE, wynosi 2,74, natomiast najlepiej dobrane parametry modelu Holta-Wintersa z sezonowością multiplikatywną są następujące: $\alpha = 0{,}38$; $\beta = 0$; $\gamma = 0{,}54$.

Bardziej precyzyjne przyglądnięcie się szeregowi czasowemu zmiennej y_1 sugeruje nam przyjęcie w celach symulacyjnych cyklu sezonowości równemu 7 dla 21 obserwacji. Dalsze działanie jest analogiczne do przedstawionego wcześniej cyklu 4-ch obserwacji. W wyniku dość długiej komputerowej realizacji optymalizacji, moduł *FC* obliczył wygasłe prognozy dla okresów 8-21 oraz prognozę na okres następny. Odbiega ona znacznie od danej rzeczywistej $y_{22} = 7{,}2$. Miara MSE = 12,94, a więc RMSE wynosi 3,6, co wskazuje na lepsze dobranie wcześniejszego cyklu 4-ro okresowego.

Rozpatrzmy teraz prognozowanie zewnętrznych usług transportowych w oparciu o w*artość usług transportowych* (y_2) *dotyczących dystrybucji w zł.* Dane z kolejnych miesięcy roku 2015 i 2016 wyszczególniono w tabeli 9.5.

Tab. 9.5. Szereg czasowy wartości usług transportowych (y_2)

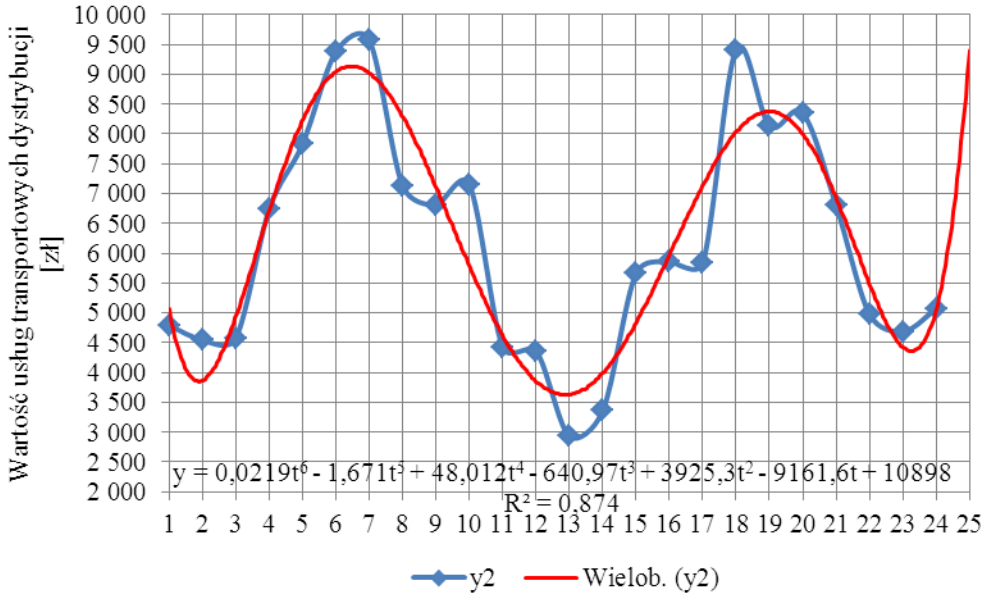
t	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
y_2	4789	4545	4581	6736	7834	9378	9575	7130	6810	7152	4425	4366	2928	3358	5662	5860	5841	9141	8137	8347	6809	4978	4687	5067

Źródło: Opracowanie własne wg - Migacz J., *Prognozowanie zewnętrznych usług transportowych* (*na podstawie Fabryki Armatur „Głuchołazy”*), s. 64.

Podejmując wstępne badanie doboru trendu jednorównaniowego okazało się, że najlepsze dopasowania wykazuje wielomian szóstego stopnia, bowiem współczynnik determinacji $R^2 = 0{,}874$:

$$\hat{y} = 0{,}0219x^6 - 1{,}6711x^5 + 48{,}013x^4 - 640{,}98x^3 + 3925{,}4x^2 - 9161{,}8x + 10898.$$

Do wyznaczenia prognozy na okres 25. skorzystano z możliwości wyświetlania na wykresie Excela równania i ustalenia prognozy na okres (n + 1) tj. 25. (zob. rysunek 9.15), która wynosi 9417 zł, a wartość danych rzeczywistych na styczeń 2017 roku była 8992 zł.



Źródło: Opracowanie własne na podstawie tabeli 10.5.

Rys. 9.15. Porównanie danych empirycznych i modelu wielomianu 6. stopnia dla y_2

Analiza ułożenia się wartości obserwacji zmiennej y_2 wskazuje na możliwość doboru cyklu sezonowości równego 11 miesięcy. Weźmy zatem 22 okresy obserwacji, dla odmiany załóżmy 3 okresy prognozowania i zastosujmy również model HWM. Po procesie iteracyjnym minimalizacji miary błędów MSE uzyskujemy najlepsze dobranie parametrów α, β, γ a w efekcie końcowym raport o początkowym fragmencie zaprezentowanym na rysunku 9.16.

07-12-2019 Month	Actual Data	Forecast by HWM	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-square
1	4789,0								
2	4545,0								
3	4581,0								
4	6736,0								
5	7834,0								
6	9378,0								
7	9575,0								
8	7130,0								
9	6810,0								
10	7152,0								
11	4425,0								
12	4366,0	4789,0	-423,0	-423,0	423,0	178929,0	9,7	-1,0	
13	2928,0	4336,2	-1408,2	-1831,2	915,6	1081044,0	28,9	-2,0	
14	3358,0	3632,5	-274,5	-2105,8	701,9	745813,9	22,0	-3,0	
15	5662,0	5131,4	530,6	-1575,2	659,1	629738,9	18,8	-2,4	0,4
16	5860,0	6288,7	-428,7	-2003,9	613,0	540554,9	16,5	-3,3	0,7
17	5840,0	7261,3	-1421,3	-3425,2	747,7	787145,5	17,8	-4,6	
18	9414,0	6659,2	2754,8	-670,5	1034,4	1758801,0	19,5	-0,6	0,4
19	8138,0	6025,5	2112,5	1442,1	1169,2	2096797,0	20,3	1,2	0,3
20	8347,0	6804,3	1542,7	2984,8	1210,7	2128270,0	20,1	2,5	0,3
21	6800,0	7988,5	-1188,5	1796,3	1208,5	2056692,0	19,8	1,5	0,4
22	4978,0	4560,2	417,8	2214,2	1136,6	1885591,0	18,8	1,9	0,5
23		5170,4							
24		4907,0							
25		4945,9							

Źródło: Opracowanie własne w module *FC*.

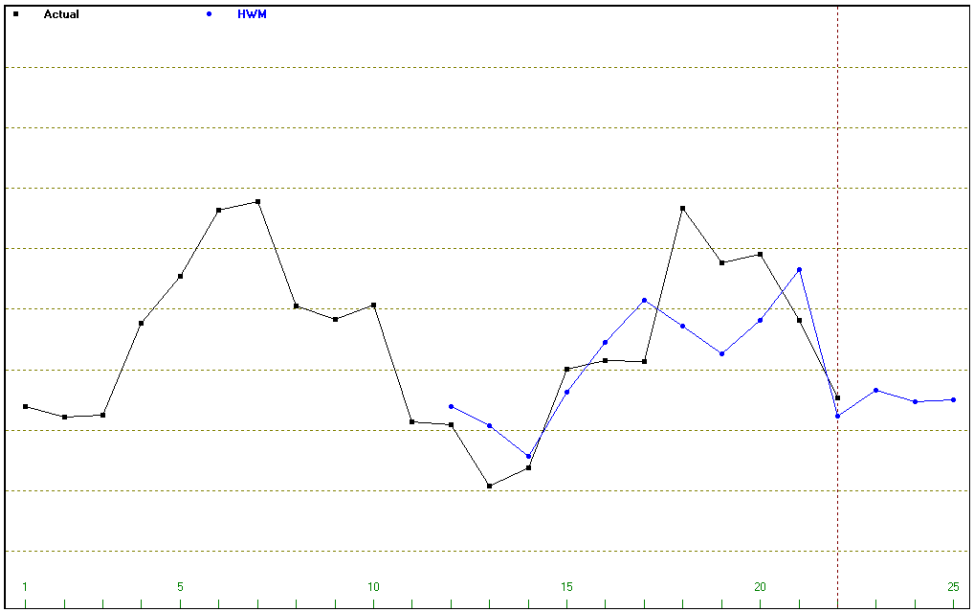
Rys. 9.16. Część początkowa raportu z procesu minimalizacji miary błędu MSE w prognozowaniu zmiennej y_2 przy cyklu sezonowości równym 11

Proponuje teraz porównać prognozy *F* z danymi rzeczywistymi *y* dla okresów 22-24 i zwrócić uwagę na prognozę na okres 25:

Okres	22	23	24	25
y	4978,0	4687,0	5066,0	-
F	4560,2	5170,4	4907,0	4945,9

W części dalszej raportu dotyczącej miar błędu oraz parametrów $\alpha = 0{,}52$, a β i γ równają się zero, natomiast MSE = 1885591, co odpowiada RMSE = 1373. Podano też współczynnik determinacji $R^2 = 0{,}5$, co wskazuje na słabe dopasowanie modelu do danych z obserwacji. Trzeba więc dalej poszukiwać lepszego dobrania cyklu sezonowości, co już dla nabrania wprawy korzystania z modułu *FC* pozostawia się Czytelnikowi.

W menu nad raportem jest zakładka, której kliknięcie generuje wykres kształtowania się danych z obserwacji oraz prognoz, w naszym przykładzie dla okresów 12-25 (zob. rysunek 9.17). Jest też informacja o wartości średniej z 22. obserwacji równej 6302,091 i odchyleniu standardowym równym 1952,934.



Źródło: Opracowanie własne w module *FC*.

Rys. 9.17. Dane rzeczywiste i modelowe zmiennej y_2

* * *

Pakiet *WinQSB*, moim zdaniem, stanowi sprawne narzędzie do adekwatnego doboru modelu do danych rzeczywistych z przeprowadzonych obserwacji, gdyż zawiera wiele metod prognozowania. Pozostaje nam jednak przeprowadzenie symulacji i optymalizacji parametrów równań poszczególnych modeli prognozowania zarówno jedno jak i wielorównaniowych.

Warto kontynuować niniejsze rozważania także w odniesieniu do grup zaopatrzenia materiałowego - w naszym przypadku odlewów, odkuwek i materiałów hutniczych. Jak już podano wcześniej zostały te grupy wyłonione w wyniku przeprowadzonej klasyfikacji ABC.

10. Estymowanie parametrów modelu ekonometrycznego produktu brutto programem *Gretl* dla województwa

10.1. Wstęp

W badaniach demograficznych występuje potrzeba znajdowania zależności formalnych między cechami przyczynowymi, a cechą uważaną jako skutek pewnej działalności w wyniku tych przyczyn. Przykładem zaprezentowanym w niniejszym materiale jest model ekonometryczny zmiennej „*Liczba bezrobotnych*" w województwie opolskim. Opracowano go korzystając z szeregu 12-okresowego - przekrojowego obejmującego zmienne oznaczone dla potrzeb tej pracy symbolami:

*X*3 - liczba bezrobotnych zrejestrowanych ogółem [tys.],
*X*9 - przeciętne zatrudnienie w przemyśle [tys.],
*X*10 - przeciętne zatrudnienie w budownictwie [tys.],
*X*49 - pobór wody na potrzeby gospodarki narodowej i ludności na 1 km^2 [dam^3],
*X*50 - zużycie energii elektrycznej w gospodarstwach domowych w miastach [GWh],
*Y*1 - asolwenci szkół zawodowych [tys.],
*Y*2 - absolwenci liceów ogólnokształcących dla młodzieży [tys.],
*Y*3 - absolwenci technikum dla młodzieży [tys.],
*Y*4 - absolwenci szkół policealnych [tys.].

Zatem cechą badaną, czyli objaśnianą jest *X3*, a cechami przyczynowymi - objaśniającymi *X*9, *X*10, *X*49, *X*50, *Y*1, *Y*2, *Y*3 i *Y*4. Dane pobrane zostały z roczników statystycznych województw wydanych przez Główny Urząd Statystyczny[157]. Procedura postępowania w zakresie tworzenia i weryfikacji zaproponowanego modelu ekonometrycznego bazuje na wcześniejszej publikacji autora dotyczącej modelowania cechy „*Liczba bezrobotnych*" w odniesieniu do całego naszego kraju[158]. Wtedy wyjściem do opracowania modelu były dane statystyczne obejmujące okres szesnastu lat (1992-2007), z których wybrano intuicyjnie wymienione zmienne pochodzące z określonych rynków roczników statystycznych, nadając im jednak krótkie symbole dla celów przetwarzania komputerowego:

- rynek pracy/aktywność ekonomiczna ludności w wieku 15 i więcej - przeciętne w roku w tys./bezrobotni (*X*3);
- przeciętne zatrudnienie w tys.: przemysł (*X*9), budownictwo (*X*10), handel i naprawy (*X*11), transport, gospodarka magazynowa i łączność (*X*12), obsługa nieruchomości i firm (*X*13), działalność usługowa komunalna, społeczna i indywidualna, pozostałe (*X*14);
- koszty pracy w zł/ogółem na jednego zatrudnionego przeciętnie miesięcznie (*X*25);
- rynek płacy/wynagrodzenia w mln/wskaźnik - przeciętne miesięczne wynagrodzenie nominalne brutto - rok poprzedni = 100 (*X*38);
- budżety gospodarstw domowych/ przeciętne miesięczne wydatki na 1 gospodarstwo domowe w zł (*X*48);
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych): zużycie wody z wodociągów w ciągu roku w hm^3 (*X*49), zużycie energii elektrycznej w ciągu roku w GWh (*X*50), zużycie gazu w sieci w ciągu roku w hm^3 (*X*51);

---

[157] Przykładowe odniesienia do źródeł danych statystycznych:
- https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-statystyczny-wojewodztw-2015,4,10.html?pdf=1 (lata 2010-2015),
- https://stat.gov.pl/obszary-tematyczne/roczniki-statystyczne/roczniki-statystyczne/rocznik-statystyczny-rzeczypospolitej-polskiej-2006,2,1.html,
- http://opole.stat.gov.pl/publikacje-i-foldery/roczniki-statystyczne/rocznik-statystyczny-wojewodztwa-opolskiego-2016,4,12.html?pdf=1.

[158] Wornalkiewicz W., *Model ekonometryczny zmiennej „Liczba bezrobotnych*", opublikowane w ramach monografii *Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne,* red. naukowa T. Pokusa, Wydawnictwo Instytut Śląski, Opole 2011.

- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci - policealne w tys. (*X*59);

- rynek mieszkaniowy/wskaźnik zawartych małżeństw na 10000 ludności (*X*78).

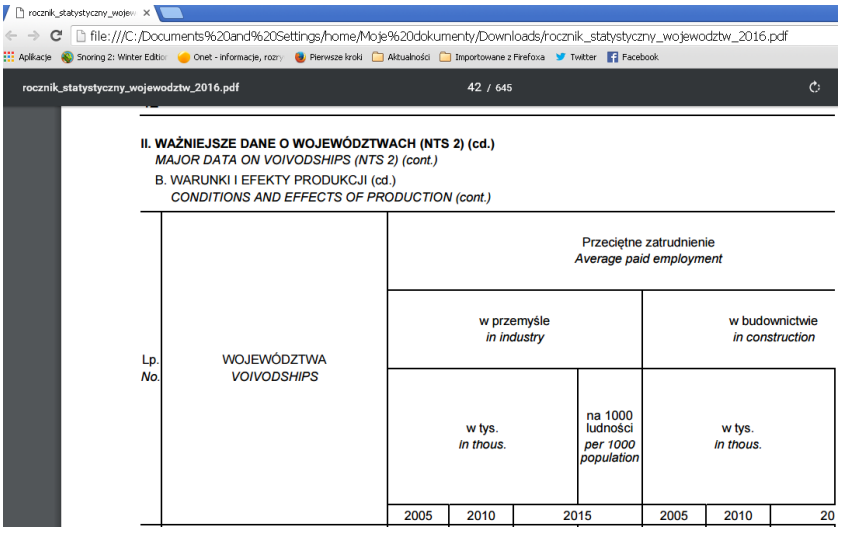
Jak już nadmieniono zmienne dobrano intuicyjnie, a ich ilość wynikała z warunku *Gretla*: ilość obserwacji > ilość parametrów (zmiennych + 1). Tak więc: 16 obserwacji > 14 zmiennych objaśniających + 1. W efekcie końcowym, po dziesięciu iteracjach z zastosowaniem metody *a posteriori* oraz sprawdzeniu współliniowości zmiennych objaśniających w programie *Gretl*, dobre dopasowanie do szeregu danych rzeczywistych cechy *X*3 dla Polski wykazały zmienne: *X*9, *X*10, *X*48, *X*50, *X*59[159]. Moim zdaniem, istnieje potrzeba zapoznania z procedurami tworzenia modeli z wykorzystaniem dorobku ekonometrii i programu *Gretl*, również przedstawicieli lokalnych urzędów prowadzących statystykę. Myśl ta była przesłaniem do niniejszej publikacji. *Gretl* jest bowiem programem bezpłatnie dostępnym w Internecie, a jego nowa wersja lansowana przez T. Kufla ma dogodny interfejs wejścia dla użytkownika. Stronę tytułową tego pakietu pobieranego z witryny *www.kufel.torun.pl* zaprezentowano na rysunku 10.1.

Źródło: Opracowanie własne.

Rys. 10.1. Strona tytułowa programu *Gretl*

10.2. Wejściowe dane statystyczne

Obecne czasy dają możliwość zestawiania danych do szeregów przekrojowych, czyli wielu zmiennych, bezpośrednio ze zasobów internetowych roczników statystycznych województw, opublikowanych i ogólnie dostępnych w formie elektronicznej. Przykład wejścia do rocznika 2016 pokazano na rysunku 10.2.



[159] Wornalkiewicz W., *Model ekonometryczny zmiennej „Liczba bezrobotnych*", op. cit., s. 183.

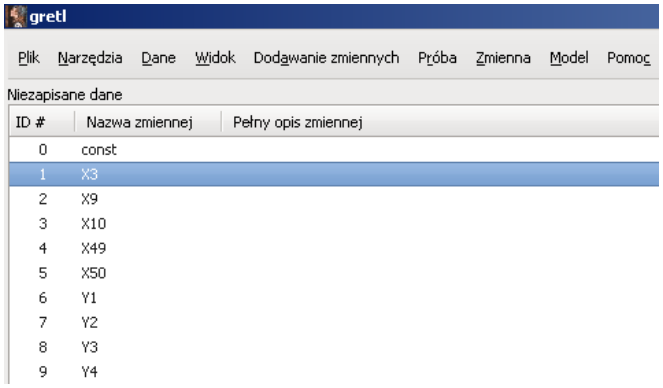
Źródło: Opracowanie własne.

Rys. 10.2. Fragment pliku „*Rocznik Statystyczny Województw 2016*”

Do badań cechy „*Liczba bezrobotnych*” we województwie opolskim przyjęto okres 2004-2015. Charakteryzował się on dość dynamicznymi zmianami w zakresie gospodarki i układów sprawozdawczości statystycznej, wynikającymi często z zaleceń Unii Europejskiej. Z tego powodu trzeba było poszukiwać dane szeregów czasowych w różnych rocznikach statystycznych na poziomie województw jak i kraju.

W dalszych rozważaniach teoretycznych posłużymy się programem *Gretl*. Po uruchomieniu programu *gretl.exe* pojawia się okno dialogowe umożliwiające wprowadzanie danych. *Gretl* sugeruje nam trzy typy danych, a mianowicie szeregi przekrojowe, czasowe oraz panelowe. Wybieramy przekrojowe, gdyż mamy 9 cech - zmiennych, których nazwy kolejno wprowadzamy wywołując zakładkę „*Dane*” (zob. rysunek 5.3). Po tej operacji mamy wstępnie zaprojektowany model o postaci ogólnej:

$$(X3) = a_0 + a_1(X9) + a_2(X10) + a_3(X49) + a_4(X50) + a_5(Y1) + a_6(Y2) + a_7(Y3) + a_8(Y4)$$



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.3. Zainicjowanie zmiennych do wprowadzania szeregów czasowych

Na rysunku 10.3 widzimy, że oprócz nazwy zmiennych możemy też wprowadzić pełny opis poszczególnych zmiennych. Jeśli mamy już utworzony zestaw zmiennych szeregu przekrojowego to w ramach menu głównego występują kolejne zakładki, a w ramach nich m.in. funkcjonalności:

Plik (otwórz dane, dołącz dane z pliku, zapisz dane, zapisz dane jako, eksport danych do pliku, wyślij plik danych e-mail, wyczyść zbiór danych);

Narzędzia (ustawienia, tablice statystyczne, wyznaczanie wartości *p*, wykresy rozkładów, wykresy funkcji, testy parametryczne, testy nieparametryczne, ziarno generatora poleceń;

Dane (zaznacz wszystko, znajdź zmienną, definiowanie lub edycja listy zmiennych, wybór zmiennych z listy, pokaż wartości, edycja wartości, dodaj obserwację);

Widok (ikony sesji, wykresy zmiennych, wykres wielokrotny, statystyki opisowe, macierz korelacji, tabela krzyżowa, główne składowe, odległość Mahalanobisa);

Dodawanie zmiennych (logarytmy dla wybranych zmiennych, kwadraty dla wybranych zmiennych, indeks obserwacji, generowanie zmiennych losowych, definiowanie nowej zmiennej, definiowanie macierzy);

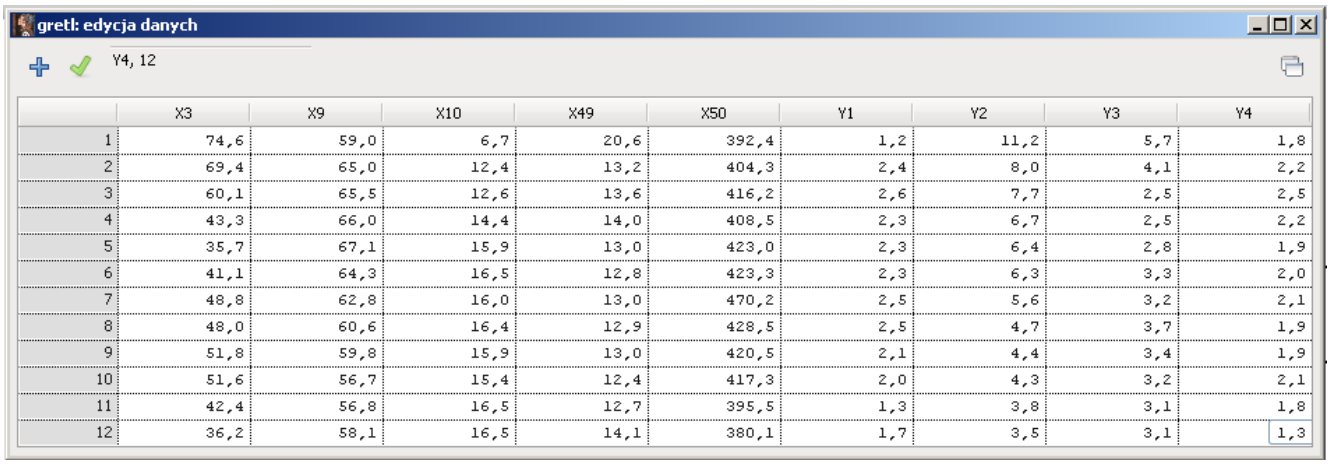
Próba (zakres prób, restrykcja dla podpróby, losowe podpróby, losowe próby z powtórzeniami, pomija obserwacje z brakującymi danymi, pokaż status danych);

Zmienna (pokaż wartości, edycja atrybutów, kod brakujących danych, statystyki opisowe, testy normalności rozkładu, wykres gęstości, wykres pudełkowy, średnie przedziałowe);

Model (klasyczna metoda najmniejszych kwadratów, metoda zmiennych instrumentalnych, inne liniowe modele, ograniczona zmienna zależna, nieliniowa metoda najmniejszych kwadratów, model równań współzależnych);

Pomoc (opis poleceń *Gretla*, opis funkcji *Gretla*, pakiety, skróty klawiszowe, przewodnik pakietu funkcji).

Wprowadźmy teraz do pliku o nazwie *Dane.gdt* dane statystyczne szeregu przekrojowego. Obejmują one jak już wspomniałem 12 okresów rocznych dla zmiennych *X*3, *X*9, *X*10, *X*49, *X*50, *Y*1-*Y*4 (zob. rysunek 10.3). Widok wprowadzonego szeregu przekrojowego w tys. pokazano na rysunku 10.4.

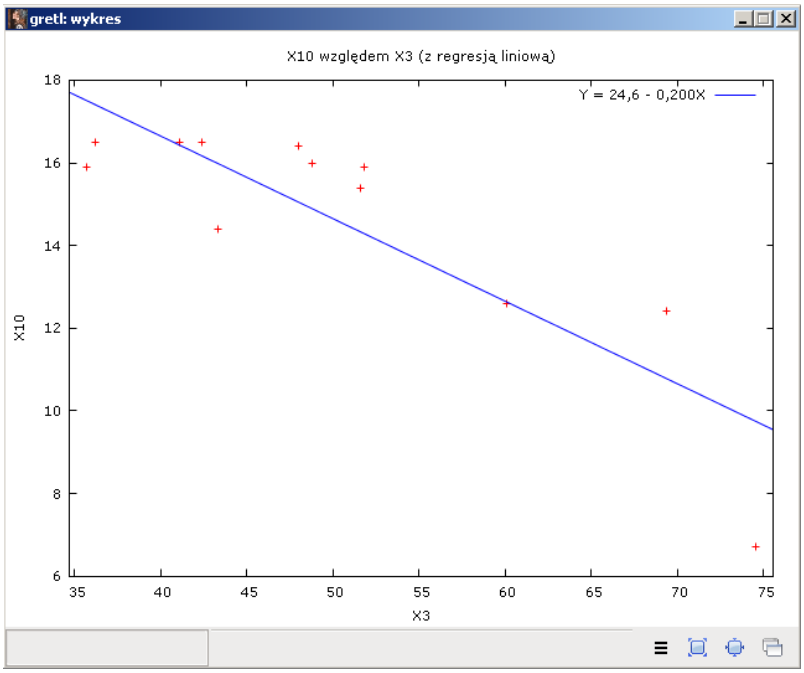


	X3	X9	X10	X49	X50	Y1	Y2	Y3	Y4
1	74,6	59,0	6,7	20,6	392,4	1,2	11,2	5,7	1,8
2	69,4	65,0	12,4	13,2	404,3	2,4	8,0	4,1	2,2
3	60,1	65,5	12,6	13,6	416,2	2,6	7,7	2,5	2,5
4	43,3	66,0	14,4	14,0	408,5	2,3	6,7	2,5	2,2
5	35,7	67,1	15,9	13,0	423,0	2,3	6,4	2,8	1,9
6	41,1	64,3	16,5	12,8	423,3	2,3	6,3	3,3	2,0
7	48,8	62,8	16,0	13,0	470,2	2,5	5,6	3,2	2,1
8	48,0	60,6	16,4	12,9	428,5	2,5	4,7	3,7	1,9
9	51,8	59,8	15,9	13,0	420,5	2,1	4,4	3,4	1,9
10	51,6	56,7	15,4	12,4	417,3	2,0	4,3	3,2	2,1
11	42,4	56,8	16,5	12,7	395,5	1,3	3,8	3,1	1,8
12	36,2	58,1	16,5	14,1	380,1	1,7	3,5	3,1	1,3

Źródło: Opracowanie własne w *Gretl*.

Rys. 10.4. Szereg przekrojowy zmiennej objaśnianej (zależnej) i zmiennych objaśniających (niezależnych)

Przykładowo skorzystajmy z funkcji (*Widok/Wykresy zmiennych/Wykres rozrzutu X-Y*) i zobaczmy czy jest zależność między wartościami zmiennej „*Liczba bezrobotnych*" a „*Liczba absolwentów technikum*". Łatwo zauważyć, że ze wzrostem liczby absolwentów technikum (*X*10) maleje liczba bezrobotnych (*X*3) - zob. rysunek 10.5. Równanie regresji liniowej podane na rysunku 10.5 jest następujące: (*X*3) = 24,6 - 0,2 (*X*10).

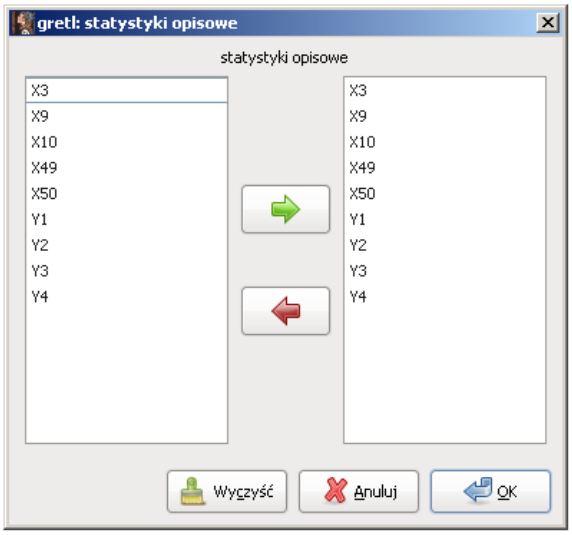


Źródło: Opracowanie własne w *Gretl*.

Rys. 10.5. Wykres liniowy zależności: (*X*3) → (*X*10)

Zachęceni tym przykładem, przejdziemy dalej do szukania zależności wielowymiarowej między zmienną zależną (*X*3) a innymi zmiennymi niezależnymi, przyjętymi intuicyjnie w tym opracowaniu do modelu ekonometrycznego. Występowanie danych w programie *Gretl* umożliwia mu

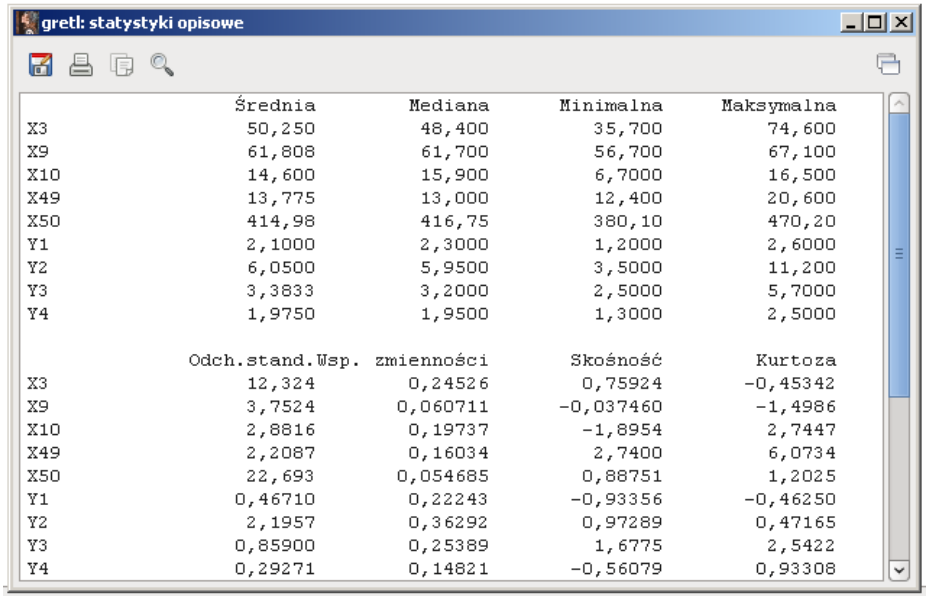
wygenerowanie statystyki opisowej dla wszystkich zmiennych lub selektywnie wybranych (zob. rysunek 10.6). Występuje możliwość wyboru opcji: standardowa, rozszerzona.



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.6. Okno dialogowe wyboru zmiennych do statystyki opisowej

Zakres statystyki rozszerzonej obejmuje dodatkowo współczynnik zmienności, skośność oraz kurtozę (zob. rysunek 10.7). Kurtoza jest miarą smukłości lub płaskości rozkładu danych szeregu zmiennej, natomiast skośność charakteryzuje niesymetryczność rozkładu zmiennej względem jej średniej.


gretl: statystyki opisowe

	Średnia	Mediana	Minimalna	Maksymalna
X3	50,250	48,400	35,700	74,600
X9	61,808	61,700	56,700	67,100
X10	14,600	15,900	6,7000	16,500
X49	13,775	13,000	12,400	20,600
X50	414,98	416,75	380,10	470,20
Y1	2,1000	2,3000	1,2000	2,6000
Y2	6,0500	5,9500	3,5000	11,200
Y3	3,3833	3,2000	2,5000	5,7000
Y4	1,9750	1,9500	1,3000	2,5000

	Odch.stand.	Wsp. zmienności	Skośność	Kurtoza
X3	12,324	0,24526	0,75924	-0,45342
X9	3,7524	0,060711	-0,037460	-1,4986
X10	2,8816	0,19737	-1,8954	2,7447
X49	2,2087	0,16034	2,7400	6,0734
X50	22,693	0,054685	0,88751	1,2025
Y1	0,46710	0,22243	-0,93356	-0,46250
Y2	2,1957	0,36292	0,97289	0,47165
Y3	0,85900	0,25389	1,6775	2,5422
Y4	0,29271	0,14821	-0,56079	0,93308

Źródło: Opracowanie własne w *Gretl*.

Rys. 10.7. Rozszerzona statystyka opisowa badanych zmiennych

Dodam jeszcze, że poszczególne miary statystyki opisowej zostały szczegółowo zdefiniowane w książce autora „*Formułowanie modeli ekonometrycznych na potrzeby zarządzania*”[160]. Skupmy jednak uwagę na współczynniku zmienności *V*, który traktowany jest jako narzędzie wstępnej selekcji zmiennych objaśniających budowanego modelu ekonometrycznego określonej cechy skutku. Tak

[160] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, Wydawnictwo Instytut Śląski, Opole 2014.

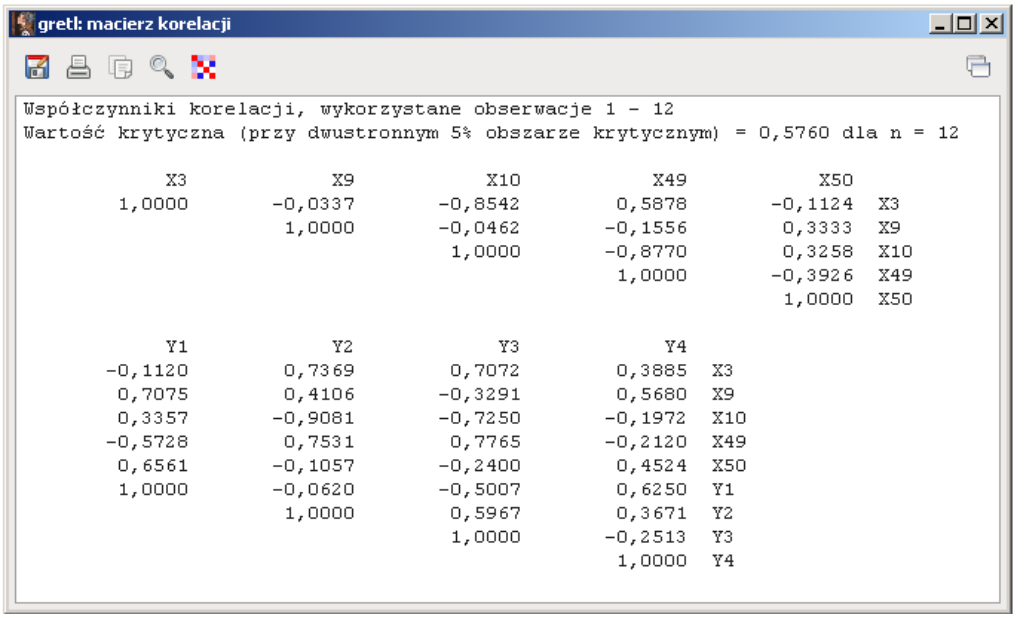
więc wyznaczenie współczynnika zmienności poszczególnych zmiennych *j* na podstawie powyższej statystyki opisowej stanowi wstępny krok zbadania wystarczającego zróżnicowania danej *j*-tej zmiennej objaśniającej. Współczynnik ten określany jest w procentach i stanowi moduł z ilorazu odchylenia standardowego i średniego poziomu wartości rozpatrywanej zmiennej.

Zmienna	*X*3	***X*9**	X10	X49	***X*50**	*Y*1	*Y*2	*Y*3	*Y*4
V_j	24,5	**6,1**	19,7	16,0	**5,5**	22,2	36,3	25,4	14,8

Z zestawienia wartości współczynnika zmienności wynika, że dla dwóch zmiennych *X*9 i *X*50 współczynnik zmienności nie przekracza wartości krytycznej *V** = 10%, czyli zmienne te z tego punktu widzenia można już wstępnie wyeliminować ze struktury projektowanego modelu ekonometrycznego. Biorąc pod uwagę tylko to kryterium model w postaci ogólnej będzie następujący:

$$(X3) = a_0 + a_1(X10) + a_2(X49) + a_3(Y1) + a_4(Y2) + a_5(Y3) + a_6(Y4).$$

Zależność między zmiennymi możemy przedstawić przy pomocy macierzy korelacji korzystając w *Gretl* z menu: *Widok/Macierz korelacji* w wyniku czego uzyskujemy dla naszych zmiennych raport pokazany na rysunku 10.8.



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.8. Rozszerzona statystyka opisowa badanych zmiennych

Dobór zmiennych objaśniających do modelu zmiennej (*X*3) „*Liczba bezrobotnych*" możemy przeprowadzić stosując analizę współczynników korelacji. W tym celu możemy skorzystać z podanej na rysunku 10.8. wartości krytycznej współczynnika korelacji *r** = 0,5760, dla 12. obserwacji, przy czym dalsze postępowanie w tej analizie jest następujące[161]:

1. Sprawdzamy, czy $|r_j| \le r^*$ i eliminujemy zmienne X_j (spośród: *X*9, *X*10, *X*49, *X*50, *Y*1-*Y*4) nieistotnie skorelowane z (X3).

2. Kierując się tym kryterium i przeglądając w macierzy korelacji pary relacji zmiennej objaśnianej *X*3 względem

$$|r_j| \le r^* = 0{,}5760$$

odpowiadających zmiennych objaśniających możemy wyeliminować cztery zmienne: *X*9, *X*50, *Y*1, *Y*4. Spośród wymienionych zmiennych, według współczynnika zmienności, eliminacji podlegały także zmienne *X*9 i *X*50.

[161] Dziechciarz J.: *Ekonometria Metody, przykłady, zadania*, op. cit.

3. Z pozostałych potencjalnych zmiennych wybieramy zmienną X_h według kryterium:

$$r_h = \max_j (r_j)$$

4. Jest nią $X10$ i badamy następnie relacje tak wybranej zmiennej z pozostałymi zmiennymi i eliminujemy te dla których $|r_{hi}| > r^*$, gdyż są one zbyt silnie skorelowane ze zmienną X_h. Okazuje się, że są nimi zmienne: $X49$, $Y2$, $Y3$. Po analizie współczynników korelacji wyeliminowano zmienne: $X9$, $X49$, $X50$, $Y1$-$Y4$. Pozostało nam wiec tylko $X10$, co sugeruje regresję o równaniu podanym na rysunku 10.5.

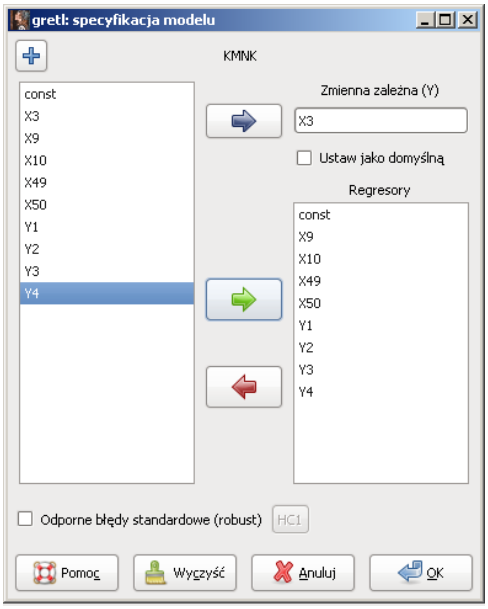
Przy dużej liczbie zmiennych objaśniających ich selekcja według procedury współczynnika korelacji jest pracochłonna. Z tego względu łatwiej jest skorzystać z metody zwanej eliminacją *a posteriorii* zawartej w programie *Gretl*.

10.3. Zastosowanie metody *a posteriori*

Powróćmy teraz do naszego wstępnego zestawu zmiennych *X3*, *X9*, *X10*, *X49*, *X50*, *Y1*-*Y4* danych szeregu przekrojowego zapisanego w pliku *Dane.gdt*. Zastosujemy do formułowania modelu klasyczną metodę najmniejszych kwadratów (KMNK) poprzez wywołanie menu:

Model/Klasyczna metoda najmniejszych kwadratów.

Definiujemy zmienną zależną *X3* oraz zmienne niezależne, czyli wcześniej wymienione *X9* do *Y4* i naciskamy OK (zob. rysunek 10.9).



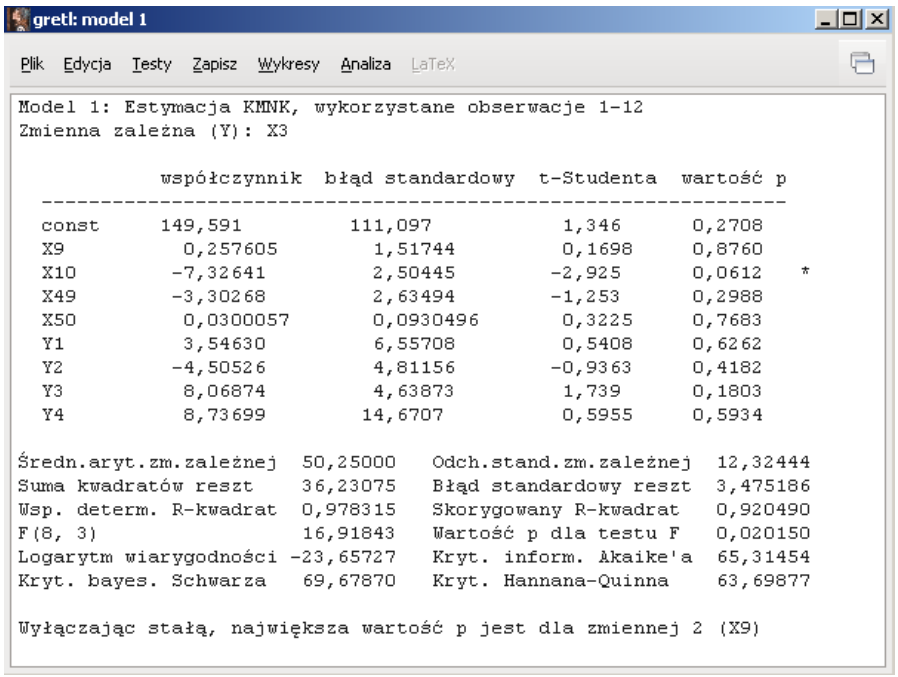
Źródło: Opracowanie własne w *Gretl*.

Rys. 10.9. Określenie zmiennych wstępnego modelu ekonometrycznego

Zwróćmy uwagę na przyjazne dla użytkownika wskazywanie zmiennych do modelu. Umożliwia to nam przeprowadzanie symulacji doboru zmiennych objaśniających. Efektem pracy programu jest *wygenerowany Model 1* wskazujący na oszacowane parametry a_0, (a_1-a_8) stojące przy zmiennych *X9*, *X10*, *X49*, *X50*, *Y1*-*Y4*, a ponadto *Gretl* podaje nam:

- błąd standardowy oszacowania danego parametru;
- statystykę *t*-Studenta;
- wartość *p*, czyli prawdopodobieństwa błędu oszacowania danego parametru.

Pod informacjami o estymowanym modelu wyszczególniona jest statystyka opisowa i zalecenie redukcji modelu o zmienną *X9*.



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.10. Oszacowanie parametrów *Modelu 1* zmiennej *X*3

Zwróćmy uwagę na wysoki skorygowany współczynnik determinacji (*Skorygowany R-kwadrat*) równy 0,920490 będący miarą dobroci dopasowania modelu wielowymiarowego liniowego do danych empirycznych. W powyższym raporcie szereg pojęć jest nam znanych z wykładów statystyki opisowej[162], jednak na szczególną uwagę proponuję zwrócić na k*ryterium informacyjne Akaike'a* (AIC) równe 65,31454 , które w literaturze określane jest według wzoru[163]:

$$AK = N \ln\left(\sum Z_n^2\right) + 2k$$

gdzie:

k - liczba parametrów modelu o *n* obserwacjach, w naszym przykładzie *k* = 9 (8 zmiennych i jedna $X_0 = 1$ przy parametrze a_0),

ΣZ_n^2 – suma kwadratów reszt.

Natomiast w *Gretl* obliczenia realizowane są według formuły:

$$AIC = -2l\left(\hat{\Theta}\right) + 2k$$

gdzie logarytm wiarygodności równy (-23,65727) oznaczany jest symbolem:

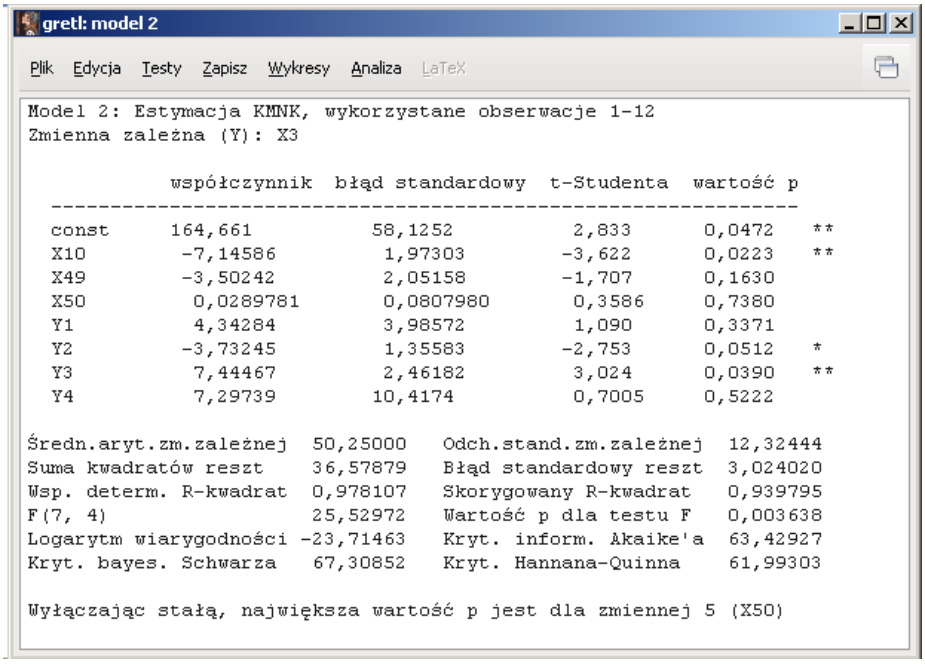
$$l\left(\hat{\Theta}\right)$$

Tak więc AIC = -2 (-23,65727) + 18 = 65,31454.

Pozostaje nam teraz korzystanie w kolejnych krokach z podpowiedzi *Gretla* na końcu raportu i budowanie coraz dokładniejszych modeli. W ten sposób uzyskujemy *Model 2*, i w tym miejscu dodam, że *Gretl* nadaje automatycznie narastająco nazwy wygenerowanych modeli ekonometrycznych (zob. rysunek 10.11).

[162] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, op. cit.

[163] Lipiec-Zajchowska (red), *Wspomaganie procesów decyzyjnych* , t. 2. *Ekonometria*. Wydawnictwo C.H. Beck Warszawa 2003.



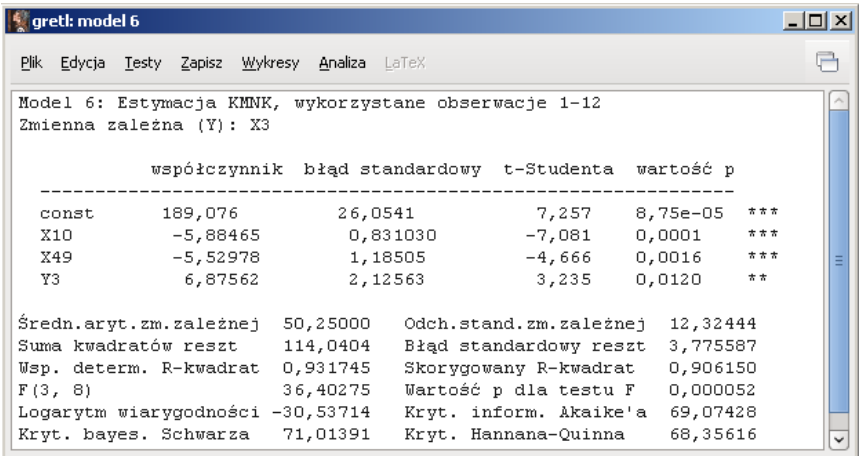
Źródło: Opracowanie własne w *Gretl*.

Rys. 10.11. Oszacowanie parametrów *Modelu 2* zmiennej *X*3 po wyeliminowaniu *X*9

Możemy tu zauważyć, że w wyniku wyeliminowania zmiennej *X*9 nastąpiło poprawienie skorygowanego współczynnika determinacji, który wynosi teraz 0,939795. Tak postępując doprowadzamy do *Modelu 6*, przy czym łącznie w kolejnych podejściach *Gretl* zaproponował nam wyeliminowanie zmiennych:

Model	1	2	3	4	5	6
Zmienna	*X*9	*X*50	*X*4	*Y*1	*Y*2	-

Model 6 uznany został przez program jako najlepszy według kryterium *p* najmniejszej wartości prawdopodobieństwa błędu oszacowania współczynnika (parametru modelu) - zob. rysunek 10.12.



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.12. *Model 6*, czyli końcowy w procedurze *a posteriori*

Ostatecznie, po zaokragleniu parametrów uzyskaliśmy model ekonometryczny:

$$(X3) = 189{,}08 - 5{,}88\ (X10) - 5{,}53\ (X49) + 6{,}88\ (Y3).$$

Program *Gretl* oprócz obliczania statystyki empirycznej *t*-Studenta podaje nam również graficznie informacje o poziomie istotności α dla obliczonego parametru, a mianowicie:

α	0,01	0,05	0,1	$\alpha > 0,1$
Oznaczenie	***	**	*	puste

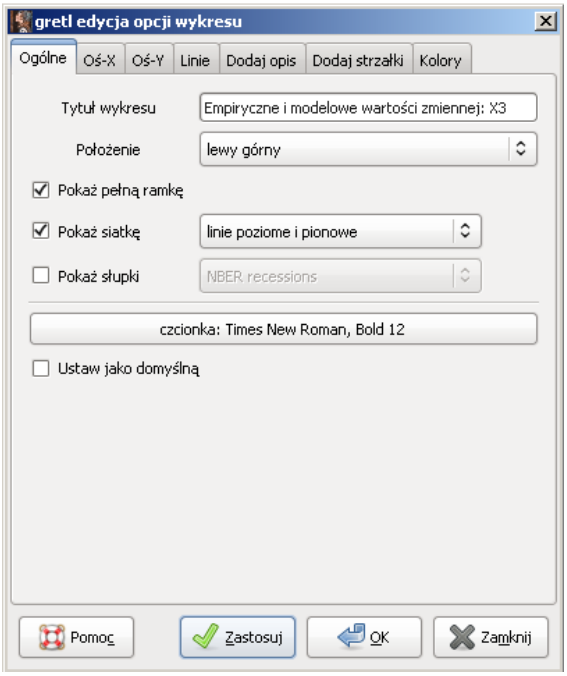
Puste oznacza brak statystycznej istotności danego parametru. W naszym przykładzie mamy trzy oraz dwie gwiazdki, czyli model jest zadawalający przy skorygowanym R^2 równym 0,906150. Przypomnijmy sobie teraz zmienne występujące w modelu końcowym:

- bezrobotni [tys.] - *X*3,
- przeciętne zatrudnienie w budownictwie [tys.] - *X*10,
- zużycie wody w gospodarstwach domowych w miastach [hm3] - *X*49,
- absolwenci technikum [tys.] - *Y*3.

Przy otwartym *Modelu 6* program *Gretl* umożliwia pokazanie na wykresie punktowym wartości z obserwacji i teoretycznych obliczonych na podstawie wygenerowanego modelu. W tym względzie korzystamy z menu:

Wykresy/Wykres empirycznych i wyrównanych/względem obserwacji.

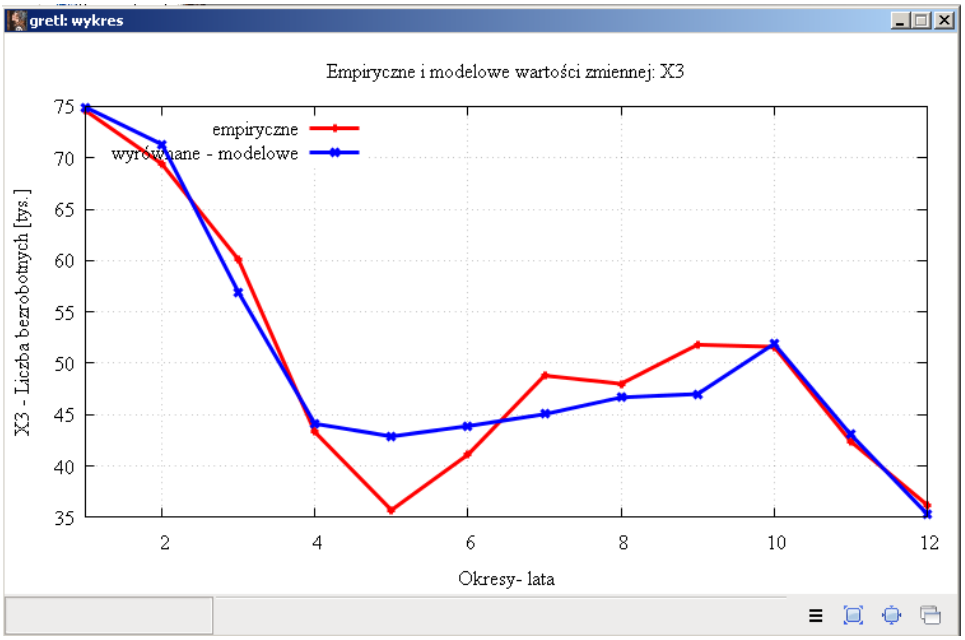
Gretl umożliwia edycję wykresu według życzenia użytkownika i w tym względzie posługujemy się oknem dialogowym (zob. rysunek 10.13).



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.13. Okno dialogowe edycji opcji wykresu

Istnieje dość bogata możliwość korekty wykresu wygenerowanego standardowo w formie punktowej. Na rysunku 10.14. widzimy wprowadzoną przez autora siatkę tła, linie łączące punkty oraz ich pogrubienia. Ponadto możemy zmienić czcionkę, tytuły osi oraz całego wykresu.



Źródło: Opracowanie własne w *Gretl*.

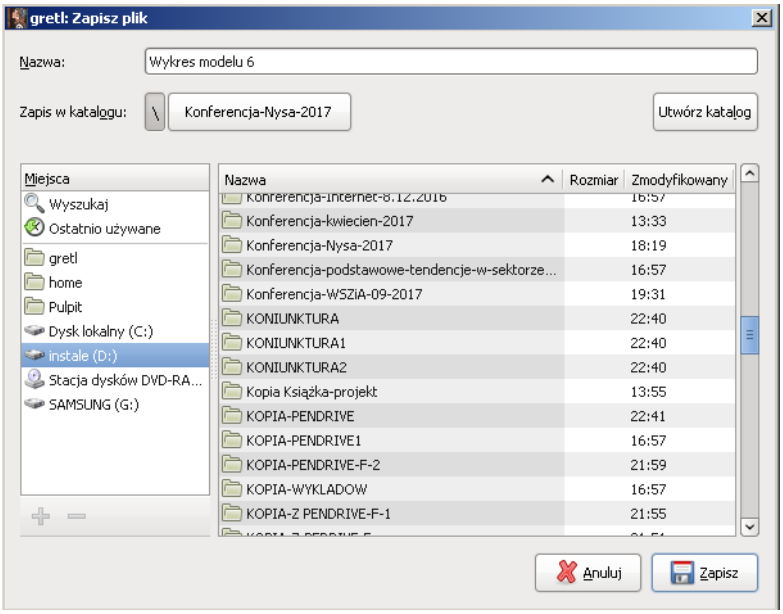
Rys. 10.14. Wartości empiryczne i modelowe zmiennej *X*3 „*Liczba bezrobotnych*”

10.4. Testy sprawdzające model

Oprócz sprawdzającego według współczynnika determinacji doboru zmiennych, *Gretl* udostępnia nam wiele testów sprawdzających dopuszczalność zastosowanej metody KMNK, co uwiarygodnia nasze postępowanie estymacyjne. W niniejszym materiale ograniczono się do oceny:

- normalności reszt;
- heteroskedastyczności, czyli zmienności wariancji resztowej;
- współliniowości zmiennych objaśniających.

Przystępując jednak do testowania wymagane jest występowanie *Modelu* 6 na ekranie monitora i dlatego dobrze jest zapisać go wcześniej w bazie danych programu (zob. rysunek 10.15).



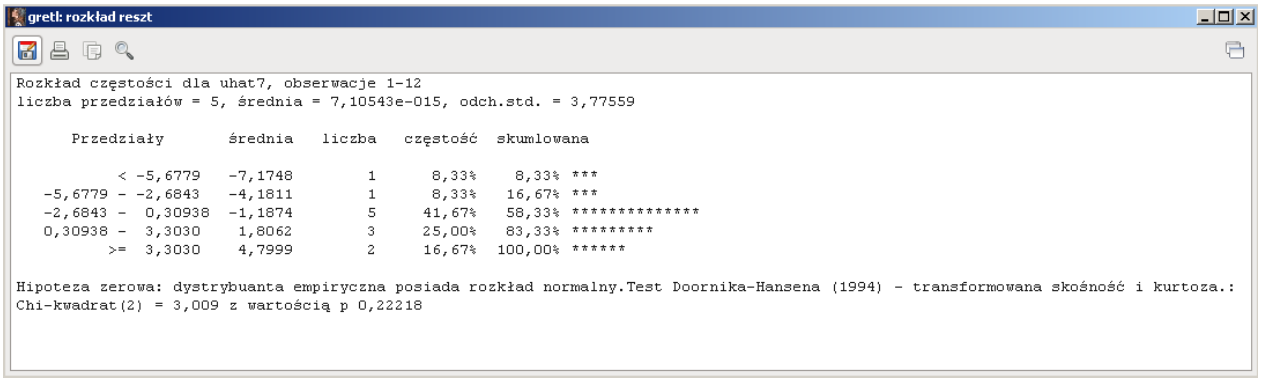
Źródło: Opracowanie własne w *Gretl*.

Rys. 10.15. Okno dialogowe zapisu *Modelu 6* na dysku użytkownika

Normalność reszt

Sprawdzając wymaganie jakim jest występowanie normalności reszt korzystamy z menu:
Testy/Test na normalność rozkładu reszt.

Gretl generuje jednocześnie rozkład częstości reszt oraz ich przedstawienie na wykresie dla 12. obserwacji (zob. rysunki 10.16 i 10.17). Liczba przedziałów jest 5, średnia = 7,10543e-015, a odchylenie standardowe reszt równa się 3,7755.



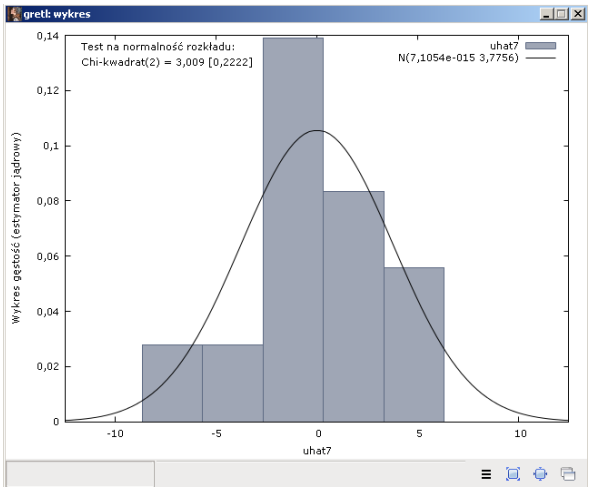
Źródło: Opracowanie własne w *Gretl.*

Rys. 10.16. Szereg rozdzielczy reszt zmiennej *X*3

Wychodzimy z hipotezy zerowej H_0*: dystrybuanta empiryczna posiada rozkład normalny*. W *Gretl* zastosowano test bazujący na skośności i kurtozie. Z wykresu odczytamy statystykę testu chi-kwadrat przy dwóch stopniach swobody:

$$\chi^2_{(2)} = 0{,}3009 \text{ z wartością prawdopodobieństwa } p = 0{,}2222.$$

Wartość krytyczna $\chi^2_{(\alpha,r)}$ przy poziomie istotności $\alpha = 0{,}05$ i dwóch stopniach swobody wynosi 5,991. Zatem spełnione jest hipoteza H_0, gdyż wartość statystyki jest mniejsza niż wartość krytyczna.



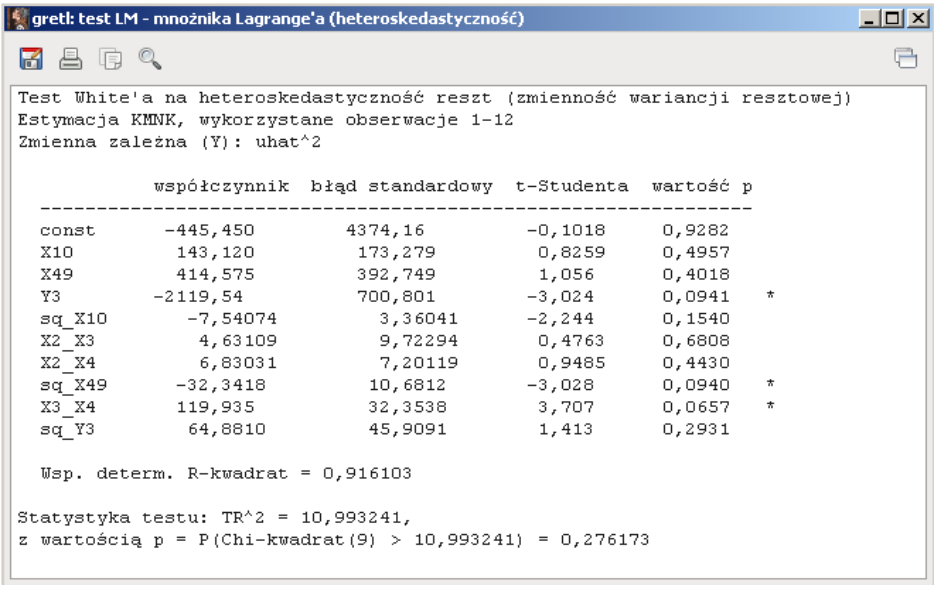
Źródło: Opracowanie własne w *Gretl.*

Rys. 10.17. Przedstawienie częstości występowania zakresów reszt zmiennej *X*3

Heteroskedastyczność

Dla oceny zmienności wariancji resztowej zastosowano w *Gretl* test White'a przeprowadzany poprzez wywołanie menu: *Testy/Test heteroskedastyczności/Test White'a.* W tym celu zakładamy na wstępie hipotezę zerową H_0*: heteroskedastyczność reszt nie występuje*. Zaprezentowany na rysunku 10.18. raport podaje statystykę testu, przy czym dla zmiennych *Modelu 6* można ją obliczyć również według wzoru: $LM = TR^2 = 12 \times 0{,}916103 = 10{,}993232$.

Test Withe'a zakłada wprowadzenie dodatkowo do modelu kwadratów zmiennych, oznaczonych przez program standardowo jako sq_*X*10, sq_*X*49, sq_*Y*3, a ponadto adekwatnych wzajemnych iloczynów zmiennych modelu. Wartość krytyczna $\chi^2_{(\alpha,r)}$ wg rozkładu χ^2 przy poziomie istotności α = 0,05, przy r = 2 stopniach swobody wynosi 5,991, czyli jest mniejsza od obliczonego LM, co potwierdza założoną wcześniej hipotezę zerową[164].



Źródło: Opracowanie własne w *Gretl*.

Rys. 10.18. Test White'a na heteroskedastyczność reszt zmiennej *X*3

Współliniowość zmiennych objaśniających

Gdy występują wysokie współczynniki korelacji między zmiennymi objaśniającymi to dodatkowo stosowane jest zbadanie współliniowości zmiennych objaśniających[165]. Sytuacja taka powoduje zaniżanie wartości statystyki *t*-Studenta w ocenie istotności parametru. Ocenę stopnia współliniowości zmiennych objaśniających w programie *Gretl* można dokonać posługując się miarą VIF *j*-tych par zmiennych określaną jako czynnik powiększania wariancji, zwany też inflacją wariancji. Zobaczmy najpierw jak kształtują się *współczynniki* korelacji między zmiennymi objaśniającymi *X*10, *X*49, *Y*3.

R_j	*X*10	*X*49	*Y*3
*X*10	1	-0,8770	-0,7250
*X*49	-0,8770	1	0,7765
*Y*3	-0,7250	0,7765	1

Są one na średnim poziomie, tak więc nie ma potrzeby przeprowadzania testu współliniowości między zmiennymi objaśniającymi naszego *Modelu 6*. Nadmienię, że miara VIF_j *rowna się* $1/(1 - R_j^2)$, przy czym Rj^2 jest współczynnikiem korelacji wielorakiej między daną zmienną *j* a pozostałymi zmiennymi modelu. Minimalna możliwa wartość VIF_j równa się 1,0. Jeśli wartość jest większa od 10 to wskazuje to na problem współliniowości, czyli rozdęcia wariancji[166].

[164] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, op. cit, s. 648.
[165] Kufel T.: *Ekonometria Rozwiązywanie problemów z wykorzystaniem programu Gretl,* op. cit., rozdz. 4.3.6.
[166] Wornalkiewicz W., *Model ekonometryczny zmiennej „liczba bezrobotnych*", op. cit., s. 184.

* * *

Analiza współczynników zmienności pozwala na usuniecie z modelu ekonometrycznego zmiennych objaśniających o charakterze prawie stacjonarnym. Zastosowanie natomiast analizy współczynników korelacji umożliwia wyeliminowanie zmiennych zbyt silnie skorelowanych ze zmienną ostatnio dobraną w procedurze tej analizy. Zastosowana wygodna procedura *a posteriori* w *Gretl* powinna być uzupełniona sprawdzeniem współliniowości zmiennych objaśniających według kryterium VIF. Postępując w ten sposób dodatkowo upewniamy się o poprawności zaprojektowanego modelu ekonometrycznego bazującego na szeregu przekrojowym. Zaprezentowane podejście dogodnego modelowania wielowymiarowego na danych modelowych województwa opolskiego i cechy „*Liczba bezrobotnych*" może być moim zdaniem z powodzeniem zastosowane w badaniach demograficznych i innych.

11. Etapy i procedury budowy modelu produktu brutto podregionów

11.1. Wprowadzenie

Produkt krajowy brutto jest końcowym rezultatem wszystkich podmiotów gospodarki narodowej. Produkt ten równa się sumie wartości dodanej brutto wszystkich sektorów własności oraz sektorów instytucjonalnych powiększonej o podatki od przedmiotów i pomniejszonej o dotacje do produktów[167]. Produkt krajowy zarówno brutto jak i netto liczony jest w cenach rynkowych (bieżących). Z dniem 1.01.1999 r. wprowadzono trójstopniowy podział terytorialny państwa (gminy, powiaty, województwa). W dniu 31.12.2006 r. było 16 województw, 314 powiatów, 65 miast na prawach powiatu oraz 2478 gmin, a ramach gmin jednostki pomocnicze - sołectwa, których liczba wynosiła 40348. Obecnie dla celów sprawozdawczych występuje 6 regionów grupujących określone województwa oraz 45 podregionów obejmujących jednostki szczebla powiatowego. Nomenklatura jednostek terytorialnych dla celów statystycznych odpowiada nomenklaturze w Unii Europejskiej.

W przykładowym zestawieniu statystycznym „*Podstawowe dane o regionach w roku* ...) występują 22 kolumny z informacjami charakteryzującymi nie tylko poszczególne regiony, lecz również w ramach nich podregiony. Spośród tych informacji wybrano takie które tworzą szeregi o długości 7 okresów (lata 2000-2006) zmiennych egzogenicznych (objaśniających). Dane czasowo-przekrojowe dotyczące podregionów dostępne są w rocznikach statystycznych lat (2000-2007). Dla realizacji procedury budowy modelu produktu brutto podregionów zmiennym nadano następujące symbole:

Y - produkt brutto (ceny bieżące) na 1 mieszkańca w roku poprzednim w zł,
X_1 - ludność na 1 km^2 powierzchni ogólnej,
X_2 - przyrost naturalny na 1000 ludności,
X_3 - saldo migracji wewnętrznych i zagranicznych na pobyt stały (na 1000 ludności),
X_4 - przestępstwa stwierdzone na 1000 ludności,
X_5 - pracujący w tys. (rolnictwo, łowiectwo i leśnictwo, rybactwo),
X_6 - pracujący w tys. (przemysł i budownictwo),
X_7 - pracujący w tys. (usługi rynkowe i nierynkowe),
X_8 - bezrobotni zarejestrowani ogółem w tys.,
X_9 - mieszkania oddane do użytkowania na 1000 ludności,
X_{10} - miejsca noclegowe w turystycznych obiektach zbiorowego zakwaterowania w tys.,
X_{11} - użytki rolne w tys. ha w roku poprzednim.

Jako zmienną objaśniającą przyjęto produkt brutto, natomiast zmiennymi objaśniającymi liniowego modelu ekonometrycznego są (X_1 - X_{11}). Dane zaczerpnięto z tabel:

- wybrane dane o podregionach (lata 2000÷2003),
- podstawowe dane o regionach (lata 2004÷2007).

Ponieważ niektóre informacje za ostatni rok podawane są na podstawie danych nie ostatecznych (a więc mogą ulec zmianie w roku następnym), dla celów badawczych zestawiono dane za okres (2000-2006) dla poszczególnych 45. podregionów rozpoczynając roku 2006 wstecz. W okresie sprawozdawczym (2000-2007) nastąpiły zmiany w strukturze administracyjnej kraju, stąd też niektóre brakujące dane dla podregionów województwa śląskiego przyjęto takie same jak dla roku sąsiedniego. Wystąpiły też zmiany w jednostkach miary. Przykładem są użytki rolne określane obecnie w tys. ha. Przy przeliczaniu na 1000 ludności w danych statystycznych przyjmowany jest

[167] *Rocznik Statystyczny Rzeczypospolitej Polskiej 2007*. Dział: *Rachunki narodowe*.

stan na 31 grudnia. Warto przytoczyć za rocznikiem statystycznym z roku 2007 niektóre pojęcia występujące w powyższych zmiennych:

- przestępstwo stwierdzone jest to zdarzenie, co do którego w zakończonym postępowaniu przygotowawczym potwierdzono, że jest przestępstwem;
- przyrost naturalny ludności stanowi różnicę miedzy liczbą urodzeń żywych i zgonów w danym okresie;
- przyrost rzeczywisty ludności jest sumą przyrostu naturalnego i salda migracji;
- migracje wewnętrzne ludności są to zmiany miejsca zamieszkania w kraju;
- migracje zagraniczne ludności są to wyjazdy poza granicę kraju zamieszkania, a także przyjazdy do kraju.

Dane o migracjach zagranicznych na popyt stały bazują na informacjach ewidencyjnych gmin (imigracja) oraz o wymeldowaniach osób wyjeżdzających z Polski (emigracja), a dane o pracujących dotyczą osób wykonujących pracę przynoszącą zarobek lub dochód. Natomiast dane o bezrobotnych zarejestrowanych obejmują osoby, które zgodnie z ustawą z dnia 21.04.2004 o promocji zatrudnienia i instytucjach rynku pracy, określane są jako bezrobotne. Przy czym pod pojęciem bezrobotnego rozumie się osobę niezatrudnioną i niewykonującą innej pracy zarobkowej, zdolną i gotową do podjęcia zatrudnienia, nieuczącą się (z wyjątkiem szkół dla dorosłych)[168].

Sformułowanie modelu ekonometrycznego dla podregionów wymaga wielu lat obserwacji, co przy dynamicznie zmieniającej się sytuacji gospodarczej wydaje się niecelowe. Z tego względu posłużono się poszczególnymi 45 jednostkami administracyjnymi (podregionami) obserwowanymi przez okresy 7-mio letnie. Powstał w ten sposób panel danych o 45 x 7 = 315 obserwacjach. Fragment tego panelu obejmującego województwo łódzkie zaprezentowano w tabeli 11.1. Jeżeli w rocznikach statystycznych użytki rolne występowały w % to dokonano przeliczenia na tys. ha w sposób następujący:

- powierzchnia ogółem w km^2,
- użytki rolne w km^2 (powierzchnia ogółem x przelicznik w %),
- przeliczenie na hektary (1 km^2 = 100 ha),
- przemnożenie przez 1000.

Tab. 11.1. Fragment panelu danych statystycznych obejmujący województwo łódzkie

Lp.	Region/Województwo/Podregion	Rok	Okres	Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	%
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
1	**centralny/łódzkie/**	2000	1	11442	110,0	-2,3	0,4	223,9	164,2	75,3	83,7	77,6	1,1	9,4	629,9	72,2
	łódzki	2001	2	12790	109,0	-2,6	0,1	269,6	164,1	75,3	83,7	86,1	1,5	5,0	629,9	72,3
		2002	3	14350	108,0	-2,8	0,4	278,8	93,6	66,4	81,0	87,4	1,6	5,2	575,5	
		2003	4	15259	108,0	-2,9	0,8	287,4	93,3	67,7	80,5	88,3	3,2	4,9	**564,1**	66,9
		2004	5	15955	108,0	-2,4	0,9	287,0	93,5	68,5	81,1	83,2	2,1	4,9	**564,1**	66,9
		2005	6	18713	107,0	-3,0	0,6	275,0	93,5	72,5	81,0	77,6	2,0	4,8	552,7	
		2006	7	19837	107,0	-2,5	0,5	256,0	93,5	76,5	82,5	65,3	1,8	4,7	552,7	
2	piotrkowsko-skierniewicki	2000	1	12254	97,0	-0,9	-1,0	216,3	177,3	76,7	79,3	75,0	1,7	8,6	611,8	66,5
		2001	2	13478	97,0	-1,3	-1,0	265,5	176,8	74,9	74,9	82,6	2,2	9,2	610,9	76,0
		2002	3	15289	86,0	-1,1	-1,1	268,5	96,5	70,2	77,8	82,3	2,3	8,2	548,1	
		2003	4	16060	95,0	-1,7	-1,0	275,0	96,1	69,6	78,1	79,6	4,9	7,0	553,8	60,2
		2004	5	17034	95,0	-1,4	-1,4	266,0	96,2	71,8	79,6	73,4	2,5	7,3	553,8	60,2
		2005	6	20040	95,0	-1,3	-1,1	246,0	96,2	72,6	81,6	66,2	2,3	7,0	514,2	
		2006	7	21657	94,7	-0,8	-1,8	257,0	96,2	73,7	82,9	56,7	2,3	7,5	514,2	
3	miasto Łódź	2000	1	20645	2694,0	-7,0	-0,9	472,4	3,9	73,5	140,7	55,5	1,5	5,7	**11,0**	37,3
		2001	2	21767	2672,0	-6,7	-0,9	450,0	3,9	72,6	141,5	61,9	2,1	4,8	10,6	36,0
		2002	3	23924	2671,0	-6,3	-1,2	489,3	2,2	64,1	135,7	65,1	1,6	4,6	23,1	
		2003	4	25072	2647,0	-6,2	-1,2	538,7	2,1	60,0	137,7	65,2	1,6	3,3	**21,2**	27,7
		2004	5	26531	2629,0	-5,8	-1,5	550,0	2,1	59,1	134,5	61,7	1,8	3,5	**21,2**	27,7
		2005	6	29104	2608,0	-6,1	-1,8	512,0	2,1	59,3	138,9	54,7	0,9	3,5	19,2	
		2006	7	30602	2593,0	-6,5	-2,4	434,0	2,1	60,9	147,9	38,7	1,6	3,6	19,2	

Źródło: Opracowanie własne na podstawie roczników statystycznych.

[168] *Rocznik Statystyczny Rzeczypospolitej Polskiej 2007*. Dział: *Rynek pracy*.

W poszczególnych podregionach występuje znaczne zróżnicowanie wysokości produktu brutto w przeliczeniu na 1. mieszkańca. Możemy to szczególnie zauważyć po wprowadzeniu wartości średniej podregionów Polski i wykonaniu wykresu.

11.2. Podejście do modelowania

W literaturze występują różne sposoby sformułowania i weryfikacji liniowego modelu ekonometrycznego opartego o dane panelowe. Najczęściej sugeruje się rozwiązania ze wspomaganiem komputerowym. Jednym z przykładów jest określenie modelu po wcześniejszym wyznaczeniu wartości średnich z obserwacji zmiennych (zob. tabela 11.2). Zabieg ten ma na celu wyeliminowanie problemu z odwracaniem wysokiego rzędu macierzy.

Tab. 11.2. Wartości średnie obserwacji zmiennych dla podregionów

Lp.	Podregion	$\bar{y}$	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$	$\bar{x}_5$	$\bar{x}_6$	$\bar{x}_7$	$\bar{x}_8$	$\bar{x}_9$	$\bar{x}_{10}$	$\bar{x}_{11}$
1	łódzki	15478	108,1	-2,6	0,5	268,2	113,7	71,7	81,9	80,8	1,9	5,6	581,3
2	piotrkowsko-skierniewicki	16544,6	94,2	-1,2	-1,2	256,3	119,3	72,8	79,2	73,7	2,6	7,8	558,1
3	miasto Łódź	25377,9	2644,9	-6,4	-1,4	492,3	2,6	64,2	139,6	57,5	1,6	4,1	17,9
4	ciechanowsko-płocki	20354,1	81,4	0,1	-1,7	282,8	80,7	44,0	61,8	64,8	2,6	2,9	535,5
5	ostrołęcko-siedlecki	14160,1	62,6	0,7	-2,3	278,8	126,0	35,2	64,9	64,1	2,1	5,3	750,9
6	warszawski	20122,1	163,4	-0,1	8,1	347,0	88,2	89,9	132,9	66,7	6,0	4,7	460,6
7	radomski	14743,6	104,6	0,5	-1,8	311,9	98,9	143,3	64,9	81,8	2,1	3,2	405,6
8	miasto Warszawa	60932,7	3271,3	-2,5	3,8	536,1	139,2	157,6	602,4	55,4	8,1	19,8	88,4
9	krakowsko-tarnowski	14488,3	189,6	0,6	1,1	262,4	145,2	91,2	116,0	83,8	3,1	6,0	427,4
10	nowosądecki	12391,9	147,7	3,9	-0,1	230,5	119,5	55,6	98,1	79,7	3,0	43,7	331,5
11	miasto Kraków	32526,4	2302,3	-1,1	1,8	608,2	2,9	71,7	178,8	32,6	6,1	14,2	20,9
12	częstochowski (północnośląski)	18488,7	177,3	-2,4	-0,5	298,3	44,6	62,4	73,8	47,5	2,3	4,4	198,1
13	bielsko-bialski (południowośląski)	21405,3	288,9	0,5	0,7	300,8	38,4	100,5	105,6	47,8	3,8	21,4	133,9
14	centralny śląski	25433,8	860,1	-1,8	-3,1	456,8	22,3	302,4	378,8	172,8	1,3	8,3	160,7
15	rybnicko-jastrzębski	22125,8	474,2	1,3	-3,8	353,1	8,3	66,3	66,2	34,7	1,8	1,9	66,1
16	bialskopodlaski	12878,1	70,5	-0,6	-2,9	306,4	55,6	24,1	41,7	30,4	1,9	6,2	326,1
17	chełmsko-zamojski	12266,1	71,6	-1,5	-2,8	248,4	128,5	27,6	58,6	56,3	1,5	6,1	582,3
18	lubelski	16052,6	123,7	-0,2	-1,3	292,4	162,5	71,3	142,6	81,5	2,4	9,0	663,3
19	rzeszowsko-tarnobrzeski	16164	154,6	1,9	-0,9	232,9	122,6	99,5	126,8	88,7	2,6	7,0	399,3
20	krośnieńsko-przemyski	12890,4	91,3	1,2	-1,7	187,5	117,2	57,8	93,8	86,5	1,9	13,2	424,2
21	białostocko-suwalski	16364,4	60,2	-0,9	-0,9	290,2	108,5	47,8	103,9	54,2	3,2	11,2	781,0
22	łomżyński	12990,3	58,9	0,9	-3,5	239,8	55,6	12,3	24,9	23,6	1,4	2,0	360,6
23	swiętokrzyski	15974,6	110,9	-1,4	-1,9	263,5	185,3	81,0	131,9	121,6	1,8	8,8	627,4
24	gorzowski	18755	62,4	1,3	-0,4	418,0	14,3	30,3	45,8	34,3	3,2	8,3	211,4
25	zielonogórski	18169,3	80,2	0,7	-1,1	409,1	20,9	45,4	71,8	61,5	2,5	10,7	289,8
26	pilski	16908,6	63,3	2,2	-1,3	330,4	27,8	35,4	37,8	35,3	1,9	4,3	309,0
27	poznański	18518	99,6	2,1	4,2	262,1	79,6	114,0	110,6	65,1	3,4	17,9	694,5
28	kaliski	16073,4	112,3	0,7	-0,6	243,9	86,2	75,0	74,3	58,9	2,2	4,0	486,0
29	koniński	16534,4	98,3	1,1	-1,2	286,8	52,6	36,5	37,8	42,4	2,2	6,0	296,1
30	miasto Poznań	42297,9	2187,4	-1,3	-2,4	667,3	4,6	64,8	157,0	18,4	5,0	7,2	29,1
31	szczeciński	21570,7	88,6	0,3	-0,2	424,6	31,9	73,1	138,0	98,6	3,2	61,3	623,1
32	koszaliński	16983,9	57,6	1,2	-1,8	372,5	22,4	35,4	60,6	72,3	2,6	54,0	410,5
33	jeleniogórsko-wałbrzyski	16728,4	129,0	-1,9	-2,2	385,1	46,8	101,4	130,8	138,4	1,5	37,4	497,2
34	legnicki	26157,6	121,3	1,1	-2,6	461,5	19,7	52,4	61,2	44,0	2,0	2,4	208,4
35	wrocławski	15930	85,1	0,9	3,6	236,4	30,2	33,6	33,8	34,8	3,4	3,2	285,9
36	miasto Wrocław	30598,4	2171,4	-2,1	1,0	586,9	2,0	50,4	140,0	30,4	5,4	7,1	56,2
37	opolski	17184,6	112,7	-0,7	-3,7	331,6	65,6	76,3	108,7	73,9	1,4	7,1	563,1
38	bydgoski	19971,7	114,6	0,5	-0,6	378,8	53,7	86,8	125,0	88,9	2,2	14,6	496,3
39	toruńsko-włocławski	17075,6	116,7	0,7	-1,1	355,1	87,1	76,7	111,0	105,0	2,5	11,9	599,8
40	słupski	15745,3	58,7	2,8	-2,1	422,8	21,4	36,2	50,3	58,0	2,6	34,1	301,8
41	gdański	15276,6	99,1	4,3	2,5	369,7	50,3	63,9	77,2	79,1	4,3	38,6	496,5
42	Gdańsk-Gdynia-Sopot	30127,1	1812,9	-0,8	-0,8	672,3	3,1	64,0	151,2	30,2	4,9	16,3	66,1
43	elbląski	15207,6	71,5	2,1	-2,7	336,9	31,8	39,0	51,2	61,7	2,1	8,6	378,6
44	olsztyński	17814,9	59,7	1,7	-1,3	396,8	30,1	42,4	78,6	62,1	3,5	22,0	486,6
45	ełcki	14091,6	45,2	2,2	-2,7	395,5	19,4	14,7	24,5	36,9	2,7	9,2	257,8

Źródło: Opracowanie własne z zastosowaniem *Excela*.

Po uśrednieniu zmiennej objaśnianej *Y* oraz zmiennych objaśniających (X_1-X_{11}) możemy zastosować funkcje Excela dla określenia wektora parametrów modelu liniowego, który w postaci ogólnej określany jest następującym równaniem macierzowym:

$$\mathbf{a} = (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathbf{T}}\mathbf{y}$$

gdzie:

X - macierz zmiennych o wymiarach N x K = 315 x 11,
$\mathbf{X}^{\mathbf{T}}$ - macierz transponowana o wymiarach 11 x 315,
$(\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-1}$ - macierz odwrotna o wymiarach 11 x 11,
y - wektor zmiennej objaśnianej o wymiarach 315 x 1.

W określeniu modelu na bazie danych panelowych podregionów korzysta się z estymatora *fixed effect* (FE), który zakłada, że efekty indywidualne b_i dotyczące poszczególnych 45. podregionów można oszacować. Wzór w postaci macierzowej na ocenę parametrów wektora **b** jest następujący[169] :

$$\hat{\boldsymbol{\beta}}_{\mathbf{FE}} = (\sum_{\mathbf{i}=1}^{\mathbf{N}} \tilde{\mathbf{X}}_{\mathbf{i}}^{\ \mathbf{T}} \tilde{\mathbf{X}}_{\mathbf{i}})^{-1} \sum_{\mathbf{i}=1}^{\mathbf{N}} (\tilde{\mathbf{X}}_{\mathbf{i}}^{\ \mathbf{T}} \tilde{\mathbf{y}}_{\mathbf{i})}$$

przy czym zastosowano tu podstawienia:

$$\tilde{\mathbf{y}}_{\mathbf{i}} = \mathbf{y}_{\mathbf{i}} - \mathbf{c}\bar{\mathbf{y}}_{\mathbf{i}} \qquad\qquad \tilde{\mathbf{X}}_{\mathbf{i}} = \mathbf{X}_{\mathbf{i}} - \mathbf{c}\bar{\mathbf{x}}_{\mathbf{i}}^{\ \mathbf{T}}$$

Wektor oszacowanych efektów $\mathbf{b_{FE}}$ dotyczy poszczególnych 11. zmiennych objaśniających. Obliczenia poprzedzone muszą być określeniem różnic wartości obserwacji od średnich. Podane w powyższym zapisie macierzowym określenie **c** to zmienna zero-jedynkowa odnosząca się tylko do określonego podregionu. Przekształcenie jako wektor $\mathbf{b_{FE}}$ umożliwia zastosowanie klasycznej metody najmniejszych kwadratów (KMNK), a ramach niej funkcji programu Excel na obszarach adresowych macierzy lub wektorów:

TRANSPONUJ(...) transpozycja macierzy **X**,
MACIERZ.ILOCZYN(...;...) – iloczyn macierzy transponowanej $\mathbf{X}^{\mathbf{T}}$ oraz macierzy zmiennych **X**,
MACIERZ.ODW(...) – macierz odwrotna macierzy powstałej z iloczynu ($\mathbf{X}^{\mathbf{T}}\mathbf{X}$).

Po określeniu wektora parametrów stojących przy zmiennych objaśniających estymator 45. efektów indywidualnych dla poszczególnych podregionów otrzymujemy ze z poniższego wzoru[170]:

$$\hat{\boldsymbol{\beta}}_{\mathbf{i}} = \bar{\mathbf{y}}_{\mathbf{i}} - \bar{\mathbf{x}}_{\mathbf{i}}^{\ \mathbf{T}} \hat{\boldsymbol{\beta}}_{\mathbf{FE}}$$

W obliczeniach oprócz funkcji rachunku macierzowego możemy skorzystać alternatywnie z funkcji REGLINP. Korzystając z tej funkcji nie przyjmujemy występowanie parametru stałego, gdyż będzie on rozczłonkowany jako efekty indywidualne wpływu danych podregionów. Sformułowanie modelu po powyższej procedurze umożliwia porównanie wartości empirycznych i modelowych zmiennej objaśnianej. Ponadto możemy obliczyć procentowy błąd względny MAPE dla 315 obserwacji *Y* dla całego panelu oraz poszczególnych obiektów (podregionów) korzystając ze wzoru:

$$\mathbf{MAPE} = \left|(\mathbf{y} - \hat{\mathbf{y}})/\mathbf{y}\right| * 100$$

Kontynuowanie manualne ze wspomaganiem komputerowym funkcjami Excela jest zbyt mozolne i z tego powodu w dalszej procedurze budowy modelu ekonometrycznego produktu brutto podregionów posłużono się programem *Gretl.*

[169] Osińska M. (red. naukowa), *Ekonometria współczesna,* Towarzystwo Naukowe Organizacji i Kierownictwa. Toruń 2007.
[170] Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2007.

11.3. Informacja o programie *Gretl*

Oprogramowanie *Gretl* (*GNU Regression Econometrics Time-Series Library*) opracowane jest w wersji polskiej. Dostępne jest na stronie internetowej http://www o symbolu *gretl.source.net* lub *kufel.torun.pl*. Obejmuje podstawowe tematy z przedmiotu ekonometria oraz prognozowanie i symulacje. Poświęcona jemu przez Kufla publikacja obejmuje szereg tematów dotyczących możliwości jego menu na bazie plików danych oferowanych w Internecie przez tego autora. Oprogramowanie wydane jest przez Fundację Wolnego Oprogramowania (*Free Software Foundation*) jest ogólnie bezpłatnie dostępne. Może być przez użytkowników modyfikowany jego kod źródłowy, jednak ryzyko w zakresie skuteczności działania funkcji ponosi sam użytkownik. Oferowana jest wersja *Gretl*-a o numerze 1.6.5 i wyższe. Procedurę instalacji pliku *gretl.install.exe* oraz plików z danymi pochodzącymi z podręczników ekonometrii podano w rozdziale 1.2 wymienionej publikacji Kufla, przy czym menu programu obejmuje:

plik - operacje na plikach;
narzędzia - funkcje odczytu wartości krytycznych oraz prezentacji graficznej rozkładów;
dane - operacje umożliwiające edycję zmiennych dla potrzeb modelu;
widok - zestaw funkcji obejmujących między innymi budowę wykresów, szacowanie statystyk;
dodawanie zmiennych - tworzenie nowych zmiennych na bazie istniejących;
próba - określenie zakresu obserwacji z pliku danych;
zmienna - udostępnianie statystyk opisowych, rozkładów i ich zgodności;
model - obszerny zestaw metod estymacji modeli ekonometrycznych, szeregów czasowych, nieliniowych, równań współzależnych, opartych o dane panelowe.

Oprogramowanie *Gretl* ma dostęp do serwera *ricardo.ecn.wfu.edu* poprzez menu:

Plik/Pliki baz danych/instalowanie baz danych (vide rozdział 2.10 publikacji Kufla).

Aby pracować z *Gretl*-em pozyskany plik należy zapisać na dysku własnego komputera. W ramach programu *Gretl* występują przykłady modeli ekonometrycznych, a dla części z nich zestaw poleceń - komend *Gretl*-a. Wbudowano również tablice statystyczne rozkładów, a mianowicie:

normalnego,
t-Studenta,
chi-kwadrat (χ^2),
F-Fishera,
Durbina-Watsona.

Ułatwieniem jest kalkulator testów statystycznych pozwalający wyznaczyć wielkość statystyki testowej i empiryczne poziomy istotności dla weryfikacji modelu. Wprowadzane informacje o parametrach dla kalkulatora testów są wartościami liczbowymi.

Oprogramowanie *Gretl*-a w zakresie ekonometrycznych modeli dla danych przekrojowych umożliwia zarówno dobór zmiennych jak i estymację parametrów klasyczną metodą najmniejszych kwadratów (KMNK) oraz przeprowadzenie etapów weryfikacji modelu. Dla doboru zmiennych objaśniających stosuje się funkcje wyznaczania macierzy korelacji. Oprócz KMNK w programie wykorzystuje się także inne metody estymacji, część z nich jest modyfikacją metody najmniejszych kwadratów. W obecnej wersji *Gretl*-a występują także modele do budowy i weryfikacji modeli bazujących na danych panelowych. Bardzo pomocne są informacje z procesu estymacji oraz specjalne funkcje *Gretl*-a do weryfikacji modelu, a służące do oceny:

istotności parametrów strukturalnych,
stopnia dopasowania modelu,
normalności rozkładu reszt,
jednorodności wariancji reszt,
liniowości postaci analitycznej,
współliniowości zmiennych objaśniających,
wpływowych obserwacji.

Pomocną cechą programu *Gretl* jest zapisanie wyników modelowania, grafiki i wykonywanych komend w formie pliku sesji. Analizę ekonometryczną procesów ekonomicznych ułatwiają uzyskane z tego programu charakterystyki, a w tym funkcja autokorelacji i korelacji cząstkowej. W pracy Kufla na uwagę zasługuje nawiązanie krótkiego wyjaśnienia teoretycznego podstawowych modeli procesów ekonomicznych z przedstawieniem na konkretnym pliku danych odpowiedniej funkcji *Gretl*-a. Daje to możliwość zrozumienia raportów programu z realizacji budowy modeli: wielomianowych trendu, wahań sezonowych, autoregresyjnych, bazujących na danych panelowych.

Na podstawie modelu trendu i sezonowości przy skorzystaniu z funkcji *Analiza/Prognoza* możemy wyznaczyć prognozy i oszacować błędy predykcji dla każdego prognozowanego okresu. Gdy nie są spełnione założenia o stosowaniu KMNK tj. występuje autokorelacja reszt lub heteroskedastyczność wariancji (*niestałość*) to wówczas stosowana jest uogólniona metoda najmniejszych kwadratów (UMNK). W warunkach autokorelacji reszt program Gretl proponuje w ramach modeli szeregów czasowych metody: Cochran'a-Orcutta, Hirdertha-Lu, Praisa-Winstena.

Istnienie odstających (*nietypowych*, *wpływowych*) obserwacji jest przyczyną występowania heteroskedastyczności wariancji resztowej. Nie spełnienie warunku stałości wariancji oceniamy stosując KMNK i test White'a, posługując się funkcją *Test/test heteroskedastyczności Gretl*-a. Gdy wynik testu wskaże na istotność zmienności wariancji to wówczas stosujemy UMNK dla lepszej estymacji parametrów. W wersji polskiej o numerze 1.6.5 można rozwiązywać również modele nieliniowe, a w ramach nich: logitowe, probitowe, tobitowe (określone na podstawie modeli logitowych i probitowych).

Godną uwagi jest możliwość określania wielowymiarowych modeli ekonometrycznych, które charakteryzują wiele zjawisk ekonomicznych. Do estymacji równań tych modeli służy podwójna metoda najmniejszych kwadratów (2MNK) wywoływana z funkcji *Modele/Inne liniowe modele/Podwójna metoda najmniejszych kwadratów*.

Niekiedy brakuje teoretycznych podstaw do wskazywania współzależności procesów, a więc rodzą się wątpliwości co do celowości modelowania strukturalnego. Na kanwie tego niepokoju powstały modele *VAR*, czyli wielorównaniowe w których zestaw zmiennych objaśniających obejmuje procesy opóźnione w czasie. Modele te były doskonalone poprzez wprowadzenie trendu i sezonowości. Dla oszacowania modelu VAR korzystamy z menu *Model/Modele szeregów czasowych/model wektorowej autoregresji*. Z danymi panelowymi, przykładowo przestrzenno-czasowymi mamy do czynienia, gdy obserwujemy dane na kilku obiektach, obszarach i okresach czasowych. Paweł Kufel jako przykład podaje „*...badanie ekonomicznej efektywności ludności ..., stopy bezrobocia w powiatach w Polsce w latach 1998÷2002*" oraz rozwiązuje ten przykład w Gretl (rozdział 12 publikacji podanej w przypisie …).

11.4. Dopasowanie modelu do danych empirycznych

Program Gretl umożliwia szybkie otrzymywanie statystyki opisowej w odniesieniu do zmiennej Y jak też do zmiennych objaśniających $X_1 \div X_{11}$, a w ramach niej określenie średniej, mediany, wartości minimalnej oraz maksymalnej, odchylenia standardowego, współczynnika zmienności, skośności i kurtozy. Na uwagę zasługuje współczynnik zmienności w %:

$$V_j = \left| \frac{S_j}{\bar{x}_j} \right| * 100$$

Wyznaczenie V_j stanowi wstępny krok zbadania wystarczającego zróżnicowania danej *j*-tej zmiennej objaśniającej. Określany jest jako moduł z ilorazu odchylenia standardowego i średniego poziomu wartości rozpatrywanej zmiennej.

Tab. 11.3. Współczynniki zmienności

Zmienna	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}
V_j	226	65	153	106	102	79	88	50	59	104	56

Źródło: Opracowanie własne w *Gretl*.

Z tabeli wynika, że dla danych panelowych współczynnik zmienności znacznie przekracza wartość krytyczną $V^* = 10\%$, czyli wszystkie zmienne pełnego panelu o 315 obserwacjach wchodzą wstępnie do budowy modelu ekonometrycznego.

W selekcji zmiennych objaśniających możemy następnie posłużyć się jedną z metod doboru zmiennych objaśniających, przykładowo analizą korelacji. Niezbędna w tym zakresie jest tablica współczynników korelacji między zmienną objaśnianą **Y** a zmiennymi objaśniającymi X_j, jak również współczynników korelacji wzajemnych miedzy zmiennymi objaśniającymi (patrz tabela 11.4). Skorzystano w tym względzie z *Gretl*-a, który numeruje obserwacje według kolejności obiektów - podregionów, a w ramach tego kolejne okresy czasowe (1:1 - 45:7).

Tab. 11.4. Współczynniki korelacji

Y	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	Y
1	**0,7994**	-0,3757	0,2317	0,6465	-0,2653	0,2443	0,7321	-0,2088	0,6093	**-0,0020**	-0,5392	Y
	1	-0,5590	0,1954	0,7204	-0,1893	0,2158	0,6571	-0,2385	0,5230	-0,0627	-0,6477	X_1
		1	0,0206	-0,3258	0,0075	-0,2053	-0,3964	-0,0760	-0,0874	0,2933	0,2210	X_2
			1	0,0651	0,0700	0,1299	0,2747	-0,0440	0,5859	0,1063	0,0042	X_3
				1	-0,4647	0,1039	0,3960	-0,1952	0,4664	0,1166	-0,6063	X_4
					1	0,1777	0,1373	0,2919	-0,1017	-0,0747	0,5152	X_5
						1	0,6309	0,5659	0,0389	0,0220	-0,0550	X_6
							1	0,3115	0,4366	0,1074	-0,2461	X_7
								1	-0,2591	0,2885	0,4271	X_8
									1	0,1298	-0,3273	X_9
										1	0,1441	X_{10}
											1	X_{11}

Źródło: Opracowanie własne w *Gretl*.

Program *Gretl* oblicza także wartość krytyczną (przy dwustronnym 5% obszarze krytycznym) która wynosi $r^* = 0{,}1105$ dla $n = 315$ obserwacji. Wartość tą możemy także obliczyć ze

$$r^* = \sqrt{\frac{t_\alpha^{\,2}}{t_\alpha^{\,2} + n - 2}} = \sqrt{\frac{(1{,}979)^2}{(1{,}979)^2 + 315 - 2}} = 0{,}1105$$

wzoru:

przy czym:

$t_a = 1{,}979$ - statystyka wyznaczona na podstawie rozkładu *t*-Studenta dla poziomu istotności $a = 0{,}05$ oraz ilości stopni swobody $iss = n - 2 = 315 - 2 = 313$.

Przeglądamy powyższą tabelę i sprawdzamy, czy współczynnik korelacji

$$\left|r_j\right| \le r^* = 0{,}1105$$

Wg tego kryterium możemy wyeliminować zmienną X_{10} o r_{10} = 0,0020 ze zbioru zmiennych objaśniających. Zmienna najsilniej skorelowaną ze zmienną Y jest zmienna X_1 o r_1 = 0,7994. Możemy posłużyć się jeszcze innymi manualnymi metodami, lecz rozpatrzenie wszystkich kombinacji zmiennych jest pracochłonne. Warto jednak choć wymienić te metody, a więc:

metoda wskaźników zmiennych objaśniających,
bazowanie na miarze błędów *ex ante*,
metoda pojemności nośników informacji,
metoda grafowa i inne.

Wróćmy jednak do naszego programu *Gretl*. Pierwszym krokiem jest przygotowanie danych panelowych. W tym względzie dogodnie jest uczynić to w *Excelu* korzystając z jednego arkusza skoroszytu. Oznaczamy zmienne począwszy od komórki A1 w pierwszym wierszu, a więc wpisujemy Y i w kolejności $X_1 \div X_{11}$. Warto zaznaczyć, że znak precyzji przecinek jest akceptowalny przez *Gretl* (w wersji programu z grudnia 2008 roku). Program ten jest prawie co kwartalnie modyfikowany oraz rozszerzany przez ekonometryków w miarę rozwoju metod ekonometrycznych. Import pliku *Excela* do *Gretl*-a następuje po kolejnym wywołaniu opcji menu:

plik; otwórz dane; import; excel; wybranie pliku (w tym przykładzie ***Dane-gretl.xls*** *z podkatalogu na dysku D:).*

Zastosowanie klasycznej metody najmniejszych kwadratów (KMNK) przy formułowaniu modelu na danych panelowych jest ograniczone ze względów merytorycznych. Zdaniem T. Kufla (str. 167 wymienionej wcześniej publikacji) możemy zastosować KMNK, gdy wprowadzimy zmienne zero-jedynkowe do poszczególnych 45 obiektów (podregionów) panelu o 315 obserwacjach, przy czym 1 - oznacza uwzględnienie danego okresu czasowego jednostki panelu (podregionu), 0 - pominięcie.

Zabieg ten stosowany jest także przez innych autorów między innymi w metodzie Kleina w której zastosowano sezonową zmienną zero-jedynkową. W naszym przypadku umożliwia uzyskanie efektów indywidualnych b_{FE} w odniesieniu do poszczególnych podregionów. Wcześniej jednak musimy dokonać dołączenia „*podmacierzy*" zero-jedynkowej do macierzy zmiennych objaśniających. Dokonujemy to poprzez polecenie: Dodawanie zmiennych/zmienne 0-1 jednostek panelu.

Kolejny krok to polecenie *Model/Klasyczna metoda najmniejszych kwadratów*. W specyfikacji zmiennych określamy zmienną objaśnianą (zależną) oraz zmienne objaśniające (niezależne) $X_1 \div X_{11}$ bez stałej „*const*", gdyż jej rolę przejmą efekty indywidualne odniesione do podregionów. Moglibyśmy już teraz zrezygnować z zmiennej X_{10}, której nieistotność wykazała wstępna analiza korelacji. Efektem pracy programu *Gretl* jest poniższy *model1* stanowiący estymację danych panelowych z wykorzystaniem 315 obserwacji. Włączono do niego 45 jednostek danych przekrojowych (podregionów) o szeregach czasowych długości 7 (lat).

Tab. 11.5. Parametry *modelu1*

Zmienna	Współczynnik	Błąd standardowy	t-Student	Wartość p	
X1	-9,43147	1,93873	-4,8648	<0,00001	***
X2	2489,94	398,217	6,2527	<0,00001	***
X3	-930,436	235,347	-3,9535	0,00010	***
X4	-4,38485	3,88392	-1,1290	0,25995	
X5	-33,9051	3,35581	-10,1034	<0,00001	***
X6	-8,46421	5,28443	-1,6017	0,11043	
X7	77,376	16,2766	4,7538	<0,00001	***
X8	-71,8521	16,6996	-4,3026	0,00002	***
X9	403,339	157,835	2,5554	0,01118	**
X10	-141,224	85,0459	-1,6606	0,09801	*
X11	-36,3181	6,24222	-5,8181	<0,00001	***
du_1	49802,3	3862,25	12,8946	<0,00001	***
du_2	44621,4	3512,42	12,7039	<0,00001	***

du_3	61595,6	7245,92	8,5007	<0,00001	***
du_4	42305,9	3409,37	12,4087	<0,00001	***
du_5	43383,3	4412,54	9,8318	<0,00001	***
du_6	44168,9	4056,78	10,8877	<0,00001	***
du_7	33819,9	2800,85	12,0749	<0,00001	***
du_8	70167,8	14011,9	5,0077	<0,00001	***
du_9	34797,3	3020,25	11,5214	<0,00001	***
du_10	24652,8	3859,02	6,3884	<0,00001	***
du_11	50856,7	6827,43	7,4489	<0,00001	***
du_12	33244	2430,98	13,6752	<0,00001	***
du_13	28033,7	2851,16	9,8324	<0,00001	***
du_14	29606	7143,92	4,1442	0,00005	***
du_15	20550,4	2646,26	7,7659	<0,00001	***
du_16	24720,8	2673,92	9,2451	<0,00001	***
du_17	40802,2	3751,68	10,8757	<0,00001	***
du_18	43016,9	4161,7	10,3364	<0,00001	***
du_19	29115,1	2902,21	10,0321	<0,00001	***
du_20	30041,1	2618,63	11,4720	<0,00001	***
du_21	48196,6	4829,41	9,9798	<0,00001	***
du_22	23776,5	2692,25	8,8315	<0,00001	***
du_23	48567,1	3937,48	12,3346	<0,00001	***
du_24	24855,6	2414,78	10,2931	<0,00001	***
du_25	28965,3	2614	11,0808	<0,00001	***
du_26	24208,9	2502,71	9,6731	<0,00001	***
du_27	45489,9	3911,89	11,6286	<0,00001	***
du_28	35148,1	3013,94	11,6618	<0,00001	***
du_29	27726	2261,49	12,2601	<0,00001	***
du_30	56659,7	6380,39	8,8803	<0,00001	***
du_31	51356,1	5468,97	9,3905	<0,00001	***
du_32	37517,2	4641,81	8,0824	<0,00001	***
du_33	47212,5	4257,06	11,0904	<0,00001	***
du_34	30772,7	2721,12	11,3089	<0,00001	***
du_35	29420,5	2170,14	13,5570	<0,00001	***
du_36	52595,5	6204,47	8,4770	<0,00001	***
du_37	38651,7	3810,65	10,1431	<0,00001	***
du_38	39295	3380,15	11,6252	<0,00001	***
du_39	41995,9	3660,14	11,4739	<0,00001	***
du_40	25112	3581,53	7,0115	<0,00001	***
du_41	33129,6	3830,45	8,6490	<0,00001	***
du_42	45174,2	5783,76	7,8105	<0,00001	***
du_43	25566,6	2814,73	9,0831	<0,00001	***
du_44	33930,2	3375,82	10,0510	<0,00001	***
du_45	19498,7	2710,71	7,1932	<0,00001	***

Źródło: Opracowanie własne w *Gretl*.

Przeglądając powyższy wydruk *modelu1* widzimy poszczególne parametry zwane współczynnikami odnoszące się do zmiennych objaśniających, a także stanowiące efekty indywidualne stojące przy zmiennych zero-jedynkowych (*du*_1 do *du*_45. Na wydruku występuje też określenie błędu standardowego oszacowania danego parametru i obliczenie statystyki *t*-Studenta według wzoru:

$$t_{obl} = \frac{a_i}{D(a_{i)}}$$

gdzie:

a_j - parametr j-tej zmiennej (współczynnik), $D(a_j)$ - błąd standardowy oszacowanego parametru.

Program sprawdza warunek:

$$|t_{obl}| > t_{\alpha}$$

przy czym: t_a - wartość krytyczna wg rozkładu t -Studenta dla poziomu istotności a.

Jeżeli warunek jest spełniony to dany parametr jest statystycznie istotny na poziomie istotności a. W tym względzie program wprowadza dogodne oznaczenia informujące:

*** - α = 0,01; ** - α = 0,05; *- α = 0,1; puste - brak statystycznej istotności (a > 0,1).

Dużym udogodnieniem jest informacja na zakończenie wydruku o treści: *Wyłączając stałą, największą wartość p jest dla zmiennej* X_4. Uwaga ta wskazuje zmienną nieistotną statystycznie dla której prawdopodobieństwo błędu jest największe i wynosi p = 0,26, a statystyka t_{obl} = 1,129, czyli jest najmniejsza. Teraz możemy przystąpić w *Gretl* do usuwania zmiennych nieistotnych statystycznie posługując się w kolejnych krokach eliminacją *a posteriori*. W tej eliminacji kierujemy się wskazaniem zmiennej na zakończeniu wydruku z kolejnego modelu. Tak postępując otrzymano *model2*, gdzie *Gretl* zaproponował usuniecie X_6, *model3* z sugestią eliminacji zmiennej X_{10}, a więc takiej jaką wskazała wcześniej wstępna analiza korelacji. Ostatnim jest zamieszczony niżej *model4*, który nie zawiera już propozycji eliminacji zmiennej objaśniającej. Tylko przy zmiennej X_9 występują dwie gwiazdki co świadczy o oszacowaniu na poziomie istotności a = 0,05.

Tab. 11.6. Parametry *modelu4*

Zmienna	Współczynnik	Błąd standardowy	t-Student	Wartość p	
X1	-9,96098	1,93834	-5,1389	<0,00001	***
X2	2583,1	380,316	6,7920	<0,00001	***
X3	-1031,14	226,898	-4,5445	<0,00001	***
X5	-34,3597	3,33367	-10,3069	<0,00001	***
X7	69,1179	15,9021	4,3465	0,00002	***
X8	-72,249	16,6168	-4,3480	0,00002	***
X9	402,411	158,493	2,5390	0,01170	**
X11	-40,4479	5,85304	-6,9106	<0,00001	***
du_1	50753,3	3714,09	13,6651	<0,00001	***
du_2	44862,5	3323,07	13,5003	<0,00001	***
du_3	61411,8	7225,55	8,4993	<0,00001	***
du_4	42936,8	2996,06	14,3311	<0,00001	***
du_5	44576,3	3949,65	11,2861	<0,00001	***
du_6	45208,5	3866,8	11,6914	<0,00001	***
du_7	32896,1	2342,31	14,0443	<0,00001	***
du_8	71474,5	13955,7	5,1215	<0,00001	***
du_9	35011	2782,9	12,5807	<0,00001	***
du_10	18970	2369,25	8,0068	<0,00001	***
du_11	48660	6585,54	7,3889	<0,00001	***
du_12	32524,9	2276,87	14,2849	<0,00001	***
du_13	24479,9	2284,64	10,7150	<0,00001	***
du_14	28044,6	7040,7	3,9832	0,00009	***
du_15	18769,1	2046,03	9,1734	<0,00001	***
du_16	23874,4	2237,98	10,6678	<0,00001	***
du_17	41493,7	3506,6	11,8330	<0,00001	***
du_18	43829,1	3835,53	11,4271	<0,00001	***
du_19	28862,3	2620,12	11,0156	<0,00001	***
du_20	29258,1	2424,43	12,0680	<0,00001	***
du_21	49117	4554,93	10,7833	<0,00001	***
du_22	23666,6	2230,06	10,6125	<0,00001	***
du_23	49287,8	3742	13,1715	<0,00001	***
du_24	22743,1	1583,36	14,3638	<0,00001	***
du_25	26977,1	1955,16	13,7978	<0,00001	***
du_26	23167,7	1807,71	12,8160	<0,00001	***
du_27	44968,6	3744,55	12,0091	<0,00001	***
du_28	35487,8	2671,93	13,2817	<0,00001	***

du_29	26708,8	1799,86	14,8393	<0,00001	***
du_30	54632,3	5943,02	9,1927	<0,00001	***
du_31	43981,4	3638,11	12,0891	<0,00001	***
du_32	29930,8	2341,18	12,7845	<0,00001	***
du_33	42608,8	3484,75	12,2272	<0,00001	***
du_34	29055,6	1715,72	16,9349	<0,00001	***
du_35	29463,8	1949,56	15,1130	<0,00001	***
du_36	51455,4	5980,38	8,6040	<0,00001	***
du_37	38584,2	3429,38	11,2511	<0,00001	***
du_38	37931,2	2986,35	12,7015	<0,00001	***
du_39	41479,5	3285,02	12,6269	<0,00001	***
du_40	19387,4	1980,46	9,7894	<0,00001	***
du_41	28157,9	2750,7	10,2366	<0,00001	***
du_42	41870,1	5251,86	7,9724	<0,00001	***
du_43	24148,1	2191,24	11,0203	<0,00001	***
du_44	31177,6	2707,83	11,5139	<0,00001	***
du_45	17183,1	1825,61	9,4122	<0,00001	***

Źródło: Opracowanie własne w *Gretl.*

Warto jeszcze zwrócić uwagę na miary pomocnicze dobroci dopasowania modelu do specyficznych danych empirycznych jakimi są dane panelowe (przestrzenno-czasowe) , a należą do nich:

współczynnik determinacji R^2,
skorygowany współczynnik determinacji $\bar{R}^2$,
kryterium informacyjne Akaike'a (AIC),
kryterium Schwarza (BIC).

Tab. 11.7. Porównanie wybranych cech statystyki opisowej modeli

	Model1	*Model2*	*Model3*	*Model4*	MIN
R^2	0,94794	0,94769	094720	0,94669	Model4
R^{-2}	0,93689	0,93682	0,93647	0,93611	Model 4
AIC	5825,00	5824,55	5825,49	5826,49	Model2
BIC	6035,15	60,30,94	6028,13	6025,38	Model4

Źródło: Opracowanie własne.

Najlepszym według czterech kryteriów jest *model4*, czyli taki jaki uzyskaliśmy w *Gretl* poprzez eliminację *a posteriori*. Na wydruku *modelu4* znajduje się także statystyka *F*, która umożliwia całościową ocenę przydatności zaprojektowanego modelu ekonometrycznego:

Statystyka F (52, 262(= 89,4765 (wartość p < 0,00001)

Mała wartość prawdopodobieństwa błędu przemawia za słusznością zróżnicowania doboru parametrów dla podregionów. Pozostaje nam jeszcze ocena wygasłych już prognoz stosując miarę MAPE (vide tabela 11.8) dla przykładowego województwa łódzkiego i mazowieckiego.

Tab. 11.8. Obliczenia pomocnicze do obliczenia miary błędu MAPE

Lp.	Region/Województwo/	Rok	Okres	y	y^	y - y^	I(y - y^)/yI *100	b_i	Średnia z kol. 8
1	2	3	4	5	6	7	8	9	10
1	**centralny/łódzkie/**	2000	1	11442	12804,6	-1362,6	11,9	50753,3	7,6
	łódzki	2001	2	12790	11895,7	894,3	7,0		
		2002	3	14350	15466,4	-1116,4	7,8		
		2003	4	15259	15811,3	-552,3	3,6		
		2004	5	15955	16960,2	-1005,2	6,3		
		2005	6	18713	16548,2	2164,8	11,6		
		2006	7	19837	18854,7	982,3	5,0		
		2006	7	19837	18854,7	982,3	5,0		
2	piotrkowsko-skierniewicki	2000	1	12254	12512,3	-258,3	2,1	44862,5	6,2
		2001	2	13478	10878,7	2599,3	19,3		
		2002	3	15289	17170,3	-1881,3	12,3		
		2003	4	16060	16471,3	-411,3	2,6		
		2004	5	17034	17241,1	-207,1	1,2		

		2005	6	20040	19371,4	668,6	3,3		
		2006	7	21657	22164,0	-507,0	2,3		
3	miasto Łódź	2000	1	20645	23161,7	-2516,7	12,2	61411,8	5,7
		2001	2	21767	24006,2	-2239,2	10,3		
		2002	3	23924	24079,7	-155,7	0,7		
		2003	4	25072	24790,4	281,6	1,1		
		2004	5	26531	26424,4	106,6	0,4		
		2005	6	29104	26694,6	2409,4	8,3		
		2006	7	30602	28489,2	2112,8	6,9		
4	**.../mazowieckie/**	2000	1	14676	18309,9	-3633,9	24,8	42936,8	13,9
	ciechanowsko-płocki	2001	2	17371	17949,9	-578,9	3,3		
		2002	3	18032	21054,6	-3022,6	16,8		
		2003	4	18036	20813,4	-2777,4	15,4		
		2004	5	19800	19775,5	24,5	0,1		
		2005	6	26813	21188,3	5624,7	21,0		
		2006	7	27751	23393,1	4357,9	15,7		
5	ostrołęcko-siedlecki	2000	1	9842	11375,7	-1533,7	15,6	44576,3	6,9
		2001	2	11449	10302,7	1146,3	10,0		
		2002	3	14008	14739,0	-731,0	5,2		
		2003	4	14009	14357,1	-348,1	2,5		
		2004	5	14200	14763,5	-563,5	4,0		
		2005	6	16930	16338,5	591,5	3,5		
		2006	7	18683	17237,7	1445,3	7,7		
6	warszawski	2000	1	14089	15899,6	-1810,6	12,9	45208,5	6,5
		2001	2	17366	15487,3	1878,7	10,8		
		2002	3	20040	18393,1	1646,9	8,2		
		2003	4	20346	20190,8	155,2	0,8		
		2004	5	20520	21423,5	-903,5	4,4		
		2005	6	22949	24427,3	-1478,3	6,4		
		2006	7	25545	25039,9	505,1	2,0		
7	Radomski	2000	1	9789	12635,8	-2846,8	29,1	32896,1	9,5
		2001	2	11827	10626,9	1200,1	10,1		
		2002	3	14879	15652,4	-773,4	5,2		
		2003	4	14680	14607,2	72,8	0,5		
		2004	5	15464	16204,5	-740,5	4,8		
		2005	6	17546	16340,2	1205,8	6,9		
		2006	7	19020	17140,9	1879,1	9,9		
8	miasto Warszawa	2000	1	48217	53589,8	-5372,8	11,1	71474,5	4,1
		2001	2	52630	51547,6	1082,4	2,1		
		2002	3	58449	61157,8	-2708,8	4,6		
		2003	4	59196	59245,8	-49,8	0,1		
		2004	5	62896	63332,3	-436,3	0,7		
		2005	6	68140	67131,0	1009,0	1,5		
		2006	7	77001	70531,4	6469,6	8,4		
		Średnia dla regionu centralnego w %:					**7,5**		

W powyższej tabeli: $y^{\wedge}$ - wartość teoretyczna obliczona z uwzględnieniem parametrów i wartości zmiennych objaśniających, b_i - efekt indywidualny podregionu.

W dotychczasowych rozważaniach przyjęliśmy *a priori*, że modelem adekwatnym jest model oparty o efekty indywidualne podregionów. Może jednak wystąpić odniesienie do efektów losowych lub wystarczającym jest model bazujący na KMNK i stałej *const*. Aby to rozstrzygnąć program *Gretl* proponuje diagnostykę realizowaną poleceniem: *Testy/test diagnostyczny panelu.* Test ten przeprowadzamy po realizacji modelu wykonanego KMNK na pełnych danych panelowych, tj. 45 jednostek w przekroju dla 7. okresów i stałej *const,* a ponadto przy występowaniu wszystkich potencjalnych zmiennych objaśniających $X_1 \div X_{11}$. Program *Gretl* proponuje wówczas następujący model:

const:	37647	(3134,2)	[0,00000]
X1:	-9,4315	(1,9387)	[0,00000]
X2:	*2489,9*	*(398,22)*	*[0,00000]*
X3	-930,44	(235,35)	[0,00010]
X4:	-4,3849	(3,8839)	[0,25995]
X5:	-33,905	(3,3558)	[0,00000]
X6:	-8,4642	(5,2844)	[0,11043]
X7:	77,376	(16,277)	[0,00000]
X8:	-71,852	(16,7)	[0,00002]
X9:	403,34	(157,84)	[0,01118]
X10:	*-141,22*	*(85,046)*	*[0,09801]*
X11:	-36,318	(6,2422)	[0,00000]

Model obejmuje współczynniki, błędy standardowe - w nawiasach, wartość *p* w nawiasach kwadratowych. *Gretl* odpowiada na pytania:

Czy model ustalony ze stałą *const* jest właściwszy niż z efektami indywidualnymi (nielosowymi) jednostek przekrojowych, czyli uwzględniający zróżnicowanie wyrazu wolnego?

Następuje badanie łącznej istotności nierówności średnich grupowych bazując na teście *F* o statystyce: *F(44, 259) = 11,5759 z wartością p 3,27261e-040* Niska wartość *p* oznacza odrzucenie hipotezy H_0, że model panelowy zbudowany wg klasycznej metody najmniejszych kwadratów jest poprawny, wobec hipotezy H_1, że model o ustalonych efektach jest właściwszy.

Czy model ustalony ze stałą *const* jest właściwszy niż z efektami losowymi?

W tym względzie program *Gretl* bazuje na teście Breuscha-Pagana o statystyce:

$$LM = 150{,}016 \text{ z wartością } p = prob(\chi^2(1) > 150{,}016) = 1{,}7198e\text{-}034$$

Biorąc pod uwagę niską wartość *p* należy odrzucić hipotezę H_0, że model wg KMNK jest poprawny, wobec hipotezy H_1, że model o losowych efektach jest właściwszy.

Mamy więc dylemat, który model efektów obrać nielosowe czy losowe. W podjęciu decyzji pomaga nam test Hausmana, proponowany w tym celu przez program *Gretl* o statystyce:

$$H = 75{,}631 \text{ z wartością } p = prob(\chi^2(11) > 75{,}631) = 1{,}02556e\text{-}011$$

Występuje tu niska wartość *p* co pozwala nam na odrzucenie hipotezy zerowej o modelu z losowymi efektami, wobec hipotezy alternatywnej o modelu z ustalonymi efektami. Tak więc możemy przypuszczać, że zbudowany model produktu brutto według cen bieżących w przeliczeniu na 1. mieszkańca dla podregionów w okresach rocznych jest poprawny.

*** * ***

W niniejszym opracowaniu pokazano możliwość szybkiego sformułowania także innych modeli miar ekonomicznych bazując na danych przestrzenno-czasowych. Może to być zalążkiem do sformułowania podstawowych charakterystyk rynków, a w przyszłości zbudowania pełnego modelu predykcji zjawisk ekonomicznych podregionów. Procedurę postępowania możemy także wykorzystać do formułowania modeli ekonometrycznych zjawisk ekonomicznych występujących w innych obiektach struktury terytorialnej, a mianowicie regionów, województw, powiatów, gmin.

12. Budowa i weryfikacja modelu ekonometrycznego w wybranych programach

12.1. Wstęp

Coraz powszechniejsze staje się stosowanie platformy programowej *R* do zagadnień statystyki i ekonometrii. Z tego względu pokazano możliwość sformułowania poleceniami języka *R* modelu ekonometrycznego bazującego na danych statystycznych zawartych w roczniku statystycznym GUS-u[171]. Ponadto przeprowadzono podstawowe przykładowe testy weryfikacji zbudowanego modelu. Oparto się na fragmencie obserwacji zmiennych z okresu 1992-2007 dotyczących liczby bezrobotnych (zmienna objaśniana) oraz zmiennych objaśniających przyjętych w modelowaniu tej zmiennej[172]. Szeregi zaktualizowano o dane z roczników ostatnich oraz poszerzono o obserwacje dotyczące lat 2008-2011 oraz oznaczono w notacji wygodnej do przetwarzania komputerowego (zob. tabela 12.1).

Tab. 12.1. Wartości obserwacji zmiennych

Rok	*t*	*X*3	*X*9	*X*10	*X*11	*X*12	*X*13	*X*14	*X*25	*X*38	*X*48	*X*49	*X*50	*X*51	*X*59	*X*78
1992	1	2394	3467,6	823,0	1042,1	794,7	476,7	287,7	295,16	138,9	404,70	1921,9	18430	4626,8	41,1	5,7
1993	2	2595	3391,8	721,9	999,4	746,5	381,1	239,3	384,25	134,8	550,29	1856,5	18206	4504,2	40,1	5,4
1994	3	2375	3361,4	678,8	996,1	728,1	366,4	235,4	515,49	134,5	708,30	1750,0	18206	4472,6	45,8	5,4
1995	4	2233	3461,1	689,2	1078,6	723,4	414,4	243,5	704,99	131,6	878,29	1648,3	18075	4775,1	59,2	5,4
1996	5	1961	3436,0	684,2	1119,7	721,9	425,3	250,5	900,58	126,7	1112,95	1565,0	19224	4258,9	70,5	5,3
1997	6	1737	3433,4	723,5	1244,8	730,7	487,4	273,3	1088,03	121,9	1300,00	1514,7	19771	4437,4	77,4	5,3
1998	7	1827	3378,7	736,9	1321,1	725,2	552,5	253,0	1277,06	115,7	1527,52	1452,6	20288	4243,5	83,3	5,4
1999	8	2641	3138,4	710,4	1318,4	686,3	583,0	286,9	1428,01	112,5	1700,54	1406,5	20800	4138,6	86,4	5,7
2000	9	2760	2955,0	661,9	1325,0	654,9	614,2	286,7	2622,76	111,1	1904,57	1360,3	21037	3750,8	78,4	5,5
2001	10	3186	2820,6	627,8	1295,6	530,4	637,3	252,1	2821,30	108,0	1897,60	1310,4	21376	3905,8	75,1	5,0
2002	11	3431	2670,5	545,5	1291,1	601,9	671,3	260,0	2930,70	102,6	1886,10	1284,3	21659	3708,1	85,7	5,0
2003	12	3329	2639,1	496,4	1308,0	586,3	699,5	261,8	3017,64	104,2	1918,71	1268,6	22052	3877,2	78,4	5,1
2004	13	3230	2663,1	453,1	1295,4	583,0	700,6	264,6	3111,04	104,0	1966,72	1229,1	22804	3779,1	98,0	5,0
2005	14	3045	2665,4	483,6	1360,7	577,7	695,8	265,2	3252,02	103,8	1954,20	1219,4	26565	3855,3	104,3	5,4
2006	15	2344	2714,3	511,5	1394,2	596,7	723,3	267,6	3421,05	104,9	2106,28	1221,5	27547	3967,0	88,8	5,9
2007	16	1619	2842,4	576,8	1482,8	621,0	780,3	277,0	3720,10	107,9	2290,78	1280,1	27713	3807,0	84,7	6,5
2008	17	1211	2890,7	671,2	1633,5	792,0	743,1	406,7	3986,50	110,1	2558,46	1212,2	28425	3792,2	80,6	6,8
2009	18	1411	2725,4	706,6	1611,5	777,5	740,1	407,4	4232,68	105,4	2706,47	1194,7	28684	3873,6	80,6	6,6
2010	19	1699	2696,1	674,4	1598,7	758,2	767,0	412,0	4388,31	103,9	2805,48	1197,9	29774	4237,3	81,5	5,9
2011	20	2146	2714,2	705,2	1580,3	787,5	775,1	412,2	4633,95	105,6	2872,24	1202,0	29383	3857,1	82,0	5,4

Źródło: Opracowanie własne na podstawie roczników statystycznych GUS-u (lata 1992-2011).

Ze względu na zmianę struktury (sekcje i działy) w roczniku 2011 wartości dla lat 2009, 2010, 2011 i zmiennych *X*12-*X*14 pogrupowano następująco:

*X*12 - transport, gospodarka magazynowa i łączność:

- transport i gospodarka magazynowa,
- informacja i komunikacja.

*X*13 - obsługa nieruchomości i firm:

- obsługa rynku nieruchomości,

[171] Opracowanie bazuje na rozdziale 2.15 książki autora pt. *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, Wydawnictwo Instytut Śląski, Opole 2015.

[172] *Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne*, red. naukowa T. Pokusa, artykuł W. Wornalkiewicz *Model ekonometryczny zmiennej „liczba bezrobotnych"*, Wydawnictwo Instytut Śląski, Opole 2011.

- zakwaterowanie i gastronomia,
- administrowanie i działalność wspierająca,
- pozostała działalność usługowa.

*X*14 - działalność usługowa komunalna, społeczna i indywidualna:

- działalność finansowa i ubezpieczeniowa,
- działalność związana z kultura, rozrywka i rekreacją.

Niniejsze opracowanie nawiązuje również do pracy autora niniejszego opracowania pt. „*Modele ekonometryczne PKB obiektów struktury terytorialnej*". Zestawiono w nim szeregi czasowe zmiennych objaśniających (*X*1-*X*79) przyjętych z poszczególnych rynków do badania ich wpływu na zmienną objaśnianą *Y* jaką jest produkt krajowy brutto, według cen bieżących w przeliczeniu na jednego mieszkańca[173]. Dane statystyczne zapisano w formie pliku arkusza kalkulacyjnego *Excel*, skąd zostały importowane do programu *Gretl* jako plik *Korelacja.gdt*. Plik ten po rozszerzeniu obserwacji o lata 2008-2011 w zakresie zmiennych *X*3, *X*9-*X*14, *X*25, *X*38, *X*48-*X*51, *X*59, *X*78 o nazwie *Korelacja-roz-X3.gdt* stanowił bazę do wstępnego sformułowania zmodyfikowanego modelu zmiennej *X*3 - „*Liczba bezrobotnych*" z zastosowaniem programu *Gretl*[174]. Dla ciągłości rozważań w tabeli 12.1 zachowano symbole wybranych zmiennych, a ponadto dla ewentualnego sprawdzenia wartości zmiennych wymieniono także „ścieżki" dojścia do danych w Rocznikach Statystycznych Rzeczypospolitej Polskiej[175]:

- rynek pracy/aktywność ekonomiczna ludności w wieku 15 i więcej - przeciętne w roku w tys./bezrobotni (*X*3);
- przeciętne zatrudnienie w tys.: przemysł (*X*9), budownictwo (*X*10), handel i naprawy (*X*11), transport, gospodarka magazynowa i łączność (*X*12), obsługa nieruchomości i firm (*X*13), działalność usługowa komunalna, społeczna i indywidualna, pozostałe (*X*14);
- koszty pracy w zł/ogółem na jednego zatrudnionego przeciętnie miesięcznie (*X*25);
- rynek płacy/wynagrodzenia w mln/wskaźnik - przeciętne miesięczne wynagrodzenie nominalne brutto - rok poprzedni = 100 (*X*38);
- budżety gospodarstw domowych/przeciętne miesięczne wydatki na 1 gospodarstwo domowe w zł (*X*48),
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych): zużycie wody z wodociągów w ciągu roku w hm^3 (*X*49), zużycie energii elektrycznej w ciągu roku w GWh (*X*50), zużycie gazu w sieci w ciągu roku w hm^3 (*X*51),
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci - policealne w tys. (*X*59),
- wskaźnik zawartych małżeństw na 1000 ludności (*X*78).

12.2. Budowa modelu ekonometrycznego w *Gretl*

Dla sformułowania modelu ekonometrycznego „*Liczby bezrobotnych*" - po rozszerzeniu bazy danych o wartości z lat 2008-2011 do 20 obserwacji, zapisanych jako plik *Gretl*-a *Korelacja-roz-X3.gdt* wywołujemy menu: *Model/klasyczna metoda najmniejszych kwadratów.* Następnie wskazujemy zmienną zależną (objaśnianą) *X*3 oraz zmienne objaśniające *X*9-*X*14, *X*25, *X*38, *X*48-*X*51, *X*59, *X*78. Rezultatem pracy programu jest raport *modelu1* zamieszczony w tabeli 12.2 obejmujący estymację parametrów modelu zwanych współczynnikami.

[173] Wornalkiewicz W., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, tab. 3.3, Wydawnictwo Instytut Śląski, Opole 2013.

[174] Początkowy model zaproponował autor w artykule *Model ekonometryczny zmiennej „liczba bezrobotnych"* w ramach monografii *Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne*, op. cit. s. 172.

[175] *Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne*, red. naukowa T. Pokusa, op. cit. s. 172.

Tab.12.2. Oszacowanie parametrów modelu *X3* dla 20 obserwacji (*model 1*)

Zmienna	*Współczynnik*	*Błąd stand.*	*Statystyka t*	*Wartość p*	
const	11037,4	6207,87	1,7780	0,13555	
*X*9	-5,18169	2,50564	-2,0680	0,09348	*
*X*10	2,30576	2,81819	0,8182	0,45048	
*X*11	2,05037	4,56698	0,4490	0,67225	
*X*12	-0,983977	3,34782	-0,2939	0,78063	
*X*13	0,599386	2,6107	0,2296	0,82751	
*X*14	-2,87111	8,04704	-0,3568	0,73581	
*X*25	-0,663503	0,506457	-1,3101	0,24712	
*X*38	82,6385	79,3045	1,0420	0,34513	
*X*48	0,250775	0,882549	0,2841	0,78769	
*X*49	0,143198	3,20708	0,0447	0,96611	
*X*50	-0,0689788	0,0748771	-0,9212	0,39920	
*X*51	-0,191685	0,423229	-0,4529	0,66959	
*X*59	14,6661	15,441	0,9498	0,38582	
*X*78	-604,484	302,113	-2,0009	0,10183	

Źródło: Opracowanie własne w programie *Gretl*.

Ponadto otrzymujemy również statystykę opisową obejmującą cechy:

- średnia arytmetyczna zmiennej zależnej = 2358,7;
- odchylenie standardowe zmiennej zależnej = 664,528;
- suma kwadratów reszt = 205105;
- błąd standardowy reszt = 202,536;
- współczynnik determinacji R^2 = 0,975555;
- skorygowany współczynnik determinacji = **0,907108;**
- statystyka F (14, 5) = 14,2527 (wartość p = 0,00419);
- logarytm wiarygodności = -120,734;
- kryterium informacyjne Akaike'a = 271,468;
- kryterium bayesowskie Schwarza = 286,404.

Na zakończenie raportu występuje podpowiedz, którą ze zmiennych usunąć dla poprawy modelu: „Wyłączając stałą największa wartość ***p*** jest dla zmiennej *X*49". Dalej w kolejnych iteracjach kierując się podpowiedziami *Gretl*-a usuwamy kolejno *X*49, *X*48, *X*13, *X*12, *X*51, *X*11, *X*14, *X*59, *X*38, *X*10, *X*50, czyli zmienne dla których prawdopodobieństwo popełnienia błędu jest duże. W efekcie końcowym procesu iteracyjnego uzyskano następujący model.

Tab. 12.3. Oszacowanie końcowe parametrów modelu

Zmienna	*Współczynnik*	*Błąd stand.*	*Statystyka t*	*Wartość p*	
const	16749	1061,19	15,7832	<0,00001	***
*X*9	-3,317	0,369783	-8,9701	<0,00001	***
*X*25	-0,668989	0,0910932	-7,3440	<0,00001	***
*X*78	-501,059	107,708	-4,6520	0,00027	***

Źródło: Opracowanie własne w programie *Gretl*.

Zwróćmy uwagę na wybrane elementy statystyki opisowej wskazujące na dobre dopasowanie modelu do danych empirycznych, a mianowicie:

- suma kwadratów reszt = 519489;
- błąd standardowy reszt = 180,189;
- współczynnik determinacji R^2 = 0,938085
- skorygowany współczynnik determinacji = **0,926476**;
- statystyka F (3, 16) = 80,8061 (wartość p < 0,00001);
- logarytm wiarygodności = -130,027;
- kryterium informacyjne Akaike'a = 268,055;
- kryterium bayesowskie Schwarza = 272,038.

W tabeli 12.2 oraz 12.3 warto zauważyć graficzne informowanie przez *Gretl* - w postaci gwiazdek o poziomie α istotności danego parametru: *** $\rightarrow \alpha = 0{,}01$; ** $\rightarrow \alpha = 0{,}05$; * $\rightarrow \alpha = 0{,}1$; puste $\rightarrow \alpha > 0{,}1$. Tak więc występuje niski poziom istotności - popełnienia błędu w oszacowaniu parametrów oraz wysoki skorygowany współczynnik determinacji $\bar{R}^2 =$ **0,926476**.

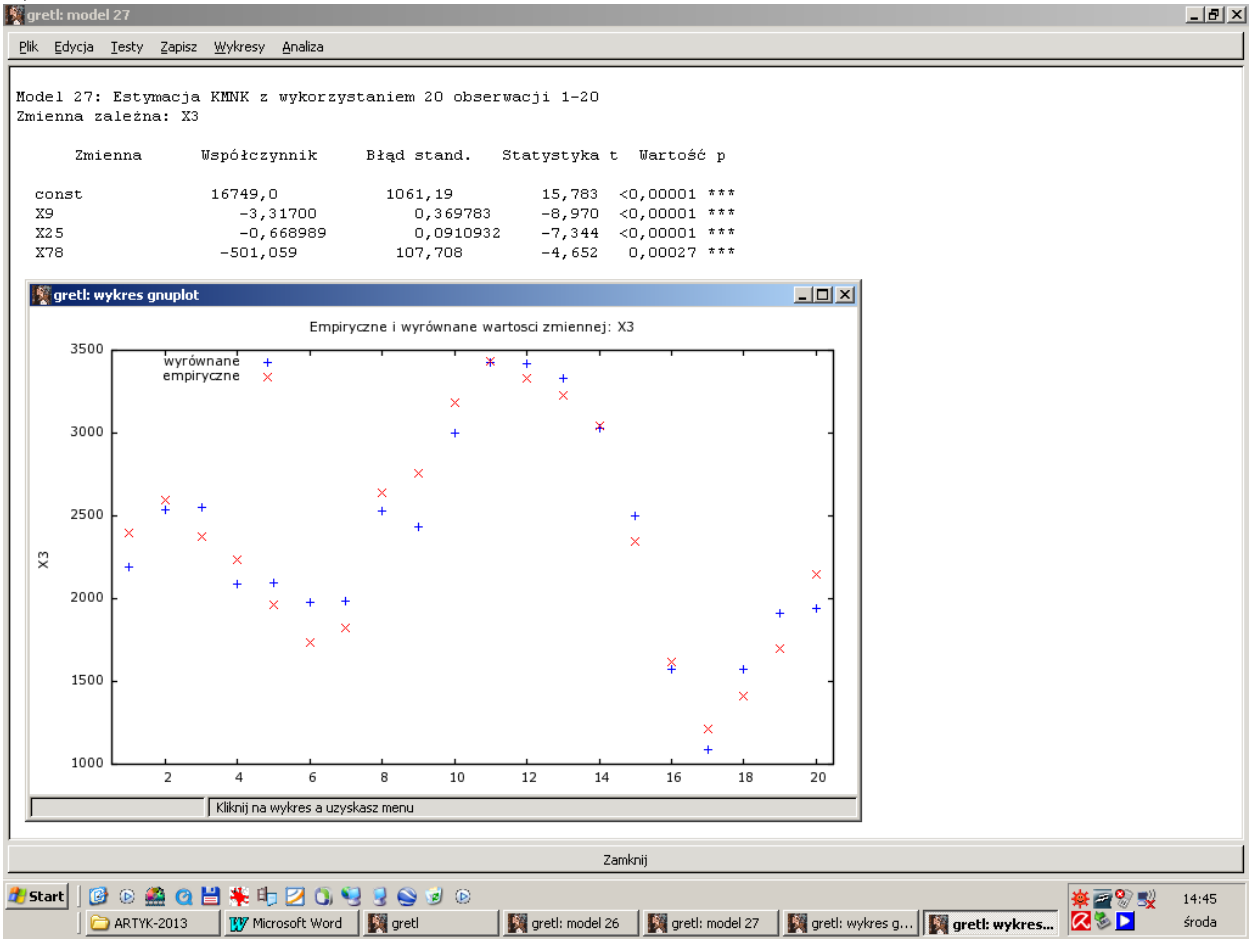
Wprowadzenie dodatkowych czterech dalszych obserwacji wpłynęło na istotną zmianę struktury modelu liniowego wielowymiarowego w stosunku do wcześniejszego przy 16. obserwacjach, który w tabeli cytowanego artykułu był następujący[176]:

$X_3\hat{} = 17753{,}3 - 4{,}33449X_9 + 3{,}19675X_{10} - 2{,}29688X_{13} - 0{,}88772X_{48} - 0{,}116366X_{50} + 16{,}1492X_{59}$

Warto tu przypomnieć zmienne występujące w naszym rozszerzonym modelu końcowym (zob. tabela 12.3):

- bezrobotni ($X3$),
- przeciętne zatrudnienie w tys: przemysł ($X9$),
- koszty pracy ogółem na 1 zatrudnionego - przeciętne miesięczne ($X25$),
- ilość zawartych małżeństw na 1000 ludności ($X78$).

Przy aktywnym modelu program *Gretl* umożliwia wykonanie wykresu empirycznych oraz wyrównanych - modelowych wartości zmiennej zależnej $X3$ względem 20 obserwacji (zob. rysunek 12.1).



Źródło: Opracowanie własne.

Rys. 12.1. Widok ekranu z raportem po estymacji parametrów oraz wykresem graficznym wartości empirycznych i modelowych

[176] Artykuł autora: *Model ekonometryczny zmiennej „Liczba bezrobotnych”*, op. cit. s. 178.

12.3. Przygotowanie do realizacji w programie *R*

Procedura postępowania mająca na celu sformułowanie pliku danych w postaci tzw. tabeli danych do potrzeb budowy modelu ekonometrycznego przykładowej zmiennej *X*3 jest następująca:

1. Skopiowanie tabeli 12.1 zamieszczonej w niniejszej pracy, a zapisanej w formacie Worda do skoroszytu *Excela* i nadanie mu nazwy *Dane.xls* umieszczając go na przykładowej ścieżce *D:\ARTYK-2013*.

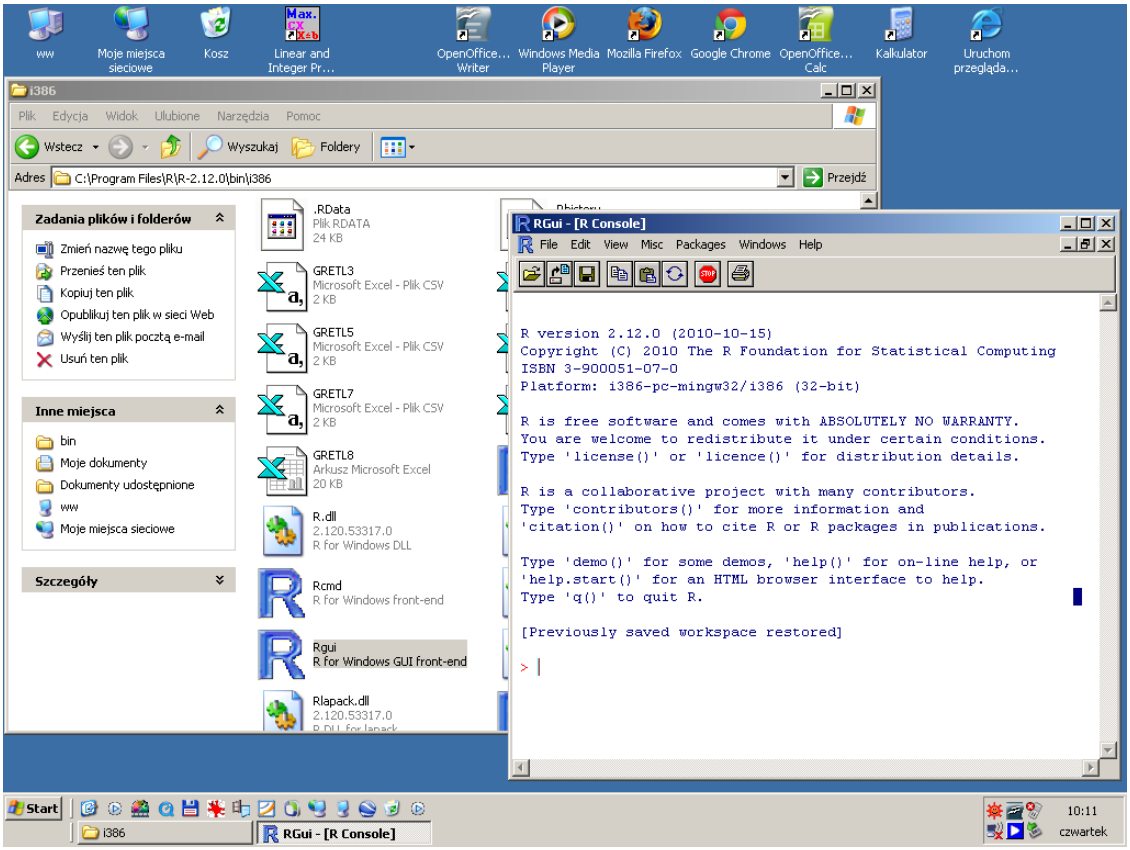
2. Pozostawienie tylko jednego arkusza w ramach skoroszytu i nadanie mu nazwy „*Dane-R*".

3. Zmniejszenie szerokości kolumn szeregów danych ze standardowej 8.43 na 7, celem zmieszczenia danych na ekranie.

4. Sformatowanie zmiennych *X*9-*X*11, *X*38, *X*48, *X*49, *X*51, *X*59, *X*78 z dokładnością do jednego miejsca po przecinku, a zmienną *X*25 dwa miejsca po przecinku. Zabieg ten wynika z faktu występowania supresji zer po skopiowaniu tabeli danych z *Worda* do *Excela*.

5. Usunięcie kolumn „*Rok*" i „*t*".

Mając już przygotowane dane przystępujemy do zainicjowania środowiska platformy programowej *R* zawierającej obszerną strukturę programów[177].



Źródło: Opracowanie własne w programie *R*.

Rys. 12.2. Obraz ekranów wywołania programu *Rgui* trybu okienkowego

Postępowanie w tym zakresie sprowadza się do pozyskania aktualnej wersji szeroko rozumianego programu *R* z Internetu i zapisanie go na ścieżce *D:\Program Files\R\R-2-12.0*[178]. W wyniku instalacji utworzone zostały podkatalogi: *bin* - z plikami wykonywalnymi; *doc* - zawierający

[177] Charakterystykę tego środowiska opisano w książce: *Wstęp do ekonometrii i badań operacyjnych* ..., wydanie II rozszerzone, Wornalkiewicz W., "*Wybrane modele ekonometryczne*", Wydawnictwo Instytut Śląski, Opole 2012, rozdz. 1.16.

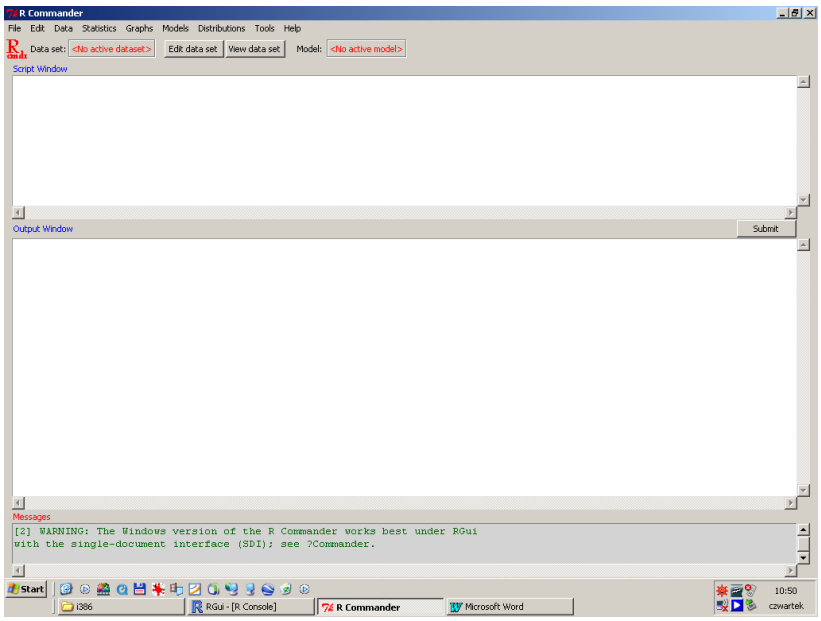
[178] Wersja R-2-12.0 była aktualna w 2012 roku w trakcie pisania wydania II rozszerzonego książki *„Wstęp do ekonometrii i badań operacyjnych ...,* op. cit.

dokumentację pakietu *R*; *library* - biblioteka z pakietami programowymi, które można później wywołać funkcją: *install.packages*.

Następnie przystępujemy do otwarcia podkatalogów *bin* i w kolejności *i386*. Tryb okienkowy wywołujemy otwierając program *Rgui.exe*. Zwróćmy uwagę na zamieszczony rysunek 12.2 z widocznymi fragmentami ekranów pulpitu, wywołanego trybu okienkowego *Rgui* (*R Console*) ze znakiem zachęty „>" do wprowadzania komend w języku *R*. Jednak wygodniejszy w początkującym kontakcie ze środowiskiem programowym *R* jest tryb nadzorowany zwany także windowsowym. Instalujemy go poleceniem - komendą: *library* (*Rcmdr*). Powoduje to przywołanie pakietu *Rcmdr* o ile jest on już wcześniej w bibliotece. Jeśli nie występuje to doinstalowujemy go z Internetu poleceniem: *install.packages („Rcmdr",dependencies=TRUE).* Opcja „*dependencies=TRUE*" oznacza pobranie również podpakietów niezbędnych do działania *Rcmdr*[179]. Alternatywnym rozwiązaniem jest skorzystanie z trybu okienkowego *Rgui* i przywołanie programu *Rcmdr* poprzez menu: *Packages/install package(s)...* Obraz ekranu trybu windowsowego z programem *Rcmdr* określony jak *R Commander* pokazano na rysunku 12.3. Zwróćmy uwagę na menu, które obejmuje między innymi grupy funkcji:

File - otwarcie i zapisanie zbioru poleceń tzw. zbioru scriptowego;
Edit - kopiowanie/wklejanie;
Data - założenie zbioru, załadowanie istniejącego, import z....;
Statistics - podstawowa statystyka opisowa, testy sumaryczne.
Graphs - możliwość wykonania wykresów różnego typu dla pojedynczej zmiennej jak i grupowych;
Models - analiza regresji oraz *Distributions* - tablice statystyczne, *Tools* - ładowanie pakietu i różne opcje;
Help - opis do poleceń programowych.

Na rysunku 12.3 widzimy trzy podokna opisowe: *Script Window* - skryptowe (tekstów poleceń), *Output Window* - realizacji - raportów programu na daną sekwencje poleceń, *Messages* - komunikatów - informacji o błędach w trakcie rozpoznania składni poleceń a następnie ich realizacji.



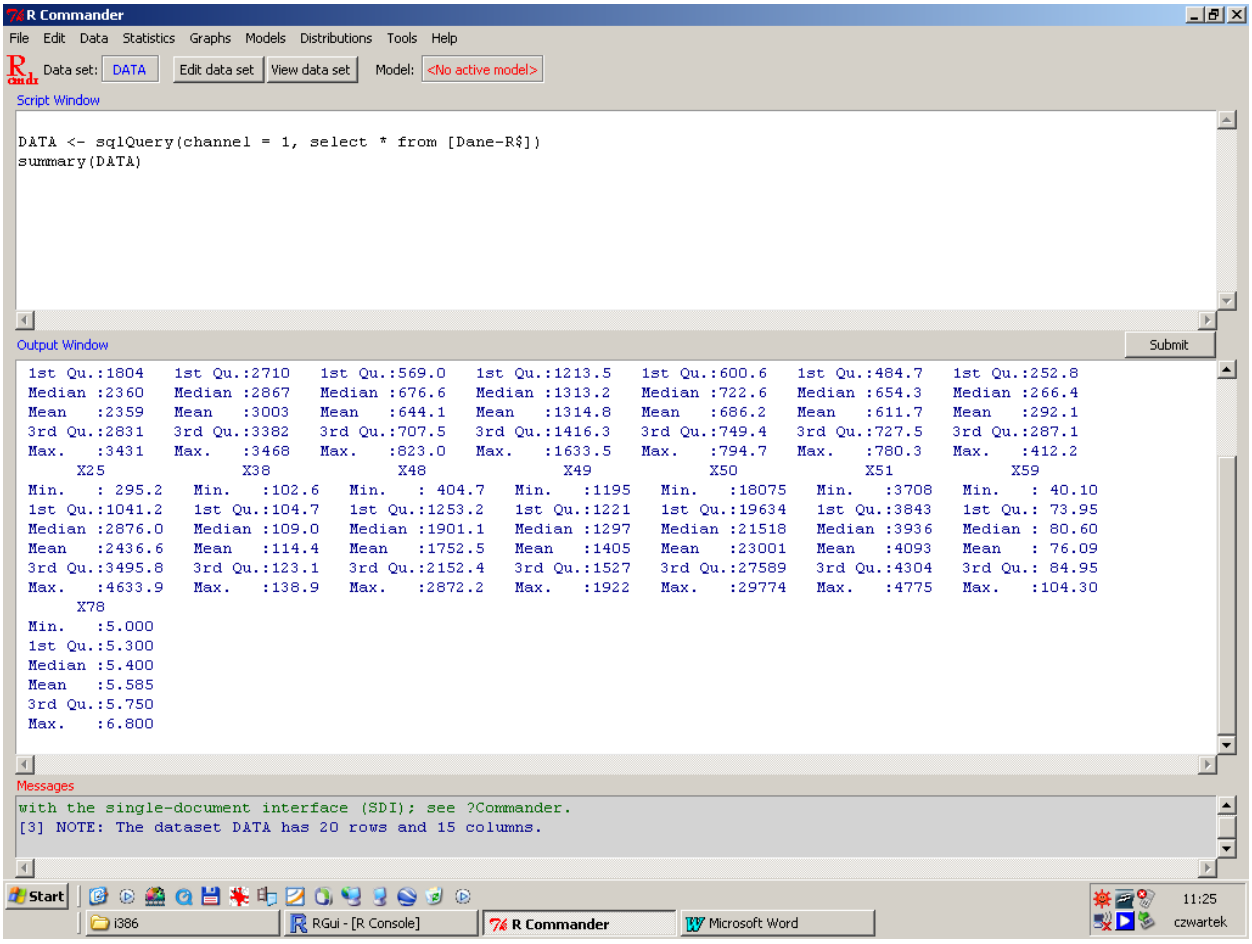
Źródło: Opracowanie własne korzystając ze środowiska programowego *R*.

Rys. 12.3. Widok ekranu trybu windowsowego

[179] Ibidem, s. 377.

Jesteśmy teraz w trybie windowsowym o dogodnym dla użytkownika interfejsie. Umożliwia to nam dokonanie importu naszego pliku danych *Dane.xls* z Excela umieszczonego w podkatalogu *D:\ARTYK-2013* korzystając z menu: *Data/Import data/from Excel, Access or dBase data set ...* Określamy nazwę obiektu jako *DATA* i OK. Pojawia się okno „*Otwieranie*" i na ścieżce *D:\ARTYK-2013* wskazujemy do pobrania plik *Dane.xls*. W pierwszym oknie tj. skryptowym pojawia się polecenie w języku *R*, gdyż program dokonuje konwersji naszych „ruchów" w menu na sekwencję języka *R*: *DATA <- sq1Query(channel = 1, select * from [Dane-R$]).*

Odpowiedzią programu w oknie wyjścia *Output Window* jest potwierdzenie. Natomiast w oknie komunikatów jest informacja, że traktowany jako tabela danych, importowany plik *Dane.xls* zawiera 20 wierszy (oprócz nagłówka) i 15 kolumn. Po zakończeniu importu mamy aktywną tabelę danych i przystępujemy przykładowo do sporządzenia raportu sumarycznego podstawowych statystyk opisowych naszych 15 zmiennych. korzystając z menu: *Statistics/Summaries/Active dataset.* Odpowiada to poleceniu: *summary(DATA).* Raport z realizacji tego polecenia ukazał się w oknie *Output Window* (zob. rysunek 12.4). Widzimy tu obliczone dla każdej zmiennej wartości z 20. obserwacji: *Min.* – minimum, *1 st Qu.* - kwartyl pierwszy, *Median* - mediana, *Mean* - średnia arytmetyczna, *3 rd Qu.* - kwartyl trzeci, *Max.* - maksimum.



Źródło: Opracowanie własne korzystając ze środowiska programowego *R*.

Rys. 12.4. Podstawowa statystyka opisowa

Zwróćmy uwagę, że wartości poszczególnych zmiennych z przecinkiem zamienia program na kropkę dziesiętną. Będąc w trybie windowsowym sporządźmy przykładowo wykres liniowy zależności

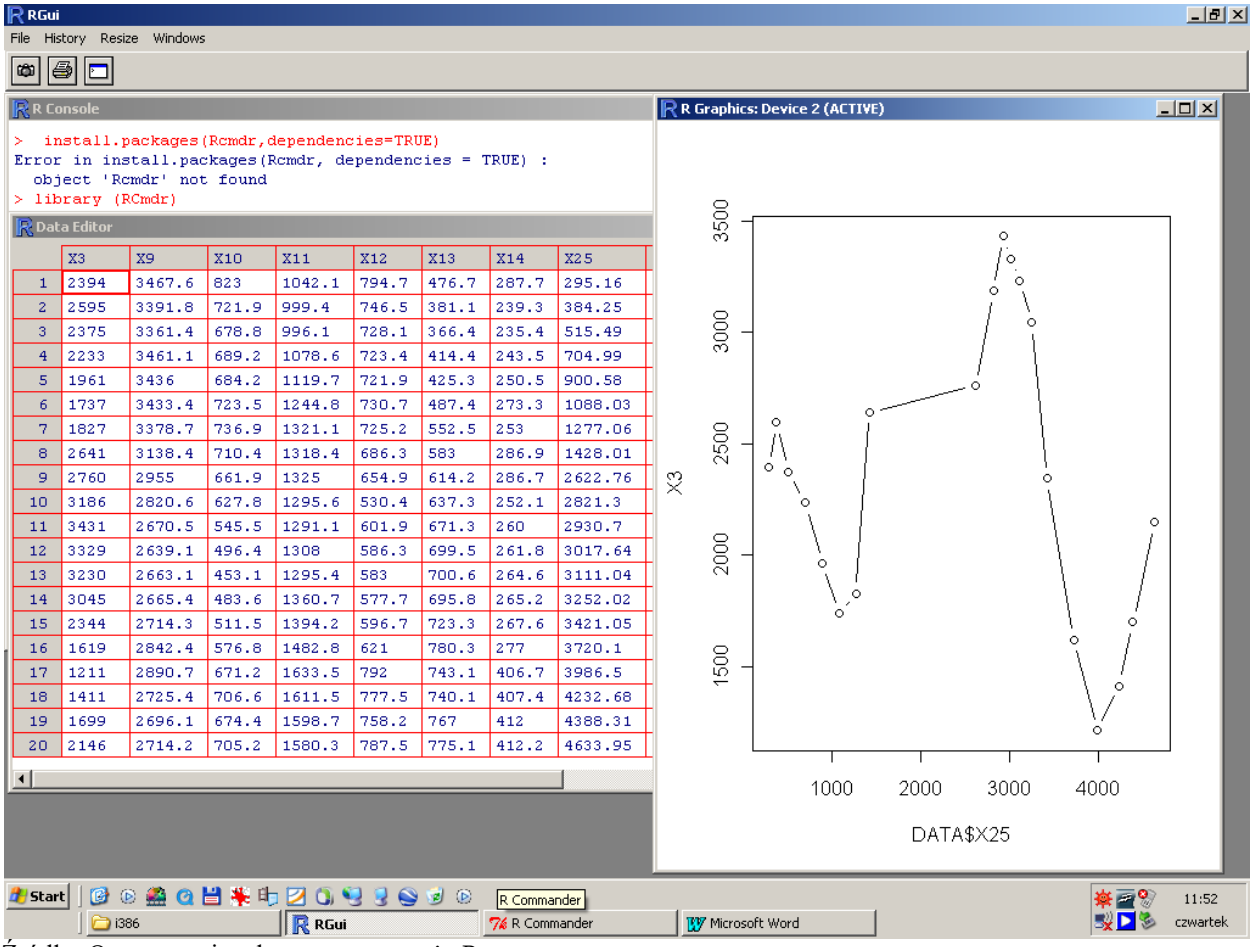
zmiennych *X*3 ~ *X*25. Ukazano go na tle fragmentu tabeli danych oraz widoku ekranu okna trybu okienkowego *Rgui* (zob. rysunek 12.5), a realizację tę umożliwia menu: *Graphs/Line graph...*

Przystąpmy teraz do oszacowania parametrów modelu regresji liniowej wielowymiarowej, poprzez wcześniejsze uaktywnienie tabeli danych przyciskiem *DATA* (zob. rysunek 12.4), a następnie wywołanie menu: *Statistics/Fit models/Linear model* ...Pojawia się okno dialogowe w którym określamy ogólnie składnię argumentu „*formuła funkcji lm dla modelu*[180]:

$$\hat{\mathbf{Y}} = \mathbf{b}_0 + \mathbf{b}_1\mathbf{X}_1 + \mathbf{b}_2\mathbf{X}_2 + \mathbf{...}$$

W odniesieniu do trzech zmiennych naszego końcowego modelu uzyskanego programem *Gretl* model przyjmuje postać:

$$\hat{\mathbf{X}}_3 = \hat{\mathbf{Y}} = \mathbf{b}_0 + \mathbf{b}_1\mathbf{X}_9 + \mathbf{b}_2\mathbf{X}_{25} + \mathbf{b}_3\mathbf{X}_{78}$$



	X3	X9	X10	X11	X12	X13	X14	X25
1	2394	3467.6	823	1042.1	794.7	476.7	287.7	295.16
2	2595	3391.8	721.9	999.4	746.5	381.1	239.3	384.25
3	2375	3361.4	678.8	996.1	728.1	366.4	235.4	515.49
4	2233	3461.1	689.2	1078.6	723.4	414.4	243.5	704.99
5	1961	3436	684.2	1119.7	721.9	425.3	250.5	900.58
6	1737	3433.4	723.5	1244.8	730.7	487.4	273.3	1088.03
7	1827	3378.7	736.9	1321.1	725.2	552.5	253	1277.06
8	2641	3138.4	710.4	1318.4	686.3	583	286.9	1428.01
9	2760	2955	661.9	1325	654.9	614.2	286.7	2622.76
10	3186	2820.6	627.8	1295.6	530.4	637.3	252.1	2821.3
11	3431	2670.5	545.5	1291.1	601.9	671.3	260	2930.7
12	3329	2639.1	496.4	1308	586.3	699.5	261.8	3017.64
13	3230	2663.1	453.1	1295.4	583	700.6	264.6	3111.04
14	3045	2665.4	483.6	1360.7	577.7	695.8	265.2	3252.02
15	2344	2714.3	511.5	1394.2	596.7	723.3	267.6	3421.05
16	1619	2842.4	576.8	1482.8	621	780.3	277	3720.1
17	1211	2890.7	671.2	1633.5	792	743.1	406.7	3986.5
18	1411	2725.4	706.6	1611.5	777.5	740.1	407.4	4232.68
19	1699	2696.1	674.4	1598.7	758.2	767	412	4388.31
20	2146	2714.2	705.2	1580.3	787.5	775.1	412.2	4633.95

Źródło: Opracowanie własne w programie *R*.

Rys. 12.5. Podstawowa statystyka opisowa i wykres zależności zmiennych *X3* → *X25*

Program *R* w oknie *Output Windows* dla podprogramu **lm** pokazuje sposób określenia opcji **formula**, określa statystykę opisową reszt modelu zmiennej zależnej *X3*, a ponadto oszacowanie parametrów modelu liniowego ze zmiennymi *X9*, *X25*, *X78*. Sposób zapisu wartości jest w notacji inżynierskiej, przykładowo: 1.675e+04 = 1,675 x 10^4 = 16750 (zob. tabela 12.4).

[180] Walesiak M., Gatnar E. (red. nauk.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009, s. 143.

Tab. 12.4. Oszacowanie parametrów dla trzech zmiennych modelu końcowego w programie *R* i dodatkowe informacje wejściowe oraz podstawowa statystyka opisowa

lm(formula = **X3 ~ X9 + X25 + X78**, data = DATA)

Residuals:

Min	1Q	Median	3Q	Max
-239.89	-155.26	12.21	128.91	323.18

Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	**1.675e+04**	1.061e+03	15.783	3.55e-11	***
X9	**-3.317e+00**	3.698e-01	-8.970	1.22e-07	***
X25	**-6.690e-01**	9.109e-02	-7.344	1.65e-06	***
X78	**-5.011e+02**	1.077e+02	-4.652	0.000266	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 180.2 on 16 degrees of freedom
Multiple R-squared: 0.9381; Adjusted R-squared: 0.9265
F-statistic: 80.81 on 3 and 16 DF; p-value: 7.009e-10

gdzie:

Intercept - stała *const*,
Estimate - współczynnik (oszacowany parametr),
Std. Error - standardowy błąd oszacowania
t value - wartość statystyki *t*-Studenta,
Pr(>|t|) - poziom istotności w odniesieniu do parametru,
Signif. Codes - oznaczenie graficzne poziomu istotności,
Residual standard error - błąd standardowy reszt,
Multiple R-squared - współczynnik determinacji R^2,
Adjusted R-squared - skorygowany współczynnik determinacji,
F-statistic - statystyka *F*,
p-value - poziom istotności modelu.

Źródło: Opracowanie własne programem **lm** pakietu *R*.

Porównanie wyników z rezultatem uzyskanym z *Gretl*-a (zob. tabela 12.3) wskazuje na identyczność estymacji i uzyskanej statystyki opisowej. Zauważmy, że w *R* występuje też sygnalizacja graficzna błędów istotności poszczególnych parametrów modelu w postaci gwiazdek.

Zobaczmy jeszcze jak radzi sobie środowisko *R* z oszacowaniem regresji zmiennej *X*3 z pozostałymi 14. zmiennymi objaśniającymi, które stanowiły wejście do *modelu 1.* zrealizowanego wcześniej programem *Gretl* (zob. tabela 12.2). Możemy to zrealizować korzystając z menu programu *R* w trybie windowsowym lub wprowadzając do okna skryptowego następujące polecenie:

```
LinearModel.3 <- lm(X3 ~ X9 + X10 + X11 + X12 + X13 + X14 + X25 + X38 + X48 + X49 + X50 + X51 + X59 + X78, data=DATA)
```

Odpowiedzią komputera jest raport z zestawieniem sumarycznym modelu regresji liniowej wielowymiarowej nazwany „*LinearModel.3*) - zob. tabela 12.5.

Tab. 12.5. Oszacowanie parametrów dla czternastu zmiennych modelu wstępnego w programie *R*, dodatkowe informacje wejściowe oraz podstawowa statystyka opisowa

```
> summary(LinearModel.3)
--------------------------------------------------------------------------------------------------------------
Call:
lm(formula = X3 ~ X9 + X10 + X11 + X12 + X13 + X14 + X25 + X38 +
    X48 + X49 + X50 + X51 + X59 + X78, data = DATA)
--------------------------------------------------------------------------------------------------------------
Residuals:
    1       2       3       4       5       6       7       8       9      10      11
-46.41  118.79 -175.16  178.07  -24.31 -120.37  -60.73  104.19  108.12  -43.20   48.21
   12      13      14      15      16      17      18      19      20
-13.98 -121.87   41.02   77.33  -55.06  106.79 -170.12  -34.03   82.72
--------------------------------------------------------------------------------------------------------------
Coefficients:
             Estimate        Std. Error      t value   Pr(>|t|)
(Intercept) 9833.04163  6662.18868    1.476     0.2000
X9          -5.31161    2.37295  -2.238   0.0754 .
X10          2.34751    2.65491   0.884   0.4171
X11          2.20456    4.46969   0.493   0.6428
X12         -0.85295    3.11914  -0.273   0.7954
X13          0.61113    2.51962   0.243   0.8180
X14         -3.76527    7.68213  -0.490   0.6448
X25         -0.63737    0.50188  -1.270   0.2600
X38         88.29286   76.31622   1.157   0.2996
X48          0.36000    0.88948   0.405   0.7024
X49          0.46933    3.17212   0.148   0.8882
X50         -0.07832    0.07615  -1.029   0.3508
X51         -0.16798    0.42598  -0.394   0.7096
X59         17.09851   15.97375   1.070   0.3334
X78       -585.81937  297.79532  -1.967   0.1063
--------------------------------------------------------------------------------------------------------------
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
--------------------------------------------------------------------------------------------------------------
Residual standard error: 199.1 on 5 degrees of freedom
Multiple R-squared: 0.9764; Adjusted R-squared: 0.9102
F-statistic: 14.76 on 14 and 5 DF; p-value: 0.003864
```

Źródło: Opracowanie własne programem **lm** pakietu *R*.

Gdy porównamy z wynikami *modelu 1*. uzyskanymi z *Gretl*-a to występują jednak nieznaczne różnice liczbowe w oszacowaniu parametrów modelu, być może jest to kwestia dokładności przeprowadzenia obliczeń. Ze względu na ramy tego opracowania sprawdźmy tylko trzy wiersze z *modelem 1*. *Gretl*-a.

Zmienna	*Współczynnik*	*Współczynnik wg Gretl-a*	*Błąd stand.*	*Statystyka t*	*Wartość p*	
const	**11037,4**	**9833,0**	6207,87	1,7780	0,13555	
*X*9	**-5,18169**	**-5,31161**	2,50564	-2,0680	0,09348	*
*X*10	**2,30576**	**2,34751**	2,81819	0,8182	0,45048	
*X*11	**2,05037**	**2,20456**	4,56698	0,4490	0,67225	

Niedogodnością w raporcie z programu *R* jest brak podpowiedzi, którą ze zmiennych (oprócz stałej) usunąć w następnej iteracji. Tu musimy się sami kierować maksymalną wartością popełnienia błędu zapisaną w kolumnie *Wartość p*. Według tabeli 12.5 jest nią zmienna *X*49 o ***p*** = **0,8882,** czyli taka

sama jak w *modelu 1.* uzyskanym z *Gretl*-a. Pozostaje nam teraz zorientowanie się jakie możliwości daje menu programu *R* w zakresie weryfikacji uzyskanego modelu końcowego.

12.4. Weryfikacja podstawowa modelu końcowego

Przyjmijmy sytuację, że etap weryfikacji modelu podanego w tabeli 12.4 realizujemy po zamknięciu systemu komputerowego. W takiej sytuacji musimy ponownie zainicjować tryb okienkowy wywołując program *Rgui* ze ścieżki: *D:\Program Files\R\Bin\i386*. Następnie instalujemy program trybu windowsowego poleceniem: *library (Rcmdr)*. Pojawia się okno *R Commander* na którym uaktywniamy odpowiednią tabelę danych przyciskiem *Select Data Set*. Zaznaczamy plik *DATA* i naciskamy *OK*. Pojawia się przycisk DATA z czcionką w kolorze niebieskim. W kolejności korzystamy z menu: *Statistics/Fit models/Linear Model ...* Wpisujemy formułę naszego modelu: *X*3 ~ *X*9 +*X*25 + *X*78 i pojawia się przycisk również ze czcionką w kolorze niebieskim o treści: *LinearModel.*1. Oznacza to uaktywnienie naszego modelu regresji liniowej o trzech zmiennych objaśniających. Możemy teraz przystąpić do weryfikacji modelu korzystając w tym względzie z pięciu testów podstawowych proponowanych przez program *R* wywołując menu *Models/Numerical diagnostics*, a mianowicie:

1. Variance-inflation factor - badanie, czy w modelu nie występuje problem współliniowości.

2. *Breuch-Pagan test for heteroscedasticity* - badanie, czy w próbie na podstawie której estymowano model zmiennej objaśnianej *X*3 nie występuje niejednorodność wariancji składników losowych.

3. Durbin-Watson test for autocorrelation - badanie czy w modelu nie występuje autokorelacja reszt pierwszego stopnia.

4. *RESET test for nonlinearity* - badanie nieliniowości modelu.

5. Bonferroni outlier test - sprawdzenie czy w zbiorze danych występują obserwacje nietypowe.

Ponadto wywołując menu: *Statistics/summaries/Shapiro-Wilk test normality* możemy sprawdzić, czy składnik losowy modelu zmiennej objaśnianej *X*3, w przybliżeniu reprezentowany przez reszty ma rozkład normalny.

Wywołanie poszczególnych poleceń menu generuje odpowiednie komendy, które są następnie zapisywane w oknie *Script Window*. Natomiast odpowiedzi programu (raporty) z kropką dziesiętną występują w oknie *Output Window* z powtórzeniem polecenia (zob. rysunek 12.6). Odnośnie badania współliniowości to polecenie i informacja wyniku testu są następujące:

Vif(LinearModel.1)

	*X*9	*X*25	*X*78
8.936837	**10.425633**	1.859964	

W modelu występuje problem współliniowości - nieznaczny, gdyż kryterium jest $VIF_2 > 10$[181]. Miara ta zwaną czynnikiem inflacji wariancji dla *j*-tego estymatora parametru obliczana jest według wzoru: gdzie: R_j^2 - współczynnik determinacji modelu, w którym zmienna X_j , jest zmienną objaśnianą, a

$$\mathbf{VIF_j} = \frac{1}{\left(1 - \mathbf{R_j^2}\right)}$$

pozostałe z naszych trzech zmiennymi objaśniającymi.

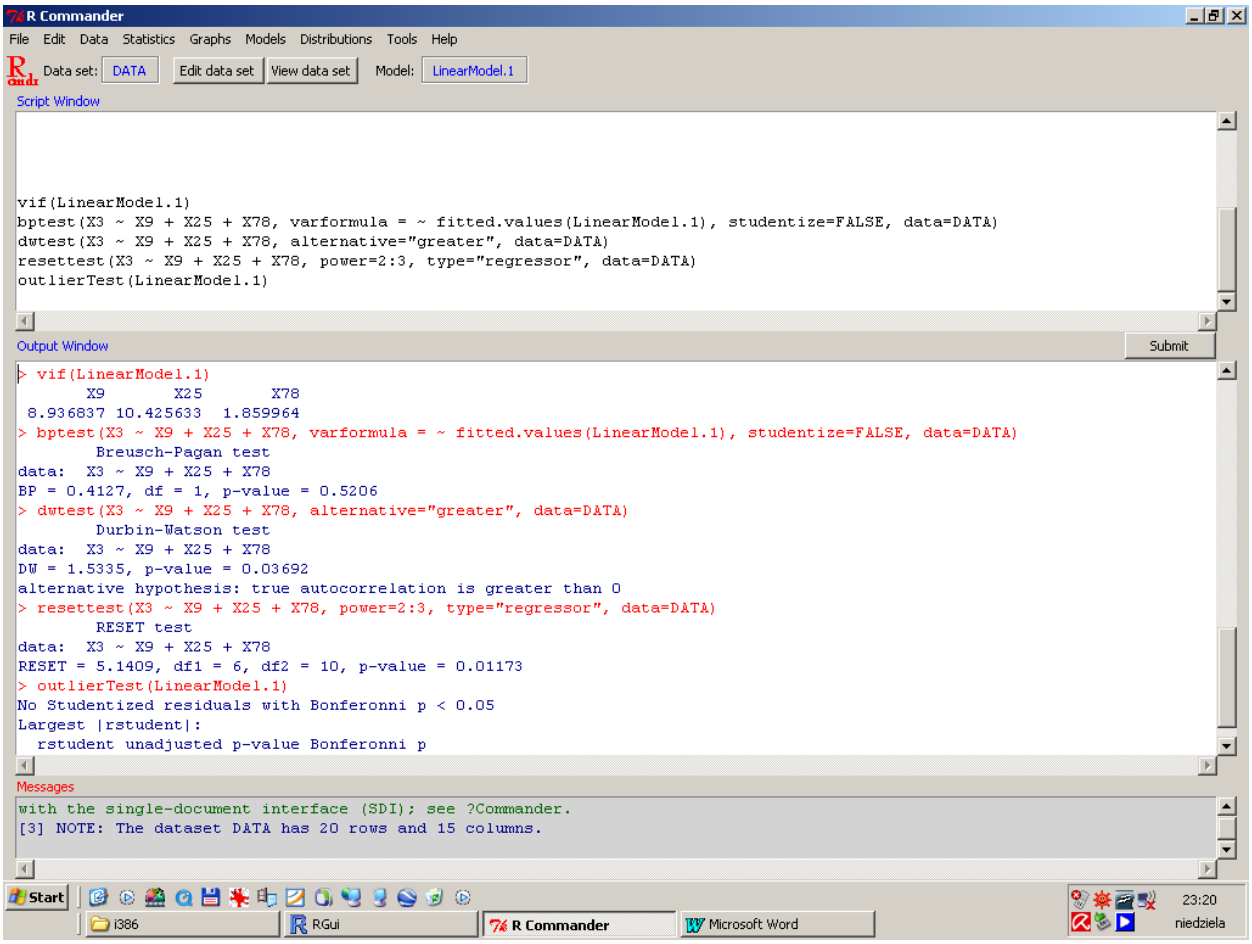
Niejednorodność wariancji składników losowych w próbie, czyli *heteroskedastyczność* występuje przeważnie w modelach bazujących na danych przekrojowych, a takim jest nasz model zmiennej *X*3. Na rysunku 12.6 występuje test Breuscha-Pagana, a polecenie i odpowiedź programu są następujące:

bptest(X3 ~ X9 + X25 + X78)

[181] Walesiak M., Gatnar E. (red. nauk.), *Statystyczna analiza danych z wykorzystaniem programu R, op. cit.*, s. 148.

Breusch-Pagan test
data: X3 ~ X9 + X25 + X78
BP = 0,4127, df = 1, p-value= **0,5206**.

Wartość *p-value* = 0,5206 ≥ α = 0,05, co wskazuje na *homoskedastyczność* składnika losowego[182], czyli nie występuje niejednorodność wariancji składników losowych.



Źródło: Opracowanie własne w programie *R*.

Rys. 12.6. Widok poleceń oraz reakcji programu R na żądania wywołane poprzez menu

W badaniu autokorelacji pierwszego stopnia lub inaczej rzędu *AR*(1) zastosowano test Durbina-Watsona przy założonym poziomie istotności α = 0,05. W raporcie występuje polecenie i następująca odpowiedź (zob. rysunek 12.6):

dwtest(X3 ~ X9 + X25 + X78, alternative=”greater”, data=DATA)
data: X3 ~ X9 + X25 + X78
DW = 1.5335, p-value = **0.03692**
alternative hypothesis: true autocorrelation is greater than 0.

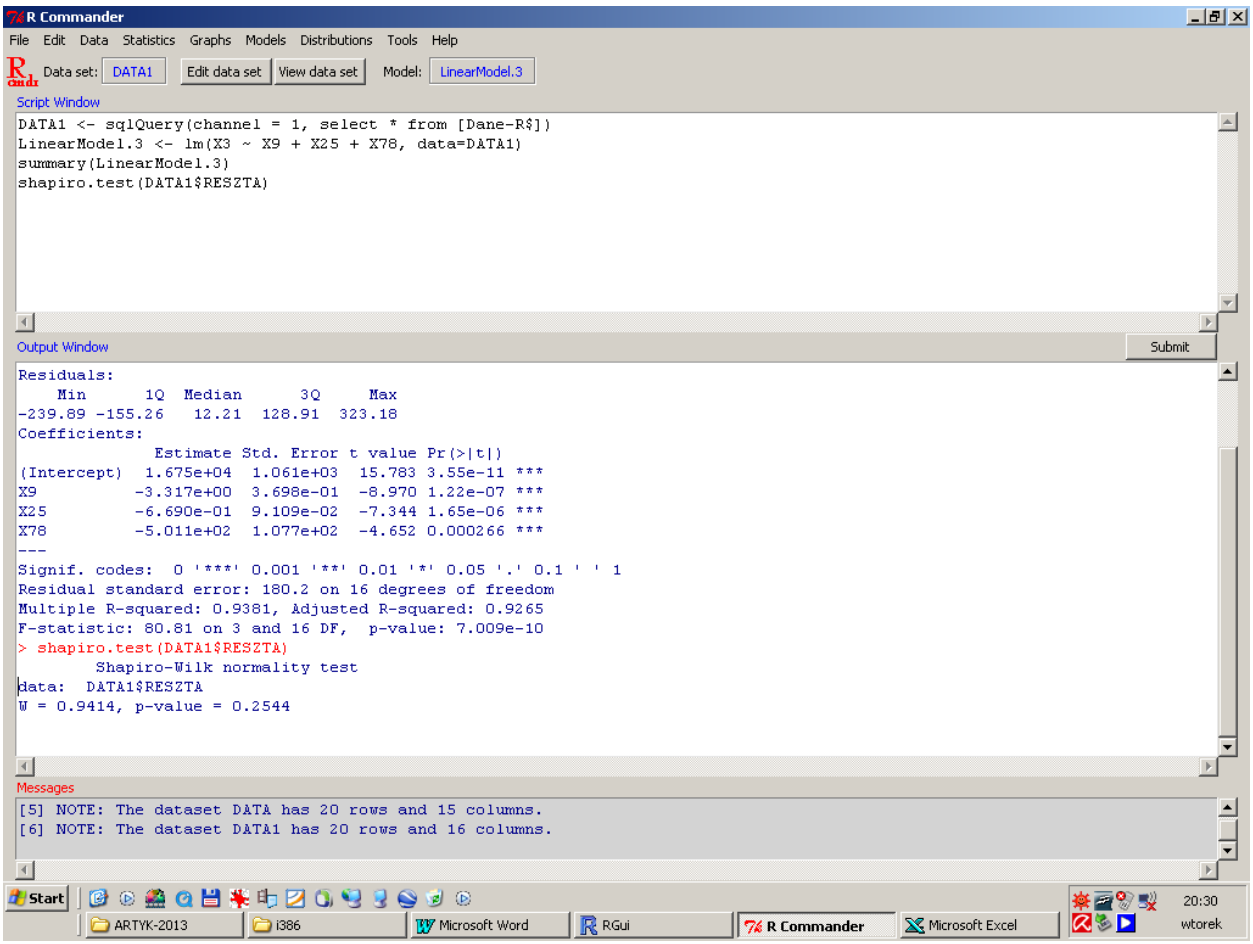
Test ten wskazuje jednak na występowanie autokorelacji reszt pierwszego stopnia, ponieważ α = 0,05 > p-value = 0,03692. Przyczyny tego zjawiska są różne, a jedną z nich jest dobór zmiennych objaśniających w okresach czasowych.

Sprawdzamy czy składnik losowy ma rozkład normalny. W tym względzie zastosujemy test Shapiro-Wilka w odniesieniu do obliczonych reszt (*X*3 – *X*3^) nazwanych w zbiorze *DATA*1 jako

[182] Ibidem, s. 149.

RESZTA. Musi to być poprzedzone obliczeniem w *Excelu* wartości teoretycznych modelu końcowego z programu *R* oraz reszt, więc przypomnijmy sobie ten model:

$$\hat{X}3 = 16750 - 3{,}317\,X9 - 0{,}669\,X25 - 501{,}1X78$$



Źródło: Opracowanie własne w programie *R*.

Rys. 12.7. Widok poleceń oraz reakcji programu *R* na żądania estymacji modelu i przeprowadzenia testu na normalność składnika losowego

Na rysunku 12.7 wymieniono polecenia wygenerowania tabeli danych *DANE*1 oraz odpowiedź programu na żądanie wykonania testu: *Shapiro-Wilk normality test* na podstawie zmiennej RESZTA przy domyślnej wartości poziomu istotności α = 0,05. Widzimy, że *p-value* = **0,2544** > α = 0,05, zatem nie ma podstaw do odrzucenia hipotezy o normalności składnika losowego.

W badaniu występowania w zbiorze danych obserwacji nietypowych, które cechują duże reszty, a po ich wyeliminowaniu następuje poprawa współczynnika determinacji (R^2) wykorzystuje się funkcję **rstudent**:

outlierTest(LinearModel.1)
No Studentized residuals with Bonferonni p < **0.05**
Largest |**rstudent**|:

rstudent	unadjusted	p-value	Bonferonni p
9	2.012186	**0.062517**	NA

Program wskazał obserwację o numerze **9** jako nietypową (*p-value* > ***p* = 0,05**).
Zachodzi więc potrzeba ponownego oszacowania parametrów modelu końcowego bez tej obserwacji. Z ciekawości porównajmy wartości tej obserwacji ze średnią - *mean* poszczególnych zmiennych modelu.

Rok	*t*	*X*3	*X*9	*X*10	*X*11	*X*12	*X*13	*X*14	*X*25	*X*38	*X*48	*X*49	*X*50	*X*51	*X*59	*X*78
Średnia	-	2358,7	3003,3	644,1	1314,9	682,2	611,7	292,1	2436,6	114,4	1752,5	1404,8	23001	4093,4	76,1	5,6
2000	9	2760	2955,0	661,9	1325,0	654,9	614,2	286,7	2622,76	111,1	1904,57	1360,3	21037	3750,8	78,4	5,5

Postępując analogicznie w wyniku kolejnych iteracji wyeliminowano jeszcze obserwacje **19**, **18** oraz **6** uzyskując w efekcie poprawę współczynnika determinacji na R^2= **0,9596** i oszacowanie parametrów modelu:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	1.583e+04	9.232e+02	17.146	8.35e-10	***
*X*9	-3.054e+00	3.292e-01	-9.277	8.01e-07	***
*X*25	-6.009e-01	8.041e-02	-7.473	7.50e-06	***
*X*78	-5.029e+02	9.754e+01	-5.156	0.000238	***

oraz raport z realizacji funkcji wyłonienia obserwacji wpływowych:

No Studentized residuals with Bonferonni p < 0.05

Largest |rstudent|:

	rstudent	unadjusted p-value	Bonferonni p
6	-1.671100	0.12288	NA

Funkcja *rstudent* nie proponuje kolejnej obserwacji jako nietypowej do usunięcia z tabeli danych.

Na zakończenie tego opracowania zobaczmy jak można uzyskać w trybie windowsowym programem *R* statystykę opisową obejmującą (średnią - *mean*, odchylenie standardowe - *sd*, kwantyle) korzystając z menu: *Statistics/Summaries/Numerical summaries ...*

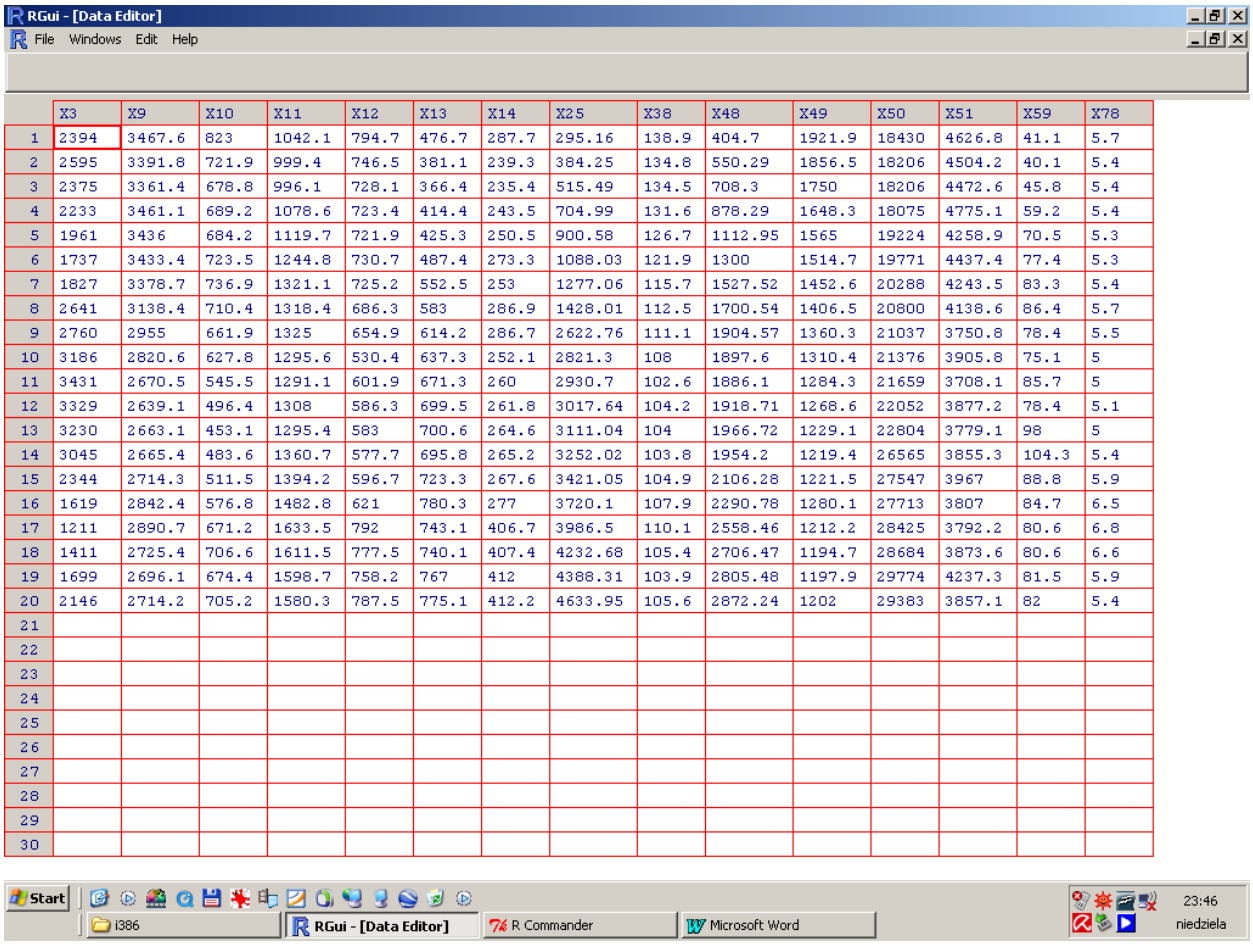
Tab. 12.6. Polecenie oraz wartości podstawowej statystyki poszczególnych zmiennych modelu wejściowego

```
> numSummary(DATA[,c("X3", "X9", "X10", "X11", "X12", "X13", "X14", "X25", "X38", "X48", "X49",
"X50", "X51", "X59",
+   "X78")], statistics=c("mean", "sd", "quantiles"), quantiles=c(0,.25,.5,.75,1))
```

	mean	sd	0%	25%	50%	75%	100%	n
*X*3	2358.700	664.527541	1211.00	1804.500	2359.500	2831.250	3431.00	20
*X*9	3003.260	334.192300	2639.10	2709.675	2866.550	3381.975	3467.60	20
*X*10	644.095	99.467289	453.10	568.975	676.600	707.550	823.00	20
*X*11	1314.850	199.076998	996.10	1213.525	1313.200	1416.350	1633.50	20
*X*12	686.195	84.393243	530.40	600.600	722.650	749.425	794.70	20
*X*13	611.720	140.931830	366.40	484.725	654.300	727.500	780.30	20
*X*14	292.145	62.042816	235.40	252.775	266.400	287.100	412.20	20
*X*25	2436.581	1465.266731	295.16	1041.168	2876.000	3495.813	4633.95	20
*X*38	114.405	12.278414	102.60	104.725	109.050	123.100	138.90	20
*X*48	1752.510	736.167686	404.70	1253.237	1901.085	2152.405	2872.24	20
*X*49	1404.800	230.951980	1194.70	1220.975	1297.350	1527.275	1921.90	20
*X*50	23000.950	4237.527425	18075.00	19634.250	21517.500	27588.500	29774.00	20
*X*51	4093.380	328.385402	3708.10	3843.225	3936.400	4303.525	4775.10	20
*X*59	76.095	17.219923	40.10	73.950	80.600	84.950	104.30	20
*X*78	5.585	0.523425	5.00	5.300	5.400	5.750	6.80	20

Źródło: Opracowanie własne w programie *R*.

Ponadto przy aktywnej tabeli danych przykładowo *DATA*, modelu regresji liniowej możemy skorzystać z polecenia **fix(DATA)** do pokazania zawartości tablicy danych wejściowych (zob. rysunek 12.8).



	X3	X9	X10	X11	X12	X13	X14	X25	X38	X48	X49	X50	X51	X59	X78
1	2394	3467.6	823	1042.1	794.7	476.7	287.7	295.16	138.9	404.7	1921.9	18430	4626.8	41.1	5.7
2	2595	3391.8	721.9	999.4	746.5	381.1	239.3	384.25	134.8	550.29	1856.5	18206	4504.2	40.1	5.4
3	2375	3361.4	678.8	996.1	728.1	366.4	235.4	515.49	134.5	708.3	1750	18206	4472.6	45.8	5.4
4	2233	3461.1	689.2	1078.6	723.4	414.4	243.5	704.99	131.6	878.29	1648.3	18075	4775.1	59.2	5.4
5	1961	3436	684.2	1119.7	721.9	425.3	250.5	900.58	126.7	1112.95	1565	19224	4258.9	70.5	5.3
6	1737	3433.4	723.5	1244.8	730.7	487.4	273.3	1088.03	121.9	1300	1514.7	19771	4437.4	77.4	5.3
7	1827	3378.7	736.9	1321.1	725.2	552.5	253	1277.06	115.7	1527.52	1452.6	20288	4243.5	83.3	5.4
8	2641	3138.4	710.4	1318.4	686.3	583	286.9	1428.01	112.5	1700.54	1406.5	20800	4138.6	86.4	5.7
9	2760	2955	661.9	1325	654.9	614.2	286.7	2622.76	111.1	1904.57	1360.3	21037	3750.8	78.4	5.5
10	3186	2820.6	627.8	1295.6	530.4	637.3	252.1	2821.3	108	1897.6	1310.4	21376	3905.8	75.1	5
11	3431	2670.5	545.5	1291.1	601.9	671.3	260	2930.7	102.6	1886.1	1284.3	21659	3708.1	85.7	5
12	3329	2639.1	496.4	1308	586.3	699.5	261.8	3017.64	104.2	1918.71	1268.6	22052	3877.2	78.4	5.1
13	3230	2663.1	453.1	1295.4	583	700.6	264.6	3111.04	104	1966.72	1229.1	22804	3779.1	98	5
14	3045	2665.4	483.6	1360.7	577.7	695.8	265.2	3252.02	103.8	1954.2	1219.4	26565	3855.3	104.3	5.4
15	2344	2714.3	511.5	1394.2	596.7	723.3	267.6	3421.05	104.9	2106.28	1221.5	27547	3967	88.8	5.9
16	1619	2842.4	576.8	1482.8	621	780.3	277	3720.1	107.9	2290.78	1280.1	27713	3807	84.7	6.5
17	1211	2890.7	671.2	1633.5	792	743.1	406.7	3986.5	110.1	2558.46	1212.2	28425	3792.2	80.6	6.8
18	1411	2725.4	706.6	1611.5	777.5	740.1	407.4	4232.68	105.4	2706.47	1194.7	28684	3873.6	80.6	6.6
19	1699	2696.1	674.4	1598.7	758.2	767	412	4388.31	103.9	2805.48	1197.9	29774	4237.3	81.5	5.9
20	2146	2714.2	705.2	1580.3	787.5	775.1	412.2	4633.95	105.6	2872.24	1202	29383	3857.1	82	5.4
21															
22															
23															
24															
25															
26															
27															
28															
29															
30															

Źródło: Opracowanie własne w programie *R*.

Rys. 12.8. Zawartość tabeli danych wejściowych

* * *

Celem niniejszego opracowania było pokazanie możliwości korzystania z trybu windowsowego programów wywołanych w środowisku pakietu programowego *R*. Odnośnie budowy oraz weryfikacji modelu ekonometrycznego *w „Statystycznej analizie danych ...*" prezentowane są przykłady realizacji w trybie scriptowym (tekstowym). Wymaga to jednak szczegółowej znajomości struktury i komend języka nazwanego również *R*.

Ponieważ opracowanie adresowane jest zwłaszcza do studentów, zwłaszcza studiów zaocznych kierunku zarządzanie, którzy łączą pracę zawodowa z nauką, szybkie i sprawne posługiwanie się rozbudowanym menu umożliwia im dogodną estymację parametrów jak i sprawdzenie poprawności zbudowanego modelu ekonometrycznego. Ponadto może stanowić alternatywne, obok innych programów komputerowych, obszerne narzędzie badań statystycznych - w różnych branżach obejmujące nowe elementy z tej dziedziny.

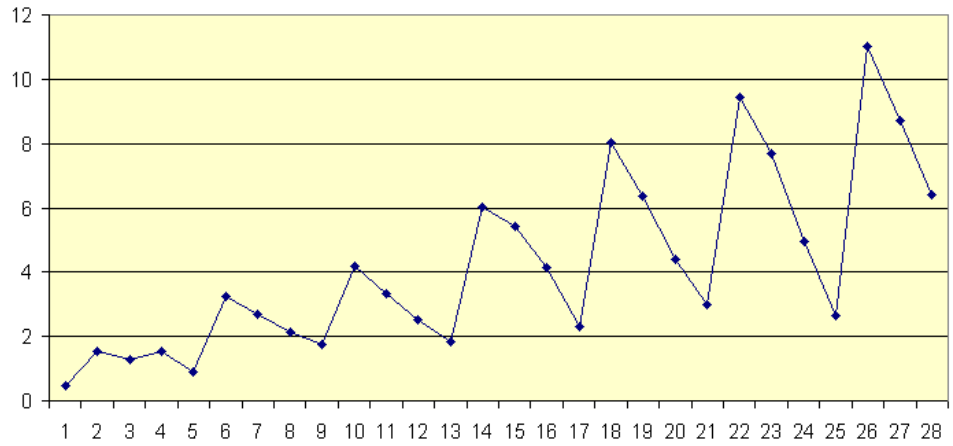
13. Prognozowanie z zastosowaniem modelu multiplikatywnego

13.1. Wstęp

Model Holta-Wintersa ze sezonowością zawiera trzy procedury wyrównywania wykładniczego z użyciem parametrów: α - wygładzanie wartości danych, β - wyrównywanie trendu, γ - wyrównywanie w celu wyodrębnienia sezonowości. Prognozowanie wielu zjawisk ekonomicznych wymaga użycia tej klasy modeli, stąd też duże zainteresowanie w Internecie równaniami oraz przykładami rozwiązań zagadnień praktycznych zwłaszcza modelami multiplikatywnymi. Model multiplikatywny jest często używanym modelem w dekompozycji szeregów czasowych. Występuje tu trend przykładowo rosnący i sezonowość np. w okresie 4. kwartałów, czy też 12. miesięcy roku. Ideę takiego trendu przedstawiono na rysunku 13.1. Model sezonowości multiplikatywnej zapisywany jest w postaci[183]:

$$Y_{n+1} = (S_n + b_n).I_{n-L+1} + z_{n+1}$$

Tak więc wartość Y_{n+1} jest równa sumie wyrównanej wykładniczo wartości S_n i wartości trendu b_n, pomnożonej przez wartość indeksu sezonowości L_{n-L+1} i z uwzględnieniem błędu z w okresie $(n + 1)$.



Źródło: http://www.ekonometria.4me.pl/modele_szeregow_czasowych.htm.

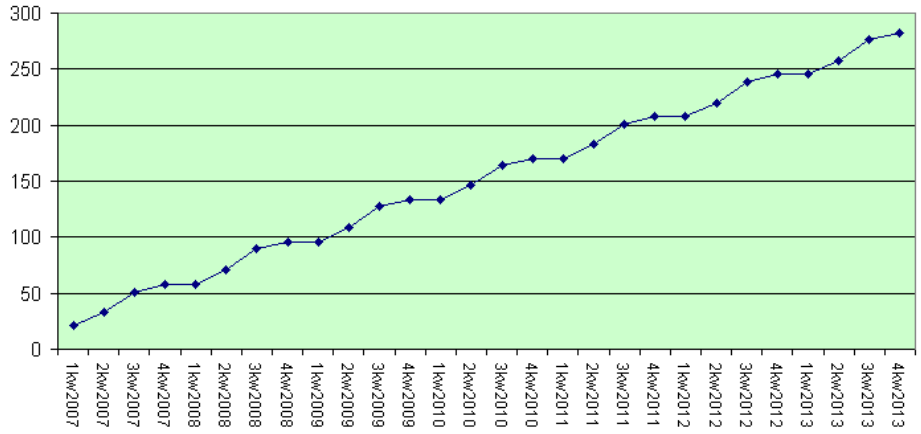
Rys. 13.1. Model multiplikatywny Holta-Wintersa

Obserwujemy tu „*rozchylające*" się wahania przypadkowe, trend oraz sezonowość. Rozróżniamy także model Holta-Wintersa z sezonowością addytywną. W modelach addytywnych ogólnie biorąc wahania przypadkowe stanowią różnicę pomiędzy wartościami empirycznymi poziomu badanego zjawiska i sumą komponentu trendu i wahań sezonowych[184]:

$$WP_t = Y_t - (T_t + WSS).$$

Wykres poglądowy modelu addytywnego z trendem i sezonowością kwartalną pokazano na rysunku 13.2. W modelu tym wahania przypadkowe ograniczają się do „*przedziału*" trendu.

[183] Lipiec-Zajchowska M., *Wspomaganie procesów decyzyjnych,* Tom II, *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.
[184] Lipiec-Zajchowska M., *Wspomaganie procesów decyzyjnych,* Tom II, *Ekonometria*, op. cit., s. 53.



Źródło: http://www.ekonometria.4me.pl/modele_szeregow_czasowych.htm.

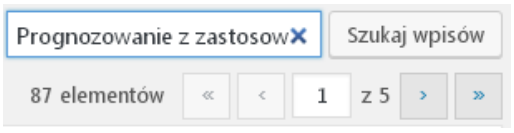
Rys. 13.2. Model addytywny Holta-Wintersa

Wpis dotyczący tematu (frazy) „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*" zamieściłem na blogu „*Zastosowanie komputera w procesach decyzyjnych*". Po wprowadzeniu frazy w cudzysłowie i naciśnięciu „*Szukaj*" system *WordPress* informuje nas o rezultacie wyszukiwania i udostępnia pełny tytuł danego wpisu oraz początek jego tekstu (zob. rysunek 13.3). Dostęp do edycji omawianego wpisu uzyskuje administrator blogu po wprowadzeniu do adresu swego identyfikatora.

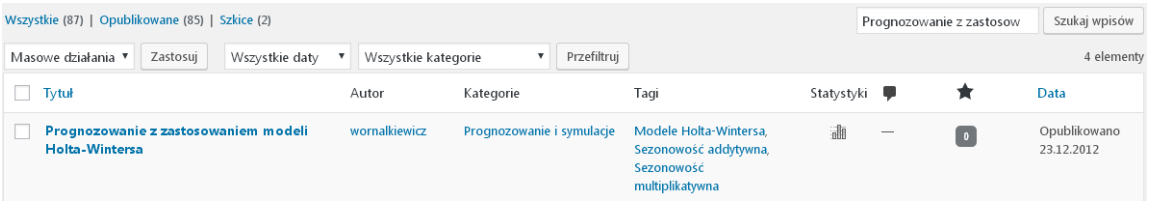
Źródło: Opracowanie własne w WordPress.

Rys. 13.3. Komunikat o o rezultacie wyszukiwania dla podanej frazy

Wyszukiwarka informuje nas ponadto o liczbie (5) występowania danej frazy we wszystkich np. 87 wpisach (elementach) na dzień korzystania z blogu:



Wywołujemy pierwszy z pięciu i udostępniona jest możliwość zaglądnięcia do statystyk prowadzonych w *WordPress* dla danego wpisu. Jedną z nich, jak już wcześniej nadmieniłem jest statystyka wpisu według lat i miesięcy (zob. rysunki 13.4 i 13.5).



Źródło: Opracowanie własne w WordPress.

Rys. 13.4. Informacja o pierwszym z pięciu wpisów występowania żądanej frazy

Są jeszcze cztery inne wpisy, gdzie występuje podana fraza wyszukiwania, ale nie są one podstawowe pod względem merytorycznym. Dodatkowo informowani jesteśmy przez system o bezpośrednim odnośniku do udostępnionego wpisu na pozycji 1/5:

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012												2	2
2013	18	12	5	20	48	46	15	12	22	40	57	66	361
2014	165	42	35	65	77	63	11	26	22	29	147	104	786
2015	150	50	80	33	76	71	22	19	35	39	85	93	753
2016	183	77	44	81	110	118	15	13	33	18			692

Źródło: Opracowanie własne w *WordPress*.

Rys. 13.5. Statystyka miesięczna odwiedzin wpisu
„*Prognozowanie z zastosowaniem modeli Holta-Wintersa*”

Zaznaczmy jeszcze, że w miesiącach styczniu i czerwcu przygotowywania się studentów do zaliczenia zwłaszcza przedmiotu „*Prognozowanie i symulacje*” statystyka wejść do omawianego wpisu pobrana z rysunku 13.5 jest następująca:

Rok/miesiąc	styczeń	czerwiec
2013	18	46
2014	165	63
2015	150	71
2016	183	118

Łatwo zauważyć trend rosnący zainteresowania wpisem „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*”.

Jak już wcześniej wspomniałem zasadniczo możemy wyróżnić trzy rodzaje modeli Holta-Wintersa: bez sezonowości, z sezonowością multiplikatywną, z sezonowością addytywną. Poznajmy teraz bliżej te trzy wymienione modele.

13.2. Model Holta-Wintersa bez sezonowości

Model bez sezonowości ma zastosowanie, gdy występuje trend i wahania przypadkowe. W modelu bez sezonowości oprócz parametru α występuje drugi parametr β odpowiednio do wygładzania wartości poziomu i trendu. Parametry te kształtują się (0-1). Równania tego modelu są następujące:

$$S_n = \alpha\, Y_n + (1- \alpha)\, (S_{n-1} + b_{n-1})$$

$$b_n = \beta\, (S_n - S_{n-1}) + (1 - \beta)\, b_{n-1}$$

$$F_{n+m} = S_n + b_n\, m$$

Wartości początkowe: $S_1 = Y_1$; $b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2$.

Przykład: Przyjmijmy $\alpha = 0{,}3$; $\beta = 0{,}1$; $Y_1 = 29$, $Y_2 = 27$; $Y_3 = 31$, $Y_4 = 35$; $b_1 = [(27 - 29) + (35 - 31)]/2 = 1$; $m = 1$; $n = 10$ obserwacji zmiennej Y_n i obliczamy elementy równań np. dla obserwacji drugiej:

$$S_2 = 0{,}3 \cdot 27 + (1 - 0{,}3)\,(29 + 1) = 29{,}1$$
$$b_2 = 0{,}1\,(29{,}1 - 29) + (1 - 0{,}1)\,1 = 0{,}91$$
$$F_2 = 29 + 1 = 30$$

Obliczamy teraz dalsze elementy równań modelu. Korzystamy z formuł *Excela* dla kolejnych dziesięciu obserwacji (zob. tabela 13.1). Obliczamy także kwadrat różnicy między wartością obserwacji a jej prognozą $(Y_n - F_n)^2$ potrzebną nam do wyznaczenia miary błędu prognozy według wzoru:

$$RMSE = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(Y_n - F_n)^2}$$

RMSE jest więc pierwiastkiem ze średniej kwadratów różnicy zmiennej prognozowanej Y_n i jej wygasłej prognozy F_n. Obliczenia w *Excelu* z widokiem komórek tego arkusza kalkulacyjnego prezentuje tabela 13.1[185].

Tab. 13.1. Obliczenie prognoz i kwadratów błędów prognoz

	A	B	C	D	E	F	G	H
1	t	Y_n	S_n	b_n	F_{n+m}	$(Y_n - F_n)^2$	m	
2	1	2	3	4	5	6	1	2
3	1	29	*29*	*1*			a	
4	2	*27*	29,1	0,91	*30*	*9*	0,3	
5	3	31	30,31	0,94	30,01	0,98	β	
6	4	35	32,37	1,05	31,25	14,09	0,1	
7	5	33	33,30	1,04	33,42	0,18	MSE	
8	6	44	37,24	1,33	34,34	93,37	26,74	
9	7	45	40,50	1,52	38,57	41,40	RMSE	
10	8	48	43,81	1,70	42,02	35,78	5,17	
11	9	51	47,16	1,87	45,51	30,09		
12	10	53	50,22	1,99	49,03	15,79		
13	11				52,20	S		
14	12				54,19	240,68		

Źródło: Opracowanie własne w *Excelu*.

W obliczeniach S_n, b_n, F_{n+m}, $(Y_n - F_n)^2$, S, MSE, RMSE zastosowano następujące formuły poprzedzone znakiem równości przykładowo do obserwacji 10:

[185] Fragment danych liczbowych oraz wzory zaczerpnięto z książki: Lipiec-Zajchrowska (red.), *Wspomaganie procesów decyzyjnych*, Tom II. *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003, rozdział 2.14. *Model Holta-Wintersa*.

J	K	L	M	N	O	P
S_{10}	b_{10}	F_{10+1}	$(Y_{10} - F_{10})^2$	S	MSE	RMSE
$G$4*B12+(1-$G$4)*(C11+D11)						
	$G$6*(C12-C11)+(1-$G$6)*D11					
		C11+D11*$G$2				
			(B12-E12)^2			
				SUMA(F4:F12)		
					F14/(A12-1)	
						PIERWIASTEK(G8)

Formuły kopiujemy z pominięciem znaku równości, gdyż wtedy stanowią tekst a nie wyrażenie do realizacji przez arkusz kalkulacyjny *Excel*. Posłużenie się formułami *Excela* może stanowić zadanie domowe symulacyjne dla studentów przedmiotu „*Prognozowanie i symulacje*" w celu doboru najlepszej wartości parametrów α i β z punktu widzenia minimum miary błędów RMSE. W tym celu sporządzamy kopię *arkusza1* skoroszytu *Tabele.xls* i próbujemy poprzez kolejne podstawianie najpierw parametru α dobierając najlepsze a później dobrać β. Rezultat naszego działania pokazano w załączeniu:

α	0,4	0,5	0,6	0,7	**0,8**	0,9
RMSE	4,53	4,12	3,89	3,79	3,78	3,85
β			0,3	**0,2**	0,1	
RMSE			3,78	3,74	3,78	

Startując z RMSE = 5,17 widzimy, że gdy dokonujemy symulacji ze skokiem 0,1 do najlepsze α = 0,8 a β = 0,2.

13.3. Model Holta-Wintersa z sezonowościa multiplikatywną

W modelu tym występuje trend, wahania przypadkowe oraz zmienna sezonowość zależna od trendu, przy czym równania tego modelu są następujące:

$$S_n = \alpha\,(Y_n/I_{n-L}) + (1 - \alpha)\,(S_{n-1} + b_{n-1})$$

$$b_n = \beta\,(S_n - S_{n-1}) + (1 - \beta)\,b_{n-1}$$

$$I_n = \gamma\,(Y_n/S_n) + (1 - \gamma)\,I_{n-L}$$

$$F_{n+m} = (S_n + b_n\,m)\,I_{n-L+m}$$

gdzie:

γ - parametr sezonowości,
m - horyzont prognozy,
I_{n-L+m} - wyrównana wartość indeksu sezonowości na okres m, np. $(n + 1)$,
L - długość cyklu sezonowości (12 - dla danych miesięcznych, 4 - dla danych kwartalnych).

Przykład**:** Mamy kwartalne popyty na lody w kolejnych czterech latach (n = 16 obserwacji)[186]. Popyt wykazuje sezonowość powiększaną w następnych latach. Podaj procedurę określenia wartości początkowych korzystając z danych zapisanych w *Excelu*.

Tab. 13.2. Kwartalne spożycie lodów (dane umowne)

[186] Dane liczbowe, wzory i procedury zaczerpnięto z książki: Lipiec-Zajchrowska (red.), *Wspomaganie procesów decyzyjnych* Tom II. *Ekonometria*, op. cit., s. 90.

	A	B	C	D	E
1	Rok	Kw. 1	Kw. 2	Kw. 3	Kw. 4
2	1	56	122	255	107
3	2	73	219	439	156
4	3	110	329	564	195
5	4	153	407	757	271

Źródło: Opracowanie własne w *Excelu*.

Procedura wstępna obliczania elementów równania jest następująca:

1. Ustalenie średniej wartości trendu na podstawie odpowiadających sobie trzech kwartałów roku drugiego i pierwszego:

$$[(73 - 56) + (219 - 122) + (439 - 255)]/16 = 18{,}625.$$

2. Wartości wyrównane:

$$S_1 = S_2 = S_3 = S_4 = (56 + 122 + 255 + 107)/4 = 135$$

3. Scentrowanie średniej:

$$[(56 \cdot 1) + (122 \cdot 2) + (255 \cdot 3) + (107 \cdot 4)]/4 = 373{,}25$$

$$373{,}25/135 = 2{,}76, \text{ w przybliżeniu } 2{,}5$$

4. Orientacyjna korekta wartości trendu b_n przykładowymi mnożnikami, przy czym suma mnożników równa się zero dla czterech kwartałów:

Kw.	b_n
1	-1,5 · 18,625 = -27,937
2	-0,5 · 18,625 = -9,312
3	0,5 · 18,625 = 9,312
4	1,5 · 18,625 = 27,937

5. Obliczenie $(S_n + b_n)$ dla czterech pierwszych kwartałów:

Kw.	$S_n + b_n$
1	135 - 27,937 = 107,063
2	135 - 9,312 = 1215,698
3	135 + 9,312 = 144,312
4	135 + 27,937 = 162,937

Na podstawie tabeli 13.2 obliczamy średnią dla kwartału pierwszego i czterech kolejnych lat:

$$(56 + 73 + 110 + 153)/4 = 98.$$

6. Wstępne indeksy sezonowości czterech pierwszych kwartałów są następujące:

Kw.	Y_n	$S_n + b_n$	$I_n = Y_n/((S_n + b_n)$
1	56	107,063	0,523
2	122	125,688	0,971
3	255	144,312	1,767
4	107	162,937	0,657

7. Ponowne ustalenie S_4 i przyjęcie $b_4 = 18{,}625$: $S_4 = Y_4/I_4 = 107/0{,}657 = 162{,}861$.

8. Przyjęcie wartości początkowych parametrów: $\alpha = 0{,}2$; $\beta = 0{,}2$; $\gamma = 0{,}1$.

Mamy już określone wartości początkowe do skorzystania z wielorównaniowego modelu Holta-Wintersa z sezonowością multiplikatywną, pora więc na zadanie: *Zastosuj funkcje Excela do wyznaczenia elementów równań modelu Holta - Wintersa oraz oblicz wygasłe prognozy dla* 16

obserwacji. Ponadto oblicz miarę błędu RMSE *wygasłych prognoz dla okresów* (5-16) *oraz oblicz prognozę przyszłą na kwartał pierwszy roku piątego.*

Tab. 13.3. Obliczenia elementów i wygasłych prognoz dla modelu multiplikatywnego

	A	B	C	D	E	F	G	H	I	J	K
1	Rok	Kw.	Y_n	S_n	b_n	$S_n + b_n$	In	F_n	$(Y_n - F_n)^2$	Okres	α
2	1	1	56			*107,063*	*0,523*			1	0,2
3	1	2	122			*125,688*	*0,971*			2	β
4	1	3	255			*144,312*	*1,737*			3	0,2
5	1	4	107	*162,861*	*18,625*	181,486	*0,657*			4	γ
6	2	1	73	**173,105**	**16,949**	**190,053**	**0,513**	**94,917**	**480,363**	5	0,1
7	2	2	219	197,151	18,368	215,519	0,985	184,542	1187,364	6	m
8	2	3	439	222,962	19,857	242,819	1,760	374,357	4178,763	7	1
9	2	4	156	241,744	19,642	261,386	0,656	159,532	12,476	8	S
10	3	1	110	252,004	17,766	269,770	0,505	134,057	578,744	9	**21414,918**
11	3	2	329	282,619	20,335	302,954	1,003	265,719	4004,546	10	MSE
12	3	3	564	306,447	21,034	327,481	1,768	533,259	945,027	11	**1784,5765**
13	3	4	195	321,452	19,828	341,280	0,651	214,772	390,950	12	RMSE
14	4	1	153	333,590	18,290	351,880	0,501	172,426	377,372	13	**42,24**
15	4	2	407	362,669	20,448	383,117	1,015	352,899	2926,967	14	
16	4	3	757	392,116	22,248	414,364	1,784	677,434	6330,691	15	
17	4	4	271	414,759	22,327	437,086	0,651	269,714	1,654	16	
18	5	1						**218,794**			

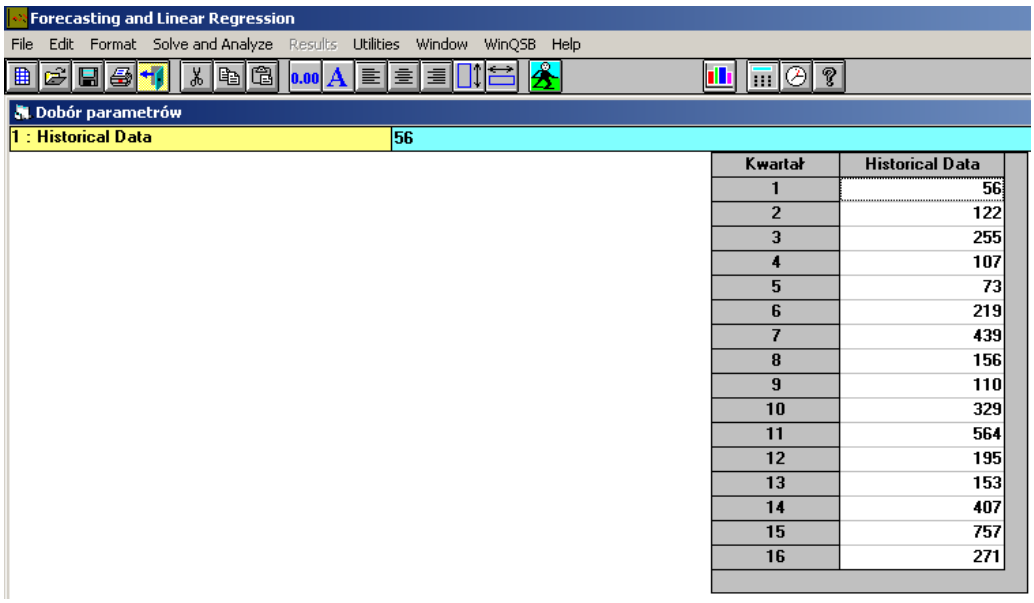
Źródło: Opracowanie własne w *Excelu*.

Pierwiastek średniego kwadratu błędu RMSE przy dokładnych obliczeniach w *Excelu* wynosi 42,24 a prognoza na okres 17. 218,794, czyli około 219. W obliczeniach S_n, b_n, $(S_n + b_n)$, I_n, F_n, $(Y_n - F_n)^2$, S, MSE, RMSE zastosowano następujące formuły poprzedzone znakiem równości, które przykładowo do obserwacji piątej są następujące:

M	N	O	P	Q	R	S	T	U	V
S_5	b_5	$S_5 + b_5$	I_5	F_5	$(Y_5 - F_5)^2$	S	MSE	RMSE	F_{17}
$SK$2*(C6/G2)+(1-$K$2)*F5									
	K$4*(D6-D5)+(1-$K$4)*E5								
		D6+E6							
			$K$6*(C6/D6)+(1-$K$6)*G2						
				(D5+E5*$K$8)*G2					
					(C6-H6)^2				
						SUMA(I6:I17)			
							K10/(J17-J5)		
								PIERWIASTEK(K12)	
									(D17+E17*$K$8)*G14

Na zakończenie podanego zestawienia formuł podano również wyrażenie na obliczenie prognozy na pierwszy kwartał roku piątego.

Kontynuujmy jeszcze nasze zadanie i zastosujmy moduł *Forecasting and Linear Regression* (*FC*) pakietu *WinQSB* dla doboru najlepszej wartości parametrów α, β, γ przy warunku minimum RMSE bez podawania wartości początkowych.

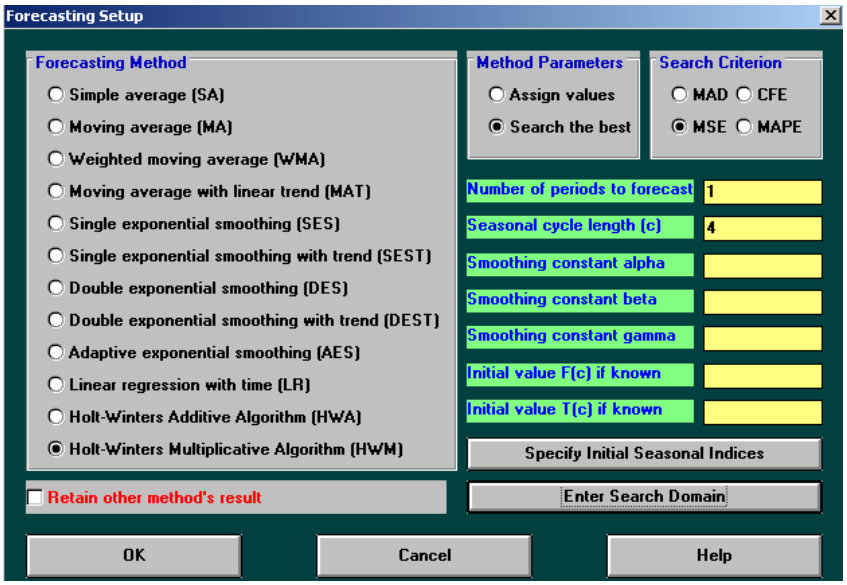


Źródło: Opracowanie własne w *FC*.

Rys. 13.6. Zdefiniowanie szeregu czasowego w module *FC*

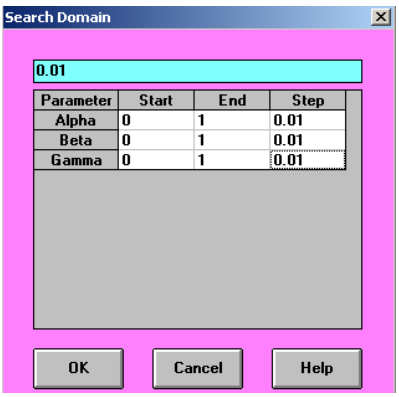
Następnie definiujemy w poszczególnych podoknach opcje metody prognozowania (zob. rysunek 13.7):

„*Forecasting Method*" - Holt-Winters Multiplicative Algorithm (HWM),
"*Method Parameters*" - Search the best,
"*Search Criterion*" - MSE,
"*Number of periods to forecast*" - 1,
"*Seasonal cycle length* (*c*)" - 4,
"*Enter Search Domain*" - 0.01 (domyślnie jest przecinek i zamieniamy go na kropkę dziesiętną (rysunek 13.8).



Źródło: Opracowanie własne w *FC*.

Rys. 13.7. Ustawienie opcji metody HWM



Źródło: Opracowanie własne w *FC*.

Rys. 13.8. Ustawienie skoku iteracyjnego doboru najlepszych wartości parametrów α, β, γ

Moduł *FC* po dłuższym czasie procesu obliczeniowego podał nam rozwiązanie problemu optymalizacyjnego pod nazwą *Forecast Result for Dobór parametrów,* w tym z dokładnością wyników do liczb całkowitych (zob. rysunek 13.9). W raporcie występują następujące miary błędów wygasłych prognoz: CFE = 106, MAD = 25, MSE = 996, MAPE = 12%. MAPE to średni procentowy absolutny błąd. Wcześniej podana miara RMSE jest pierwiastkiem z MSE i wynosi 31,56, tak więc jest mniejsza od wyliczonej formułami Excela (RMSE = 42,24) przy realizacji według parametrów α = 0,2, β = 0,2, γ = 0,1**.** Program ustalił jako najlepsze (*Search the best*) parametry:

Alpha=0,27
Beta=1
Gamma=0,99
F(0)=135

Widzimy, że przyjął również wyjściową wartość średnią przy cyklu czterech kwartałów dla pierwszego roku F(0) = 135. Ponadto założył prognozę na okres piaty równą obserwacji (Y_1 = 56).

12-28-2016 Kwartał	Actual Data	Forecast by HWM	Forecast Error	CFE	MAD	MSE	MAPE (%)
1	56						
2	122						
3	255						
4	107						
5	73	56	17	17	17	289	23
6	219	142	77	94	47	3109	29
7	439	405	34	128	43	2467	22
8	156	205	-49	80	44	2440	24
9	110	132	-22	58	40	2046	23
10	329	319	10	68	35	1722	20
11	564	557	7	75	31	1484	17
12	195	190	5	81	28	1302	16
13	153	136	17	98	26	1189	15
14	407	432	-25	72	26	1134	14
15	757	736	21	94	26	1072	13
16	271	258	13	106	25	996	12
17		202					

Źródło: Opracowanie własne w *WinQSB* - moduł *FC*.

Rys. 13.9. Raport modułu FC realizacji zadania „*Dobór parametrów*”

13.4. Model Holta-Wintersa z sezonowością addytywną

W omawianym blogu w tekście wpisu „*Prognozowanie z zastosowanie modeli Holta-Wintersa*” zamieszczono również równania modelu z sezonowością addytywną, które są następujące[187]:

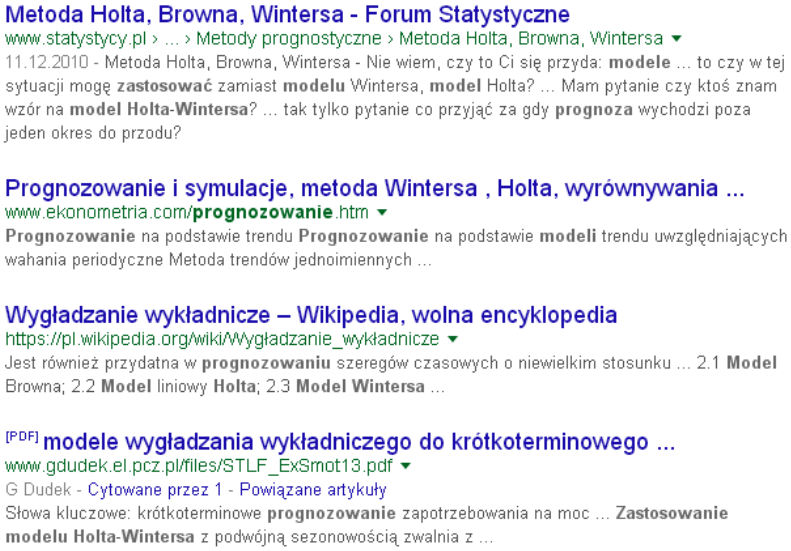
$$S_n = \alpha\,(Y_n - I_{n-L}) + (1 - \alpha)\,(S_{n-1} + b_{n-1})$$
$$b_n = \beta\,(S_n - S_{n-1}) + (1 - \beta)\,b_{n-1}$$
$$I_n = \gamma\,(Y_n - S_n) + (1 - \gamma)\,I_{n-L}$$
$$F_{n+m} = S_n + b_n \cdot m + I_{n-L+\mathrm{m}}$$

Zwróćmy uwagę we wzorze na S_n, że model Holta-Wintersa z sezonowością addytywną stanowi analogię do modelu multiplikatywnego, jednak występuje odejmowanie (zamiast dzielenia) indeksu sezonowości I_{n-L} jako to miało miejsce w modelu multiplikatywnym. Procedury korzystania z formuł *Excela*, a także z modułu *FC* pakietu *WinQSB* są podobne.

* * *

Zauważyłem, że zagadnienie prognozowania z zastosowaniem modeli Holta-Wintersa zbudza największe zainteresowanie internautów, będących gośćmi blogu „*Zastosowanie komputera w procesach decyzyjnych*”. Wiele pozycji - wpisów dla wymienionej wcześniej frazy bez cudzysłowu, bo aż 1710, znalazło się w dniu wyszukiwania w zasobach Google. Rozpatrzmy tylko stronę pierwszą o 10. wpisach. oraz otwórzmy publikacje internetowe na tej stronie w formacie PDF, a jest ich cztery. Zapoznajmy się w skrócie z ich zawartością, tj. otwórzmy publikacje formatu PDF (zob. rysunek 13.10 oraz 13.11). Po udostępnieniu obszernej listy rezultatu ogólnoświatowa wyszukiwarka jaką jest Google proponuje nam jeszcze wyszukiwanie podobne według fraz pisanych małymi literami:

- metoda wintersa przykład,
- metoda wintersa *excel*,
- metoda holta wintersa *excel*,
- metoda holta przykład,
- model holta wintersa *excel*,
- model wintersa przykład,
- model addytywny i multiplikatywny,
- metoda browna.



Źródło: Opracowanie własne w Google.

[187] Dane liczbowe, wzory i procedury zaczerpnięto z książki: Lipiec-Zajchrowska (red.), *Wspomaganie procesów decyzyjnyc*h Tom II. *Ekonometria*, op. cit., s. 94.

Rys. 13.10. Pierwszy fragment pozycji na pierwszej stronie listy rezultatu wyszukiwania

Źródło: Opracowanie własne w Google.

Rys. 13.11. Drugi fragment pozycji na pierwszej stronie listy rezultatu wyszukiwania

Strona 1/6 publikacji internetowej „*Modele wygładzania wykładniczego do krótkoterminowego prognozowania obciążeń systemów elektroenergetycznych*" podaje streszczenie oraz słowa kluczowe w tym artykule, a autor Grzegorz Dudek podaje:

"*Przedstawiono modele do sporządzania prognoz zapotrzebowania na moc elektryczną. Szeregi czasowe charakteryzują się wieloma wahaniami sezonowymi, co komplikuje model prognostyczny. W celu uproszczenia problemu szeregi czasowe zdekomponowano na szeregi obciążeń w tych samych chwilach doby, co pozwoliło zastosować model wygładzania wykładniczego z pojedynczą sezonowością. Inny rodzaj dekompozycji z wykorzystaniem regresji lokalnej (LOESS) umożliwia użycie modelu o mniejszej złożoności bez składnika sezonowego. Zastosowanie modelu Holta-Wintersa z podwójną sezonowością zwalnia z potrzeby dekompozycji szeregu czasowego. Dokładność proponowanych metod porównano na przykładach aplikacyjnych z dokładnością modeli ARIMA i modelu opartego na sieci neuronowej*".

W publikacji internetowej „*Oprogramowanie na podstawie modeli Holta-Wintersa dla pełnych i niepełnych danych*" Uniwersytetu Ekonomicznego we Wrocławiu na stronie 3/17 znajduje się streszczenie i słowa kluczowe tego artykułu. Autorzy Maria Szmuksta-Zawadzka oraz Jan Zawadzki informuje nas: „*Praca poświęcona jest praktycznym aspektom modeli Holta-Wintersa w prognozowaniu zmiennych wykazujących wahania sezonowe na podstawie pełnych szeregów czasowych oraz szeregów z lukami w danych. Autorzy sformułowali przesłanki wyboru równań do celów prognozowania inter- i ekstrapolacyjnego. Ze względu na zbliżone oceny mierników* $MAPE^{\alpha,\beta,\gamma}$ *do prognozowania należy wybierać kilkanaście modeli różniących się dość znacznie kombinacjami*

stałych wygładzania. Do budowy prognoz ex ante powinny zostać wybrane spośród nich modele charakteryzujące minimalnymi ocenami błędów prognoz ex post".

Kolejna publikacja „*Model Wintersa*" w formacie PDF ma charakter prezentacji - wykładu akademickiego z tego zakresu. Na stronie ¼ zdefiniowano pojęcia występujące w modelu Wintersa.

Otwórzmy teraz artykuł internetowy „*Zastosowanie wybranych modeli adaptacyjnych w prognozowaniu brakujących danych w szeregach ze złożoną sezonowością dla luk niesystematycznych*". Ci sami autorzy co już wcześniej omówiony artykuł poświecili swoją publikację wykorzystaniu wybranych modeli wyrównywania wykładniczego Browna i Holta-Wintersa i streszczają go następująco: „*Prognozy wyjściowe będą budowane na podstawie szeregów oczyszczonych ze sezonowości. Prognozy końcowe uwzględniające wahania sezonowe, będą sumami prognoz wyjściowych i składników sezonowości lub iloczynami prognoz tego rodzaju i wskaźników sezonowości*".

Została nam jeszcze publikacja Macieja Oesterreicha pt. „*Wykorzystanie modelu Holta-Wintersa oraz metod bootstrapowych w prognozowaniu na podstawie zmiennych ekonomicznych z wahaniami sezonowymi"*a we wstępie tej pracy czytamy: „*Modele wyrównywania wykładniczego należą do nieklasycznych modeli szeregu czasowego najczęściej wykorzystywanych w prognozowaniu zmiennych z wahaniami sezonowymi. Związane jest to m.in. z ich elastycznością, tzn. z szybką reakcją na zmiany trendu oraz wahań okresowych. Dodatkową zaleta jest dostępność programów obliczeniowych, pozwalających na szybkie wyznaczanie optymalnych wartości stałych wyrównywania i budowę prognoz. Niekiedy jednak dokładność prognoz dotyczących tych zmiennych jest niewystarczająca. Zwiększenie ich efektywności jest możliwe dzięki wykorzystaniu technik i metod symulacyjnych, w tym metod bootstrapowych*".

14. Modelowanie ekonometryczne z użyciem algebry liniowej macierzy

14.1. Wstęp

W modelowaniu ekonometrycznym regresji liniowej wielowymiarowej w odniesieniu do zadań praktycznych występują długie szeregi obserwacji zmiennych. Zmienną objaśniającą traktować możemy jako wektor kolumnowy, a obserwacje zmiennych objaśniających jako macierz. Dla zaprezentowania formułowania modelu liniowego wielowymiarowego w oparciu o wektor oraz macierz obserwacji określonego zjawiska ekonomicznego posługujemy się równaniem macierzowym:

$$\boldsymbol{a} = (\mathbf{X}^T\mathbf{X})^{-1}\mathrm{X}^T\mathbf{y}$$

gdzie:

a - wektor współczynników stojących przy poszczególnych zmiennych objaśniających,
X - macierz wartości zmiennych objaśniających,
$\mathbf{X}^T$ - macierz transponowana macierzy **X**,
y - wektor obserwacji zmiennej objaśnianej.

Widzimy, że zachodzi potrzeba wprowadzenia do określonego środowiska programowego wektora **y** oraz macierzy **X**. Dokonujemy zatem transpozycji macierzy **X** uzyskując macierz transponowaną $\mathbf{X}^T$. W kolejności musimy wykonać następujące operacje na macierzach:

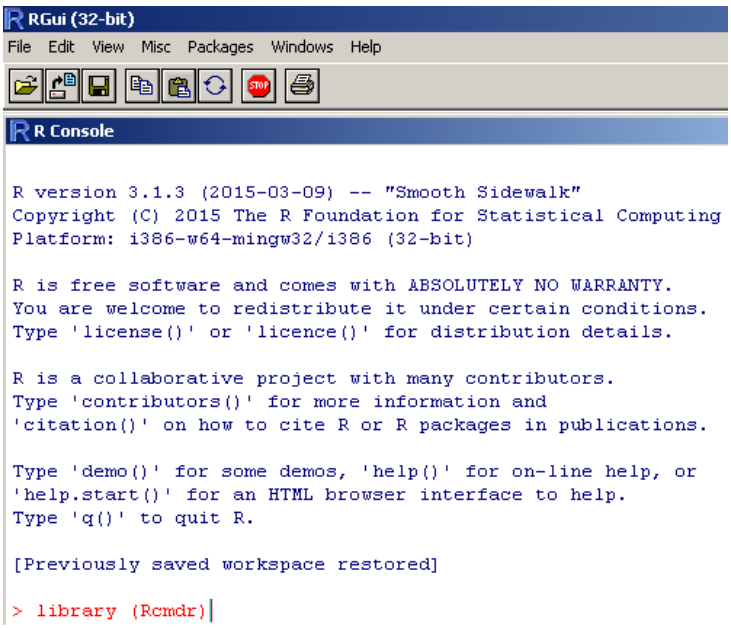
- pomnożyć macierz transponowaną $\mathbf{X}^T$ i macierz **X** uzyskując w wyniku macierz $[\mathbf{X}^T\mathbf{X}]$,
- wyznaczyć macierz odwrotną $[\mathbf{X}^T\mathbf{X}]^{-1}$,
- pomnożyć macierz transponowaną $\mathbf{X}^T$ i wektor **y** uzyskując macierz $[\mathbf{X}^T\mathbf{y}]$,
- pomnożyć macierze wynikowe $[\mathbf{X}^T\mathbf{X}]^{-1}$ i $[\mathbf{X}^T\mathbf{y}]$ uzyskując w efekcie końcowym wektor współczynników **a**.

Rozwiązanie komputerowe możemy uzyskać różnymi programami. Zademonstrowano to na przykładzie funkcji **REGLINP()** i funkcji macierzowych *Excela* oraz przy zastosowaniu funkcji coraz popularniejszego programu *R* stosowanego do metod ilościowych. Spośród funkcji macierzowych *Excela* do naszego równania macierzowego niezbędne są: **MACIERZ.ILOCZYN()** oraz **MACIERZ.ODW()**. Do transpozycji stosuje się menu *Narzędzia główne Excela* w zakresie kopiowania, a następnie wklejenia specjalne z opcją *Transpozycja*. Program *R* do naszych potrzeb proponuje funkcje[188]:

- formułowania wektorów zmiennych i sklejania ich w macierze;
- operator mnożenia macierzy (*);
- funkcje utworzenia macierzy odwrotnej oraz jej sprawdzenia;
- pomocnicze funkcje do obliczenia wyznaczników, wymiaru macierzy, wektora przekątnej diagonalnej;
- funkcje do składania nowych macierzy z innych obiektów tj. macierzy i wektorów.

Skorzystanie z funkcjonalności pakietu *R* wymaga pobrania z serwera lokalnego programu Rcmdr, o ile nie jest jeszcze zainstalowany na naszym komputerze. Sprawdzamy to komendą **library** (**Rcmdr**) - zobacz okno wejściowe pakietu *R* w trybie RGui pokazane na rysunku 14.1.

[188] W opracowaniu zabazowano na formułach funkcji programu *R* zamieszczonych w książce: Kopczewska K., Kopczewski T., Wójcik P., *Metody ilościowe w R: aplikacje ekonomiczne i finansowe,* CeDeWu, Warszawa 2009, rozdz. 17.3. Podstawowe operacje na macierzach.

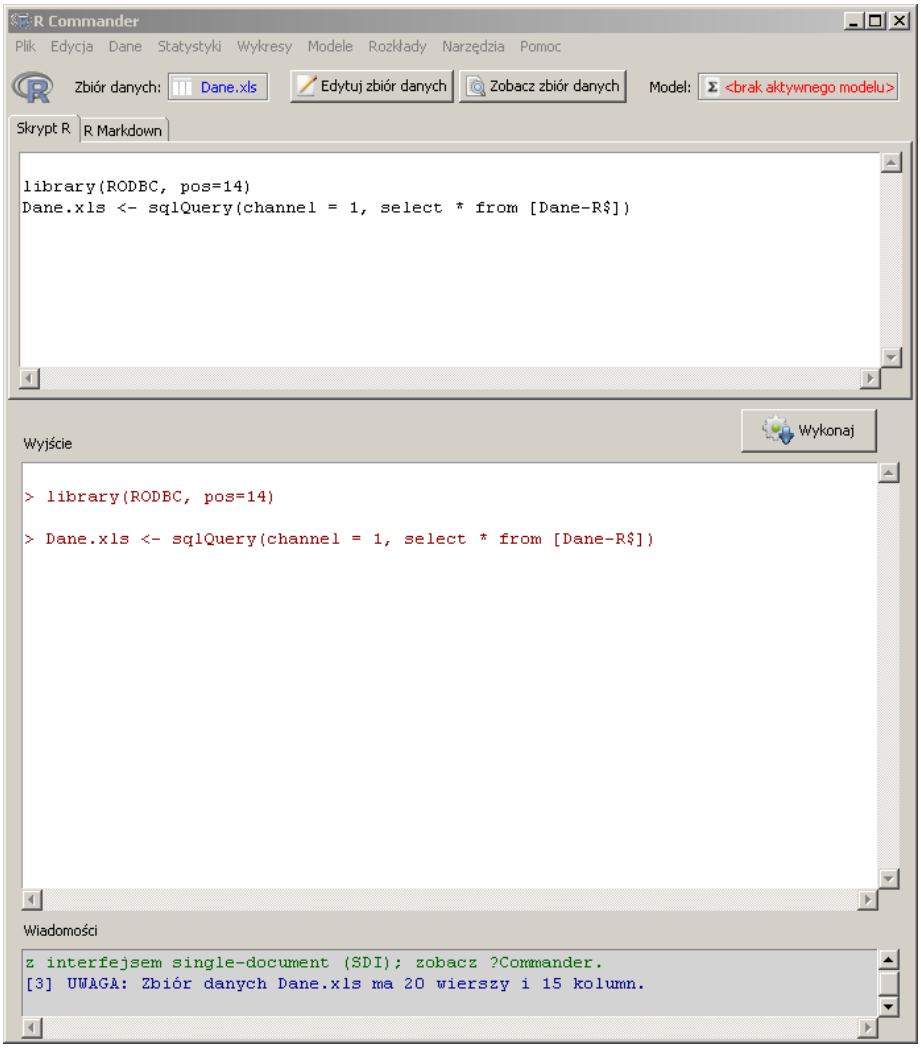


Źródło: Opracowanie własne.

Rys. 14.1. Zapytanie programowe dotyczące zainstalowania programu Rcmdr w bibliotece

Wywołanie programu Rcmdr skutkuje zmianą trybu współpracy ze skryptowej na windowsową (okienkową) obsługiwaną przez program R Commander. Na ekranie pojawiają się trzy podokna: *SkryptR*, *Wyjście*, *Wiadomości*. Korzystamy z menu i możemy obserwować generowanie przez *R* kodu źródłowego w podoknie *SkryptR*. Po akceptacji menu występuje reakcja programu w podoknie *Wyjście*. Uwagi o poprawności lub usterkach pojawiają się w podoknie *Wiadomości*. Tryb windowsowy stosowany jest tylko do zawężonej funkcjonalności *R*, jest jednak pomocny do nauki pisania skryptów (kodu źródłowego) i symulacji zachowania pakietu *R* przy wywołaniu określonej funkcji menu.

Menu główne *R Commander* obejmuje grupy funkcji: *Plik*, *Edycja*, *Dane*, *Statystyki*, *Wykresy*, *Modele*, *Rozkłady*, *Narzędzia*, *Pomoc* (zob. rysunek 14.2). W zakresie zbioru danych *R Commander* informuje o bieżącej aktywności zbioru oraz daje możliwość jego edycji lub podglądu. Pojawia się również informacja o zastosowanym modelu w danej sesji obliczeniowej.



Źródło: Opracowanie własne z zastosowaniem programu *R*

Rys. 14.2. Okno trybu windowsowego

14.2. Estymacja parametrów modelu zmiennej X_3 w środowisku *R*

Do naszej pracy testowej funkcji macierzowych skorzystamy ze zbioru autora niniejszej pracy założonego wcześniej podczas pisania książki *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*[189]. W tej publikacji efektem prac z udziałem programu *Gretl* był model:

$$x_3 = a_0 + a_1 x_9 + a_2 x_{25} + a_3 x_{78}$$

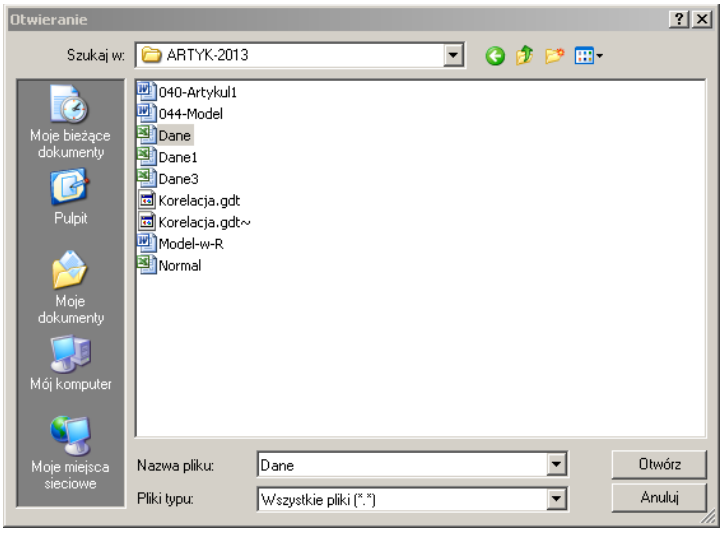
gdzie:

x_3 - liczba bezrobotnych;
x_9 - przeciętne zatrudnienie, w tys.: przemysł;
x_{25} - koszty pracy na jednego zatrudnionego, przeciętne miesięczne,;
x_{78} - liczba zawartych małżeństw na 1000 ludności.

Wymieniony model o współczynniku determinacji R^2 = 0,938085 (dobroci dopasowania modelu do danych empirycznych) powstał w wyniku wielokrotnego powtórzenia procedury *a posteriori* w programie *Gretl* spośród 78 zmiennych, przy czym zmienną objaśnianą jest X_3. Wymieniona procedura polega na realizacji zadania poprzez eliminowanie zmiennych o największym prawdopodobieństwie popełnienia błędu przez ich udział w modelu. Sygnalizowane jest to przez *Gretl* w postaci pustej informacji lub jako jedna gwiazdka, dwie gwiazdki. Występowanie trzech

[189] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, część II, Wydawnictwo Instytut Śląski, Opole 2015, rozdz. 2.15. *Budowa i weryfikacja modelu ekonometrycznego w programie R.*

gwiazdek wskazuje na właściwe dobranie określonej zmiennej do modelu ekonometrycznego. Zmienne dotyczące lat 1992-2011 pobrane zostały z „*Roczników statystycznych*" GUS-u. Z pozycji menu programu *R Commander* wywołujemy zbiór *Dane.xls*, który zapisany jest na ścieżce: *D:\ARTYK-2013* (zob. rysunku 14.3).



Źródło: Opracowanie własne.

Rys. 14.3. Wywołanie pliku źródłowego *Dane.xls*

Korzystamy z menu: *Dane/Importuj danych/z pliku Excel… Dalsze stosowanie* pliku *Dane.xls* w *R Commander* wymaga jego uaktywnienia stosując menu: *Dane/Aktywny zbiór danych/Wybierz aktywny zbiór danych/Dane.xls* i OK. Po wczytaniu pliku *Dane.xls* następuje reakcja programu *R* w podoknach trybu windowsowego. Po akceptacji biblioteki i wywołanie zbioru danych, co zaznaczone jest w *Zbiór danych: Dane.xls* program informuje nas w *Wiadomości* o wymiarach zbioru pośredniego kolejnej iteracji jeszcze przed ostateczną selekcją *a posteriori* obejmującego 20 wierszy i 15 kolumn zmiennych (zob. rysunek 14.2). Dane liczbowe zmiennych (lata 1992-2011) przedstawiono na rysunku 14.4.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	X3	X9	X10	X11	X12	X13	X14	X25	X38	X48	X49	X50	X51	X59	X78
2	2394	3467,6	823,0	1042,1	794,7	476,7	287,7	295,16	138,9	404,7	1921,9	18430	4626,8	41,1	5,7
3	2595	3391,8	721,9	999,4	746,5	381,1	239,3	384,25	134,8	550,3	1856,5	18206	4504,2	40,1	5,4
4	2375	3361,4	678,8	996,1	728,1	366,4	235,4	515,49	134,5	708,3	1750,0	18206	4472,6	45,8	5,4
5	2233	3461,1	689,2	1078,6	723,4	414,4	243,5	704,99	131,6	878,3	1648,3	18075	4775,1	59,2	5,4
6	1961	3436,0	684,2	1119,7	721,9	425,3	250,5	900,58	126,7	1113,0	1565,0	19224	4258,9	70,5	5,3
7	1737	3433,4	723,5	1244,8	730,7	487,4	273,3	1088,03	121,9	1300,0	1514,7	19771	4437,4	77,4	5,3
8	1827	3378,7	736,9	1321,1	725,2	552,5	253,0	1277,06	115,7	1527,5	1452,6	20288	4243,5	83,3	5,4
9	2641	3138,4	710,4	1318,4	686,3	583,0	286,9	1428,01	112,5	1700,5	1406,5	20800	4138,6	86,4	5,7
10	2760	2955,0	661,9	1325,0	654,9	614,2	286,7	2622,76	111,1	1904,6	1360,3	21037	3750,8	78,4	5,5
11	3186	2820,6	627,8	1295,6	530,4	637,3	252,1	2821,30	108,0	1897,6	1310,4	21376	3905,8	75,1	5,0
12	3431	2670,5	545,5	1291,1	601,9	671,3	260,0	2930,70	102,6	1886,1	1284,3	21659	3708,1	85,7	5,0
13	3329	2639,1	496,4	1308,0	586,3	699,5	261,8	3017,64	104,2	1918,7	1268,6	22052	3877,2	78,4	5,1
14	3230	2663,1	453,1	1295,4	583,0	700,6	264,6	3111,04	104,0	1966,7	1229,1	22804	3779,1	98,0	5,0
15	3045	2665,4	483,6	1360,7	577,7	695,8	265,2	3252,02	103,8	1954,2	1219,4	26565	3855,3	104,3	5,4
16	2344	2714,3	511,5	1394,2	596,7	723,3	267,6	3421,05	104,9	2106,3	1221,5	27547	3967,0	88,8	5,9
17	1619	2842,4	576,8	1482,8	621,0	780,3	277,0	3720,10	107,9	2290,8	1280,1	27713	3807,0	84,7	6,5
18	1211	2890,7	671,2	1633,5	792,0	743,1	406,7	3986,50	110,1	2558,5	1212,2	28425	3792,2	80,6	6,8
19	1411	2725,4	706,6	1611,5	777,5	740,1	407,4	4232,68	105,4	2706,5	1194,7	28684	3873,6	80,6	6,6
20	1699	2696,1	674,4	1598,7	758,2	767,0	412,0	4388,31	103,9	2805,5	1197,9	29774	4237,3	81,5	5,9
21	2146	2714,2	705,2	1580,3	787,5	775,1	412,2	4633,95	105,6	2872,2	1202,0	29383	3857,1	82,0	5,4

Źródło: Opracowanie własne w *Excelu*.

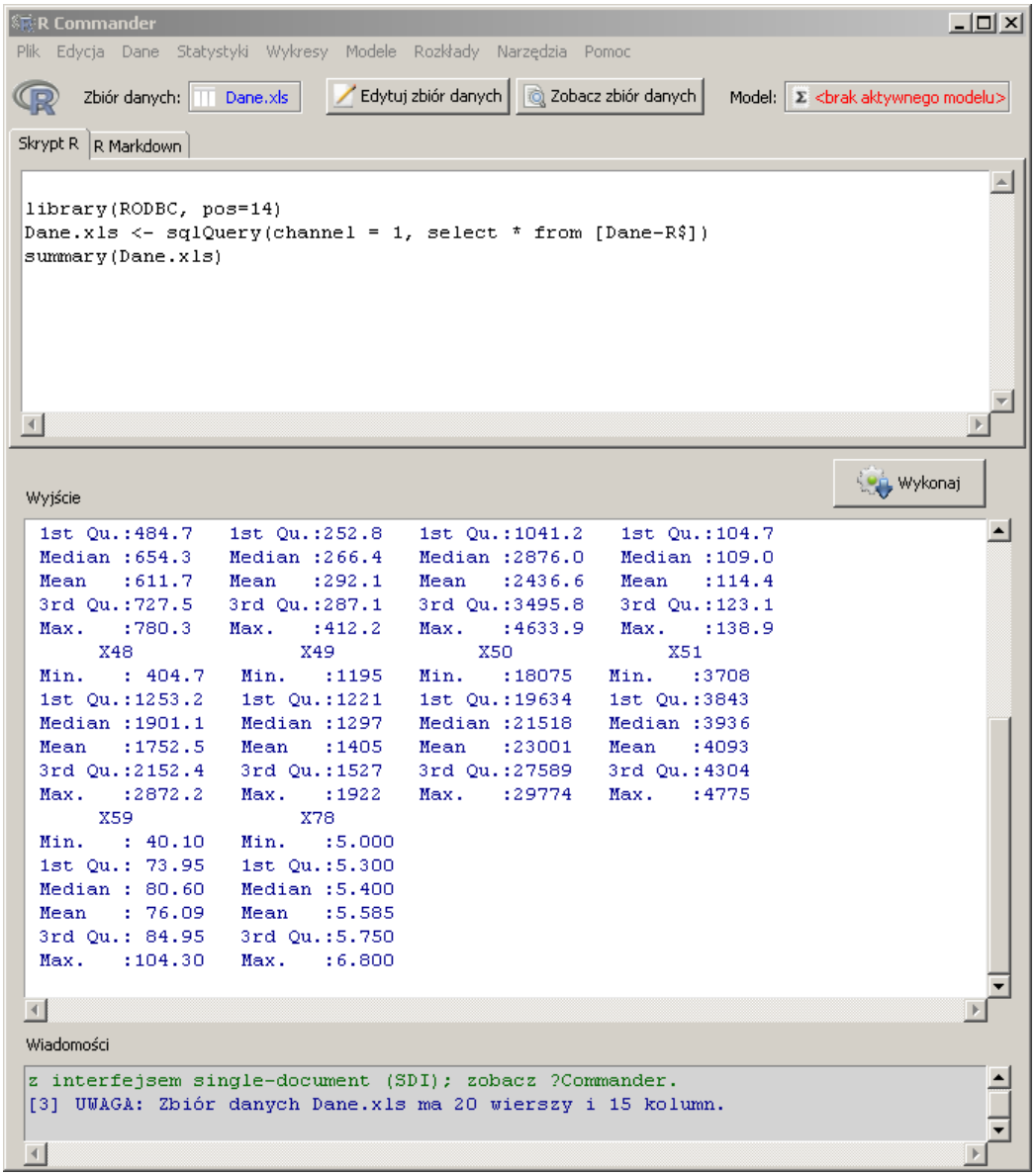
Rys. 14.4. Dane źródłowe pliku *Dane.xls*

Mając aktywną tabelę danych w programie *R Commander* korzystamy z menu *Statystyki/Podsumowania/Aktywny zbiór danych* w celu wykonania raportu statystyki opisowej, który

generowany jest w oknie skryptowym komendą **summary(DATA)**. Statystyka ta ukazuje się w oknie *Wyjście* i obejmuje następujące cechy naszych 15 zmiennych[190]:

(*Min*) - minimum,
(1 *st Qu*.) - kwartyl 1.,
(*Median*) - mediana,
(*Mean*) - średnia arytmetyczna,
(3 *rd Qu*.) - kwartyl 3.,
(*Max*.) - maksimum.

Wartości cech podawane są z kropką dziesiętna zamiast przecinka. Fragment widoku statystyki opisowej pokazano na rysunku 14.5.



Źródło: Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządz*ania, Wydawnictwo Instytut Śląski, Opole 2015, wykres 2.15.4.

Rys. 14.5. Odpowiedz programu *R Commander* na menu *Statystyka/Podsumowania/Aktywny zbiór danych*

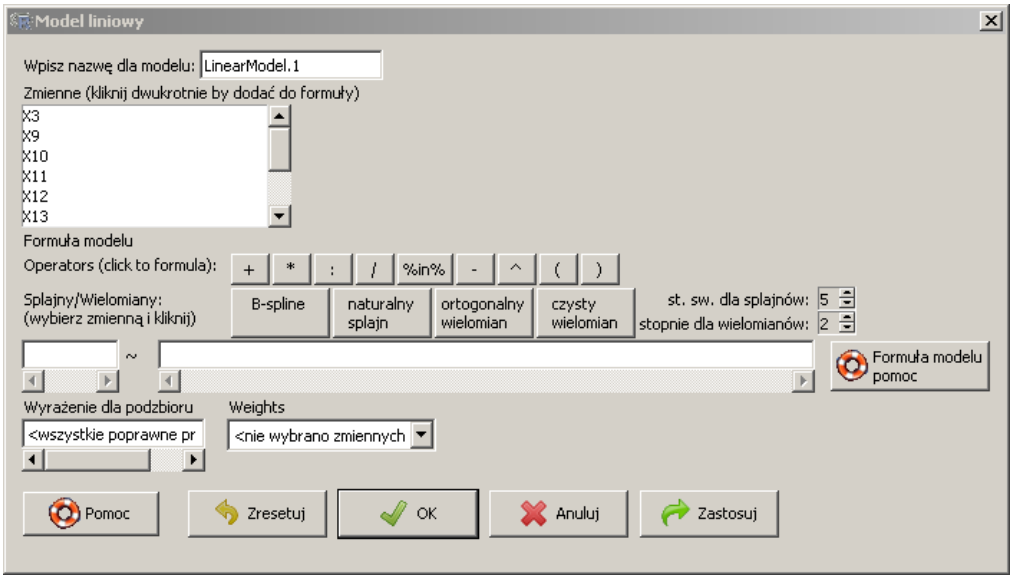
[190] Ta część opracowania nawiązuje do podrozdziału 2.15.3. *Przygotowanie do realizacji w programie R*, książki: Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, op. cit.

Następnie korzystamy z menu: *Modele* programu *R Commander* i inicjujemy sporządzanie modelu liniowego zależności zmiennej X_3 od wyselekcjonowanych przez *Gretl* zmiennych objaśniających X_9, X_{25}, X_{78}. Pojawia się okno dialogowe w którym podajemy opcje modelu o n/w postaci ogólnej, w tym zmienne uczestniczące w formule regresji wielowymiarowej z uwzględnieniem operatorów „*" oraz „+":

$$X3 = a_0 + a_1\ x\ X9 + a_2\ x\ X25 + a_3\ x\ X78$$

Zmienne dla potrzeb komputera zapisane są jako *X*3, *X*9, *X*25, *X*78 (zob. rysunek 14.6). Potwierdzenie podanych opcji przyciskiem OK, powoduje odpowiednie reakcje w podoknach windowsowych, a w tym wygenerowanie komend kodu źródłowego:

```
LinearModel.1<-lm(X3~X9+X25+X78, dane=Dane.xls)
Summary(LinearModel.1)
```



Źródło: Opracowanie własne w środowisku *R Commandera*.

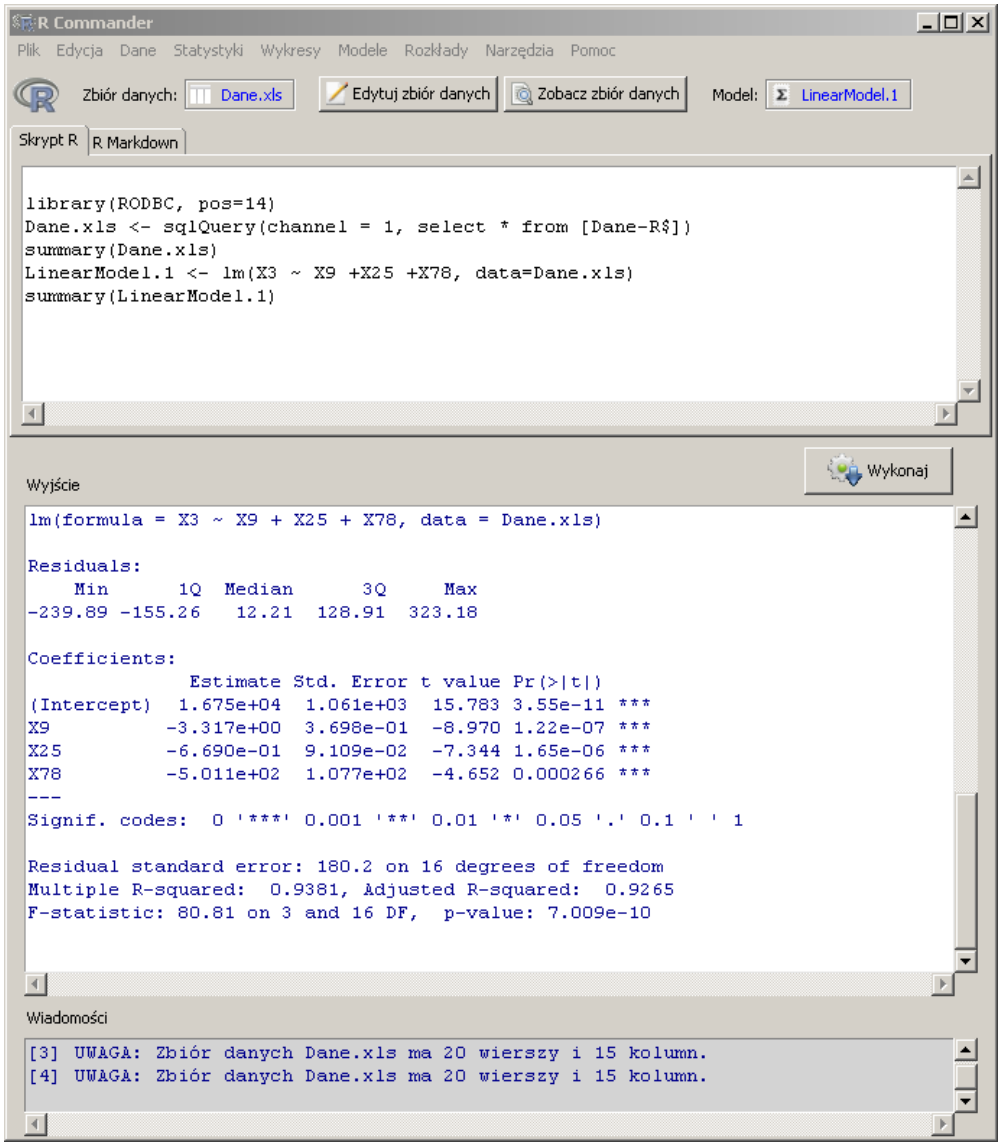
Rys. 14.6. Okno dialogowe deklaracji modelu liniowego wielowymiarowego

Odpowiedzią komendy **lm()** oraz **summary()** jest oszacowanie współczynników modelu zmiennej X_3 w postaci notacji komputerowej w formie inżynierskiej, przy czym *e* = 10:

$$X3 = 1.675e{+}04 - 3.317e{+}00\ X9 - 6.690e{-}01\ X25 - 5.011e{+}02\ X78$$

co odpowiada zapisowi: $x_3 = 16750 - 3{,}317\ x_9 - 0{,}669\ x_{25} - 0{,}05011\ x_{78}$.

Zwróćmy uwagę na dobre dopasowanie modelu do danych empirycznych zmiennej X_3 wynoszące: $R^2 = 0{,}9381$. Ponadto dobór zmiennych jest najlepszy z możliwych - świadczą o tym trzy gwiazdki (***) przy wierszach estymacji zmiennych. W oknie windowsowym pojawił się model: *LinearModel.1* (zob. rysunek 14.7).



Źródło: Opracowanie własne programem *R*.

Rys. 14.7. Wyniki pracy komend **lm()** i **summary()**

Alternatywnym podejściem - jak już nadmieniono - jest wpisywanie komend do okna skryptowego RGui. Ze względu na pamiętanie przez *R* naszych operacji i utworzonych obiektów tworzymy kolejny model o nazwie „*LinearModel.2*" wpisując podobnie jak wcześniej komendę:

LinearModel2<-**lm**(X3~X9+X25+X78,data=Dane.xls).

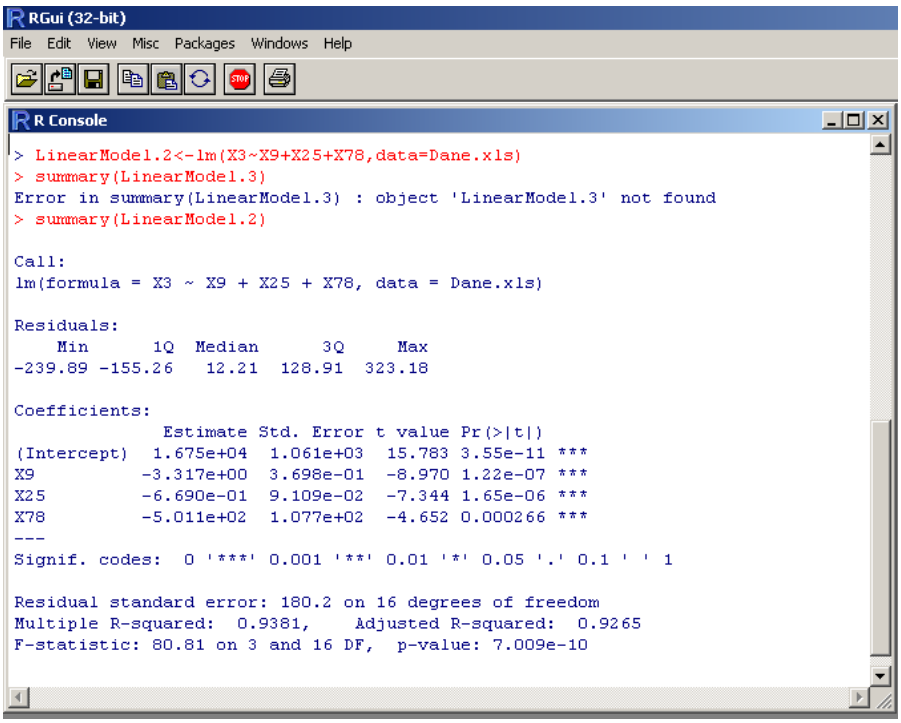
Program wykonuje obliczenia, a ich efekt zobaczymy po wywołaniu komendy:

summary(LinearModel.2).

Komenda **lm()** określa statystykę opisową błędów między modelem, a danymi rzeczywistymi zmiennej *X*3 w wierszu *Residual*. Zauważmy również reakcję programu *R* na wprowadzoną celowo błędną nazwę modelu „*LinearModel3*" oraz wywołanie komendy jako:

lm(formula = X3 ~ X9 + X25 + X78, data = Dane.xls).

Efekt estymacji poszczególnych parametrów modelu zmiennej X_3 jest identyczny jak wcześniej wykonany w trybie okienkowym pod *R Commanderem* (zob. rysunek 14.8).



Źródło: Opracowanie własne programem *R*.

Ryc. 14.8. Wyniki pracy komend **lm()** i **summary()** w trybie skryptowym (okno RGui)

14.3. Zastosowanie funkcji macierzowych programu *R*

W prezentacji operacji algebry macierzowej wykorzystamy szeregi obserwacji wynikające z naszego ostatniego wysublimowanego modelu zależności zmiennej objaśnianej X_3 od zmiennych objaśniających $\boldsymbol{X}_9$, $\boldsymbol{X}_{25}$ oraz $\boldsymbol{X}_{78}$. Inspiracją do naszych działań będzie równanie macierzowe, które już we wstępie zostało przytoczone: $\mathbf{a} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ Dla wygody korzystania z szeregów danych zapisano je w kolejnym arkuszu pliku *Dane.xls* (zob. rysunek 14.9).

Na początek skorzystamy z trybu skryptowego programu *R* i w oknie RGui zapisywać będziemy komendy operacji macierzowych wynikające z podanego wcześniej równania macierzowego. Tym razem testujemy wprowadzenie macierzy jako zbioru 20 wektorów odpowiadających zmiennym: *X*9, *X*25, *X*78.

	A	B	C	D
1	X3	X9	X25	X78
2	2394	3467,6	295,16	5,7
3	2595	3391,8	384,25	5,4
4	2375	3361,4	515,49	5,4
5	2233	3461,1	704,99	5,4
6	1961	3436,0	900,58	5,3
7	1737	3433,4	1088,03	5,3
8	1827	3378,7	1277,06	5,4
9	2641	3138,4	1428,01	5,7
10	2760	2955,0	2622,76	5,5
11	3186	2820,6	2821,30	5,0
12	3431	2670,5	2930,70	5,0
13	3329	2639,1	3017,64	5,1
14	3230	2663,1	3111,04	5,0
15	3045	2665,4	3252,02	5,4
16	2344	2714,3	3421,05	5,9
17	1619	2842,4	3720,10	6,5
18	1211	2890,7	3986,50	6,8
19	1411	2725,4	4232,68	6,6
20	1699	2696,1	4388,31	5,9
21	2146	2714,2	4633,95	5,4

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.9. Szeregi obserwacji zmiennych modelu wynikowego zmiennej $\boldsymbol{X}_3$

Sekwencja kodu źródłowego (skryptu *R*), dla macierzy nazwanej jako MAT1 jest następująca:

```
> # definiowanie macierzy MAT1 i jej elementów
> MAT1[1,]<-c(3467.6,295.16,5.7)
> MAT1[2,]<-c(3391.8,384.25,5.4)
> MAT1[3,]<-c(3361.4,515.49,5.4)
> MAT1[4,]<-c(3461.1,704.99,5.4)
> MAT1[5,]<-c(3436.0,900.58,5.3)
> MAT1[6,]<-c(3433.4,1088.03,5.3)
> MAT1[7,]<-c(3378.7,1277.06,5.4)
> MAT1[8,]<-c(3138.4,1428.01,5.7)
> MAT1[9,]<-c(2955.0,2622.76,5.5)
> MAT1[10,]<-c(2820.6,2821.30,5.0)
> MAT1[11,]<-c(2670.5,2930.70,5.0)
> MAT1[12,]<-c(2639.1,3017.64,5.1)
> MAT1[13,]<-c(2663.1,3111.04,5.0)
> MAT1[14,]<-c(2665.4,3252.02,5.4)
> MAT1[15,]<-c(2714.3,3421.05,5.9)
> MAT1[16,]<-c(2842.4,3720.10,6.5)
> MAT1[17,]<-c(2890.7,3986.50,6.8)
> MAT1[18,]<-c(2725.4,4232.68,6.6)
> MAT1[19,]<-c(2696.1,4388.31,5.9)
> MAT1[20,]<-c(2714.2,4633.95,5.4)
```

Dla wstępnego testowania oraz dla wykorzystania modelu do ewentualnej optymalizacji jako funkcja celu w zadaniu decyzyjnym, wprowadzone wektory nie obejmują stałej a_0 tożsamościowo równej „1" przy zmiennej nazywanej jako X_0. Poszczególne wektory „c" macierzy **MAT1** zapisujemy wierszami od 1-20, zastępując przecinki kropką dziesiętną, oddzielając dane przecinkiem. Kolejne komendy akceptujemy klawiszem *Enter*. Dla wyświetlenia wprowadzonych danych wystarczy w linii komend wpisać tylko nazwę macierzy **MAT1**[191].

```
> MAT1 # wyświetlenie wyniku
        [,1]    [,2] [,3]
 [1,] 3467.6  295.16  5.7
 [2,] 3391.8  384.25  5.4
 [3,] 3361.4  515.49  5.4
 [4,] 3461.1  704.99  5.4
 [5,] 3436.0  900.58  5.3
 [6,] 3433.4 1088.03  5.3
 [7,] 3378.7 1277.06  5.4
 [8,] 3138.4 1428.01  5.7
 [9,] 2955.0 2622.76  5.5
[10,] 2820.6 2821.30  5.0
[11,] 2670.5 2930.70  5.0
[12,] 2639.1 3017.64  5.1
[13,] 2663.1 3111.04  5.0
[14,] 2665.4 3252.02  5.4
[15,] 2714.3 3421.05  5.9
[16,] 2842.4 3720.10  6.5
[17,] 2890.7 3986.50  6.8
[18,] 2725.4 4232.68  6.6
[19,] 2696.1 4388.31  5.9
[20,] 2714.2 4633.95  5.4
```

Zauważmy, sposób notacji w nawiasach kwadratowych kolejnych wierszy oraz kolumn, poprzez wstawianie przecinka, gdy dalsze kolumny/wiersze są takie jak w danym wierszu/kolumnie np. [1,], [1,]. Pierwszym naszym krokiem jest transpozycja (czyli zamiana wierszy na kolumny) macierzy **MAT1** stosując komendę **t()**. Efekt przeprowadzenia transpozycji uzyskujemy poprzez wpisanie w linii komend **MAT**1 pełniacego rolę komendy.

[191] Wprowadzenie danych oraz formułowanie funkcji operacji macierzowych opracowano przez analogię do danych modelowych zamieszczonych w książce: Kopczewska K., Kopczewski T., Wójcik P., *Metody ilościowe w R: Aplikacje ekonomiczne i finansowe,* op. cit., strony: 584-588.

```
> # definiuje MAT1 jako przetransponowaną macierz MAT1
> MAT2<-t(MAT1)
> MAT2
        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
[1,] 3467.60 3391.80 3361.40 3461.10 3436.00 3433.40 3378.70 3138.40 2955.00
[2,]  295.16  384.25  515.49  704.99  900.58 1088.03 1277.06 1428.01 2622.76
[3,]    5.70    5.40    5.40    5.40    5.30    5.30    5.40    5.70    5.50
      [,10]  [,11]   [,12]   [,13]   [,14]   [,15]  [,16]  [,17]   [,18]
[1,] 2820.6 2670.5 2639.10 2663.10 2665.40 2714.30 2842.4 2890.7 2725.40
[2,] 2821.3 2930.7 3017.64 3111.04 3252.02 3421.05 3720.1 3986.5 4232.68
[3,]    5.0    5.0    5.10    5.00    5.40    5.90    6.5    6.8    6.60
       [,19]   [,20]
[1,] 2696.10 2714.20
[2,] 4388.31 4633.95
[3,]    5.90    5.40
```

Macierze **MAT1** i **MAT2** odpowiadają w równaniu macierzowym macierzom **X** oraz $\mathbf{X}^T$. Musimy teraz dokonać przemnożenia macierzy **MAT2** i **MAT1** uzyskując macierz **MAT3**, czyli $[\mathbf{X}^T\mathbf{X}]$. W wyniku mnożenia macierzy o wymiarach $(\mathbf{X}^T\mathbf{X})$ = (3 x 20) (20 x 3) następuje redukcja do (3 x 3) w macierzy **MAT**3:

```
> # macierz MAT3 jest efektem mnożenia macierzy MAT2 i MAT1
> MAT3<-MAT2%*%MAT1
> MAT3
             [,1]        [,2]      [,3]
[1,] 182513417.9 137905436.2 335038.33
[2,] 137905436.2 159531664.7 277937.71
[3,]    335038.3    277937.7    629.05
```

Jak widać, mnożenie macierzy dokonujemy podając we właściwej kolejności nazwy macierzy **MAT2**, **MAT**1 oraz operator mnożenia „*****” otoczony znakami „%”.

Przystępujemy teraz do mnożenia macierzy transponowanej **MAT2** i wektora zmiennej objaśnianej X_3 uzyskując macierz **MAT4**. Wymaga to jednak wcześniejszego wprowadzenia danych zmiennej X_3 dla obserwacji (1-20). Zauważmy reakcje programu *R* na nie podanie na początku linii tekst symbolu „**#**”. W wyniku redukcji wymiarów: (3 x 20) x (20 x1) powstaje wektor kolumnowy - macierz **MAT4** o wymiarach (3 x 1). Obraz wyniku uzyskujemy przez wpisanie w linii komend **MAT4**.

```
> y<-c(2394,2595,2375,2233,1961,1737,1827,2641,2760,3186,3431,3329,3230,3045,2344,1619,1211,1411,1699,2146)
> macierz MAT4 jako wynik z przemnożenia macierzy MAT2 i wektora y
Error: unexpected symbol in "macierz MAT4"
> # macierz MAT4 jako wynik z przemnożenia macierzy MAT2 i wektora y
> MAT4<-MAT2%*%y
> MAT4
            [,1]
[1,] 140502253.8
[2,] 112784218.0
[3,]    258409.8
```

Mamy już składniki z przemnożeń macierzy, co odpowiada w równaniu macierzowym $(\mathbf{X}^T\mathbf{X})^{-1}$ oraz $\mathbf{X}^T\mathbf{y}$, przechodzimy więc do wyznaczenia komendą **solve()** programu R macierzy odwrotnej **MAT.odw** odpowiadającej $(\mathbf{X}^T\mathbf{X})^{-1}$.

```
> # wyszukiwanie macierzy odwrotnej do macierzy MAT3
> MAT.odw<-solve(MAT3)
> MAT.odw
                [,1]          [,2]          [,3]
[1,]  7.834649e-07  2.160214e-07 -0.0005127275
[2,]  2.160214e-07  8.678951e-08 -0.0001534020
[3,] -5.127275e-04 -1.534020e-04  0.3424522271
```

Wektor parametrów zwanych także współczynnikami modelu regresji wielowymiarowej zmiennej X_3 uzyskujemy poprzez kolejne przemnożenie, ale tym razem macierzy **MAT.odw**, tj. $(\mathbf{X}^T\mathbf{X})$-1 i **MAT4**, czyli $\mathbf{X}^T\mathbf{y}$:

```
> # przemnożenie macierzy odwrotnej MAT.odw przez MAT4
> a<-MAT.odw%*%MAT4
> a
              [,1]
[1,]     1.9485652
[2,]     0.4994009
[3,] -847.6854880
```

Zatem model zmiennej X_3 - bez stałej a_0 jest następujący:

$$\widehat{x_3} = 1{,}95\ x\ x_9 + 0{,}5\ x\ x_{25} - 847{,}69\ x\ x_{78}$$

Długa jest procedura korzystania z równania macierzowego zanim dojdziemy do modelu regresji wielowymiarowej, przydatna owszem w przypadku macierzy o dużych wymiarach. Dla małych wymiarów macierzy (m x n) możemy z powodzeniem zastosować funkcję *REGLINP Excela*:

a_3	a_2	a_1	
-847,6855	0,4994	1,94857	parametry
416,40456	0,20963	0,62983	błąd standardowy
0,9280663	711,566	#N/D!	R^2; S_e
73,109623	17	#N/D!	F; iss
111052097	8607557	#N/D!	ESS; RSS

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.10. Oszacowanie parametrów modelu bez stałej funkcją *REGLINP Excela*

Uzyskaliśmy zatem zgodność wyników z modelem wygenerowanym przez komendę **lm()** programu *R*, a poszczególne określenia wymienione na rysunku 14.10 oznaczają:

- parametry (współczynniki) - a_1, a_2, a_3 - modelu odpowiadające kolejnym zmiennym objaśniającym, w naszym przykładzie X_9, X_{25}, X_{78};
- błąd standardowy – dotyczy błędów standardowych oszacowania odpowiadających im parametrów;
- R^2 - współczynnik determinacji;
- S_e - odchylenie standardowe reszt modelu;
- F - statystyka Fishera;
- *iss* - ilość stopni swobody;
- *ESS* - estymowana suma kwadratów;
- *RSS* - resztowa suma kwadratów;
- *#N/D!* - pole nie występujące.

14.4. Wprowadzenie parametru stałego do równania liniowego modelu ekonometrycznego

Przetestujmy teraz funkcję *REGLINP* oraz komendę **lm()** programu *R* w zakresie estymacji parametrów modelu ekonometrycznego z parametrem a_0. Formuła funkcji *REGLINP* jest następująca:

=REGLINP(zakres zmiennej objaśnianej; zakres macierzy zmiennych objaśniających; potwierdzenie stałej słowem „PRAWDA”; potwierdzenie potrzeby statystyki opisowej też słowem „PRAWDA”).

Dla naszych danych zapisanych w arkuszu kalkulacyjnym *Excela* (zob. rysunek 14.9) powyższa formuła przybiera postać:

=REGLINP(A2:A21;B2;D21;PRAWDA;PRAWDA).

Tą funkcję potwierdzamy naciskając jednocześnie 3 klawisze: <Ctrl>+<Shift>+<Enter>. Efekt naszego działania prezentuje rysunek 14.11, a uzyskany model przy zaokrągleniu wyników do dwóch miejsc po przecinku zmiennej X_3 jest następujący:

$$x_3 = 16749 - 3{,}32\ x\ x_9 - 0{,}67\ x\ x_{25} - 501{,}06\ x\ x_{78}$$

a_3	a_2	a_1	a_0	
-501,0594	-0,669	-3,317	16749	parametry
107,70829	0,09109	0,36978	1061,19	błąd standardowy
0,9380849	180,189	#N/D!	#N/D!	R^2; S_e
80,806115	16	#N/D!	#N/D!	F; iss
7870851,4	519489	#N/D!	#N/D!	ESS; RSS

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.11. Oszacowanie parametrów modelu ze stałą funkcją *REGLINP Excela*

Wykonajmy teraz estymację parametrów modelu ze stałą komendami programu *R*. Wymaga to kopiowania komend z poprzedniej wersji skryptu zachowanej w naszym katalogu np. na ścieżce: D:\Artykul-R-macierz1. Wprowadzimy do 20 wektorów macierzy **MAT1A** definiowanej komendą **matrix()** liczbę 1 wyrażająca parametr przy zmiennej X_0 tożamościowo równej też jeden. W ten sposób macierz składająca się z 20 wektorów „*sklejona*" komendą **matrix()** będzie o wymiarach (20 x 4) - zob. sekwencja kodu źródłowego podana w języku *R*:

```
> # skorygowane definiowanie macierzy MAT1A i jej elementów
> MAT1A<-matrix(0,20,4)
> MAT1A[1,]<-c(1,3467.6,295.16,5.7)
> MAT1A[2,]<-c(1,3391.8,384.25,5.4)
> MAT1A[3,]<-c(1,3361.4,515.49,5.4)
> MAT1A[4,]<-c(1,3461.1,704.99,5.4)
> MAT1A[5,]<-c(1,3436.0,900.58,5.3)
> MAT1A[6,]<-c(1,3433.4,1088.03,5.3)
> MAT1A[7,]<-c(1,3378.7,1277.06,5.4)
> MAT1A[8,]<-c(1,3138.4,1428.01,5.7)
> MAT1A[9,]<-c(1,2955.0,2622.76,5.5)
> MAT1A[10,]<-c(1,2820.6,2821.30,5.0)
> MAT1A[11,]<-c(1,2670.5,2930.70,5.0)
> MAT1A[12,]<-c(1,2639.1,3017.64,5.1)
> MAT1A[13,]<-c(1,2663.1,3111.04,5.0)
> MAT1A[14,]<-c(1,2665.4,3252.02,5.4)
> MAT1A[15,]<-c(1,2714.3,3421.05,5.9)
> MAT1A[16,]<-c(1,2842.4,3720.10,6.5)
> MAT1A[17,]<-c(1,2890.7,3986.50,6.8)
> MAT1A[18,]<-c(1,2725.4,4232.68,6.6)
> MAT1A[19,]<-c(1,2696.1,4388.31,5.9)
> MAT1A[20,]<-c(1,2714.2,4633.95,5.4)
```

Sprawdzenie poprawności uzyskamy czytając wprowadzone dane poprzez wpisanie w linii komend nazwy macierzy **MAT1A**.

```
> MAT1A # wyświetlenie wyniku skorygowanej macierzy
      [,1]   [,2]    [,3] [,4]
 [1,]    1 3467.6  295.16  5.7
 [2,]    1 3391.8  384.25  5.4
 [3,]    1 3361.4  515.49  5.4
 [4,]    1 3461.1  704.99  5.4
 [5,]    1 3436.0  900.58  5.3
 [6,]    1 3433.4 1088.03  5.3
 [7,]    1 3378.7 1277.06  5.4
 [8,]    1 3138.4 1428.01  5.7
 [9,]    1 2955.0 2622.76  5.5
[10,]    1 2820.6 2821.30  5.0
[11,]    1 2670.5 2930.70  5.0
[12,]    1 2639.1 3017.64  5.1
[13,]    1 2663.1 3111.04  5.0
[14,]    1 2665.4 3252.02  5.4
[15,]    1 2714.3 3421.05  5.9
[16,]    1 2842.4 3720.10  6.5
[17,]    1 2890.7 3986.50  6.8
[18,]    1 2725.4 4232.68  6.6
[19,]    1 2696.1 4388.31  5.9
[20,]    1 2714.2 4633.95  5.4
```

Mając zmodyfikowaną macierz **MAT1A** powtarzamy naszą już sprawdzoną procedurę wynikająca z równania macierzowego. W pierwszej kolejności dokonujemy przetransformowania macierzy **MAT1A** w macierz **MAT2A**:

```
> # definiuje MAT2A jako przetransponowaną macierz MAT1A
> MAT2A<-t(MAT1A)
> MAT2A
        [,1]    [,2]    [,3]    [,4]    [,5]    [,6]    [,7]    [,8]    [,9]
[1,]    1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00    1.00
[2,] 3467.60 3391.80 3361.40 3461.10 3436.00 3433.40 3378.70 3138.40 2955.00
[3,]  295.16  384.25  515.49  704.99  900.58 1088.03 1277.06 1428.01 2622.76
[4,]    5.70    5.40    5.40    5.40    5.30    5.30    5.40    5.70    5.50
       [,10]   [,11]   [,12]   [,13]   [,14]   [,15]  [,16]  [,17]   [,18]
[1,]    1.0     1.0     1.00    1.00    1.00    1.00    1.0    1.0     1.00
[2,] 2820.6 2670.5 2639.10 2663.10 2665.40 2714.30 2842.4 2890.7 2725.40
[3,] 2821.3 2930.7 3017.64 3111.04 3252.02 3421.05 3720.1 3986.5 4232.68
[4,]    5.0     5.0     5.10    5.00    5.40    5.90    6.5    6.8     6.60
       [,19]   [,20]
[1,]    1.00    1.00
[2,] 2696.10 2714.20
[3,] 4388.31 4633.95
[4,]    5.90    5.40
```

Następna macierz **MAT3A** o wymiarach (4 x 4) jest rezultatem mnożenia macierzy **MAT2A** i **MAT1A**. Kopiujemy do linii komend nasz wektor zmiennej objaśnianej **y** nazywając go teraz jako **y1**, a następnie mnożymy macierz transponowaną **MAT2A** i wektor **y1** uzyskując macierz **MAT4A** o wymiarach (4 x 1):

```
> y1<-c(2394,2595,2375,2233,1961,1737,1827,2641,2760,3186,3431,3329,3230,3045,2344,1619,1211,1411,1699,2146)
> # macierz MAT4A jako wynik z przemnożenia macierzy MAT2A i wektora y1
> MAT4A<-MAT2A%*%y1
> MAT4A
            [,1]
[1,]     47174.0
[2,] 140502253.8
[3,] 112784218.0
[4,]    258409.8
```

Określamy teraz macierz odwrotną **MAT.odwA** do macierzy **MAT3A** o wymiarach (4 x 4) korzystając z komendy **solve()**. Sprawdźmy jeszcze poprawność liczbową uzyskanej macierzy odwrotnej poprzez zastosowanie komendy **round()** mnożącej macierz **MAT3A** i macierz odwrotną ***MAT.odwA***.

```
> round(MAT3A%*%MAT.odwA,10)
      [,1] [,2] [,3] [,4]
[1,]  1e+00    0    0    0
[2,] -1e-10    1    0    0
[3,]  0e+00    0    1    0
[4,]  0e+00    0    0    1
> round(MAT3A%*%MAT.odwA)
     [,1] [,2] [,3] [,4]
[1,]    1    0    0    0
[2,]    0    1    0    0
[3,]    0    0    1    0
[4,]    0    0    0    1
```

Podanie jako opcji „10” w komendzie **round()** powoduje zapisanie pierwszej kolumny w postaci inżynierskiej, którą to postać już wcześniej wyjaśniałem. W efekcie mnożenia uzyskaliśmy macierz jednostkową, co jest potwierdzenie poprawności macierzy odwrotnej, która na głównej przekątnej zwanej diagonalną ma same jedynki. Dla otrzymania wektora parametrów **a1**, zgodnie z równaniem macierzowym musimy przemnożyć macierze **MAT.odwA** i **MAT4A**.

```
> # przemnożenie macierzy odwrotnej MAT.odwA przez MAT4A
> a1<-MAT.odwA%*%MAT4A
> a1
                [,1]
[1,] 16748.9894707
[2,]    -3.3170046
[3,]    -0.6689888
[4,]  -501.0594262
```

Zauważamy, że parametry modelu ekonometrycznego zmiennej X_3 są identyczne jak dla modelu uzyskanego funkcją *REGLINP Excela*. Sprawdźmy teraz, czy szybciej można uzyskać wektor **a1** stosując algebrę macierzową zawartą w *Excelu*.

14.5. Zastosowanie funkcji macierzowych *Excela*

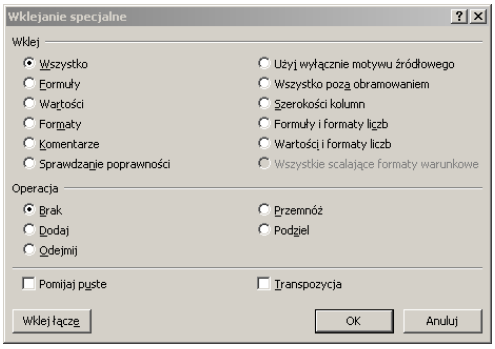
Dla zademonstrowania możliwości funkcji macierzowych *Excela* takich jak **MACIERZ.ILOCZYN()** oraz **MACIERZ.ODW()** przygotowano dane na kolejnym arkuszu skoroszytu *Dane.xls* z uwzględnieniem także zmiennej X_0 (zob. rysunek 14.12).

	A	B	C	D	E	F
1	X3		X0	X9	X25	X78
2	2394		1	3467,6	295,16	5,7
3	2595		1	3391,8	384,25	5,4
4	2375		1	3361,4	515,49	5,4
5	2233		1	3461,1	704,99	5,4
6	1961		1	3436,0	900,58	5,3
7	1737		1	3433,4	1088,03	5,3
8	1827		1	3378,7	1277,06	5,4
9	2641		1	3138,4	1428,01	5,7
10	2760		1	2955,0	2622,76	5,5
11	3186		1	2820,6	2821,30	5,0
12	3431		1	2670,5	2930,70	5,0
13	3329		1	2639,1	3017,64	5,1
14	3230		1	2663,1	3111,04	5,0
15	3045		1	2665,4	3252,02	5,4
16	2344		1	2714,3	3421,05	5,9
17	1619		1	2842,4	3720,10	6,5
18	1211		1	2890,7	3986,50	6,8
19	1411		1	2725,4	4232,68	6,6
20	1699		1	2696,1	4388,31	5,9
21	2146		1	2714,2	4633,95	5,4

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.12. Przygotowanie danych do oszacowania parametrów modelu funkcjami macierzowymi *Excela*

Rozpoczynamy od transpozycji macierzy zmiennych objaśniających X_0, X_9, X_{25}, X_{78} stosując kopiowanie, a następnie opcję *Transpozycja* w oknie dialogowym *Wklej specjalnie* (zob. rysunek 14.13).



Źródło: Opracowanie własne w *Excelu*.

Rys. 14.13. Okno dialogowe *Wklej* specjalnie z opcją *Transpozycja*

Rezultat transpozycji pokazano na fragmencie widoku ekranu *Excela* (zob. rysunek 14.14).

23	X0	1	1	1	1	1	1
24	X9	3467,6	3391,8	3361,4	3461,1	3436,0	3433,4
25	X25	295,16	384,25	515,49	704,99	900,58	1088,03
26	X78	5,7	5,4	5,4	5,4	5,3	5,3

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.14. Fragment macierzy $\mathbf{X}^T$

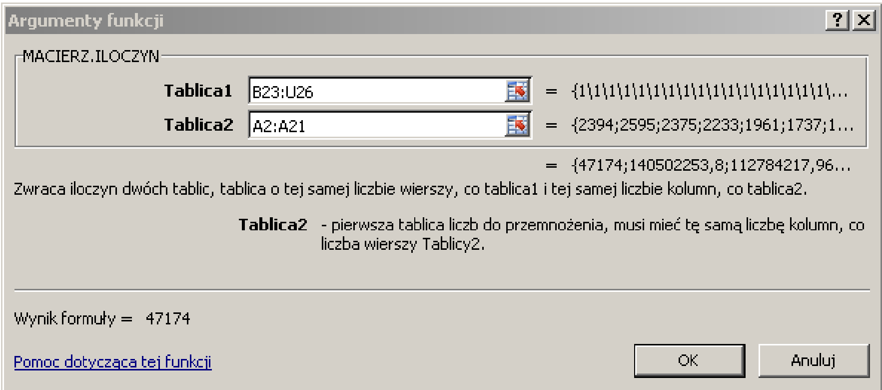
Kolejne operacje macierzowe na obszarach wektora (A2:A21) oraz macierzy podstawowej **X** (obszar C2:F21), macierzy transponowanej $\mathbf{X}^T$ (B23:U23), macierzy odwrotnej $(\mathbf{X}^T\mathbf{X})^{-1}$ (obszar H2:K5) wynikających z rysunków 14.9 oraz 14.14 realizowane są formułami *Excela*:

=MACIERZ.ILOCZYN(B23:U26;C2:F21)

=MACIERZ.ODW(H2:K5)

=MACIERZ.ILOCZYN(B23:U26;A2:A21)

Przyjrzyjmy się jeszcze oknu dialogowemu definiowania funkcji **MACIERZ.ILOCZYN(),** gdzie np. w podoknie *Tablica 1* (obszar B23:U26) podajemy pierwszą mnożona macierz a w podoknie *Tablica 2* - obszar wektora A2:A21 (zob. rysunek 14.15).



Źródło: Opracowanie własne w *Excelu*.

Rys. 14.15. Okno dialogowe mnożenia macierzy

Rezultat wykonania czterech operacji macierzowych w *Excelu* pokazano na rysunku 14.16 stanowiącej fragment arkusza pliku *Dane.xls*, który obejmuje:

- macierz $[\mathbf{X}^T\mathbf{X}]$ o wymiarach $(m \times n) = (4 \times 4)$,
- macierz odwrotną $[\mathbf{X}^T\mathbf{X}]^{-1}$, również o wymiarach (4 x 4),
- wektor kolumnowy wynikający z mnożenia $(\mathbf{X}^T\mathbf{y})$ o wymiarach (4 x 1),
- wektor wierszowy parametrów modelu liniowego wielowymiarowego stanowiącego efekt mnożenia macierzy $\{[\mathbf{X}^T\mathbf{X}\}\text{-}1 * \mathbf{X}^T\mathbf{y}]\}$ o wymiarach (4 x 1).

Trzeba zwrócić uwagę na sposób zapisu parametrów (od prawej do lewej strony) tj. od $a_0 \div a_3$. Wyniki podane w zaokrągleniu do dwóch miejsc po przecinku są zgodne z realizacją operacji macierzowych wykonaną wcześniej programem *R*.

H	I	J	K	L
				$\mathbf{X}^T\mathbf{X}$
20	60065,2	48731,6	111,7	
60065,20	182513417,92	137905436,17	335038,33	
48731,62	137905436,2	159531664,7	277937,7	
111,7	335038,33	277937,712	629,05	
				$(\mathbf{X}^T\mathbf{X})^{-1}$
34,68426028	-0,010904084	-0,002419533	0,717802621	
-0,010904084	4,21151E-06	9,76677E-07	-0,000738391	
-0,002419533	9,76677E-07	2,55573E-07	-0,000203475	
0,717802621	-0,000738391	-0,000203475	0,357307398	
47174	$\mathbf{X}^T\mathbf{y}$			
140502253,8				
112784218				
258409,8				
a_3	a_2	a_1	a_0	
-501,06	-0,67	-3,32	16748,99	a1

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.16. Fragment arkusza *Excela* z realizacją operacji równania macierzowego

Ostatnia operacja równania macierzowego $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ wykonana została formułą:

=MACIERZ.ILOCZYN(H7:K10;H12:H15)

Sprawdźmy jeszcze nasze wyniki stosując funkcję REGLINP, ze stałą lecz bez statystyki opisowej, przy czym format zapisu jest jednak inny - pionowo od a_0 do a_3.

16748,99	a_0
-3,32	a_1
-0,67	a_2
-501,06	a_3

14.6. Inne operacje w *R* na macierzach

Macierz odwrotną możemy wyznaczyć również jako iloczyn odwrotności wyznacznika macierzy i transponowanej macierzy dopełnień algebraicznych. Komendą obliczania wyznacznika jest **det()**, dzielenie wykonujemy stosując operator „/”.

```
> det(MAT3A)
[1] 8.502284e+14
> 1/8.502284e+14
[1] 1.176155e-15
```

Program *R* podał nam wynik w formacie inżynierskim np. 8.502284e+14, gdzie jak już nadmieniono, *e* = 10. Spróbujmy teraz zrealizować wspomniany sposób określenia parametrów w *Excelu* (zob. rysunek 14.17.

	A	B	C	D	E
1	**20**	60065,2	48731,62	111,7	$[\mathbf{X}^T\mathbf{X}]$
2	60065,2	182513418	137905436	335038,33	
3	48731,62	137905436	159531665	277937,712	
4	111,7	335038,33	277937,712	629,05	
5	8,5023E+14	det$[\mathbf{X}^T\mathbf{X}]$	**1,1762E-15**	1/det$[\mathbf{X}^T\mathbf{X}]$	
6	2,949E+16	9,271E+12	-2,057E+12	-6,103E+14	**[D']**
7	9,271E+12	3580741641	-830398906	-6,278E+11	
8	-2,057E+12	-830398906	217295662	1,73E+11	
9	-6,103E+14	-6,278E+11	1,73E+11	3,0379E+14	
10	1	2	3	4	$(-1)^{i+j}$
11	1	-1	1	-1	1
12	-1	1	-1	1	2
13	1	-1	1	-1	3
14	-1	1	-1	1	4
15					
16	2,949E+16	-9,271E+12	-2,057E+12	6,103E+14	**[D]**
17	-9,271E+12	3580741641	830398906	-6,278E+11	
18	-2,057E+12	830398906	217295662	-1,73E+11	
19	6,103E+14	-6,278E+11	-1,73E+11	3,0379E+14	

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.17. Fragment arkusza *Excela* z realizacją operacji macierzy odwrotnej $(\mathbf{X}^T\mathbf{X})^{-1}$

Wyznacznik macierzy **X** zamieszczony w komórce A5 obliczamy korzystając z formuły:

=WYZNACZNIK.MACIERZY(A1:D4).

Następnie określimy odwrotność wyznacznika macierzy formułą: (=1/A5). Pozostaje nam teraz wyznaczenie elementów macierzy dopełnień algebraicznych obliczanych np. jako iloczyn:

$(-1)^{i+j}$ x wyznacznik macierzy

wyłonionej z macierzy wyjściowej, po odjęciu od niej kolumny *j* oraz wiersza, *i* występowania danego elementu w macierzy dopełnień algebraicznych np. dla A1:

=WYZNACZNIK.MACIERZY(B2:D4)

=A6*A11

Na rysunku 31.17, wstępna macierz **D**' powstała z wyznaczników, bez uwzględnienia współczynnika znaku (-1^{i+j}). Dalsza realizacja wymaga zapisania macierzy dopełnień algebraicznych **D** tylko jako wartości komórek, bez formuł wewnętrznych (zob. rysunek 14.18).

21	2,949E+16	-9,271E+12	-2,057E+12	6,103E+14	$[\mathbf{D}_{wart}]$
22	-9,271E+12	3580741641	830398906	-6,278E+11	
23	-2,057E+12	830398906	217295662	-1,73E+11	
24	6,103E+14	-6,278E+11	-1,73E+11	3,0379E+14	
25					
26	2,949E+16	-9,271E+12	-2,057E+12	6,103E+14	$[\mathbf{D}^T]$
27	-9,271E+12	3580741641	830398906	-6,278E+11	
28	-2,057E+12	830398906	217295662	-1,73E+11	
29	6,103E+14	-6,278E+11	-1,73E+11	3,0379E+14	
30					
31	34,6842603	-0,0109041	-0,0024195	0,7178026	$[\mathbf{X}^T\mathbf{X}]^{-1}$
32	-0,0109041	0,0000042	0,0000010	-0,0007384	
33	-0,0024195	0,0000010	0,0000003	-0,0002035	
34	0,7178026	-0,0007384	-0,0002035	0,3573074	

Źródło: Opracowanie własne w *Excelu*.

Rys. 14.18. Fragment arkusza *Excela* z realizacją operacji macierzy dopełnień algebraicznych

Jak już wspomniano macierz odwrotna jest iloczynem odwrotności wyznacznika macierzy (komórka $C$5) na rysunku 14.17 i macierzy dopełnień algebraicznych. Musimy zatem dokonać transpozycji macierzy $[\mathbf{D}_{wart}]$ uzyskując macierz $[\mathbf{D}^T]$. Przemnożenie komórki $C$5 i poszczególnych komórek macierzy $[\mathbf{D}^T]$ np.

=$C$5*A26

daje nam w efekcie końcowym macierz odwrotną postaci $[\mathbf{X}^T\mathbf{X}]^{-1}$.

Spośród jeszcze innych komend języka *R* dotyczących operacji macierzowych trzeba wymienić komendy pomocnicze: **dim()** - określającą wymiary macierzy oraz komendę **diag()** pokazującą wartości na przekątnej diagonalnej.

```
> dim(MAT3A)
[1] 4 4
> diag(MAT3A)
[1]         20.00 182513417.92 159531664.68        629.05
```

Łączenie wybranych obiektów (macierzy/wektorów) ma zastosowanie do określenia wyznaczników w macierzy dopełnień algebraicznych, gdzie trzeba wykreślić „domyślnie" określoną kolumnę lub wiersz rozpatrywanej macierzy. Podano teraz parę przykładów kompletowania nowych poprzez wyłączenie kolumn/wierszy czy też dołączenie nowego wektora. W tym względzie do „*spinania*" w całość bez określonej kolumny stosujemy komendę **cbind()** (zob. poniższy skrypt w języku *R*). Jak już wspomniano kolumny oznaczamy jako (, *przedział dalszych kolumn*). Jeśli w kolejności wyłączymy wiersz komendą **rbind()** uzyskamy macierz do obliczenia wyznacznika elementu macierzy dopełnień algebraicznych. Sekwencję kodu źródłowego wymienionych operacji prezentuje poniższy skrypt:

```
> # utworzenie macierzy MAT5B bez kolumny pierwszej macierzy MAT3A
> MAT5B<-cbind(MAT3A[,2:4])
> MAT5B
            [,1]         [,2]      [,3]
[1,]     60065.2     48731.62    111.70
[2,] 182513417.9 137905436.17 335038.33
[3,] 137905436.2 159531664.68 277937.71
[4,]    335038.3    277937.71    629.05
> # utworzenie macierzy MAT6B bez wiersza pierwszego macierzy MAT5B
> MAT6B<-rbind(MAT5B[2:4,])
> MAT6B
            [,1]        [,2]      [,3]
[1,] 182513417.9 137905436.2 335038.33
[2,] 137905436.2 159531664.7 277937.71
[3,]    335038.3    277937.7    629.05
```

```
> det(MAT6B)
[1] 2.948954e+16
```

Powyższe dotyczyło elementu macierzy będącego w pierwszej kolumnie i pierwszym wierszu, a teraz spróbujmy obliczyć element macierzy [**D'**] dla wiersza drugiego i kolumny drugiej macierzy [$\mathbf{X}^T\mathbf{X}$]:

```
> MAT5C<-cbind(MAT3A[,1],MAT3A[,3:4])
> MAT5C
         [,1]         [,2]      [,3]
[1,]    20.00     48731.62    111.70
[2,] 60065.20 137905436.17 335038.33
[3,] 48731.62 159531664.68 277937.71
[4,]   111.70    277937.71    629.05
> MAT6C<-rbind(MAT3A[1,],MAT3A[3:4,])
> MAT6C
         [,1]        [,2]         [,3]      [,4]
[1,]    20.00     60065.2     48731.62    111.70
[2,] 48731.62 137905436.2 159531664.68 277937.71
[3,]   111.70    335038.3    277937.71    629.05
```

Zauważmy, że składamy nową macierz wyłączając kolumnę drugą przez podanie pierwszej kolumny i dalszych po kolumnie drugiej (MAT3A[,1], MAT3A[,3:4]) i podobnie postępujemy z wyłączeniem wiersza drugiego. Przetestujmy jeszcze dodanie wektora kolumnowego **a2** oraz wierszowego do określonej macierzy np. **MAT3**A:

```
> a2<-c(1,2,3,4)
> a2
[1] 1 2 3 4
> MAT7C<-cbind(MAT3A[,1:4],a2)
> MAT7C
                                                    a2
[1,]    20.00     60065.2     48731.62    111.70  1
[2,] 60065.20 182513417.9 137905436.17 335038.33  2
[3,] 48731.62 137905436.2 159531664.68 277937.71  3
[4,]   111.70    335038.3    277937.71    629.05  4
> MAT8C<-rbind(MAT3A[1:4,],a2)
> MAT8C
       [,1]        [,2]         [,3]      [,4]
      20.00     60065.2     48731.62    111.70
   60065.20 182513417.9 137905436.17 335038.33
   48731.62 137905436.2 159531664.68 277937.71
     111.70    335038.3    277937.71    629.05
a2     1.00         2.0         3.00      4.00
```

* * *

Opracowanie niniejsze skierowane jest w szczególności do studentów kierunków ekonomicznych zamiłowanych w modelowaniu ekonometrycznym zjawisk gospodarczych oraz prognozowaniu i wyznaczaniu funkcji celu w optymalizowanych zadaniach decyzyjnych. Chociaż ogólnie znajomość funkcjonalności *Excela* występuje, to jednak realizowanie operacji macierzowych jest mało znane lub wcale. Godne uwagi jest posługiwanie się komendami programu *R*. Nabranie bowiem wprawy w pisanie skryptów w języku również nazwanym *R*, daje dużą satysfakcję i jest bardzo przydatne w pisaniu prac dyplomowych jak i rozważaniach naukowych z obszaru statystyki ekonometrii i badań operacyjnych.

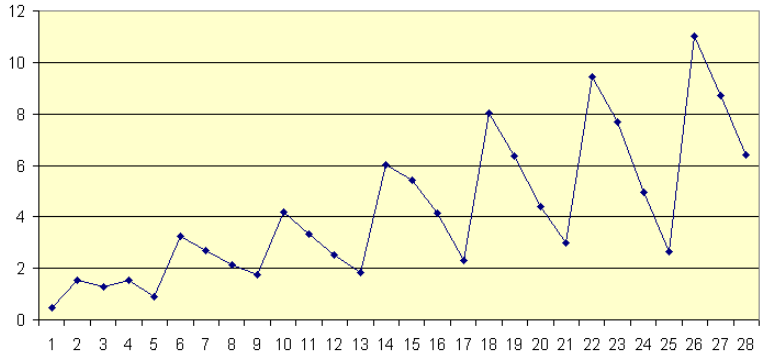
15. Prognozowanie z zastosowaniem modeli Holta-Wintersa

15.1. Wstęp

W niniejszym materiale podkreślono znaczenie modelu multiplikatywnego w dekompozycji szeregu czasowego. Ukazano zainteresowanie internautów tematem prognozowania z zastosowaniem tego modelu prowadzonego przez system *WordPress* w ramach blogu „*Zastosowanie komputera w procesach decyzyjnych*". Załączona statystyka miesięczna podkreśla trend rosnący wejść do wpisu „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*" zwłaszcza w miesiącu styczniu i czerwcu. Zamieszczono procedury obliczeniowe modelu Holta-Wintersa z wykorzystaniem formuł *Excela* w celu określenia wygasłych i przyszłych prognoz danego zjawiska. Procedury te poparte przykładami odnoszą się do modelu bez sezonowości oraz ze sezonowością multiplikatywną. Przy doborze optymalnych parametrów wyrównywania wykładniczego kierowano się miarą RMSE.

Model Holta-Wintersa ze sezonowością zawiera trzy procedury wyrównywania wykładniczego z użyciem parametrów: α - wygładzanie wartości danych, β - wyrównywanie trendu, γ - wyrównywanie w celu wyodrębnienia sezonowości. Prognozowanie wielu zjawisk ekonomicznych wymaga użycia tej klasy modeli, stąd też duże zainteresowanie w Internecie równaniami oraz przykładami rozwiązań zagadnień praktycznych zwłaszcza z użyciem modeli multiplikatywnych. Model multiplikatywny jest często używanym modelem w dekompozycji szeregów czasowych. Występuje tu trend przykładowo rosnący i sezonowość np. w okresie 4 kwartałów, czy też 12. miesięcy roku. Wartość Y_{n+1} jest równa sumie wyrównanej wykładniczo wartości S_n i wartości trendu b_n, pomnożonej przez wartość indeksu sezonowości L_{n-L+1} i z uwzględnieniem błędu z w okresie $(n + 1)$. Ideę takiego trendu przedstawiono na rysunku 15.1. Obserwujemy tu „rozchylające" się wahania przypadkowe, trend oraz sezonowość Według Lipiec-Zajchowskiej model sezonowości multiplikatywnej zapisać można w postaci[192]:

$$Y_{n+1} = (S_n + b_n).I_{n-L+1} + z_{n+1}$$



Źródło: http://www.ekonometria.4me.pl/modele_szeregow_czasowych.htm.

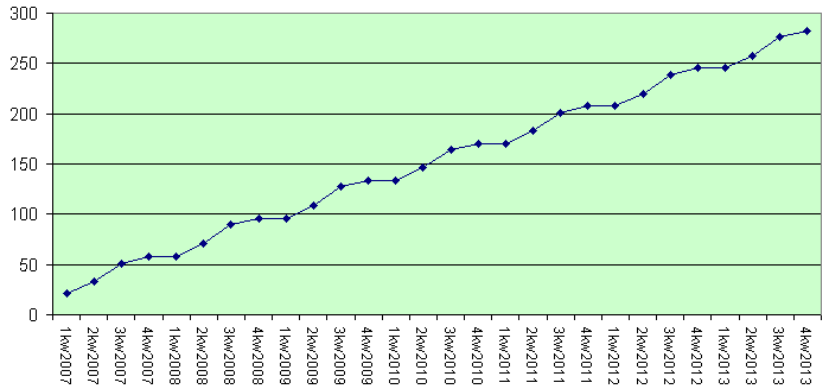
Rys. 15.1. Model multiplikatywny Holta-Wintersa

Występuje także model Holta-Wintersa z sezonowością addytywną. W modelach addytywnych ogólnie biorąc wahania przypadkowe stanowią różnicę pomiędzy wartościami empirycznymi poziomu badanego zjawiska i sumą komponentu trendu i wahań sezonowych[193]:

[192] Lipiec-Zajchowska M., *Wspomaganie procesów decyzyjnych,* Tom II, *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.
[193] Lipiec-Zajchowska M., *Wspomaganie procesów decyzyjnych,* Tom II, *Ekonometria*, op. cit., s. 53.

$$WP_t = Y_t - (T_t + WSS).$$

Wykres poglądowy modelu addytywnego z trendem i sezonowością kwartalną pokazano na rysunku 15.2. W modelu tym wahania przypadkowe ograniczają się do „przedziału" wokół trendu.



Źródło: http://www.ekonometria.4me.pl/modele_szeregow_czasowych.htm.

Rys. 15.2. Model addytywny Holta-Wintersa

15.2. Wywołanie i statystyka wpisu

Wyszukujemy frazę w *Prognozowanie z zastosowaniem modeli Holta-Wintersa* w blogu „*Zastosowanie komputera w procesach decyzyjnych*" (zob. rysunek 15.3). Fraza ta jak i następne zamieszczone w dalszych rozdziałach stanowią jednocześnie tytuły wpisów w podanym blogu, który aktualnie ma tytuły zakładek w języku angielskim.

Źródło: Opracowanie własne w WordPress.

Rys. 15.3. Wprowadzenie frazy wyszukiwania wpisu „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*"

Po naciśnięciu „*Szukaj*" system *WordPress* informuje nas o rezultacie wyszukiwania i udostępnia pełny tytuł danego wpisu oraz początek jego tekstu (zob. rysunek 15.4). Dostęp do edycji omawianego wpisu uzyskuje administrator blogu po wprowadzeniu do adresu swego identyfikatora „*wp-admin*":

https://wornalkiewicz.wordpress.com/wp-admin/

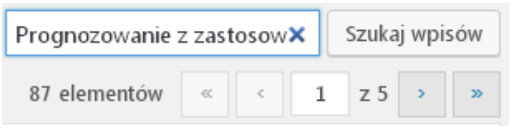
a następnie przejściu do opcji:

https://wornalkiewicz.wordpress.com/wp-admin/edit.php .

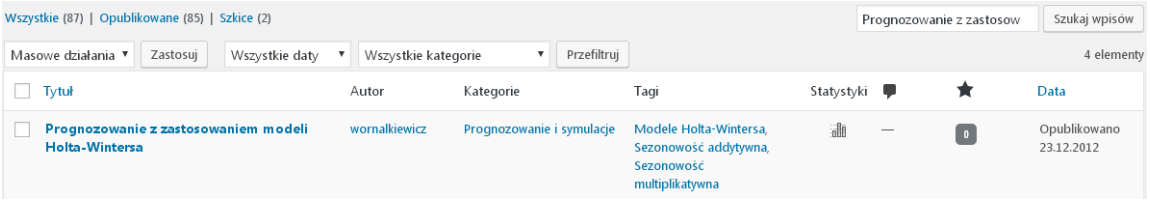
Źródło: Opracowanie własne w WordPress.

Rys. 15.4. Komunikat o o rezultacie wyszukiwania dla podanej frazy

Wyszukiwarka informuje nas ponadto o liczbie (5) występowania danej frazy we wszystkich np. 87. wpisach (elementach) na dzień korzystania z blogu:



Wywołujemy pierwszy z pięciu i udostępniona jest możliwość zaglądnięcia do statystyk prowadzonych w *WordPress* dla danego wpisu. Jedną z nich, jak już wcześniej nadmieniłem jest statystyka wpisu według lat i miesięcy (zob. rysunek 15.5 oraz 15.6).



Źródło: Opracowanie własne w *WordPress*.

Rys. 15.5. Informacja o pierwszym z pięciu wpisów występowania żądanej frazy

Są jeszcze cztery inne wpisy, gdzie występuje podana fraza wyszukiwania, ale nie są one podstawowe pod względem merytorycznym. Dodatkowo informowani jesteśmy przez system o bezpośrednim odnośniku do udostępnionego wpisu na pozycji 1/5:

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012												2	2
2013	18	12	5	20	48	46	15	12	22	40	57	66	361
2014	165	42	35	65	77	63	11	26	22	29	147	104	786
2015	150	50	80	33	76	71	22	19	35	39	85	93	753
2016	183	77	44	81	110	118	15	13	33	18			692

Źródło: Opracowanie własne w WordPress.

Rys. 15.6. Statystyka miesięczna odwiedzin wpisu „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*”

Zaznaczmy jeszcze, że w miesiącach styczniu i czerwcu przygotowywania się studentów do zaliczenia zwłaszcza przedmiotu „*Prognozowanie i symulacje*" statystyka wejść do omawianego wpisu pobrana z rysunku 15.6 jest następująca:

Rok/miesiąc	styczeń	czerwiec
2013	18	46
2014	165	63
2015	150	71
2016	183	118

Łatwo zauważyć trend rosnący zainteresowania wpisem „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*"

15.3. Tekst wpisu

Krótka informacja o zawartości wpisu była już podana w rozdziale 15.2, tu jednak udostępniono pełną jej treść. Jest to zresztą zgodne z życzeniem studentów, którzy chcieli utrwalić atrakcyjne ich zdaniem teksty w formie drukowanej uzupełnione komentarzami autora niniejszego opracowania. Jak już wcześniej wspomniałem zasadniczo możemy wyróżnić trzy rodzaje modeli Holta-Wintersa: bez sezonowości, z sezonowością multiplikatywną, z sezonowością addytywną. Poznajmy teraz bliżej te trzy wymienione modele:

1. *Model Holta-Wintersa bez sezonowości*. Model bez sezonowości ma zastosowanie, gdy występuje trend i wahania przypadkowe. W modelu bez sezonowości oprócz parametru α występuje drugi parametr β odpowiednio do wygładzania wartości poziomu i trendu. Parametry te kształtują się (0-1). Równania tego modelu są następujące:

$$S_n = \alpha\, Y_n + (1- \alpha)\, (S_{n-1} + b_{n-1})$$

$$b_n = \beta\, (S_n - S_{n-1}) + (1 - \beta)\, b_{n-1}$$

$$F_{n+m} = S_n + b_n\, m$$

Wartości początkowe: $S_1 = Y_1$; $b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2$.

Przykład: Przyjmijmy $\alpha = 0{,}3$; $\beta = 0{,}1$; $Y_1 = 29$, $Y_2 = 27$; $Y_3 = 31$, $Y_4 = 35$; $b_1 = [(27 – 29) + (35 - 31)]/2 = 1$; $m = 1$; $n = 10$ obserwacji zmiennej Y_n i obliczamy elementy równań np. dla obserwacji drugiej:

$$S_2 = 0{,}3 \cdot 27 + (1 - 0{,}3)\, (29 + 1) = 29{,}1$$

$$b_2 = 0{,}1\, (29{,}1 - 29) + (1 - 0{,}1)\, 1 = 0{,}91$$

$$F_2 = 29 + 1 = 30$$

Obliczamy teraz dalsze elementy równań modelu. Korzystamy z formuł *Excela* dla kolejnych dziesięciu obserwacji. Obliczamy także kwadrat różnicy między wartością obserwacji a jej prognozą $(Y_n - F_n)^2$ potrzebną nam do wyznaczenia miary błędu prognozy według wzoru na RMSE, który jest pierwiastkiem ze średniej kwadratów różnicy zmiennej prognozowanej Y_n i jej wygasłej prognozy F_n. Obliczenia w *Excelu* z widokiem komórek tego arkusza kalkulacyjnego prezentuje tabela 15.1[194].

[194] Fragment danych liczbowych oraz wzory zaczerpnięto z książki: Lipiec-Zajchrowska (red.), *Wspomaganie procesów decyzyjny*ch, Tom II. *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003, rozdział 2.14. *Model Holta-Wintersa*.

$$RMSE = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n-1}(Y_n - F_n)^2}$$

Tab. 15.1

Obliczenie prognoz i kwadratów błędów prognoz

	A	B	C	D	E	F	G	H
1	t	Y_n	S_n	b_n	F_{n+m}	$(Y_n - F_n)^2$	m	
2	1	2	3	4	5	6	1	2
3	1	29	*29*	*1*			a	
4	2	*27*	29,1	0,91	*30*	*9*	0,3	
5	3	31	30,31	0,94	30,01	0,98	β	
6	4	35	32,37	1,05	31,25	14,09	0,1	
7	5	33	33,30	1,04	33,42	0,18	MSE	
8	6	44	37,24	1,33	34,34	93,37	26,74	
9	7	45	40,50	1,52	38,57	41,40	RMSE	
10	8	48	43,81	1,70	42,02	35,78	5,17	
11	9	51	47,16	1,87	45,51	30,09		
12	10	53	50,22	1,99	49,03	15,79		
13	11				52,20	S		
14	12				54,19	240,68		

Źródło: Opracowanie własne w *Excelu*.

W obliczeniach S_n, b_n, F_{n+m}, $(Y_n - F_n)^2$, S, MSE, RMSE zastosowano następujące formuły poprzedzone znakiem równości przykładowo do obserwacji 10:

J	K	L	M	N	O	P
S_{10}	b_{10}	F_{10+1}	$(Y_{10} - F_{10})^2$	S	MSE	RMSE
$G$4*B12+(1-$G$4)*(C11+D11)						
	$G$6*(C12-C11)+(1-$G$6)*D11					
		C11+D11*$G$2				
			(B12-E12)^2			
				SUMA(F4:F12)		
					F14/(A12-1)	
						PIERWIASTEK(G8)

Formuły kopiujemy z pominięciem znaku równości, gdyż wtedy stanowią tekst a nie wyrażenie do realizacji przez arkusz kalkulacyjny *Excel*. Posłużenie się formułami *Excela* może stanowić zadanie domowe symulacyjne dla studentów przedmiotu „*Prognozowanie i symulacje*" w celu doboru najlepszej wartości parametrów α i β z punktu widzenia minimum miary błędów RMSE. W tym celu sporządzamy kopię *arkusza1* skoroszytu *Tabele.xls* i próbujemy poprzez kolejne podstawianie najpierw parametru α dobierając najlepsze a później dobrać β. Rezultat naszego działania pokazano w załączeniu:

α	0,4	0,5	0,6	0,7	**0,8**	0,9
RMSE	4,53	4,12	3,89	3,79	3,78	3,85
β			0,3	**0,2**	0,1	
RMSE			3,78	3,74	3,78	

Startując z RMSE = 5,17 widzimy, że gdy dokonujemy symulacji ze skokiem 0,1 do najlepsze α = 0,8 a β = 0,2.

2. *Model Holta-Wintersa z sezonowościa multiplikatywną*. W modelu tym występuje trend, wahania przypadkowe oraz zmienna sezonowość zależna od trendu, przy czym równania tego modelu są następujące:

$S_n = \alpha\,(Y_n/I_{n-L}) + (1 - \alpha)\,(S_{n-1} + b_{n-1})$

$b_n = \beta\,(S_n - S_{n-1}) + (1 - \beta)\,b_{n-1}$

$I_n = \gamma\,(Y_n/S_n) + (1 - \gamma)\,I_{n-L}$

$F_{n+m} = (S_n + b_n\,m)\,I_{n-L+m}$

gdzie:

γ - parametr sezonowości;

m - horyzont prognozy;

I_{n-L+m} - wyrównana wartość indeksu sezonowości na okres m, np. $(n + 1)$;

L - długość cyklu sezonowości (12 - dla danych miesięcznych, 4 - dla danych kwartalnych).

Przykład: Mamy kwartalne popyty na lody w kolejnych czterech latach (n = 16 obserwacji)[195]. Popyt wykazuje sezonowość powiększaną w następnych latach. Podaj procedurę określenia wartości początkowych korzystając z danych zapisanych w *Excelu.*

Tab. 15.2

Kwartalne spożycie lodów (dane umowne)

	A	B	C	D	E
1	Rok	Kw. 1	Kw. 2	Kw. 3	Kw. 4
2	1	56	122	255	107
3	2	73	219	439	156
4	3	110	329	564	195
5	4	153	407	757	271

Źródło: Opracowanie własne w *Excelu*.

Procedura wstępna obliczania elementów równania jest następująca:

1. Ustalenie średniej wartości trendu na podstawie odpowiadających sobie trzech kwartałów roku drugiego i pierwszego: [(73 - 56) + (219 - 122) + (439 - 255)]/16 = 18,625.

2. Wartości wyrównane: $S_1 = S_2 = S_3 = S_4$ = (56 + 122 + 255 + 107)/4 = 135.

3. Scentrowanie średniej: [(56 · 1) + (122 · 2) + (255 · 3) + (107 · 4)]/4 = 373,25 373,25/135 = 2,76, w przybliżeniu 2,5.

4. Orientacyjna korekta wartości trendu b_n przykładowymi mnożnikami, przy czym suma mnożników równa się zero dla czterech kwartałów.

Kw.	b_n
1	-1,5 · 18,625 = -27,937
2	-0,5 · 18,625 = -9,312
3	0,5 · 18,625 = 9,312
4	1,5 · 18,625 = 27,937

5. Obliczenie $(S_n + b_n)$ dla czterech pierwszych kwartałów.

[195] Dane liczbowe, wzory i procedury zaczerpnięto z książki: Lipiec-Zajchrowska (red.), *Wspomaganie procesów decyzyjnych* Tom II. *Ekonometria*, op. cit., s. 90.

Kw.	$S_n + b_n$
1	135 – 27,937 = 107,063
2	135 – 9,312 = 1215,698
3	135 + 9,312 = 144,312
4	135 + 27,937 = 162,937

Na podstawie tabeli 15.2 obliczamy średnią dla kwartału pierwszego i czterech kolejnych lat:

$$(56 + 73 + 110 + 153)/4 = 98.$$

6. Wstępne indeksy sezonowości czterech pierwszych kwartałów są następujące:

Kw.	Y_n	$S_n + b_n$	$I_n = Y_n/((S_n + b_n)$
1	56	107,063	0,523
2	122	125,688	0,971
3	255	144,312	1,767
4	107	162,937	0,657

7. Ponowne ustalenie S_4 i przyjęcie $b_4 = 18{,}625$: $S_4 = Y_4/I_4 = 107/0{,}657 = 162{,}861$.

8. Przyjęcie wartości początkowych parametrów: $\alpha = 0{,}2$; $\beta = 0{,}2$; $\gamma = 0{,}1$.

Mamy już określone wartości początkowe do skorzystania z wielorównaniowego modelu Holta-Wintersa z sezonowością multiplikatywną, pora więc na zadanie: *Zastosuj funkcje Excela do wyznaczenia elementów równań modelu Holta - Wintersa oraz oblicz wygasłe prognozy dla* 16 *obserwacji*. *Ponadto oblicz miarę błędu* RMSE *wygasłych prognoz dla okresów* (5-16) *oraz oblicz prognozę przyszłą na kwartał pierwszy roku piątego*.

Tab. 15.3

Obliczenia elementów i wygasłych prognoz dla modelu multiplikatywnego

	A	B	C	D	E	F	G	H	I	J	K
1	Rok	Kw.	Y_n	S_n	b_n	$S_n + b_n$	*In*	F_n	$(Y_n - F_n)^2$	Okres	α
2	1	1	56			*107,063*	*0,523*			1	0,2
3	1	2	122			*125,688*	*0,971*			2	β
4	1	3	255			*144,312*	*1,737*			3	0,2
5	1	4	107	*162,861*	*18,625*	181,486	*0,657*			4	γ
6	2	1	73	**173,105**	**16,949**	**190,053**	**0,513**	**94,917**	**480,363**	5	0,1
7	2	2	219	197,151	18,368	215,519	0,985	184,542	1187,364	6	m
8	2	3	439	222,962	19,857	242,819	1,760	374,357	4178,763	7	1
9	2	4	156	241,744	19,642	261,386	0,656	159,532	12,476	8	S
10	3	1	110	252,004	17,766	269,770	0,505	134,057	578,744	9	**21414,918**
11	3	2	329	282,619	20,335	302,954	1,003	265,719	4004,546	10	MSE
12	3	3	564	306,447	21,034	327,481	1,768	533,259	945,027	11	**1784,5765**
13	3	4	195	321,452	19,828	341,280	0,651	214,772	390,950	12	RMSE
14	4	1	153	333,590	18,290	351,880	0,501	172,426	377,372	13	**42,24**
15	4	2	407	362,669	20,448	383,117	1,015	352,899	2926,967	14	
16	4	3	757	392,116	22,248	414,364	1,784	677,434	6330,691	15	
17	4	4	271	414,759	22,327	437,086	0,651	269,714	1,654	16	
18	5	1						**218,794**			

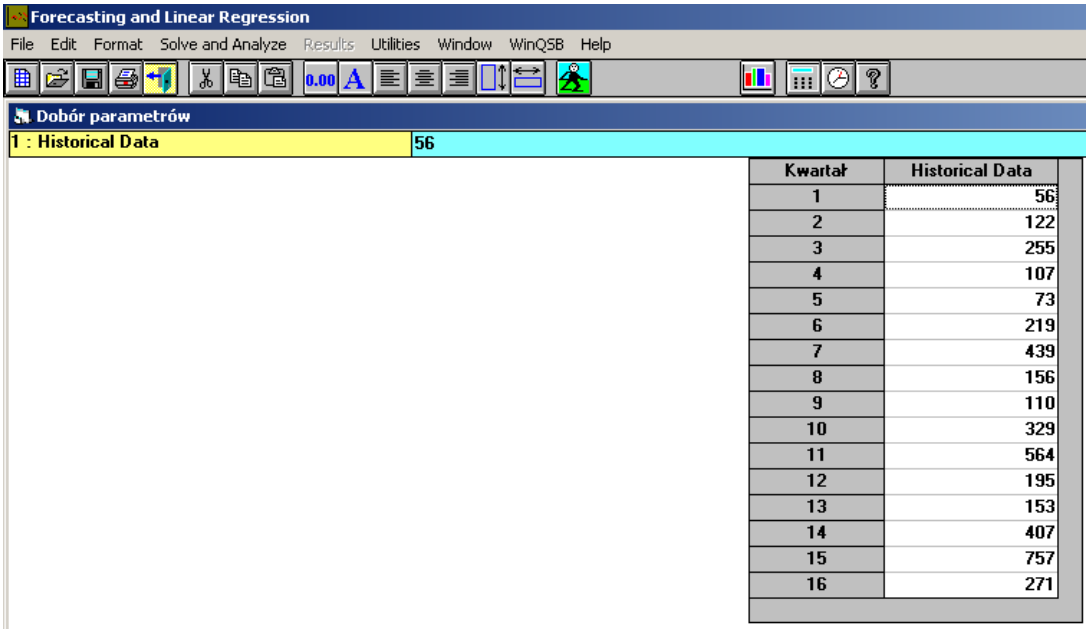
Źródło: Opracowanie własne w *Excelu*.

Pierwiastek średniego kwadratu błędu RMSE przy dokładnych obliczeniach w *Excelu* wynosi 42,24 a prognoza na okres 17. Wynosi 218,794, czyli około 219. W obliczeniach S_n, b_n, $(S_n + b_n)$, I_n, F_n, $(Y_n - F_n)^2$, S, MSE, RMSE zastosowano następujące formuły poprzedzone znakiem równości, które przykładowo do obserwacji piątej są następujące:

M	N	O	P	Q	R	S	T	U	V
S_5	b_5	$S_5 + b_5$	I_5	F_5	$(Y_5 - F_5)^2$	S	MSE	RMSE	F_{17}
$K$2*(C6/G2)+(1-$K$2)*F5									
	K$4*(D6-D5)+(1-$K$4)*E5								
		D6+E6							
			$K$6*(C6/D6)+(1-$K$6)*G2						
				(D5+E5*$K$8)*G2					
					(C6-H6)^2				
						SUMA(I6:I17)			
							K10/(J17-J5)		
								PIERWIASTEK(K12)	
									(D17+E17*$K$8)*G14

Na zakończenie podanego zestawienia formuł podano również wyrażenie na obliczenie prognozy na pierwszy kwartał roku piątego.

Kontynuujmy jeszcze nasze zadanie i zastosujmy moduł *Forecasting and Linear Regression* (FC) pakietu *WinQSB* dla doboru najlepszej wartości parametrów α, β, γ przy warunku minimum RMSE bez podawania wartości początkowych.



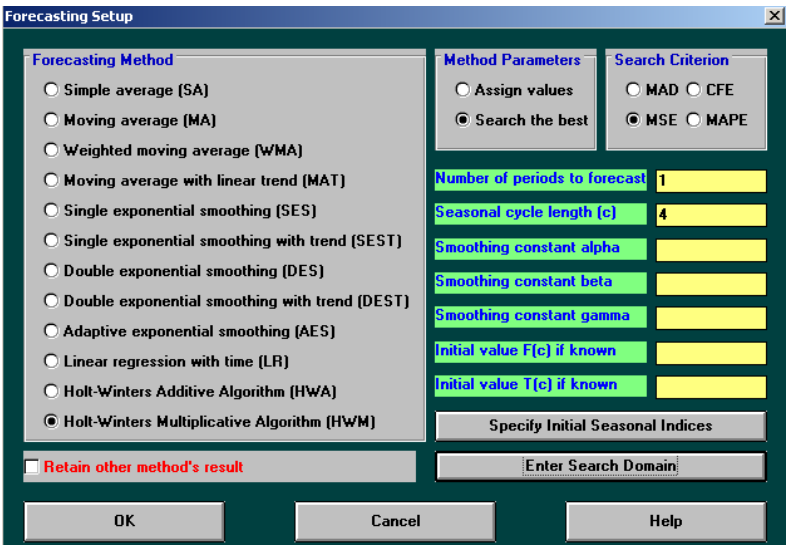
Kwartał	Historical Data
1	56
2	122
3	255
4	107
5	73
6	219
7	439
8	156
9	110
10	329
11	564
12	195
13	153
14	407
15	757
16	271

Źródło: Opracowanie własne w FC.

Rys. 15.7. Zdefiniowanie szeregu czasowego w module FC

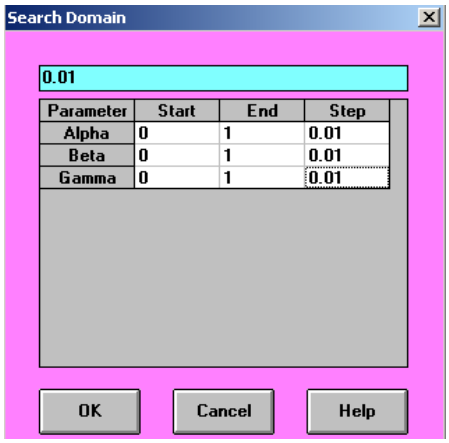
Następnie definiujemy w poszczególnych podoknach opcje metody prognozowania (zob. rysunek 15.8):

„*Forecasting Method*" - Holt-Winters Multiplicative Algorithm (HWM),
"*Method Parameters*" - Search the best,
"*Search Criterion*" - MSE,
"*Number of periods to forecast*" - 1,
"*Seasonal cycle length* (*c*)" - 4,
"*Enter Search Domain*" - 0.01 (domyślnie jest przecinek i zamieniamy go na kropkę dziesiętną (rys. 15.9).



Źródło: Opracowanie własne w FC.

Rys. 15.8. Ustawienie opcji metody HWM



Źródło: Opracowanie własne w FC.

Rys. 15.9. Ustawienie skoku iteracyjnego doboru najlepszych wartości parametrów α, β, γ

Moduł FC po dłuższym okresie procesu obliczeniowego podał nam rozwiązanie problemu optymalizacyjnego pod nazwą *Forecast Result for Dobór parametrów,* w tym z zaogrągleniem wyników do liczb całkowitych (zob. rysunek 15.10). W raporcie występują następujące miary błędów wygasłych prognoz: CFE = 106, MAD = 25, MSE = 996, MAPE = 12%. MAPE to średni procentowy absolutny błąd. Wcześniej podana miara RMSE jest pierwiastkiem z MSE i wynosi **31,56**, tak więc jest mniejsza od wyliczonej formułami *Excela* (RMSE = 42,24) przy realizacji według parametrów α = 0,2, β = 0,2, γ = 0,1. Program ustalił jako najlepsze (*Search the best*) parametry:

Alpha=0,27
Beta=1
Gamma=0,99
F(0)=135

Widzimy, że przyjął również wyjściową wartość średnią przy cyklu 4. kwartałów dla pierwszego roku F(0) = 135. Ponadto założył prognozę na okres piaty równą obserwacji pierwszej (Y_1 = 56).

12-28-2016 Kwartał	Actual Data	Forecast by HWM	Forecast Error	CFE	MAD	MSE	MAPE (%)
1	56						
2	122						
3	255						
4	107						
5	73	56	17	17	17	289	23
6	219	142	77	94	47	3109	29
7	439	405	34	128	43	2467	22
8	156	205	-49	80	44	2440	24
9	110	132	-22	58	40	2046	23
10	329	319	10	68	35	1722	20
11	564	557	7	75	31	1484	17
12	195	190	5	81	28	1302	16
13	153	136	17	98	26	1189	15
14	407	432	-25	72	26	1134	14
15	757	736	21	94	26	1072	13
16	271	258	13	106	25	996	12
17		202					

Źródło: Opracowanie własne w WinQSB - moduł FC.

Rys. 15.10. Raport modułu FC realizacji zadania „*Dobór parametrów*”

3. *Model Holta-Wintersa z sezonowością addytywną*

W omawianym blogu w tekście wpisu „Prognozowanie z zastosowanie modeli Holta-Wintersa” zamieszczono również równania modelu z sezonowością addytywną, które są następujące[196]:

$$S_n = \alpha\,(Y_n - I_{n-L}) + (1 - \alpha)\,(S_{n-1} + b_{n-1})$$

$$b_n = \beta\,(S_n - S_{n-1}) + (1 - \beta)\,b_{n-1}$$

$$I_n = \gamma\,(Y_n - S_n) + (1 - \gamma)\,I_{n-L}$$

$$F_{n+m} = S_n + b_n \cdot m + I_{n-L+m}$$

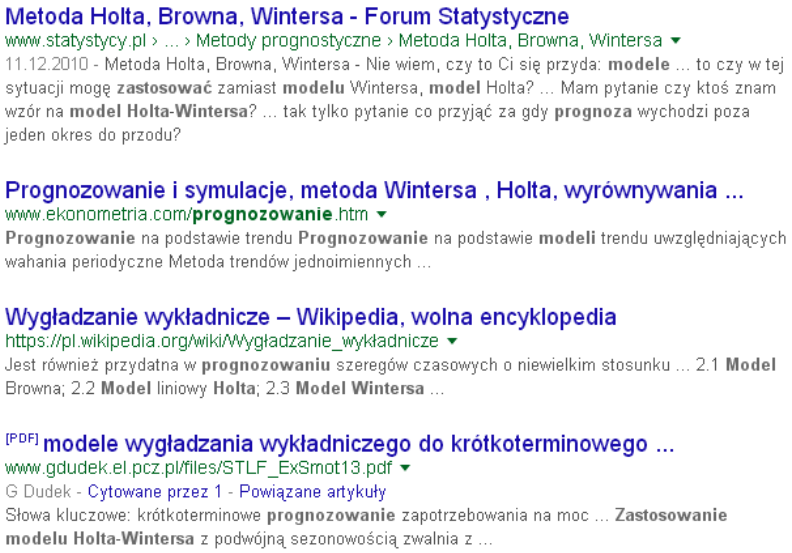
Zwróćmy uwagę we wzorze na S_n, że model Holta-Wintersa z sezonowością addytywną stanowi analogię do modelu multiplikatywnego, jednak występuje odejmowanie (zamiast dzielenia) indeksu sezonowości I_{n-L} jako to miało miejsce w modelu multiplikatywnym. Procedury korzystania z formuł Excela, a także z modułu FC pakietu *WinQSB* są podobne.

15.3. Problematyka frazy *Prognozowanie z zastosowaniem modeli Holta-Wintersa*

Zagadnienie prognozowania z zastosowaniem modeli Holta-Wintersa zbudza największe zainteresowanie internautów, będących gośćmi blogu „*Zastosowanie komputera w procesach decyzyjnych*”. Wiele pozycji - wpisów dla wymienionej wcześniej frazy bez cudzysłowia, bo aż 1710, znalazło się w dniu wyszukiwania w zasobach Google. Rozpatrzmy tylko stronę pierwszą o 10. wpisach. oraz otwórzmy publikacje internetowe na tej stronie w formacie PDF, a jest ich cztery. Zapoznajmy się w skrócie z ich zawartością, tj. otwórzmy publikacje formatu PDF (zob. rysunki 15.11 oraz 15.12). Po udostępnieniu obszernej listy rezultatu ogólnoświatowa wyszukiwarka jaką jest *Google* proponuje nam jeszcze wyszukiwanie podobne według fraz pisanych małymi literami:

- metoda wintersa przykład,
- metoda wintersa excel,
- metoda holta wintersa excel,
- metoda holta przykład,
- model holta wintersa excel,
- model wintersa przykład,
- model addytywny i multiplikatywny,
- metoda Browna.

[196] Dane liczbowe, wzory i procedury zaczerpnięto z książki: Lipiec-Zajchrowska (red.), *Wspomaganie procesów decyzyjnyc*h Tom II. *Ekonometria*, op. cit., s. 94.



Źródło: Opracowanie własne w Google.

Rys. 1.5.11. Pierwszy fragment pozycji na pierwszej stronie listy rezultatu wyszukiwania

Źródło: Opracowanie własne w *Google*.

Rys. 15.12. Drugi fragment pozycji na pierwszej stronie listy rezultatu wyszukiwania

Strona 1/6 publikacji internetowej „*Modele wygładzania wykładniczego do krótkoterminowego prognozowania obciążeń systemów elektroenergetycznych*" podaje streszczenie oraz słowa kluczowe w tym artykule, a autor Grzegorz Dudek podaje:

"*Przedstawiono modele do sporządzania prognoz zapotrzebowania na moc elektryczną. Szeregi czasowe charakteryzują się wieloma wahaniami sezonowymi, co komplikuje model prognostyczny. W celu uproszczenia problemu szeregi czasowe zdekomponowano na szeregi obciążeń w tych samych chwilach doby, co pozwoliło zastosować model wygładzania wykładniczego*

z pojedynczą sezonowością. Inny rodzaj dekompozycji z wykorzystaniem regresji lokalnej (LOESS) umożliwia użycie modelu o mniejszej złożoności bez składnika sezonowego. Zastosowanie modelu Holta-Wintersa z podwójną sezonowością zwalnia z potrzeby dekompozycji szeregu czasowego. Dokładność proponowanych metod porównano na przykładach aplikacyjnych z dokładnością modeli ARIMA i modelu opartego na sieci neuronowej".

W publikacji internetowej „*Oprogramowanie na podstawie modeli Holta-Wintersa dla pełnych i niepełnych danych*" Uniwesytetu Ekonomicznego we Wrocławiu na stronie 3/17 znajduje się streszczenie i słowa kluczowe tego artykułu. Autorzy Maria Szmuksta-Zawadzka oraz Jan Zawadzki informuje nas: „*Praca poświęcona jest praktycznym aspektom modeli Holta-Wintersa w prognozowaniu zmiennych wykazujących wahania sezonowe na podstawie pełnych szeregów czasowych oraz szeregów z lukami w danych. Autorzy sformułowali przesłanki wyboru równań do celów prognozowania inter- i ekstrapolacyjnego. Ze względu na zbliżone oceny mierników $MAPE^{\alpha, \beta, \gamma}$ do prognozowania należy wybierać kilkanaście modeli różniących się dość znacznie kombinacjami stałych wygładzania. Do budowy prognoz ex ante powinny zostać wybrane spośród nich modele charakteryzujące minimalnymi ocenami błedów prognoz ex post*".

Kolejna publikacja „*Model Wintersa*" w formacie PDF ma charakter prezentacji wykładu akademickiego z tego zakresu. Na stronie ¼ zdefiniowano pojęcia występujące w modelu Wintersa.

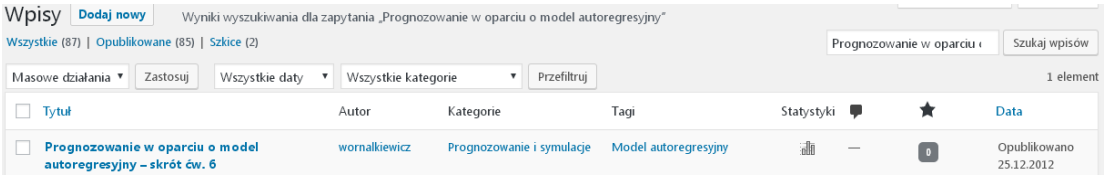
Otwórzmy teraz artykuł internetowy „*Zastosowanie wybranych modeli adaptacyjnych w prognozowaniu brakujących danych w szeregach ze złożoną sezonowością dla luk niesystematycznych*". Ci sami autorzy co już wcześniej omówiony artykuł poświecili wykorzystaniu wybranych modeli wyrównywania wykładniczego Browna i Holta-Wintersa i streszczają go następująco: „*Prognozy wyjściowe będą budowane na podstawie szeregów oczyszczonych ze sezonowości. Prognozy końcowe uwzględniające wahania sezonowe, będą sumami prognoz wyjściowych i składników sezonowości lub iloczynami prognoz tego rodzaju i wskaźników sezonowości*".

Została nam jeszcze publikacja Macieja Oesterreicha pt. „*Wykorzystanie modelu Holta-Wintersa oraz metod bootstrapowych w prognozowaniu na podstawie zmiennych ekonomicznych z wahaniami sezonowymi"*, a we wstępie tej pracy czytamy: „*Modele wyrównywania wykładniczego należą do nieklasycznych modeli szeregu czasowego najczęściej wykorzystywanych w prognozowaniu zmiennych z wahaniami sezonowymi. Związane jest to m.in. z ich elastycznością, tzn. z szybką reakcją na zmiany trendu oraz wahań okresowych. Dodatkową zaleta jest dostępność programów obliczeniowych, pozwalających na szybkie wyznaczanie optymalnych wartości stałych wyrównywania i budowę prognoz. Niekiedy jednak dokładność prognoz dotyczących tych zmiennych jest niewystarczająca. Zwiększenie ich efektywności jest możliwe dzięki wykorzystaniu technik i metod symulacyjnych, w tym metod bootstrapowych*".

16. Prognozowanie w oparciu o model autoregresyjny

16.1. Wywołanie i statystyka wpisu

W blogu „Zastosowanie komputera w procesach decyzyjnych" występuje tylko jeden wpis o temacie „*Prognozowanie w oparciu o model autoregresyjny*" wprowadzony w grudniu 2012 roku. Jest to skrót przeprowadzonego przez autora ćwiczenia 6. w ramach prowadzonych zajęć na uczelni (zob. rysunek 16.1).



Źródło: Opracowanie własne w WordPress.

Rys. 16.1. Wynik wyszukiwania dla zapytania „*Prognozowanie w oparciu o model autoregresyjny*"

Dla tego - zdaje się mało interesującego tematu - statystyka ilościowa wejść w układzie lat i miesięcy prowadzona przez system *WordPress* jest znacząca (zob. rysunek 16.2).

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2013	9	9	3	8	25	23	8	5	9	10	15	13	137
2014	46	16	15	15	34	51	6	14	23	16	29	35	300
2015	120	40	26	25	76	73	14	10	10	25	62	46	527
2016	128	43	21	53	89	72	20	11	23	12			472

Źródło: Opracowanie własne w WordPress.

Rys. 16.2. Statystyka miesięczna zainteresowania wpisem „*Prognozowanie w oparciu o model autoregresyjny*"

Widzimy na podstawie kolumny „*W sumie*", że mimo corocznego spadku studentów na kierunkach związanych z metodami ilościowymi zainteresowanie się tematem „*Prognozowanie w oparciu o model autoregresyjny*" rośnie z roku na rok. Ponadto zwróćmy uwagę na miesiące styczeń i czerwiec przygotowania się studentów do zaliczenia przedmiotów z zakresu metod ilościowych:

Rok/miesiąc	styczeń	czerwiec
2013	9	23
2014	46	51
2015	120	73
2016	128	72

Tekst wpisu Mieliśmy przypadki modeli szeregów czasowych bazujących na tylko jednej zmiennej objaśniającej jaką były okresy czasowe. Teraz rozpatrzymy specyficzny przypadek regresji tj. zależności danej zmiennej objaśnianej od siebie samej, lecz przesuniętej o 1, 2, ..., *m* okresów. Model taki przykładowo w odniesieniu do autoregresji trzeciego rzędu określony jest równaniem:

$$\hat{Y} = a_0 + a_1 Y(-1) + a_2 Y(-2) + a_3 Y(-3).$$

Skorzystajmy z danych umownych dotyczących 26. obserwacji zmiennej zależnej DZGD (depozyty złotowe gospodarstw domowych) - zob. tabela 16.1[197].

Tab. 16.1

Wartości umowne szeregu DZGD

n	DZGD	DZGD(-1)	DZGD(-2)	DZGD(-3)
1	0,95			
2	0,97	0,95		
3	1,22	0,97	0,95	
4	1,64	1,22	0,97	0,95
5	1,04	1,64	1,22	0,97
6	1,30	1,04	1,64	1,22
7	0,90	1,30	1,04	1,64
8	2,00	0,90	1,30	1,04
9	2,20	2,00	0,90	1,30
10	1,00	2,20	2,00	0,90
11	2,10	1,00	2,20	2,00
12	3,00	2,10	1,00	2,20
13	3,58	3,00	2,10	1,00
14	2,50	3,58	3,00	2,10
15	2,30	2,50	3,58	3,00
16	4,50	2,30	2,50	3,58
17	4,00	4,50	2,30	2,50
18	3,20	4,00	4,50	2,30
19	3,80	3,20	4,00	4,50
20	5,80	3,80	3,20	4,00
21	5,00	5,80	3,80	3,20
22	5,20	5,00	5,80	3,80
23	4,50	5,20	5,00	5,80
24	6,50	4,50	5,20	5,00
25	6,00	6,50	4,50	5,20
26	6,31	6,00	6,50	4,50

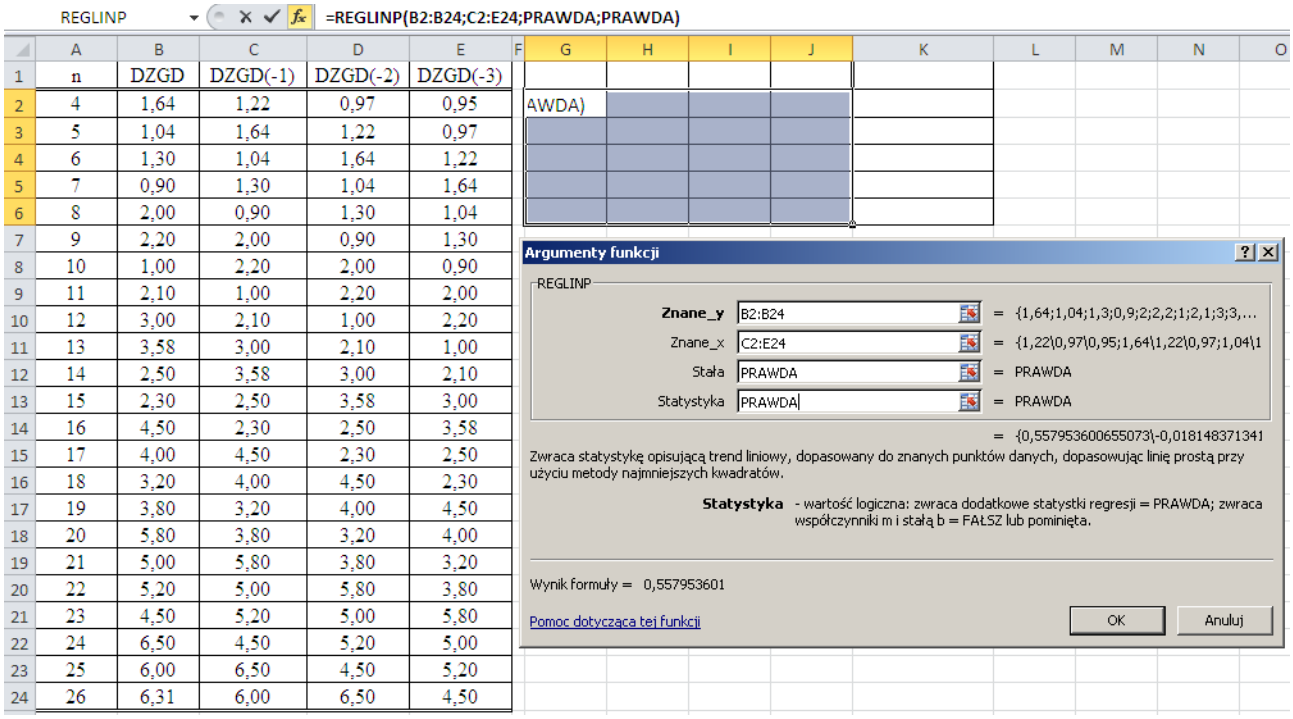
Źródło: Opracowanie własne z częściowym wykorzystaniem fragmentu szeregu DZGD (tablica 3.9, Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych. Tom* II *Ekonometria*, op. cit.

Budujemy model na 23. obserwacjach (4-26), gdyż jest to pełny komplet informacji do estymacji, przy czym dogodnie jest oszacować parametry modelu liniowego wielowymiarowego funkcją REGLINP *Excela*: (=REGLINP(zakres zmiennej *Y*; macierz zmiennych *X*, PRAWDA; PRAWDA))

gdzie:

Y - DZGD (wiersze 4-26), *X* - wiersze (4-26) kolumn DZGD(-1) do DZGD(-3), pierwsze "PRAWDA" - potwierdzenie stałej w równaniu liniowym, drugie "PRAWDA" - żądanie statystyki opisowej.

[197] Opracowanie stanowi analogie do tabeli 3.9, Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych. Tom* II. *Ekonometria*, op. cit.



	A	B	C	D	E
1	n	DZGD	DZGD(-1)	DZGD(-2)	DZGD(-3)
2	4	1,64	1,22	0,97	0,95
3	5	1,04	1,64	1,22	0,97
4	6	1,30	1,04	1,64	1,22
5	7	0,90	1,30	1,04	1,64
6	8	2,00	0,90	1,30	1,04
7	9	2,20	2,00	0,90	1,30
8	10	1,00	2,20	2,00	0,90
9	11	2,10	1,00	2,20	2,00
10	12	3,00	2,10	1,00	2,20
11	13	3,58	3,00	2,10	1,00
12	14	2,50	3,58	3,00	2,10
13	15	2,30	2,50	3,58	3,00
14	16	4,50	2,30	2,50	3,58
15	17	4,00	4,50	2,30	2,50
16	18	3,20	4,00	4,50	2,30
17	19	3,80	3,20	4,00	4,50
18	20	5,80	3,80	3,20	4,00
19	21	5,00	5,80	3,80	3,20
20	22	5,20	5,00	5,80	3,80
21	23	4,50	5,20	5,00	5,80
22	24	6,50	4,50	5,20	5,00
23	25	6,00	6,50	4,50	5,20
24	26	6,31	6,00	6,50	4,50

Źródło: Opracowanie własne w *Excelu*.

Rys. 16.4. Definiowanie argumentów funkcji REGLINP

Naciskamy jednocześnie trzy klawisze (Shift - lewy, Ctrl - lewy, Enter) i rezultat pracy REGLINP jest następujący:

G	H	I	J
0,5580	-0,0181	0,5081	0,3214
0,2132	0,2280	0,1966	0,3915
0,8066	0,8546	#N/D!	#N/D!
26,4203	19,0000	#N/D!	#N/D!
57,8895	13,8769	#N/D!	#N/D!

Źródło: Opracowanie własne w *Excelu*.

Rys. 16.5. Obliczenia wykonane przez REGLINP

Pola oznaczone jako (#N/D) nie występują. Pozostaje nam teraz opisanie we własnym zakresie powyższego zestawienia wyników wygenerowanych przez funkcję REGLINP dla modelu liniowego autoregresyjnego.

G	H	I	J	K
a_3	a_2	a_1	a_0	Opis
0,5580	-0,0181	0,5081	0,3214	parametry (a_i)
0,2132	0,2280	0,1966	0,3915	błąd standardowy oszacowania parametru S_{a_i}
0,8066	0,8546	#N/D!	#N/D!	R^2; S_e
26,4203	19,0000	#N/D!	#N/D!	F; $iss = n - m - 1 = n - k$
57,8895	13,8769	#N/D!	#N/D!	ESS; RSS

gdzie:

n - liczba obserwacji,
m - liczba zmiennych objaśniających,
R^2 - współczynnik determinacji,

S_e - odchylenie standardowe reszt,
F - statystyka Fishera-Snedecora,
iss - ilość stopni swobody,
ESS - wyjaśniona przez model suma kwadratów,
RSS - resztowa suma kwadratów.

Źródło: Opracowanie własne w *Excelu*.

Rys. 16.5. Opis tabeli wyników funkcji REGLINP

Zatem uzyskaliśmy model:

$$\hat{Y} = 0{,}3214 + 0{,}5081\ DZGD(-1) - 0{,}0181\ DZGD(-2) + 0{,}5580\ DZGD(-3)$$

Zanim zastosujemy ten model do określenia prognozy na okres (N = 24) musimy zbadać istotność statystyczną określonych parametrów modelu testem *t*-Studenta:

$$It_{obl}I = Ia_i/S_{ai}I > t_\alpha \text{ dla } \alpha = 0{,}05,\ iss = 23 - k = 23 - 4 = 19,\ t_\alpha = 2{,}019.$$

Stawiamy hipotezę zerową: $H_0 = a_i = 0$ istnieje statystyczna istotność danego parametru, wobec hipotezy alternatywnej $H_1: a_i \neq 0$. W odniesieniu do poszczególnych parametrów naszego modelu uzyskujemy następujące wyniki testu *t*-Studenta:

a_0	S_{a0}	$It_{a0}I$	Test
0,324	0,3915	0,8208	nieistotny
a_1	S_{a1}	$It_{a1}I$	Test
0,5081	0,1966	2,5844	istotny
a_2	S_{a2}	$It_{a2}I$	Test
-0,0181	0,2280	-0,0794	nieistotny
a_3	S_{a3}	$It_{a3}I$	Test
0,5580	0,2132	2,6173	istotny

Tak więc po teście *t*-Studenta, pozostawieniu mimo wszystko stałej a_0 a usunięciu zmiennej DZGD(-2) i zaokrągleniu parametrów uzyskujemy równanie liniowe modelu:

$$\hat{Y} = 0{,}3214 + 0{,}51\ DZGD(-1) + 0{,}56\ DZGD(-3)$$

Weryfikacja modelu do prognozowania wymaga sprawdzenia jeszcze innymi testami. Tak więc przekonajmy się o normalności składników resztowych testem JBT (Jargu'e-Bery):

$$JBT = (N - k)/6\ [(S^2 + 0{,}25\ (K - 3)^2$$

$$H_0 = F_N(e_t);\quad H_1 \neq F_N(e_t)$$

Hipoteza zerowa jest spełniona, gdy: $JBT_{obl} < \chi^2_{0{,}05;\,(2)}$. Zastosowano rozkład χ^2 przy α = 0,05 i liczbie stopni swobody *iss* = 2, $\chi^2_{0{,}05;\,(2)} = 5{,}991$.

Obliczamy w *Excelu* wartości teoretyczne obserwacji oraz reszty $e_i = Y - \hat{Y}$. Następnie korzystamy z możliwości wygenerowania szerszej statystyki wywołując menu:

Excel-Analiza danych-Statystyka opisowa.

Uzyskujemy następujący raport:

Średnia	0,258
Błąd standardowy	0,166
Mediana	0,340

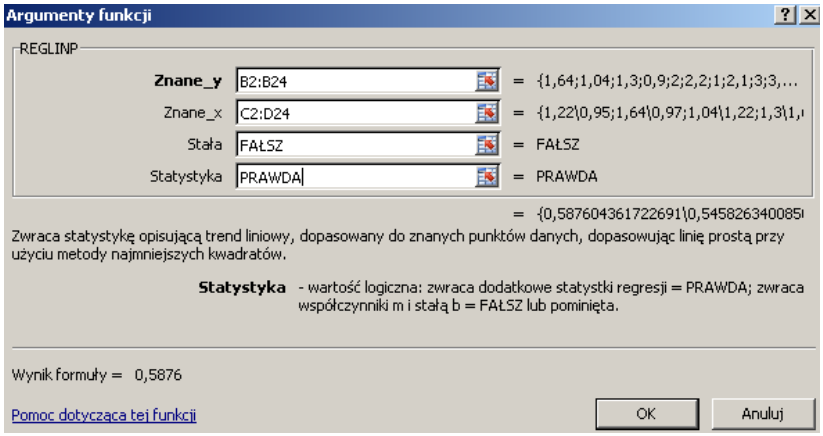
Odchylenie standardowe próby	0,795
Wariancja próby	0,632
Kurtoza (K)	-0,554
Skośność	-0,017
Zakres	3,020
Minimum	-1,400
Maksimum	1,620
Suma	5,930

$JBT_{obl} = (23 - 2)/6\ [(-0{,}017)^2 + 25\ (-0{,}554 - 3)^2] = 11{,}053$

Tak więc: $JBT_{obl} > \chi^2_{0{,}05;\,(2)}$, zatem składniki resztowe nie mają rozkładu normalnego.

Jako kolejne zadanie dla Czytelnika proponuję: *Dokonaj ponownej estymacji parametrów modelu zawierającego tylko zmienne* DZGD(-1) *oraz* DZGD(-3) *co wynikało z wcześniejszego testu t-Studenta. Następnie ponownie zbadaj istotność parametrów oraz normalność składnika resztowego.*

Tak więc w określeniu parametrów takiego modelu autoregresyjnego bazujemy na obserwacjach (4-26) oraz zmiennych DZGD, DZGD(-1), DZGD(-3), wycinając nieistotną zmienną DZGD(2) i stałą a_0. Stosując funkcję REGLINP *Excela* podajemy nowe argumenty tej funkcji (zob. rysunek 16.6).



Źródło: Opracowanie własne w *Excelu*.

Rys. 16.6. Skorygowany zestaw argumentów funkcji REGLINP bez stałej i statystyki opisowej

L	M	N
0,5876	0,5458	a_i
0,1713	0,1480	S_{ai}
0,9576	0,8272	R^2; S_e
237,0692	21,0000	F; iss
324,4342	14,3695	ESS; RSS

Źródło: Opracowanie własne w *Excelu*.

Rys. 16.7. Tabela wyników funkcji REGLINP z opisem autora

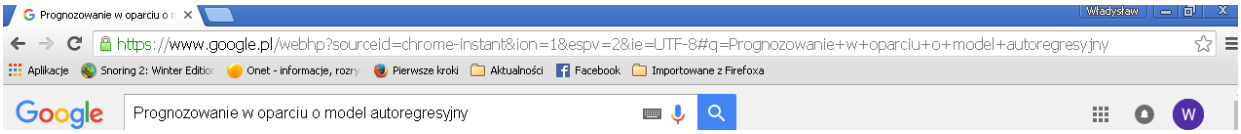
W rezultacie uzyskaliśmy model ze zmiennymi DZGD(-1), DZGD(-3) i bez stałej a_0:

$$\hat{Y} = 0{,}5458\ DZGD(-1) + 0{,}5876\ DZGD(-3)$$

Model ten w porównaniu ze wcześniejszym wykazuje wyższy współczynnik dopasowania modelu do danych empirycznych (determinacji) $R^2 = 0{,}9576$. Ponowne zbadanie istotności parametrów modelu oraz normalności składnika resztowego jest analogiczne do już zamieszczonego dla trzech zmiennych DZGD(1), DZGD(2), DZGD(3), pozostawia się to jednak do wykonania przez Czytelników.

16.2 Informacja na temat *Prognozowanie w oparciu o model autoregresyjny*

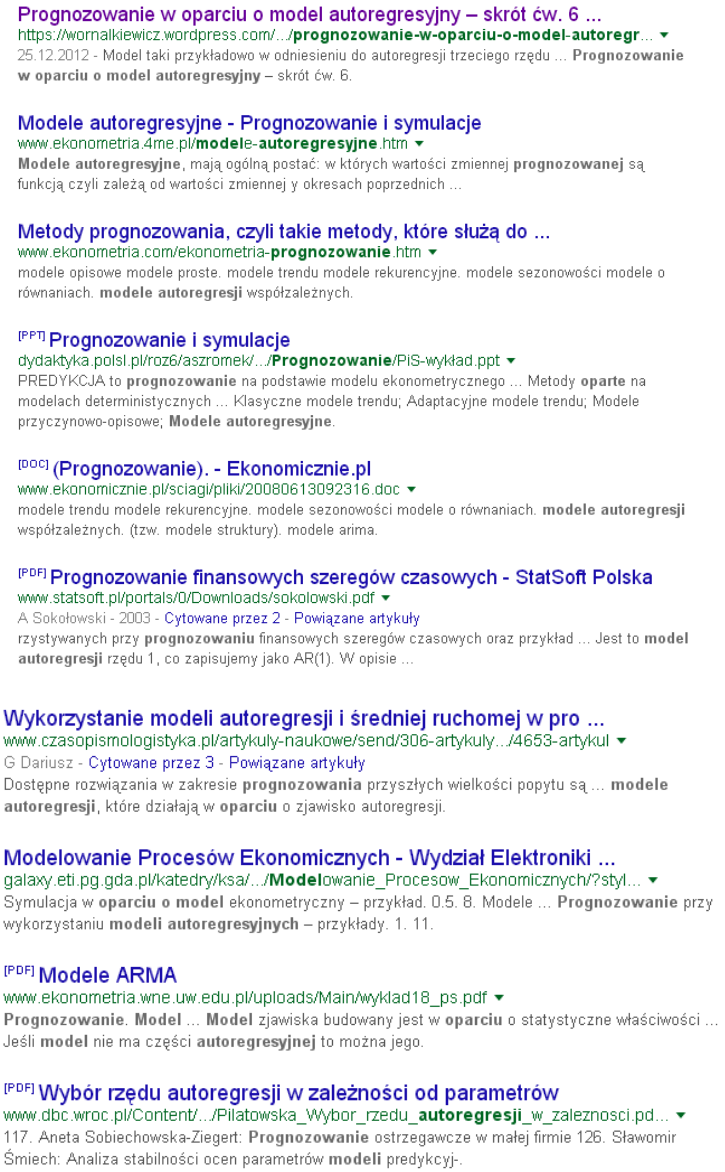
Wyszukujemy wpisów obejmujących frazę *Prognozowanie w oparciu o model autoregresyjny* w archiwach dyskowych wyszukiwarki Google (zob. rysunek 16.8). Uzyskaliśmy aż 5550 wyników w czasie 0,26 sekundy, co świadczy o popularności tego tematu wśród interesujących się prognozowanie i symulacją.



Źródło: Opracowanie własne w Google.

Rys. 16.8. Podanie frazy wyszukiwania *Prognozowanie w oparciu o model autoregresyjny*

Podobnie jak poprzednio rozpatrzmy tylko pierwszą stronę na której występują trzy publikacje o formacie PDF, więc warto zapoznać się choć pokrótce z ich zawartością (zob. rysunek 16.9).



Źródło: Opracowanie własne w Google.

Rys. 16.9. Komunikaty o wpisach na pierwszej strony listy rezultatu po podaniu frazy *Prognozowanie w oparciu o model autoregresyjny*

Google sugeruje nam jeszcze wyszukiwanie podobne do zadanego zapytania a mianowiciewedług fraz:

- model autoregresyjny *ar*,
- model autoregresyjny interpretacja,
- model wykładniczo autoregresyjny,
- autoregresja excel,
- model autoregresyjny ekonometria,
- model sredniej ruchomej,
- prognozowanie na podstawie trendu liniowego,
- prognozowanie ekonometryczne.

Otwórzmy teraz artykuł „*Prognozowanie finansowych szeregów czasowych StatSoft Polska”.* Artykuł ten autorstwa Andrzeja Sokołowskiego z Akademii Ekonomicznej w Krakowie Zakład Statystyki na stronie 1/10 informuje nas o jego zawartości a mianowicie:

„*W opracowaniu tym przedstawiono pewną grupę podstawowych modeli wykorzystywanych przy prognozowaniu finansowych szeregow czasowych oraz przykład analizy konkretnego szeregu czasowego. Składnik losowy obecny we wszystkich zjawiskach ekonomicznych jest reprezentowany przez proces stochastyczny, czyli ciąg zmiennych losowych o jednakowych rozkładach prawdopodobieństwa, zależnych od nielosowego parametru t, który reprezentuje czas. W literaturze przedmiotu i zastosowaniach spotyka się wiele klas modeli wykorzystywanych do opisu i prognozowania zjawisk finansowych. Rozsądne przedstawienie ich w jednej, krótkiej prezentacji jest niemożliwe, dlatego wybrano tylko (i tak dość liczną) pewną grupę modeli określoną przez nastepujace warunki:*

- zmienna czasowa jest zmienna skokową, a wić modelowane dane dotyczą równoległych momentów lub okresu czasu,

- rozpatrujemy kształtowanie się tylko jednej wielkości, czyli mamy do czynienia z jednowymiarowymi procesami stochastycznymi,

- prezentujemy tu tylko modele liniowe, czyli takie, w których wielkość zjawiska powiązana jest funkcją liniową z impulsami losowymi”.

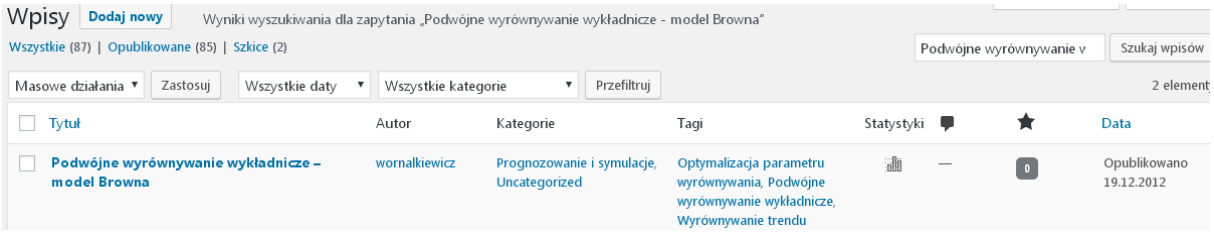
Obszerny artykuł o 69. stronach to „*Modele ARMA*”. Spotykamy tu zmiankę, że modele tego typu mogą zawierać także część autoregresyjną. Kolejny artykuł mający 25. stron wchodzacy w skład monografii *Econometric* 4(38) z roku 2012 to „*Wybór rzędu autoregresji w zależności od parametrów modelu generującego*”. Autorką jest Mariola Piłatowska z Uniwersytetu Mikołaja Kopernika w Toruniu. Wertujac tą publikację warto zainteresowac się jej streszczeniem na stronie 6/25, które jest następujące:

„W artykule pokazano zachowanie się różnych kryteriów wyboru modelu AR (kryteria informacyjne: AIC, BIC, HQ oraz sekwencyjną metodę badania istotności współczynników autoregresji) przy założeniu róznych wartości parametrów modelu autoregresyjnego i róznych wielkosci próby. Wskazano też na dużą przydatność skorygowanego kryterium AIC (AIC_c), rzadko stosowanego w polskich badaniach, do wyboru rzędu autoregresji, szczególnie w małych probach. Podkreślono, że wybór rzędu autoregresji można rozpatrywać w kontekscie wyboru prawdziwego modelu generującego (prawdziwego rzędu autoregresji), jak również w kontekscie wyboru najlepszego modelu prognostycznego (czyli wyboru rzędu modelu AR, który dałby prognozy o najmniejszych błędach prognoz). Rozważania zilustrowano analizą symulacyjną”.

17. Podwójne wyrównywanie wykładnicze - model Browna

17.1. Wywołanie i statystyka wpisu

W ramach omawianego blogu system *WordPress* znalazł dwa wpisy odpowiadajace zadanej frazie „*Podwójne wyrównywanie wykładnicze - model Browna*". Pierwszy z nich zawiera ćwiczenie piąte w ramach przedmiotu „*Badania operacyjne*" (zob. rysunek 17.1).



Źródło: Opracowanie własne w *WordPress*.

Rys. 17.1. Komunikat o występowaniu pierwszego z wpisów udostępnionych po zapytaniu „*Podwójne wyrównywanie wykładnicze - model Browna*"

Dostęp bezpośredni do tego wpisu uzyskujemy poprzez adres:

https://wornalkiewicz.wordpress.com/wp-admin/edit.php?s=Podwójne+wyrównywanie+wykładnicze+-+model+Browna&post_status=all&post_type=post&

Po wyborze opcji *Edytuj* oprócz adresu system podaje w skróconej formie odnośnik do aktualizacji tekstu danego wpisu:

Statystykę wpisu założonego w grudniu 2012 roku według miesięcy i lat pokazana została na rysunku 17.2.

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2013	8	12	10	6	30	25	2	3	7	5	7	4	119
2014	29	23	9	13	10	23	1	0	6	6	32	13	165
2015	51	21	10	23	13	15	5	1	8	19	27	38	231
2016	58	42	17	27	17	16	3	3	7	4			194

Źródło: Opracowanie własne w *WordPress*.

Rys. 17.2. Statystyka wpisu „*Podwójne wyrównywanie wykładnicze - model Browna*"

W miesiącu styczniu kolejnych lat obserwujemy wzrost liczby odwiedzających:

2013	2014	2015	2016
8	29	51	58

17.2. Tekst wpisu

Po wyborze opcji **Edytuj** *WordPress* udostępnia tekst wpisu *Podwójne wyrównywanie wykładnicze - model Browna* (*skrót ćwiczenia* 5). Dla zaprezentowania modelu Browna w blogu

podano przykład szeregu czasowego dziesięciu obserwacji zmiennej Y_n w celu określenia prognozy okresu 11 i 12[198], który jest następujący:

t	1	2	3	4	5	6	7	8	9	10
y_t	29	27	31	35	33	44	45	48	51	53

W ramach przykładu określamy na podstawie podanych dalej wzorów, dla każdego okresu (1-10) i parametru wyrównującego $\alpha = 0{,}3$, elementy modelu Browna w których a_n, b_n - parametry, Y_n - wartość obserwacji dla okresu n, S'_{n-1}, S''_{n-1} - odpowiednio pojedynczo, podwójnie wyrównywana wartość dla okresu poprzedniego (n -1):

a) Pojedynczo wyrównana wartość szeregu zmiennej w okresie n dla parametru wyrównywania $\alpha = 0{,}3$

$$S'_n = \alpha\, Y_n + (1 - \alpha)\,(S'_{n-1}).$$

b) Podwójne wyrównana wartość szeregu zmiennej

$$S''_n = \alpha\, S'_n + (1 - \alpha)\,(S''_{n-1}).$$

c) Wyrównana wykładniczo wartość zmiennej dla prognozy

$$a_n = S'_n + (S'_n - S''_n) = 2\,S'_n - S''_n.$$

d) Wyrównana wykładniczo wartość trendu dla prognozy:

$$b_n = [\alpha/(1 - \alpha)]\,(S'_n - S''_n).$$

Przyjmijmy wartości początkowe: $S'_1 = S''_1 = Y_1 = 29$; $a_1 = Y_1 = 29$; $b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2 = [(27\text{-}29) + (35\text{-}31)]/2 = 1$. Określamy teraz błąd wygasłych prognoz stosując miarę RMSE, czyli pierwiastek ze średniej sumy kwadratów odchyleń obserwacji od prognozy. Dla naszego przykładu sumujemy od t = 2-10. Po obliczeniu parametrów (a_{10}), (b_{10}) określamy prognozy na okresy 11 i 12 stosując wzór:

$$F_{(n+m)} = a_{10} + b_{10} \cdot m,$$ przy czym m odpowiednio równa się 1 oraz 2.

Obliczenia przeprowadzone zostały w *Excelu* (zob. tabela 17.1).

Tab. 17.1

Obliczenia pomocnicze oraz określenie prognoz modelem Browna

	A	B	C	D	E	F	G	H	I
1	t	Y_n	S'_n	S''_n	a_n	b_n	F_n	$(Y_n - F_n)^2$	α
2	1	29	29,00	29,00	29,00	1,00			0,3
3	2	27	28,40	29,00	27,80	-0,26	30,00	9,00	$1 - \alpha$
4	3	31	29,18	28,82	29,54	0,15	27,54	11,95	**0,7**
5	4	35	30,93	28,93	32,92	0,86	29,69	28,15	$\alpha/(1-\alpha)$
6	5	33	31,55	29,53	33,57	0,87	33,78	0,61	**0,428571**
7	6	44	35,28	30,13	40,43	2,21	34,44	91,49	m
8	7	45	38,20	31,68	44,72	2,79	42,64	5,56	1
9	8	48	41,14	33,63	48,64	3,22	47,51	0,24	2
10	9	51	44,10	35,89	52,31	3,52	51,86	0,74	S
11	10	53	**46,77**	**38,35**	**55,19**	**3,61**	**55,83**	**8,00**	**155,74**
12	11						**58,79**		S/i
13	12						**62,40**		**19,47**
14									RMSE
15									**4,41**

gdzie: MSE = 19,47.

Źródło: Opracowanie własne w *Excelu*.

Czcionka pogrubioną zaznaczono komórki arkusza dla których podano formuły *Excela* pokazane na rysunku 17.3.

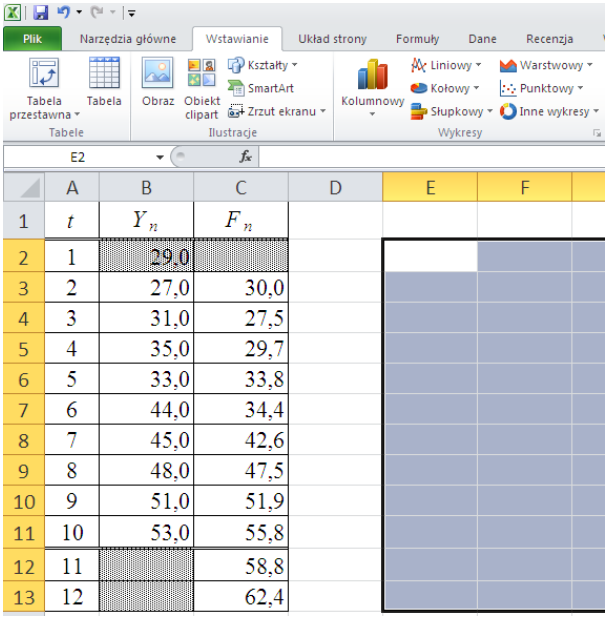
[198] Podane dziesięć liczb oraz zamieszczone wzory pochodzą z rozdziału 2.13, książki: Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom II. *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.

K	L	M	N	O	P	Q	R	S	T	U
S'_{10}	S''_{10}	a_{10}	b_{10}	F_{10}	$(Y_{10} - F_{10})^2$	$1 - \alpha$	$\alpha/(1-\alpha)$	S	S/i	RMSE
$I$2*B11+(1-$I$2)*C10										
	$I$2*C10+(1-$I$2)*D10									
		2*C11-D11								
			$I$6*(C11-D11)							
				E10+F10*$I$8						
					(B11-G11)^2					
						1-I2				
							$I$2/(1-$I$2)			
								SUMA(H3:H11)		
									I11/(A11-A3)	
										PIERWIASTEK(I13)

Źródło: Opracowanie własne w *Excelu*.

Rys 17.3. Przykład zastosowanych formuł do wyliczenia elementów metody Browna

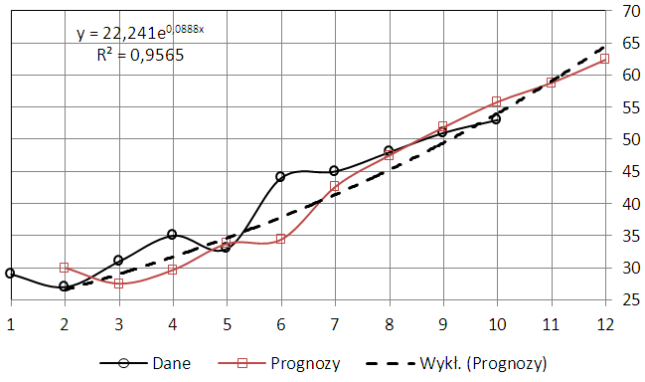
Możemy jeszcze przedstawić na wykresie punktowym dane rzeczywiste i teoretyczne prognoz otrzymane z prognozowania modelem Browna.



	A	B	C	D	E	F
1	t	Y_n	F_n			
2	1	29,0				
3	2	27,0	30,0			
4	3	31,0	27,5			
5	4	35,0	29,7			
6	5	33,0	33,8			
7	6	44,0	34,4			
8	7	45,0	42,6			
9	8	48,0	47,5			
10	9	51,0	51,9			
11	10	53,0	55,8			
12	11		58,8			
13	12		62,4			

Źródło: Opracowanie własne w *Excelu*.

Rys. 17.4. Etap początkowy przygotowania danych i wyboru wykresu punktowego dla Y_n i F_n



Źródło: Opracowanie własne w *Excelu*.

Rys. 17.5. Dane rzeczywiste, prognozy wygasłe i przyszłe na okresy 11 i 12 oraz trend wykładniczy modelu

Zwróćmy uwagę na wysokie dopasowanie modelu do danych rzeczywistych - współczynnik determinacji $R^2 = 0{,}9565$.

A teraz z ciekawości dokonajmy symulacji parametru wyrównującego α dobierając spośród przedziału < 0-1> wartość najlepszą modułem FC z punktu widzenia minimalnej wartości kryterium MSE. W tym celu definiujemy dane i parametry wejściowe jako:

Forecasting Method (metoda prognozowania) - *Double exponential smoothing with trend* (podwójne wyrównywanie wykładnicze z trendem),

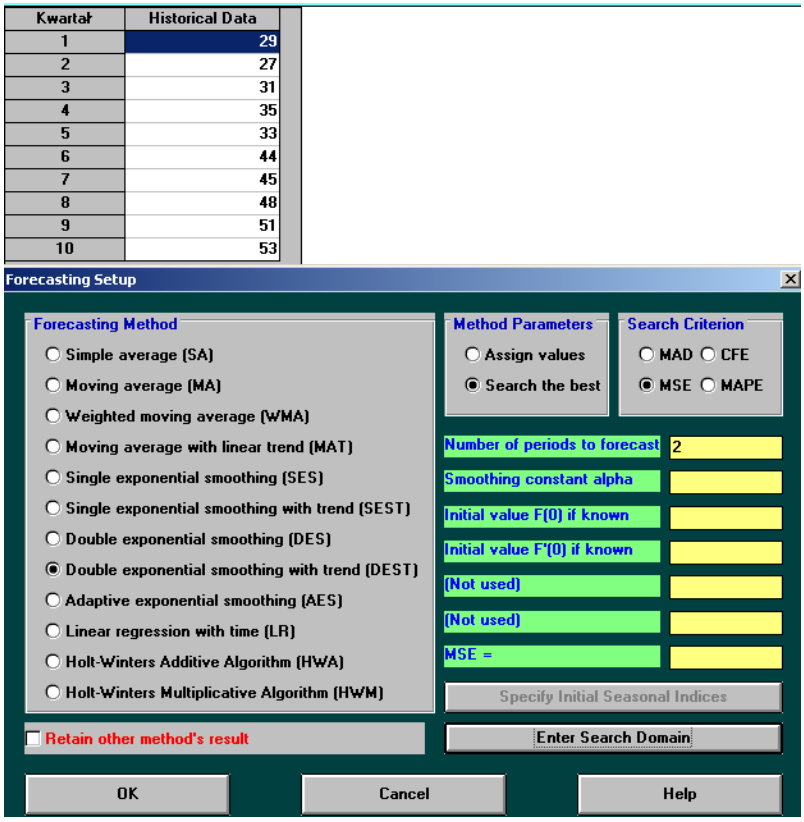
Method Parameters (metoda doboru parametrów) - Search the best (najlepsza z punktu widzenia kryterium),

Search Criterion (kryteria wyszukiwania) - MSE (pierwiastek z RMSE),

Number of periods to forecast (liczba okresów prognozowania),

Enter Search Domain (skok iteracji – 0.01, podawany z kropką dziesiętną).

Zobrazowanie podanych parametrów przedstawiono na rysunku 17.6.



Źródło: Opracowanie własne w *WinQSB*.

Rys. 17.6. Ustawienie metody i parametrów prognozowania w module FC

Uzyskaliśmy rezultat (zob. rysunek 17.9) z założoną dokładnością do jednego miejsca po przecinku. Porównajmy teraz wyniki z formuł Excela oraz z realizacji automatycznej z optymalizacją parametru wyrównywania α.

Wynik	F_{11}	F_{12}	MSE	RMSE	α
Formuły Excela/Moduł *FC*	58,8/56,2	62,4/59,1	19,47/**15,1**	4,41/3,89	0,3/0,48

Tak więc poprzez optymalny dobór α uzyskaliśmy aż o 12 % zmniejszenie błędu RMSE.

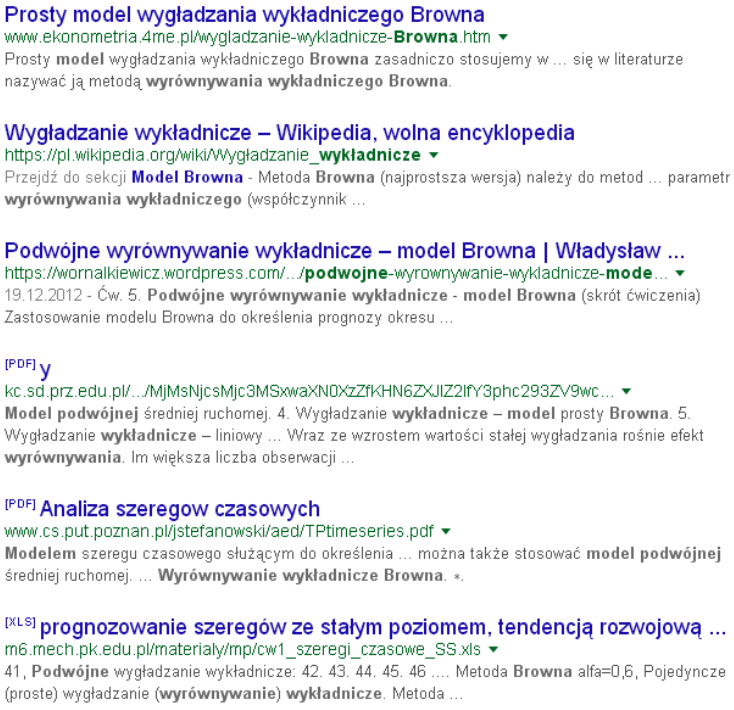
01-02-2017 Kwartał	Actual Data	Forecast by DEST	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-square
1	29,0								
2	27,0	29,0	-2,0	-2,0	2,0	4,0	7,4	-1,0	
3	31,0	27,1	3,9	1,9	3,0	9,7	10,0	0,6	0,5
4	35,0	30,4	4,6	6,5	3,5	13,6	11,1	1,9	0,6
5	33,0	35,3	-2,3	4,3	3,2	11,4	10,0	1,3	
6	44,0	34,6	9,4	13,7	4,4	26,8	12,3	3,1	0,5
7	45,0	44,6	0,4	14,1	3,8	22,4	10,4	3,7	0,9
8	48,0	48,1	-0,1	13,9	3,2	19,2	8,9	4,3	
9	51,0	51,2	-0,2	13,7	2,9	16,8	7,9	4,8	
10	53,0	54,2	-1,2	12,5	2,7	15,1	7,3	4,6	
11		56,2							
12		59,1							
CFE		12,5							
MAD		2,7							
MSE		15,1							
MAPE		7,3							
Trk.Signal		4,6							
R-square		1,2							
		Alpha=0,48							
		F(0)=29							
		F'(0)=29							

Źródło: Opracowanie własne w module FC pakietu *WinQSB*.

Rys. 17.7. Tabela wyników optymalnego doboru parametru α w FC

17.3. Problematyka frazy *Podwójne wyrównywanie wykładnicze - model Browna*

Przy podaniu frazy *Podwójne wyrównywanie wykładnicze - model Browna* system *Google* wygenerował listę zawierającą 156 pozycji w czasie 0,39 sekundy. Na stronie pierwszej znalazły się wpisy zamieszczone na rysunkach 17.8 i 17.9.



Źródło: Opracowanie własne w *Google*.

Rys. 17.8. Część pierwsza wpisów na pierwszej stronie listy rezultatu wyszukiwania frazy *Podwójne wyrównywanie wykładnicze - model Browna*

Źródło: Opracowanie własne w *Google*.

Rys. 17.9. Część druga wpisów na pierwszej stronie listy rezultatu wyszukiwania frazy *Podwójne wyrównywanie wykładnicze - model Browna*

Występują tu dwie publikacje formatu PDF, przy czym wyszukiwarka zachęca nas do skorzystania jeszcze z fraz podobnych pisanych małymi literami do początkowo zadanej, które są następujące:

- model holta,
- model browna w Excelu,
- model holta logistyka,
- model wintersa,
- model adaptacyjny holta,
- metoda wygładzania wykładniczego excel,
- **paraboliczny model browna**,
- wygładzanie wykładnicze zadania.

Pojawił się moim zdaniem rzadko spotykany w literaturze *paraboliczny model Browna* do poznania którego zachęcam Czytelnika. Publikacja pierwsza formatu PDF ma nazwę „y" i dotyczy prostego modelu wygładzania wykładniczego Browna. Autor informuje nas, że model może być stosowany do prognozowania zjawisk, które charakteryzują się prawie stałym poziomem oraz niewielkimi wahaniami przypadkowymi i definiuje go następująco:

$$y_t^* = \propto y_{t-1}(1-\propto)y_{t-1}^*, \quad \propto \in (0,1)$$

$$y_t^*, y_{t-1}^* \; prognoza \; wyznaczona \; na \; okres \; t \; i \; (t-1)$$

$$y_{t-1} \; wielkość \; zmiennej \; w \; okresie \; (t-1)$$

α - parametr wygładzania wyznaczony eksperymentalnie.

Występuje tu nieco inne oznaczenie prognozy, bo z gwiazdką (*), a wcześniej autorzy podawali z daszkiem (^). Jako wartość prognozy y_1^*, niezbędną do konstrukcji modelu, przyjmuje się najczęściej wartość początkową zmiennej prognozowanej (y_1) lub średnią arytmetyczną z kilku pierwszych obserwacji.

Druga publikacja tego samego formatu to *Analiza szeregów czasowych* i obejmuje 42 strony, przy czym na stronie 20/42 podjęto temat „*Wyrównywanie wykładnicze Browna*". Wychodzi się ze wzoru identycznego jak w publikacji wcześniejszej, który może być przekształcony do postaci:

$$y_t^* = y_{t-1}^* + (1-\propto)(y_{t-1} - y_{t-1}^*).$$

W artykule autor rozważył przykład o dwóch różnych wartościach $\alpha = 0{,}2$ i $\alpha = 0{,}8$. Wyniki końcowe obliczeń bazujące na wymienionym wzorze wykonałem w *Excelu* (zob. rysunek 17.10).

	A	B	C	D	E	F	G	H	I
1			a = 0,2			a = 0,8			α
2	t	Y_t	F_t	$Y_t - F_t$	$(Y_t - F_t)^2$	F_t	$Y_t - F_t$	$(Y_t - F_t)^2$	0,2
3	1	15	15,00	0,00	0,00	15,00	0,00	0,00	β
4	2	17	15,00	2,00	4,00	15,00	2,00	4,00	0,8
5	3	12	15,40	-3,40	11,56	16,60	-4,60	21,16	
6	4	16	14,72	1,28	1,64	12,92	3,08	9,49	
7	5	15	14,98	0,02	0,00	15,38	-0,38	0,15	
8	6	11	14,98	-3,98	15,85	15,08	-4,08	16,62	
9	7	18	14,18	3,82	14,56	11,82	6,18	38,25	
10	8	17	14,95	2,05	4,21	16,76	0,24	0,06	
11	9	13	15,36	-2,36	5,56	16,95	-3,95	15,62	
12	10	16	14,89	1,11	1,24	13,79	2,21	4,88	
13	11		15,11	Σ	58,62	15,56	Σ	110,23	

Źródło: Opracowanie własne w *Excelu* bazujące na publikacji internetowej „*Analiza szeregów czasowych*”

Rys. 17.10. Wyniki metody wyrównywania wykładniczego

Zauważmy, że suma kwadratów różnic zmiennej prognozowanej Y_t i prognoz F_t^* dla wygasłych prognoz wyraźnie jest wyższa przy α = 0,8. Dla bliżej zainteresowanych tym przykładem podaję zastosowane formuły w arkuszu kalkulacyjnym *Excel* (zob. rysunek 17.11).

K	L	M	N
F_{10}	$Y_{10} - F_{10}$	$(Y_{10} - F_{10})^2$	Σ
C11+$I$2*(B11-C11)			
	B12-C12		
		D12^2	
			SUMA(E3:E12)

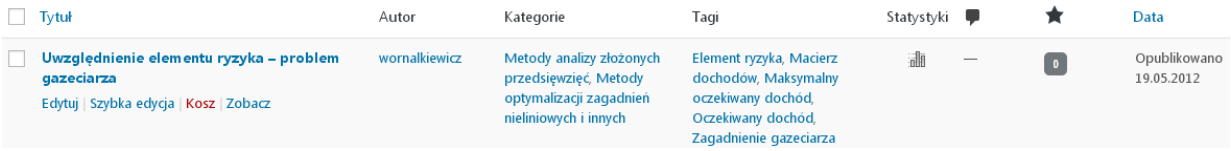
Źródło: Opracowanie własne w *Excelu* „*Analiza szeregów czasowych*”

Rys. 17.11. Zastosowane formuły obliczeń w metodzie wyrównywania wykładniczego

18. Uwzględnienie elementu ryzyka - problem gazeciarza

18.1. Tekst wpisu

_Wpis o podanym tytule (frazie) zamieściłem na blogu w maju 2012 roku (zob. rys. 17.1).



Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Uwzględnienie elementu ryzyka – problem gazeciarza Edytuj | Szybka edycja | Kosz | Zobacz	wornalkiewicz	Metody analizy złożonych przedsięwzięć, Metody optymalizacji zagadnień nieliniowych i innych	Element ryzyka, Macierz dochodów, Maksymalny oczekiwany dochód, Oczekiwany dochód, Zagadnienie gazeciarza	—	0	Opublikowano 19.05.2012

Źródło: Opracowanie własne w *WordPress*.

Rys. 18.1. Komunikat o występowaniu w omawianym blogu wpisu „*Uwzględnienie elementu ryzyka - problem gazeciarza*”

Natomiast statystykę miesięczną wejść internautów do tekstu wpisu zamieszczono na rysunku 18.2.

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012					3	3	4	2	6	1	2	6	27
2013	35	13	7	19	22	13	0	1	8	6	7	5	136
2014	12	10	2	9	9	6	0	0	2	7	12	5	74
2015	15	3	4	0	12	18	2	0	1	5	4	12	76
2016	11	4	1	9	4	7	1	3	3	3			46

Źródło: Opracowanie własne w *WordPress*.

Rys. 18.2. Statystyka miesięczna zainteresowania wpisem „*Uwzględnienie elementu ryzyka - problem gazeciarza*”

Na koniec roku 2016 w kolumnie „*W sumie*” było 55 wejść. Oprócz roku 2013, gdy było jeszcze znaczne zainteresowanie studentów kierunkiem *Ekonometria* obserwowany jest wyraźny skok sumarycznych wejść, a potem występuje prawie stacjonarne (74, 76, 55).

Zaznaczone małą czcionką fragmenty tego wpisu pochodzą z książki: Duczmal M. *Metody optymalizacji w zarzadzaniu*. Ponadto w opracowaniu całości bazowano na pracy zbiorowej: *Przykłady i zadania z badań operacyjnych i ekonometrii*[199].

Książka A jest reklamowana na stronie WWW danej księgarni internetowej. Księgarnia ta wysyła zamówione książki odbiorcom. Co tydzień zamawia je w określonej partii w hurtowni, ponosząc ryzyko związane z popytem y. Obserwowany jest cotygodniowy popyt na książkę A oznaczony jako zmienna Y, który w kolejnych 12. tygodniach kwartału był następujący: 12 10 12 15 10 8 12 18 15 12 8 10.

Księgarnia internetowa kupuje książkę A po $c_1 = 40$ zł, a sprzedaje po $c_2 = 46$ zł. Po tygodniu ma możliwość - jeśli nie ma nabywców - zwrócić ją do hurtowni po 30 zł, ponosząc stratę $c_3 = 10$ zł. Zysk ze sprzedaży książki $b = c_2 - c_1 = 6$ zł. Pojawia się problem w jakich ilościach, co tydzień, najlepiej zamawiać książkę A, aby oczekiwany dochód był maksymalny.

W niniejszej pracy problem gazeciarza rozwiązano z wykorzystaniem formuł arkusza kalkulacyjnego *Excel* rozpoczynając od wyznaczenia rozkładu prawdopodobieństwa $p(y)$ na podstawie częstości wartości popytów obliczonej jako stosunek występowania danej liczby popytu n

[199] Anholcer M., Gaspars H., Owczarkowski A., *Przykłady i zadania z badań operacyjnych i ekonometrii*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2003, rozdz. 7.1.

do N = 12 tygodni kwartału. Potencjalną podaż oznaczono jako z, a dochód jako funkcja podaży i popytu obliczany jest na podstawie wzoru:

$$g(z, y) = bz - s(z - y) = 6z - 10(z - y)$$

gdzie: b - zysk ze sprzedaży jednej książki, s - strata, z - podaż, y - popyt.

Możemy teraz sporządzić w *Excelu* macierz dochodów dla kolejnych wariantów podaży i popytu Występuje jednak sytuacja, że dochód nie rośnie jeśli potencjalny popyt przekracza podaż i pozostaje na poziomie podaży (zob. rysunek 18.3). Ponadto na podanym rysunku obliczono średnią ważoną dochodów i ich prawdopodobieństw według przykładu jak dla podaży z = 12:

$$d(12) = 32 * 0{,}1667 + 52 * 0{,}25 + 72 * 0{,}3333 + 72 * 0{,}1667 + 72 * 0{,}0833 = 60{,}3.$$

W obliczeniach zastosowano formuły *Excela* zamieszczone na rysunku 18.4, które jak już nadmieniłem muszą być poprzedzane znakiem (=). Zwróćmy uwagę, że największy oczekiwany średni dochód tygodniowo jest dla podaży z = 12, bo wynosi 60,3.

	A	B	C	D	E	F	G	H
1	y	8	10	12	15	18		N
2	n	2	3	4	2	1		12
3	$p(y)$	**0,1667**	0,2500	0,3333	0,1667	0,0833		**1**
4	z	$g(z, y)$					$d(z)$	b
5	8	**48**	**48**	48	48	48	48,0	6
6	10	40	**60**	60	60	60	56,7	s
7	12	32	52	**72**	72	72	**60,3**	10
8	15	20	40	60	**90**	90	55,8	
9	18	8	28	48	78	**108**	46,3	

Źródło: Opracowanie własne w *Excelu*.

Rys. 18.3. Obliczenia prawdopodobieństw, dochodu oraz oczekiwanego dochodu

Cecha	Formuła
$p(y=8)$	B2/$H$2
$g(z=8, y=8)$	$H$5*A5-(A5-$B$1)*$H$7
$g(z=8, y=10)$	B5
$g(z=10, y=10)$	$H$5*A6-(A6-$C$1)*$H$7
$g(z=12, y=12)$	$H$5*A7-(A7-D1)*$H$7
$g(z=15, y=15)$	$H$5*A8-$H$7*(A8-$E$1)
$g(z=18, y=18)$	H5*A9-(A9-F1)*H7
$d(z)$	B7*$B$3+C7*$C$3+D7*$D$3+E7*$E$3+F7*$F$3

Źródło: Opracowanie własne w *Excelu*.

Rys. 18.4. Przykładowe formuły zastosowane w tabeli - rysunek 18.3

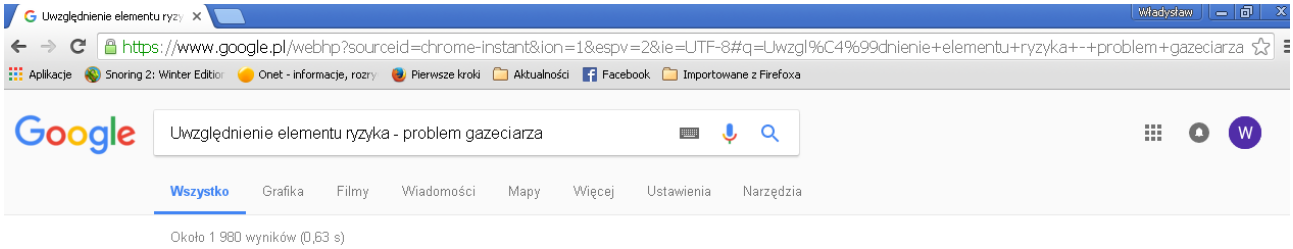
18.2. Problematyka frazy *Uwzględnienie elementu ryzyka - problem gazeciarza*

Dla tak długiej wprowadzonej frazy w cudzysłowie system *Google* wygenerował tylko cztery wyniki w czasie 0,52 sekundy, w tym wpis autora blogu w systemie W*ordPress* występujący na pierwszej pozycji raportu (zob. rysunek 18.5).

Źródło: Opracowanie własne w *Excelu*.

Rys. 18.5. Raport wyszukiwania dla frazy
„*Uwzględnienie elementu ryzyka - problem gazeciarza*”

W raporcie występuje jednak zastrzeżenie, że pokazano najbardziej trafne wyniki i pominięto kilka pozycji podobnych do czterech już wymienionych, z których jednak nie skorzystamy, gdyż nie są formatu PDF. Z ciekawości sprawdźmy ile wyników da nam wyszukiwarka, gdy wprowadzimy frazę bez cudzysłowie (zob. rysunek 18.6).



Źródło: Opracowanie własne w *WordPress*.

Rys. 18.6. Wywołanie listy rezultatu po wprowadzeniu frazy
Uwzględnienie elementu ryzyka - problem gazeciarza

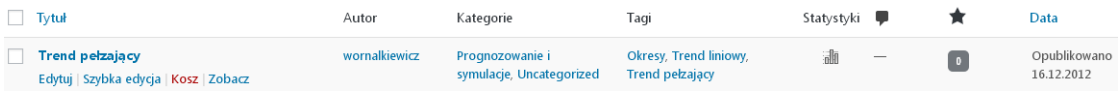
Otrzymaliśmy teraz 1980 wyników w czasie 0,63 sekundy.

19. Trend pełzający

19.1. Tekst wpisu

Wpis „*Trend pełzający*” wprowadziłem do blogu „*Zastosowanie komputera w procesach decyzyjnych*” w grudniu 2012 roku. Po dostępie do tego wpisu (zob. rysunek 19.1) możemy kliknąć na ikonę *Statystyki* i wygenerować następujące zestawienia:

- wykres słupkowy „gości” - odwiedzających,
- tabela odwiedzin według miesięcy,
- tabela odwiedzin „*Średnio na dzień*”,
- tabela odwiedzin w dniach tygodnia w okresie trwania blogu.



Źródło: Opracowanie własne w *WordPress*.

Rys. 19.1. Komunikat o występowaniu wpisu „*Trend pełzający*”

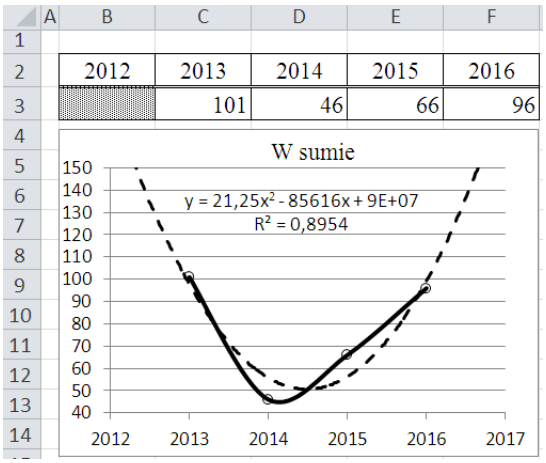
Podobnie jak wcześniej zainteresujmy się zestawieniem w układzie lat i miesięcy prowadzenia blogu (zob. rysunek 19.2). Na zakończenie roku 2016 ilość wejść internautów wynosiła 96.

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012												2	2
2013	17	8	2	15	13	14	1	8	2	3	10	8	101
2014	9	3	1	3	12	3	0	1	1	1	4	8	46
2015	15	1	2	2	7	9	0	1	2	4	7	16	66
2016	26	1	3	9	13	8	1	1	6	0			68

Źródło: Opracowanie własne w *WordPress*.

Rys. 19.2. Statystyka miesięczna wejść do wpisu „*Trend pełzający*”

Skorzystajmy z *Excela* dla wykonania wykresu trendu wielomianowego 2-go stopnia (zob. rysunek 19.3). Rok 2012 pominięto ze względu na czas wprowadzenia omawianego wpisu i tylko dwóch zainteresowanych.



Źródło: Opracowanie własne w *Excelu*.

Rys. 19.3. Wykres wielomianowy statystyki według lat wejść do wpisu „*Trend pełzający*” z prognozą na rok 2017

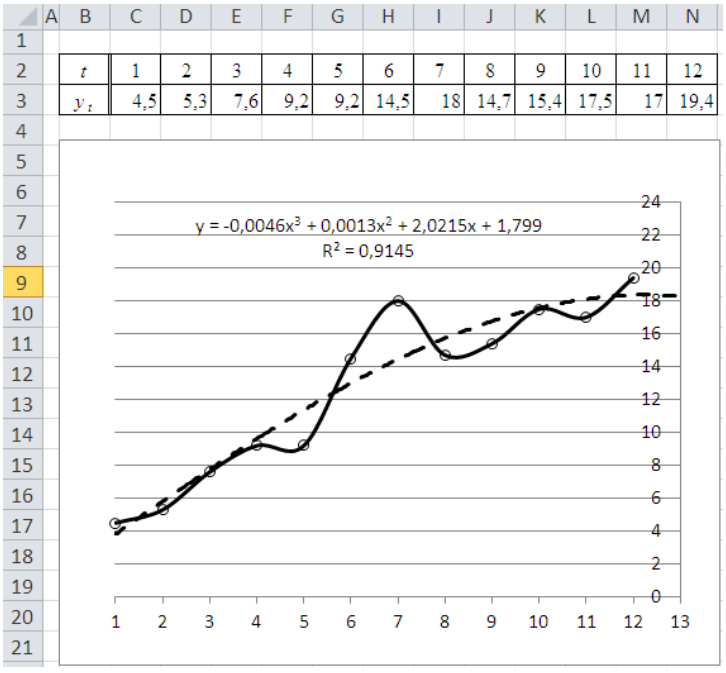
Tekst wpisu do omawianego blogu stanowi skrót ćwiczenia 4. przeprowadzonego przez autora niniejszej pracy na wykładzie z ekonometrii[200].

Mamy szereg czasowy obserwacji dla okresów *t*. Określamy długość segmentu ciągu stanowiącego *k* okresów. Dla każdego ciągu wyznaczamy trend liniowy korzystając z klasycznej metody najmniejszych kwadratów. Następnie obliczamy wartości teoretyczne dla każdego z okresów na podstawie równań trendów liniowych ciągów. Mamy (n - k +1) ciągów, przy czym n to ilość obserwacji szeregu czasowego. Dla okresu pierwszego wystąpi jedna wartość obliczenia, dla drugiego dwie, bo z ciągu pierwszego i drugiego. Natomiast dla ciągu trzeciego, aż po (n - k +1) ciąg po trzy wartości teoretyczne dla każdego danego okresu. W kolejności obliczamy średnie arytmetyczne z wartości teoretycznych dla poszczególnych n okresów. Prognozę na okres (n + 1) ustalamy na podstawie równania trendu dla ostatniego ciągu tj. (n - k +1). Możemy także określić prognozę dla okresu (n +1) budując równanie trendu liniowego na podstawie średnich arytmetycznych dla okresów (n - 2, n - 1, n).

Proponuję przeprowadzić obliczenia dla trendu pełzającego z wyznaczeniem prognozy na okres 13. na podstawie średnich arytmetycznych wartości teoretycznych trendów liniowych dla okresów 10, 11, 12 posługując się funkcją REGLINP oraz funkcjami elementarnymi *Excela*. Skorzystajmy z następującego szeregu czasowym 12 obserwacji:

t	1	2	3	4	5	6	7	8	9	10	11	12
y_t	4,5	5,3	7,6	9,2	9,2	14,5	18	14,7	15,4	17,5	17	19,4

Prognozę na okres trzynasty możemy też określić na podstawie trendu wielomianowego 3-go stopnia uzyskanego w *Excelu* (zob. rysunek 19.4). Zwróćmy uwagę na wysokie dopasowanie danych empirycznych do modelowych, gdyż współczynnik determinacji R^2 = 0,9145.



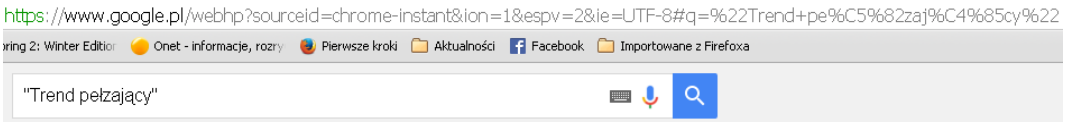
Źródło: Opracowanie własne w *Excelu*.

Rys. 19.4. Wykres wielomianowy 3-go stopnia dla danych przykładu z prognozą na rok 2017

[200] Pełne rozwiązanie przykładu znajduje się w rozdziale 2.4.6.2 książki: Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, Wydawnictwo Instytut Śląski, Opole 2015.

19.2. Problematyka frazy *Trend pełzający*

Po wprowadzeniu zapytania „*Trend pełzający*” do wyszukiwarki *Google* (zob. rysunek 19.5) uzyskano 601 wyników w czasie 0,56 s.

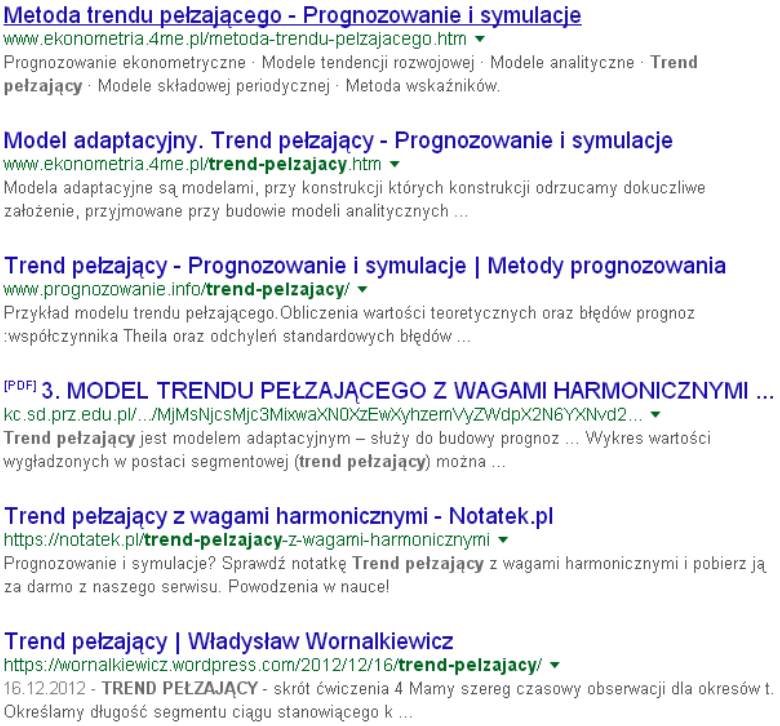


Źródło: Opracowanie własne w Excelu.

Rys. 19.5. Zapytanie skierowane do *Google*

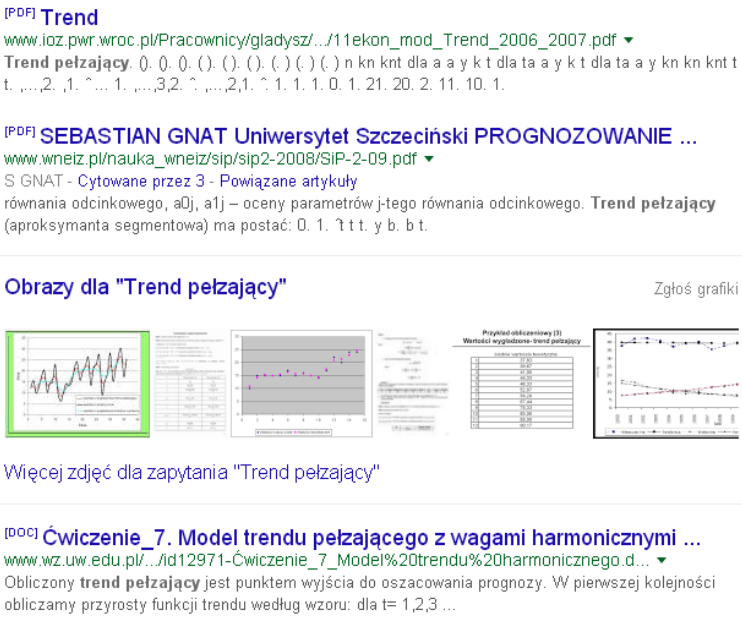
Na pierwszej stronie listy rezultatu udostępnionej przez wyszukiwarkę znalazły się wpisy zamieszczone na rysunku 19.6. Czytając skróty wpisów zauważamy także informację o trendzie pełzającym stanowiące wpis do blogu prowadzonego przez autora tego opracowania w systemie *WordPress*. W zestawieniu występują poszukiwane przez nas trzy publikacje w formacie wydruku, czyli PDF. Na końcu listy jako rezultatu wyszukiwania dla podanej frazy w cudzysłowie system sugeruje nam jeszcze możliwość skorzystania z następującego wyszukiwania podobnego pisanego małymi literami do „*Trend pełzający*”:

- trend pełzający excel,
- trend pełzający wikipedia,
- metoda wyrównywania wykładniczego,
- metoda wag harmonicznych zadania,
- wagi harmoniczne tablica,
- model adaptacyjny,
- model trendu pełzającego,
- metoda holta.



Źródło: Opracowanie własne w *Google*.

Rys. 19.6. Pierwszy fragment pierwszej strony listy rezultatu na zapytanie „*Trend pełzający*”



Źródło: Opracowanie własne w *Excelu*.

Rys. 19.7. Drugi fragment pierwszej strony listy rezultatu wyszukiwania dla frazy „*Trend pełzający*”

Na stronie 3/16 opracowania „*Model trendu pełzającego z wagami harmonicznymi*” zamieszczono krótki przykład konstrukcji modelu pełzającego dla okresu $k = 3$. Trend pełzający stosowany jest do prognoz krótkoterminowych, a w budowie modelu według tej publikacji wyróżniamy dwa etapy:

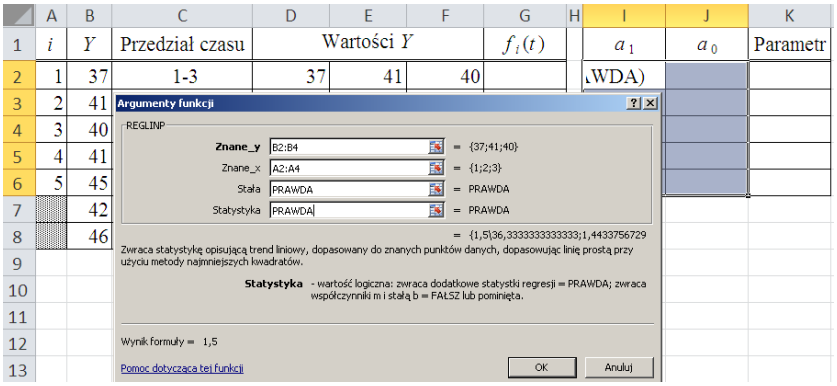
1. Ustalenie wartości stałej wygładzania $k < n$; wyższa wartość tej stałej powoduje większe wygładzanie szeregu, czyli słabsze reagowanie na zmiany zachodzące w szeregu.

2. Oszacowanie na podstawie kolejnych fragmentów szeregu o długości k, liniowych funkcji trendu.

Skupiono się zatem tylko na trendzie pełzającym bez odniesienia się do wag harmonicznych. Obliczenia zrealizowano na przykładzie szeregu czasowego danych umownych 7. obserwacji:

t	1	2	3	4	5	6	7
y_t	37	41	40	41	45	42	46

Wzór na funkcję np. pierwszego trendu segmentowego jest następujący: $f_1(t) = a_1 t + b_1$. Niech $k = 3$, to wtedy wartości dla segmentów trendu pełzającego możemy obliczyć korzystając z funkcji REGLINP Excela (zob. rysunek 19.8), który dla pierwszego segmentu obejmującego przedział czasu (1-3) jest postaci: (= REGLINP(B2:B4;A2:A4;PRAWDA;PRAWDA)).



Źródło: Opracowanie własne w *Excelu* według danych omawianej publikacji.

Rys. 19.8. Skorzystanie z REGLINP() dla określenia parametrów a_0 i a_1 funkcji segmentowej trendu pełzającego

I	J	K
a_1	a_0	Cechy
1,50	36,33	parametry modelu liniowego
1,44	3,12	błąd standardowy oszacowania parametrów
0,52	2,04	R^2, S_e
1,08	1,00	F, *iss*
4,50	4,17	ESS, RSS

gdzie:

R^2 - współczynnik determinacji,
S_e - odchylenie standardowe reszt,
F - statystyka Fishera,
iss - ilość stopni swobody (n - m - 1),
ESS - wyjaśniona prze model suma kwadratów,
RSS - resztowa suma kwadratów.

Źródło: Opracowanie własne w *Excelu*.

Rys. 19.9. Parametry modelu liniowego pierwszego segmentu trendu pełzającego ze statystyką opisową

Wykonując analogicznie obliczenia parametrów modelu trendu liniowego dla kolejnych segmentów trendu pełzającego otrzymujemy modele zamieszczone na rysunku 19.9.

	A	B	C	D	E	F	G
1	$i = t$	Y	Przedział czasu	Wartości Y			$f_i(t)$
2	1	37	1-3	37	41	40	$f_1 = 36,33 + 1,5 \cdot t$
3	2	41	2-4	41	40	41	$f_2 = 40,67 + 0 \cdot t$
4	3	40	3-5	40	41	45	$f_3 = 32,00 + 2,5 \cdot t$
5	4	41	4-6	41	45	42	$f_4 = 40,17 + 0,5 \cdot t$
6	5	45	5-7	45	42	46	$f_5 = 41,33 + 0,5 \cdot t$
7	6	42					
8	7	46					

Źródło: Opracowanie własne w *Excelu* na podstawie omawianej publikacji.

Rys. 19.9. Modele pięciu segmentów trendu pełzającego

Drugą publikacją o formacie PDF jest wpis pt. „*Trend*", w którym na stronie 7/19 podano równania trendu pełzającego dla poszczególnych segmentów, przy czym postaci tych równań są następujące:

$$\hat{y}_{t(1)} = a_{0(1)} + a_{1(1)} * t \quad dla \quad t = 1, 2, \dots, k$$
$$\hat{y}_{t(2)} = a_{0(2)} + a_{1(2)} * t \quad dla \quad t = 2, 3, \dots, k+1$$
……………………………………………………..……..
$$\hat{y}_{t(n-k+1)} = a_{0(n-k+1)} + a_{1(n-k+1)} * t \quad dla \quad t = n-k+1, n-k+2, \dots, n.$$

Kolejna publikacja Sebastiana Gnata z Uniwersytetu Szczecińskiego to opracowanie „*Prognozowanie dochodów ze sprzedaży tygodników lokalnych - wybrane podejście*". Na stronie 1/16 podkreślono znaczenie prognozowania oraz podano kilka określeń pojęcia prognozowania według różnych autorów. W tekście znajduje się też nawiązanie do trendu pełzającego tworzenia równań odcinkowych i ocena parametrów *j*-tego równania odcinkowego.

20. Informacja o programie *WinQSB*

20.1. Tekst wpisu

Wpis ten opublikowałem w blogu „*Zastosowanie komputera w procesach decyzyjnych*” w marcu 2012 roku (zob. rysunek 20.1).

Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Informacja o programie WinQSB	wornalkiewicz	Materiały pomocnicze	Modele Holta-Wintersa, Prognozowanie liniowe i całkowitoliczbowe, Prognozowanie liniowe i regresja, Sezonowość multiplikatywna	—	0	Opublikowano 03.03.2012

Źródło: Opracowanie własne w *Excelu* na podstawie omawianej publikacji.

Rys. 20.1. Informacja o występowaniu wpisu w ramach omawianego blogu

Statystykę tego wpisu według miesięcy i lat zaprezentowano na rysunku 20.2.

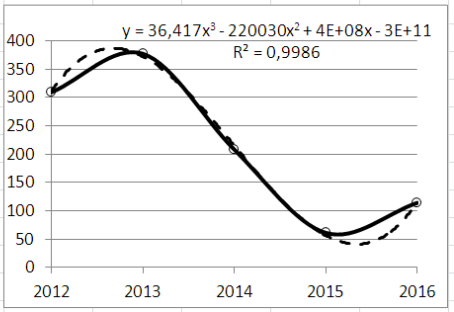
	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012			37	50	68	51	16	4	7	23	20	33	309
2013	83	15	36	23	21	17	2	13	9	66	69	23	377
2014	51	28	46	18	11	8	1	3	4	9	25	4	208
2015	24	4	5	8	1	6	0	2	2	2	4	3	61
2016	18	12	6	7	6	1	3	1	4	10			68

Źródło: Opracowanie własne w *Excelu* na podstawie omawianej publikacji.

Rys. 20.2. Statystyka miesięczna liczby wejść do wpisu „*Informacja o programie* WinQSB”

Rok 2012 jest niepełny, gdyż wpis wprowadzono w marcu, a trzeba dodać, że na zakończenie roku 2016 było 114 wejść internautów. Zobaczmy teraz jak prezentują się te dane statystyczne na wykresie wielomianowym sporządzonym w *Excelu.*

	2012	2013	2014	2015	2016
	309	377	208	61	114



Źródło: Opracowanie własne w *Excelu* na podstawie omawianej publikacji.

Rys. 20.3. Wykres wielomianowy liczby wejść do wpisu „*Informacja o programie* WinQSB”

Pakiet oprogramowania *WinQSB* (*Windows Quantitave System for Business*) służy do rozwiązywania zadań między innymi z programowania liniowego i całkowitoliczbowego[201]. Dostęp w Internecie uzyskać możemy w różny sposób, przykładowo wywołując hasło *WinQSB.exe* w *Google*. Menu modułu „*Programowanie liniowe i całkowitoliczbowe*" obejmuje funkcje:

- *File* (plik): nowy problem, otwarcie istniejącego problemu, zapisywanie problemu, wyjście z programu;
- *Edit* (edycja): wycięcie, kopiowanie, wklejenie;
- Format: format zapisu liczbowego, czcionka, wyrównanie do lewej, wyśrodkowanie, wyrównanie do prawej, wysokość wiersza, szerokość kolumny;
- *Solve and Analyze* (rozwiązanie i analiza): rozwiązanie, rozwiązanie krok po kroku, metoda graficzna, analiza parametryczna;
- *Results*: wyniki, grafy;
- *Utilities* (pomocnicze): kalkulator, zegar, pomoc.

Pakiet *WinQSB* dotyczy przede wszystkim zagadnień z badań operacyjnych. Natomiast z obszaru ekonometrii na uwagę zasługuje moduł "*Prognozowanie i liniowa regresja*" (FC) umożliwiający estymację modeli regresji i modeli trendu. Korzystając z modułu FC otrzymujemy raporty z oszacowaniem parametrów, wygasłymi oraz przewidywanymi na żądanie określonymi prognozami w zakresie modeli:

- regresji (jedno i wielowymiarowej), autoregresji;
- średnich (prostych i ruchomych);
- ważonych średnich ruchomych, podwójnych średnich ruchomych;
- prostego oraz adaptacyjnego wyrównywania wykładniczego;
- model Browna (podwójnego wyrównywania wykładniczego).

Ponadto możemy określić prognozy poprzez wyrównywanie trendu z zastosowaniem modeli Holta-Wintersa. Jak już nadmieniłem moduł FC umożliwia nam określenie parametrów wyrównywania lub doboru najlepszej ich wartości z punktu minimalizacji błędu RMSE wygasłych już prognoz - opcja - *Search the best*. Właśnie wymieniona opcja jest szczególnie przydatna, gdy określamy programy dla modelu Holta-Wintersa z sezonowością multiplikatywną lub addytywną. Musimy wtedy rozpatrzyć modele z wieloma kombinacjami parametrów wyrównywania: zmiennej α, trendu β, sezonowości γ, z krokiem co 0,01 w przedziale 0-1.

Opis pakietu programowego wspomagającego podejmowanie decyzji *WinQSB* znajdujemy także w Aneksie tomu III publikacji pod redakcją Lipiec-Zajchowskiej i obejmuje on między innymi podstawowe moduły[202]:

- prognozowanie liniowe i całkowitoliczbowe,
- programowanie dynamiczne,
- problem wyboru najkrótszej trasy,
- problem załadunku,
- sterowanie produkcją i zapasami,
- drzewo decyzyjne,
- analiza przedsięwzięć (metoda ścieżki krytycznej CPM, metoda PERT).

Zaprezentowane przykłady dotyczą głównie obszaru badań operacyjnych, chociaż omówiono także zastosowanie modułu prognozowanie. Jak już wiemy moduł ten stanowi narzędzie do formułowania na podstawie szeregu czasowego zmiennej objaśnianej parametrów różnych modeli, a także wygasłych i przewidywanych prognoz.

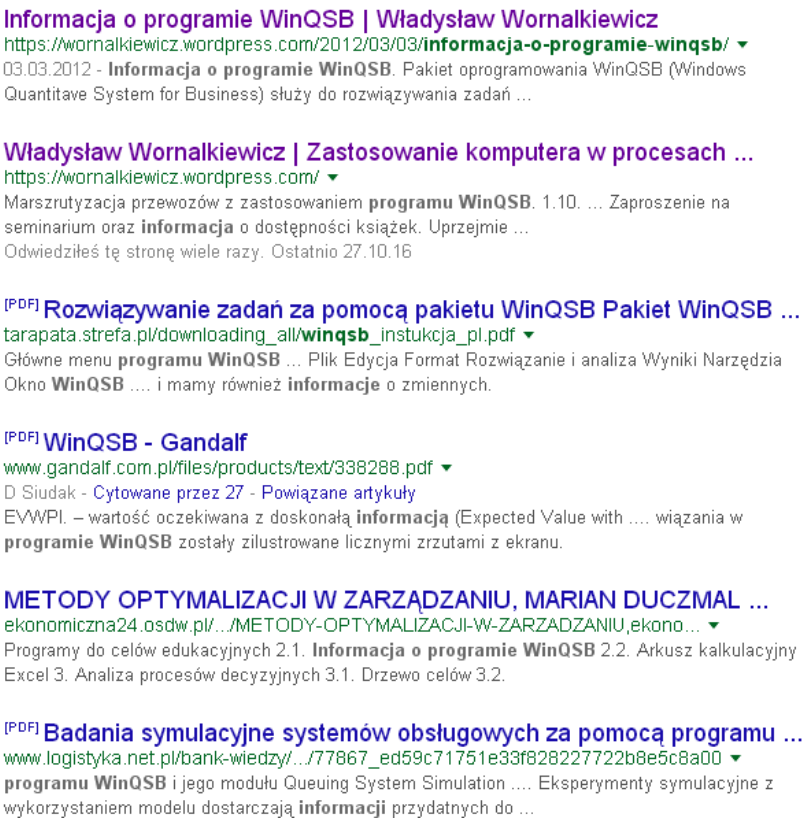
---

[201] Niniejszy opis bazuje na tekście zamieszczonym w książce „*Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*" o której informacja jest w zakładce *Publikacje* mego blogu *Zastosowanie komputera w procesach decyzyjnych*.

[202] Lipiec-Zajchowska (red.): *Wspomaganie procesów decyzyjnych*. Tom III. *Badania operacyjne*. Wydawnictwo C.H. Beck Warszawa 2003.

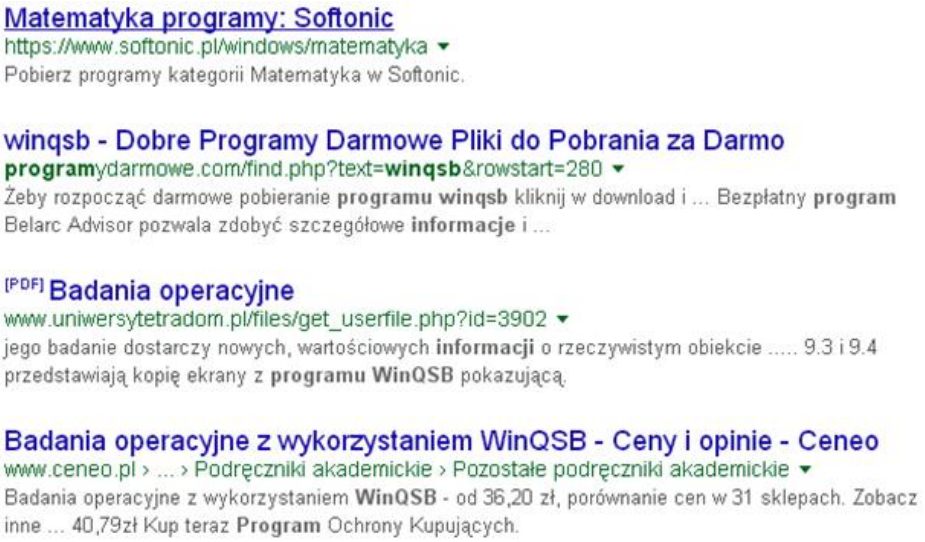
20.2. Problematyka frazy *Informacja o programie* WinQSB

Przy podaniu zapytania „*Informacja o programie WinQSB*” wyszukiwarka *Google* udostępniła 582 pozycje listy rezultatu w czasie 0,52 sekundy (zob. rysunek 20.4).



Źródło: Opracowanie własne w *Excelu* na podstawie omawianej publikacji.

Rys. 20.4. Fragment pierwszy pierwszej strony listy rezultatu wyszukiwania frazy *Informacja o programie* WinQSB



Źródło: Opracowanie własne w Excelu na podstawie omawianej publikacji.

Rys. 20.5. Fragment drugi pierwszej strony listy rezultatu wyszukiwania frazy *Informacja o programie WinQSB*

Zadowolenie budzi umieszczenie na pierwszym i drugim miejscu listy adnotacji o umiejscowieniu szukanego tematu w ramach blogu autora niniejszej pracy. Występują na pierwszej stronie cztery publikacje szersze o formacie PDF. Jako wyszukiwania podobne *Google* proponuje frazy:

- winqsb download,
- winqsb instrukcja,
- winqsb windows 10,
- winqsb windows 7.

Na stronie 1/34 publikacji „*Rozwiązywanie zadań za pomocą pakietu WinQSB*" autorstwa Iwony Staniec podano publiczną domenę, z której w czasie korzystania z Internetu, można było pobrać opis tego bardzo przydatnego pakietu programowego o wielu modułach, o którym już dość często wspominałem. Interesujące jest przy tym przetłumaczenie występujących w *WinQSB* nazw modułów (zob. rysunek 20.6).

Źródło: Opracowanie własne w *Google* na podstawie omawianej publikacji.

Rys. 20.6. Moduły pakietu *WinQSB*

Strona 6/20 publikacji „*WinQSB - Gandalf*" dotyczy książki Dariusza Siutaka „*Badania operacyjne z wykorzystaniem WinQSB*". Szczegółowe poznanie funkcjonalności poszczególnych modułów na pewno przyda się z zarówno w pracy dydaktycznej jak i w zastosowaniach praktycznych. Otwórzmy teraz wpis „*Badania symulacyjne systemów obsługiwanych za pomocą programu WinQSB i jego modułu Queuing System Simulation*". Na stronie pierwszej tego opracowania autorstwa Marka Glinku w streszczeniu czytamy:

„W artykule zarysowano możliwości wykorzystania programu WinQSB QSS do modelowania sieci obsługowych. Celem takiego modelowania może być usprawnienie obsługi dokonywanej w ramach sieci. Sieć składa się z węzłów powiązanych ze sobą. Wyniki symulacji mogą być podstawą do działań usprawniających prace poszczególnych węzłów obsługiwanych jak i całej sieci. Technika symulacyjna w tych i innych przypadkach jest bardzo wygodna gdyż nie ingeruje bezpośrednio w proces obsługi logistycznej jak i w same obiekty. Zwrócono uwagę na jakość modelu symulacyjnego i konieczność jego weryfikowania celem potwierdzenia zgodności".

Kolejna publikacja „*Badania operacyjne*" Marka Glinki z Wydział Transportu i Elektrotechniki Uniwersytetu Technologiczno-Humanistycznego w Radomiu ma 36 stron i stanowi fragmenty wykładu obejmującego tematy:

- geneza badań operacyjnych,
- zakres stosowania badan operacyjnych,

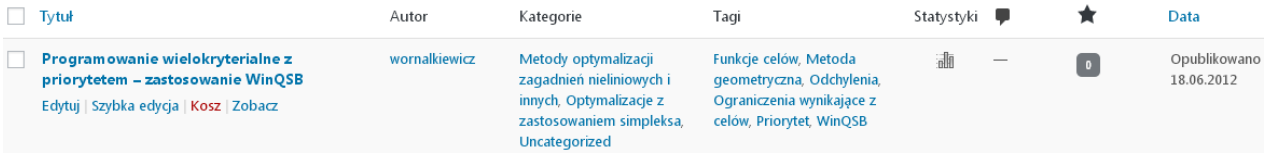
- metodyka badań operacyjnych,

- model jest narzędziem badawczym,
- struktura modeli normatywnych zagadnień decyzyjnych,
- postacie modeli zadań programowania liniowego,
- klasyfikacja modeli w badaniach operacyjnych,
- uwagi ogólne o rozwiązywaniu modeli matematycznych zagadnień decyzyjnych,
- przykład modelowania zagadnienia przepływu prądu,
- geometryczne przedstawienie modeli rozwiązań zadań programowania liniowego,
- simpleks – uniwersalna metoda rozwiazywania modeli liniowych,
- modelowanie zagadnienia transportowego.

21. Programowanie wielokryterialne z priorytetem - zastosowanie *WinQSB*

21.1. Tekst wpisu

Podjęliśmy już wcześniej temat programowania dwukryterialnego, a tu tematyka programowania *wielokryterialnego z priorytetem i to z zastosowaniem odpowiedniego modułu pakietu* WinQSB. Wpis o podanym tytule został opublikowany na blogu „*Zastosowanie komputera w procesach decyzyjnych*" w marcu 2013 roku (zob. rysunek 21.1).

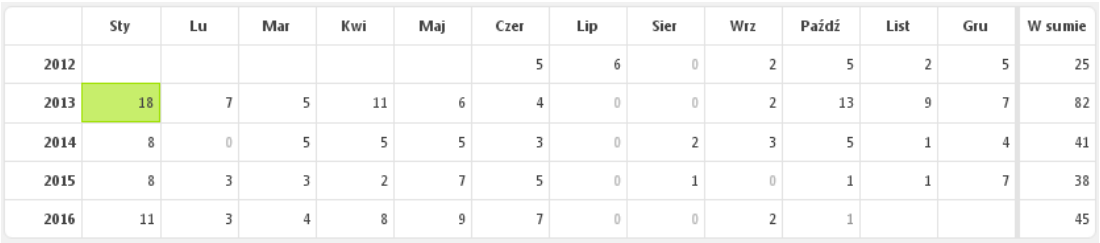


Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Programowanie wielokryterialne z priorytetem – zastosowanie WinQSB Edytuj | Szybka edycja | Kosz | Zobacz	wornalkiewicz	Metody optymalizacji zagadnień nieliniowych i innych, Optymalizacje z zastosowaniem simpleksa, Uncategorized	Funkcje celów, Metoda geometryczna, Odchylenia, Ograniczenia wynikające z celów, Priorytet, WinQSB	—	0	Opublikowano 18.06.2012

Źródło: Opracowanie własne w *WordPress*.

Rys. 20.1. Komunikat o występowaniu wpisu „*Programowanie wielkokryterialne z priorytetem - zastosowanie WinQSB*"

Statystykę według miesięcy i lat korzystających z blogu przedstawiono na rysunku 21.2, przy czym na zakończenie roku 2016 było 46 wejść internautów (zob. rysunek 21.2). Jest jeszcze stosunkowo małe zainteresowanie tym moim zdaniem merytorycznie ciekawym tekstem, a szkoda, bo programowanie wielokryterialne z priorytetem, przy zastosowaniu programu komputerowego, powinno możliwie szybko torować sobie drogę w zastosowaniach praktycznych.

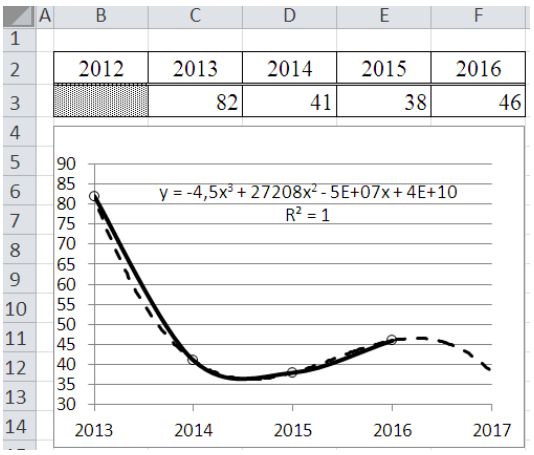


	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012						5	6	0	2	5	2	5	25
2013	18	7	5	11	6	4	0	0	2	13	9	7	82
2014	8	0	5	5	5	3	0	2	3	5	1	4	41
2015	8	3	3	2	7	5	0	1	0	1	1	7	38
2016	11	3	4	8	9	7	0	0	2	1			45

Źródło: Opracowanie własne w *WordPress*.

Rys. 20.2. Statystyka miesięczna wejść do wpisu „*Programowanie wielkokryterialne z priorytetem - zastosowanie WinQSB*"

Proponuję teraz spojrzeć na wykres trendu trendu wielomianowego 3-go stopnia sporządzonego bez niepełnego roku 2012 lecz tylko na podstawie lat 2013-2016 i prognozą na rok 2017 (zob. rysunek 21.3). Widzimy tu bardzo dobre dopasowanie danych rzeczywistych do linii trendu, bowiem współczynnik determinacji R^2 został określny bliski 1.



Źródło: Opracowanie własne w *Excelu*.

Rys. 21.3. Trend wielomianowy wpisu „*Programowanie wielkokryterialne z priorytetem - zastosowanie WinQSB*”

W opracowaniu tego skróconego wpisu internetowego do blogu zabazowano na rozdziale 19 „*Zastosowanie algorytmu simpleks w WinQSB do programowania wielokryterialnego z priorytetem*”[203]. W wymienionym rozdziale zaprezentowano rozwiązanie geometryczne zadania z trzema funkcjami celowymi o nadanych priorytetach, przy czym wystąpiło tam wspomaganie manualne tabelami odchyleń, które możemy wykonać w *Excelu*. Jednak w rozwiązaniu analitycznym powierzymy funkcje obliczania odchyleń i decydowania o wyborze rozwiązania programowi *WinQSB*. Kontynuując nasz przykład (z wpisu dotyczącego rozwiązania sposobem geometrycznym) o którym była mowa w podrozdziale 1.12 oprócz ograniczeń podstawowych - prymalnych wystąpią również dodatkowo 3 ograniczenia wynikające z funkcji celów decydenta traktowanych jako równości. Natomiast funkcje celów zadań decyzyjnych zmierzają do minimum odchyleń zarówno dodatnich $y_i(+)$ jak i ujemnych $y_i(-)$[204].

Rozpoczynamy obliczenia dla celów według ich kolejności priorytetów (1-3). Postępowanie możemy traktować jako "*rekurencyjne*", gdyż wartość istotnej zmiennej odchyleń jest przenoszona do kolejnego zadania decyzyjnego o niższym priorytecie. Model matematyczny zadania decyzyjnego celu trzeciego o priorytecie 1 jest następujący:

Funkcja celu: FC: $5y_3(+) \rightarrow \min$

Ograniczenia podstawowe:

C_1: $6x_1 + 12x_2 \leq 1\,800$

C_2: $5x_1 + 5x_2 \leq 900$

C_3: $5x_1 + 2x_2 \geq 120$

Ograniczenia wynikające z równań funkcji celów:

C_4: $400x_1 + 600x_2 - y_1(+) + y_1(-) = 90\,000$

C_5: $5x_1 + 10x_2 - y_2(+) + y_2(-) = 1\,600$

C_6: $300x_1 + 400x_2 - y_3(+) + y_3(-) = 70\,000$

Warunki brzegowe: $x_1, x_2, y_1(+), y_1(-), y_2(+), y_2(-), y_3(+), y_3(-) \geq 0$.

Sformułowanie tego zadania decyzyjnego w module LP-ILP programu *WinQSB* zaprezentowano na rysunku 21.4.

[203] Duczmal M., *Metody optymalizacji w zarządzaniu*, Wydawnictwo Instytut Śląski, Opole 2014.

[204] W opracowaniu oparto się na procedurze postępowania zamieszczonej na stronie internetowej: *dydaktyka.polsl/kwmomkm/wielokryterialne.pdf*.

Variable -->	X1	X2	y1(+)	y1(-)	y2(+)	y2(-)	y3(+)	y3(-)	Direction	R. H. S.
Minimize							5			
C1	6	12							<=	1800
C2	5	5							<=	900
C3	5	2							>=	120
C4	400	600	-1	1					=	90000
C5	5	10			-1	1			=	1600
C6	300	400					-1	1	=	70000
LowerBound	0	0	0	0	0	0	0	0		
UpperBound	M	M	M	M	M	M	M	M		
VariableType	Continuous	Continuous	Continuous	Continuous	Continuous	Continuous	Continuous	Continuous		

Źródło: Opracowanie własne w *WinQSB*.

Rys. 21.4. Sformułowanie zadania decyzyjnego w module LP-ILP

Natomiast rozwiązanie zadania decyzyjnego pokazano na rysunku 21.5, którego wyniki są następujące:

$x_1 = 90$ $y_1(+) = 0$ $y_2(+) = 0$ $y_3(+) = 0$
$x_2 = 90$ $y_1(-) = 0$ $y_2(-) = 250$ $y_3(-) = 7\,000$
FC: $5y_3(+) = 5 \cdot 0 = 0$.

Program *WinQSB* sygnalizuje nam także istnienie rozwiązania alternatywnego, a gdy z niego skorzystamy to otrzymamy wyniki:

$x_1 = 60$ $y_1(+) = 6\,000$ $y_2(+) = 0$ $y_3(+) = 0$
$x_2 = 120$ $y_1(-) = 0$ $y_2(-) = 100$ $y_3(-) = 4\,000$
Fc: $5y_3(+) = 5 \cdot 0 = 0$.

	14:31:30		Sunday	January	08	2017		
	Decision Variable	Solution Value	Unit Cost or Profit c(j)	Total Contribution	Reduced Cost	Basis Status	Allowable Min. c(j)	Allowable Max. c(j)
1	X1	90	0	0	0	basic	0	0
2	X2	90	0	0	0	basic	0	0
3	y1(+)	0	0	0	0	at bound	0	M
4	y1(-)	0	0	0	0	at bound	0	M
5	y2(+)	0	0	0	0	at bound	0	M
6	y2(-)	250	0	0	0	basic	0	0
7	y3(+)	0	5	0	5	at bound	0	M
8	y3(-)	7000	0	0	0	basic	0	0
	Objective	Function	(Min.) =	0	(Note:	Alternate	Solution	Exists!!)
	Constraint	Left Hand Side	Direction	Right Hand Side	Slack or Surplus	Shadow Price	Allowable Min. RHS	Allowable Max. RHS
1	C1	1620	<=	1800	180	0	1620	M
2	C2	900	<=	900	0	0	750	1125
3	C3	630	>=	120	510	0	-M	630
4	C4	90000	=	90000	0	0	72000	96000
5	C5	1600	=	1600	0	0	1350	M
6	C6	70000	=	70000	0	0	63000	M

Źródło: Opracowanie własne w WinQSB.

Rys. 21.5. Rozwiązanie zadania decyzyjnego

Ze względu na obszerność i dużą złożoność dalszej procedury obliczeń odsyłam zainteresowanego Czytelnika do wspomnianego już rozdziału 19. w książce „*Metody optymalizacji w zarzadzaniu*”. W rozdziale tym, szczegółowo podano kolejne kroki postępowania, w sytuacji rozwiazywania następujących decyzyjnych: zadanie trzecie o priorytecie 1, zadanie drugie o priorytecie 3, zadanie pierwsze o priorytecie 2.

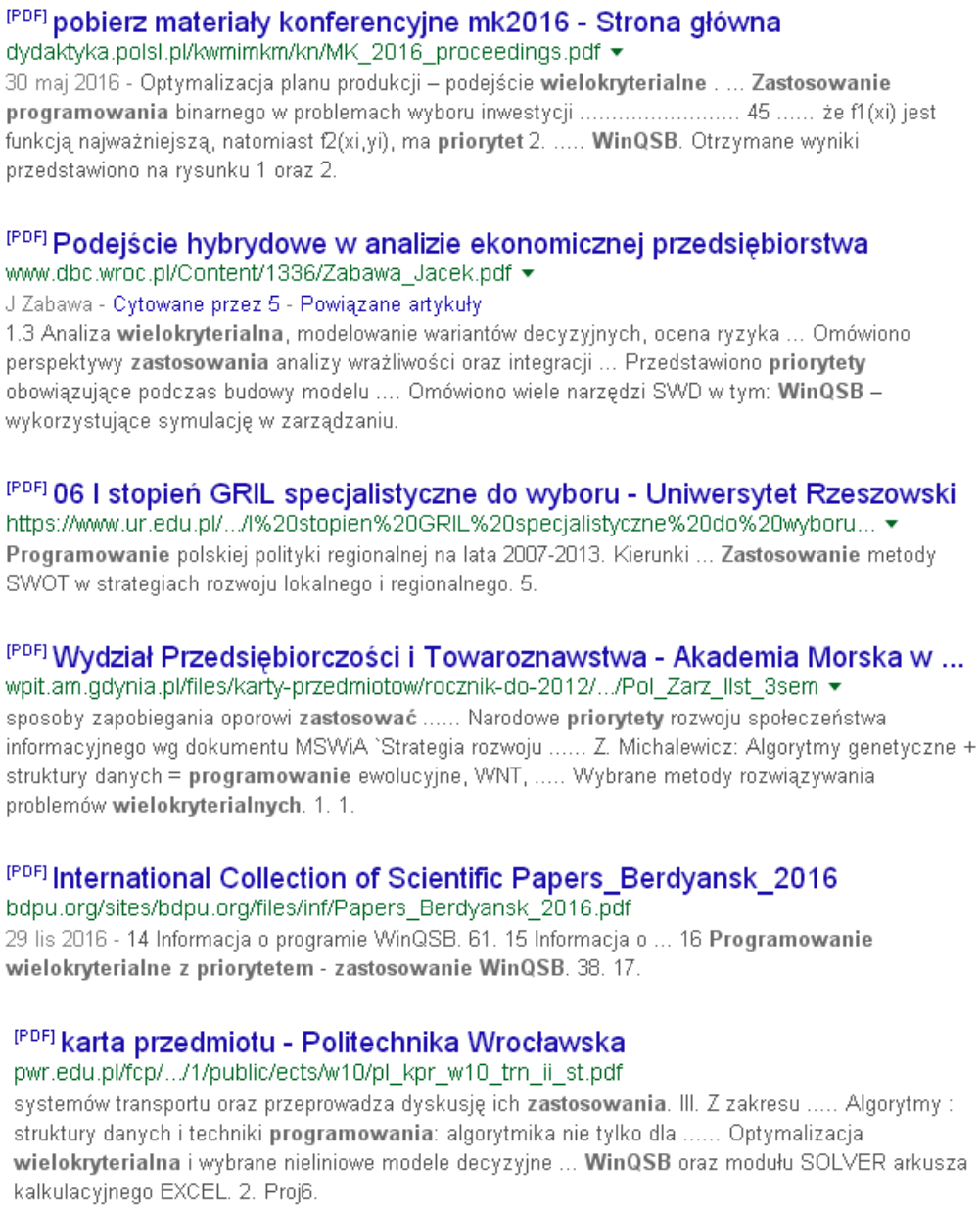
21.2. Problematyka frazy *Programowanie wielokryterialne - zastosowanie WinQSB*

Po podaniu tej tak obszernie zdefiniowanej frazy w cudzysłowie uzyskano 4 wyniki w czasie 0,44 sekundy i są to tylko wpisy autora niniejszej pracy na blogu „*Zastosowanie komputera w procesach decyzyjnych*” (zob. rysunek 21.6).

Źródło: Opracowanie własne w *Google*.

Rys. 21.6. Raport na zapytanie „*Programowanie wielokryterialne z priorytetem - zastosowanie WinQSB*”

Na zakończenie otrzymanej listy rezultatu *Googl*e proponuje nam wyszukiwanie z uwzględnieniem pominiętych wyników, które ta wyszukiwarka uznała za mniej trafne do podanej frazy. Skorzystajmy więc z tej propozycji, otrzymaliśmy 11 wyników i to tylko stanowiące wpisy do omawianego blogu. W tej sytuacji dajemy zapytanie w formie wcześniej wymienionej frazy bez cudzysłowie. Teraz otrzymaliśmy 25 wyników, w tym 6 o formacie PDF (zob. rysunek 21.7).



Źródło: Opracowanie własne w *Google*.

Rys. 21.7. Wybrane fragmenty komunikatów o wpisach formatu PDF w zasobach *Google* po zapytaniu *Programowanie wielokryterialne z priorytetem - zastosowanie WinQSB*

Pozostaje nam teraz bardziej szczegółowe zainteresowanie się publikacjami wymienionymi na rysunki 21.7. Pierwsza obszerna monografia (106 stron) zawiera 24 artykuły opublikowane z okazji Studenckiej Konferencji Naukowej „*Metody Komputerowe*", która odbyła się 30 maja 2016 roku w Centrum Edukacyjno-Kongresowym Politechniki Śląskiej w Gliwicach. Artykuły dotyczą zastosowania metod komputerowych w zakresie:

- wspomagania komputerowego prac inżynierskich,
- wytrzymałości materiałów,
- mikromechaniki i termodynamiki,
- badań doświadczalnych,
- badań operacyjnych.

Druga publikacja Jacka Zabawy z Wydziału Informatyki i Zarządzania Politechniki Wrocławskiej zawiera 351 stron. Autor w swojej pracy doktorskiej pt. „*Podejście hybrydowe w analizie ekonomicznej przedsiębiorstwa*" zajmuje się zagadnieniami:

- analiza decyzyjna;
- modelowanie i symulacja;
- zarządzanie wiedzą, sztuczna inteligencja, systemy wspomagania decyzji;
- systemy hybrydowe;
- założenia hybrydowego SWD Ekanwin;
- projektowanie interfejsu systemu *Ekanwin* dla potrzeb wizualizacji i wnioskowania;
- koncepcja, projekt i zastosowanie języka sterowania eksperymentem LEKS;
- zastosowanie jezyka sterowania eksperymentem LEKS; analiza czynnikowa.

We *Wstępie* opracowania autor informuje, że przedstawił wyniki badań nad uwarunkowaniami integracji technik inteligentnych do potrzeb analizy ekonomicznej, w hybrydowych systemach podejmowania decyzji (SWD). W analizie ekonomicznej stosowane są metody ilościowe oraz jakościowe. Jacek Zabawa sformułował tezę, że połączenie różnych metod oraz technik zwiększy możliwości poznawcze i wartość procedur podejmowania decyzji. Autor jako podstawowe moduły informatycznego systemu hybrydowego uznał:

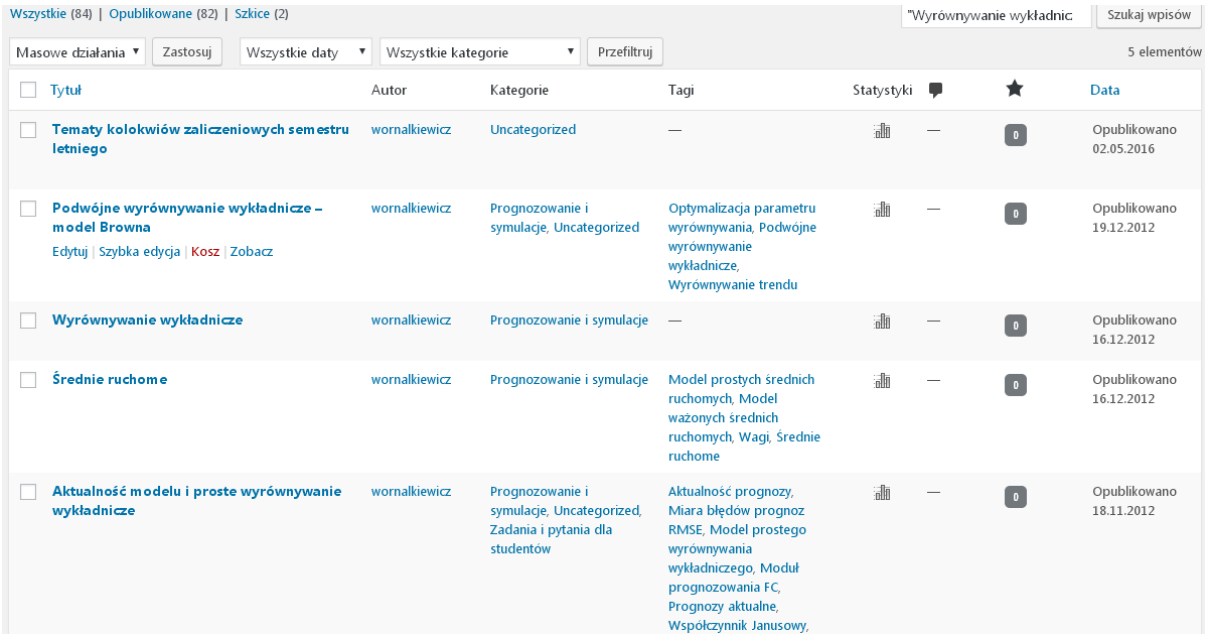
- model symulacyjnego obiektu,
- sterowanie eksperymentem,
- interfejs komunikacji między aparatami pojeciowymi,
- jezyk planowania eksperymentu.

W wyniku prac nad hybrydą powstała platforma informatyczna z hybrydowym systemem wspomagania decyzji w zakresie analizy finansowej przedsiębiorstwa. Moduł języka planowania eksperymentu pozwala na prowadzenie analiz czynnikowych poszukujących rozwiązań optymalnych. Trzecia publikacja zawiera 51 stron. Są to sylabusy różnych przedmiotów, a w tym *Komputerowe wspomaganie podejmowania decyzji*. Opracowanie czwarte wymienione we wcześniejszej liście wygenerowanej przez *Google* ma 248 stron i obejmuje karty opisu przedmiotów wykładanych na Akademii Morskiej w Gdyni, a wśród nich *Analiza finansowa*. Kolejna piąta praca to monografia obejmująca 420 stron, a w niej występuje artykuł autora niniejszej pracy. Publikacja 6. obejmuje 317 stron i są to również karty przedmiotów na kierunku *Transport* Politechniki Wrocławskiej np. *Ekonomia*.

22. Wyrównywanie wykładnicze

22.1. Tekst wpisu

Odniesienia do frazy „*Wyrównywanie wykładnicze*” występuje w pięciu wpisach wprowadzonych do blogu „*Zastosowanie komputera w procesach decyzyjnych*”, tj. czterech w roku 3012 i jednego w 2016 (zob. rysunek 22.1). W maju 2016 r. wymieniona fraza znalazła się przy okazji podania prze zemnie tematów kolokwiów zaliczeniowych semestru letniego.



Wszystkie (84) | Opublikowane (82) | Szkice (2) — "Wyrównywanie wykładnic — Szukaj wpisów

Masowe działania — Zastosuj — Wszystkie daty — Wszystkie kategorie — Przefiltruj — 5 elementów

Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Tematy kolokwiów zaliczeniowych semestru letniego	wornalkiewicz	Uncategorized	—	—	0	Opublikowano 02.05.2016
Podwójne wyrównywanie wykładnicze – model Browna Edytuj | Szybka edycja | Kosz | Zobacz	wornalkiewicz	Prognozowanie i symulacje, Uncategorized	Optymalizacja parametru wyrównywania, Podwójne wyrównywanie wykładnicze, Wyrównywanie trendu	—	0	Opublikowano 19.12.2012
Wyrównywanie wykładnicze	wornalkiewicz	Prognozowanie i symulacje	—	—	0	Opublikowano 16.12.2012
Średnie ruchome	wornalkiewicz	Prognozowanie i symulacje	Model prostych średnich ruchomych, Model ważonych średnich ruchomych, Wagi, Średnie ruchome	—	0	Opublikowano 16.12.2012
Aktualność modelu i proste wyrównywanie wykładnicze	wornalkiewicz	Prognozowanie i symulacje, Uncategorized, Zadania i pytania dla studentów	Aktualność prognozy, Miara błędów prognoz RMSE, Model prostego wyrównywania wykładniczego, Moduł prognozowania FC, Prognozy aktualne, Współczynnik Janusowy,	—	0	Opublikowano 18.11.2012

Źródło: Opracowanie własne w *Wordpress*.

Rys. 22.1. Wpisy na omawianym blogu z występowaniem frazy „*Wyrównywanie wykładnicze*”

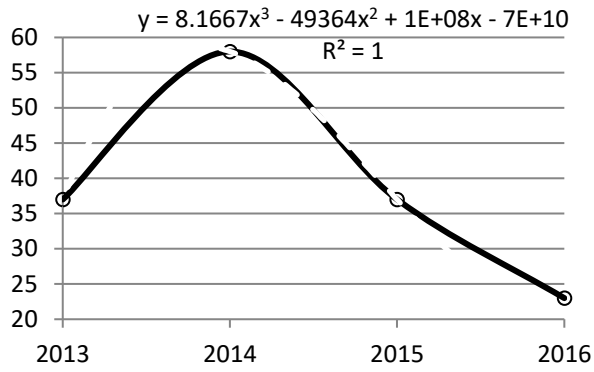
Wśród tytułów wpisów wymienionych na rysunku 22.1 znajduje się także wpis zgodny z podaną frazą. Statystykę tego wpisu według miesięcy i lat zaprezentowano na rysunku 22.2. Na zakończenie roku 2016 były 23 wejścia internautów.

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2013	2	1	2	4	10	6	0	3	3	3	1	2	37
2014	11	4	7	2	8	11	1	0	6	2	3	3	58
2015	16	4	5	4	2	0	1	1	0	2	1	1	37
2016	5	0	1	5	4	1	1	1	0	2			20

Źródło: Statystyka wpisu w *Wordpress*.

Rys. 22.2. Liczba wejść do wpisu „*Wyrównywanie wykładnicze*”

W latach 2013-2016 trend wielomianowy trzeciego stopnia daje dobre dopasowanie do danych rzeczywistych (zob. rysunek 22.3).



Źródło: Opracowanie własne w *Excelu*.

Rys. 22.3. Trend wielomianowy liczby wejść do wpisu „*Wyrównywanie wykładnicze*"

Proste wyrównywanie wykładnicze ma konstrukcję średniej ważonej z parametrem α. Prognoza na kolejny okres ustalana jest na podstawie modelu:

$$F_{n+1} = (\alpha Y_{n-1}) + (1 - \alpha) F_{n-1}.$$

Jako wartość początkową prognozy przyjmujemy wartość obserwacji dla okresu pierwszego, a α inicjujemy np. jako 0,2. Poprzez kolejne zmienianie α (dogodnie jest to wykonywać w arkuszu kalkulacyjnym *Excela*), doprowadzamy do minimum RMSE. Przykładowo, gdy α = 0,35 to RMSE = 1,99, czyli błąd wygasłych prognoz jest najmniejszy z dotychczasowych modeli.

W modelu adaptacyjnym wyrównywania wykładniczego występują dwa parametry α i β. Parametr α w kolejnych okresach (oprócz pierwszego) jest zastępowalny sygnałem adaptacyjnym TSTn. Parametr ten stanowi moduł z ilorazu (SAD_n/MAD_n). Konstrukcja SAD_n i MAD_n bazuje na parametrze β i stanowi co do wag analogię do modelu prostego wyrównywania wykładniczego. Jako wartość początkową dla okresu pierwszego w tym modelu przyjmujemy $TST_1 = \alpha = 0{,}2$, $\beta = 0{,}2$, $SAD_1 = 0$, $MAD_1 = Y_2 - Y_1$, czyli różnica wartości obserwacji okresu drugiego i pierwszego. Prognoza F_{n+1}, czyli 13. obliczana jest według wzoru:

$$F_{13} = F_{12} + TST_{12} (Y_{12} - F_{12}),$$

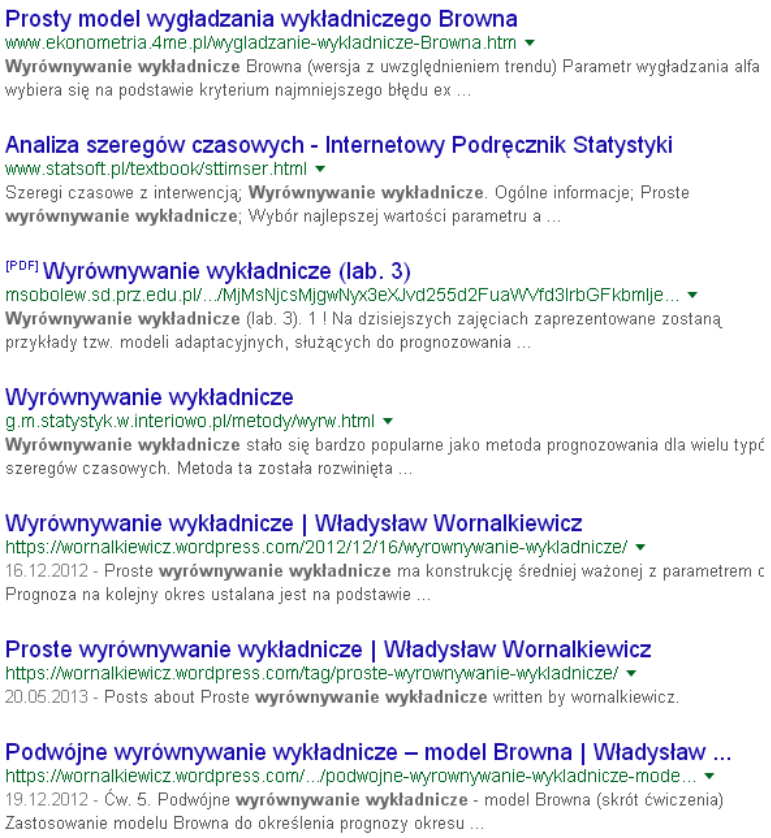
gdzie: Y_{12} - wartość obserwacji dla okresu 12.

Obliczone RMSE dla okresów (3-12) wynosi 2,62. Wzory oraz dokładniejsze obliczenia oraz wykresy punktowe obserwacji i wartości teoretycznych znajdzie Czytelnik tego blogu w podrozdziale "*Modele oparte na szeregach czasowych*" książki "*Wstęp do ekonometrii i badań operacyjnych ...*: wydanej w 2010 roku przez Instytut Śląski w Opolu. W ustalaniu najlepszych wartości parametrów modeli wyrównywania wykładniczego wygodnie jest stosować moduł "*Forecasting*" stanowiący część programu *WinQSB* bezpłatnie dostępnego w Internecie. Po wyborze określonej metody jest tam do wyboru opcja "*The best*", która umożliwia komputerowe dobranie wartości danego parametru ze względu na przyjętą miarę wygasłych prognoz np. MSE. Stosowany bowiem wcześniej RMSE w tym module stanowi pierwiastek z miary wygasłych prognoz zwanej MSE. Na kanwie dotychczasowych rozważań zrodziło się zadanie domowe o następującej treści: Bazując na artykule "*Dobór metod i parametrów w budowaniu modeli ekonometrycznych*" autora tego blogu zamieszczonym w monografii "*Ekonomiczno-społeczne problemy współczesnego zarządzania i komunikacji"* wydanej przez Instytut Śląski w Opolu w roku 2008:

- opracuj referat dotyczący metod i sposobu dobierania najlepszych wartości parametrów,
- zainstaluj program WinQSB i rozpatrz poszczególne metody prognozowania,
- przetestuj możliwość doboru parametru α na przykładzie danych z wykładu i przyjętego modelu prostego wyrównywania wykładniczego.

22.2. Odniesienia frazy we wpisach wyszukiwarki *Google*

Po wprowadzeniu frazy w cudzysłowie wyszukiwarka wygenerowała raport zawierający 492 wyniki w czasie 0,42 s.. Tak jak poprzednio interesują nas jednak wpisy występujące na stronie pierwszej i to szersze w formacie PDF (zob. rysunek 22.4).



Źródło: Opracowanie własne w *Google*.

Rys. 22.4. Zasygnalizowanie wpisów po zapytaniu „*Wyrównywanie wykładnicze*” (część 1/2)

Źródło: Opracowanie własne w *Google*.

Rys. 22.5. Zasygnalizowanie wpisów po zapytaniu „*Wyrównywanie wykładnicze*” (część 2/2)

Wyszukiwarka proponuje nam jeszcze wyszukiwanie podobne według fraz pisanych małymi literami:

- wygładzanie wykładnicze,
- wygładzanie wykładnicze przykład,
- wygładzanie wykładnicze excel,

- model holta,
- model browna logistyka,
- model browna exel,
- model multiplikatywny,
- model holta przykład.

Występuje tylko jedna 2-wu stronnicowa publikacja o formacie PDF, a mianowicie „*Wyrównywanie wykładnicze* (*lab*. 3). Autor prezentuje przykłady modeli adaptacyjnych do prognozowania wielkości przyszłych obserwacji w szeregach czasowych, a w szczególności modelom wyrównywania wykładniczego. Nawiązuje do programu STATISTICA w którym zaimplementowano 12 rodzajów modeli wyrównywania wykładniczego, będących kombinacjami modeli bez trendu, z trendem oraz z wahaniami sezonowymi lub bez nich. Omawia 6 miar błędów prognoz obliczanych w wymienionym programie względem wartości empirycznych.

23. Aktualność modelu i proste wyrównywanie wykładnicze

23.1. Tekst wpisu

W tym wpisie opublikowanym w listopadzie 2012 roku podano sposób sprawdzenia ważności formuły do określania prognoz krótkoterminowych w oparciu o zasadę wyrównywania wykładniczego szeregu czasowego obserwacji. Dostęp do podglądu oraz edycji tekstu wpisu „*Aktualność modelu i proste wyrównywanie wykładnicze*" w ramach omawianego blogu pokazano na rysunku 23.1.

Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Aktualność modelu i proste wyrównywanie wykładnicze	wornalkiewicz	Prognozowanie i symulacje, Uncategorized, Zadania i pytania dla studentów	Aktualność prognozy, Miara błędów prognoz RMSE, Model prostego wyrównywania wykładniczego, Moduł prognozowania FC, Prognozy aktualne, Współczynnik Janusowy, Wygasłe prognozy, Średni błąd kwadratowy wygasłych prognoz	—	0	Opublikowano 18.11.2012

Źródło: Opracowanie własne w *Wordpress*.

Rys. 23.1. Informacja o występowaniu wpisu w blogu „*Zastosowanie komputera w procesach decyzyjnych*"

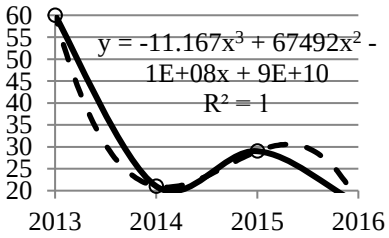
W okresie lat 2012-2016 statystyka miesięczna wejść do omawianego wpisu wykazuje tendencję malejąca (zob. rysunek 23.2).

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2012											3	6	9
2013	10	20	7	9	3	6	0	1	3	0	1	0	60
2014	3	4	1	1	0	1	1	0	3	2	3	2	21
2015	9	2	2	1	3	4	2	1	1	1	1	2	29
2016	4	0	1	4	3	1	0	0	1	2			16

Źródło: Opracowanie własne w *Wordpress*.

Rys. 23.2. Statystyka miesięczna wejść do wpisu „*Aktualność modelu i proste wyrównywanie wykładnicze*"

Jeszcze mało wyrobione jest wśród decydentów przekonanie, ze opracowana funkcja prognozowania bazująca na stałym szeregu czasowym z przeszłości jest tylko czasowo aktualna. Wpływ bowiem wygasających wartości obserwacji „*starzeje się*". Trzeba co jakiś czas, a najlepiej po każdym minionym okresie wyznaczać komputerowo parametry nowego modelu i później stosować go do bieżącego prognozowania wartości zjawiska. Sporządźmy wykres wielomianowy 3-go stopnia wejść w latach 2012- 2016, przy czym na koniec roku 2016 liczba wejść wynosiła 17 (zob. rysunek 23.3). Widzimy, ze model dobrze dopasowuje się do danych rzeczywistych z obserwacji.



Źródło: Opracowanie własne w *Excelu*.

Rys. 23.3. Wykres wielomianowy wejść do „*Aktualność modelu i proste*

wyrównywanie wykładnicze”

W ramach omawianego blogu, tekst wpisu stanowi skrót ćwiczenia 3 na kierunku *Zarządzanie* w ramach przedmiotu „*Prognozowanie i symulacje*”. Aktualność modelu ekonometrycznego do prognozowania sprawdzamy współczynnikiem Janusowym J^2 o postaci:

$$J^2 = \frac{\frac{1}{T-n}\sum_{t=n+1}^{T}(y_t - y_{tP})^2}{\frac{1}{n}\sum_{t=1}^{n}(y_t - \hat{y}_t)^2}$$

gdzie: T - okres prognozy, n - liczba okresów obserwacji, y_t - obserwacja zmiennej w okresie t, y_{tP} - wartość prognozy określona na podstawie modelu w okresie prognozowanym t, $\hat{y}_t$ - wartość prognozy w okresie opracowanego modelu.

Współczynnik Janusowy stanowi zatem iloraz średniego kwadratu reszt (błędów prognoz) z okresów prognozowanych do średniego kwadratu reszt wygasłych już prognoz okresów na podstawie których sporządzono model trendu liniowego klasyczną metodą najmniejszych kwadratów.

Model regresji liniowej możemy określić funkcją *Excela* np. na podstawie danych umownych podanych w kolumnach A i B tabeli 23.1.

D13 =D12/A11

	A	B	C	D
1	t	y	y^	$(y-y^)^2$
2	1	280	288,09	65,45
3	2	320	310,18	96,43
4	3	340	332,27	59,75
5	4	360	354,36	31,81
6	5	370	376,45	41,60
7	6	390	398,54	72,93
8	7	415	420,63	31,70
9	8	440	442,72	7,40
10	9	460	464,81	23,14
11	10	500	486,90	171,61
12			Suma:	601,82
13			Średnia:	60,18

Źródło: Opracowanie własne w *Excelu*.

Rys. 23.1. Dane umowne, modelowe oraz kwadraty błędów (reszt)

Przy wyznaczeniu parametrów modelu korzystamy z funkcji:

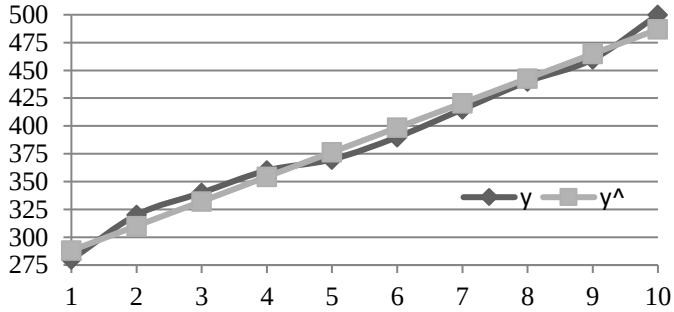
=REGLINP(B2:B11;A2:A11;PRAWDA;PRAWDA).

F	G	H
a_1	a_0	Cechy
22,09	266,00	parametry modelu liniowego
0,95	5,93	blad standardowy oszacowania parametrów
0,99	8,67	R^2, S_e
535,19	8,00	F, iss
40260,68	601,82	ESS, RSS

Określenie parametrów modelu umożliwiło obliczenie wygasłych już prognoz, oznaczonych w *Excelu* jako y^ , oraz kwadratów błędów dla okresów (1-10) - zob. rysunek 23.1. Suma kwadratów wygasłych prognoz RSS wynosi 601,82, a średni kwadrat błędów 10. obserwacji równa się 60,18.

W obliczeniu wartości w kolumnie C i D (zob. rysunek 23.1) skorzystano odpowiednio z formuł: =266+22,09*A2 oraz =(B2-C2)^2.

Zastosowano też formułę do obliczenia sumy i średniej: =SUMA(D2:D11), =D12/A11. Zobaczmy teraz jak na wspólnym wykresie układają się dane umowne i modelowe. Łatwo zauważyć dobre dopasowanie modelu do danych rzeczywistych (umownych).



Źródło: Opracowanie własne w *Excelu*.

Rys. 23.2. Wykres danych umownych i modelowych

Obliczymy teraz w sposób analogiczny prognozy, kwadraty błędów prognoz oraz kolejne średnie kwadraty błędów prognoz dla założonego dość długiego przedziału okresów prognozowanych *T* (11-20). Przyjmujemy jednak, że znamy już wartości empiryczne obserwacji po czasie *t* = 10 (zob. rysunek 23.3).

E3 =((SUMA($D$2:D3)/(A3-$F$2))/$F$4)

	A	B	C	D	E	F
1	T	y	$y^$	$(y-y^)^2$	J^2	n
2	11	507	508,99	3,96		**10**
3	12	533	531,08	3,69	0,06	$Średnia_{(1-10)}$
4	13	550	553,17	10,05	0,10	**60,18**
5	14	580	575,26	22,47	0,17	
6	15	600	597,35	7,02	0,16	
7	16	610	619,44	89,11	0,38	
8	17	640	641,53	2,34	0,33	
9	18	650	663,62	185,50	0,67	
10	19	670	685,71	246,80	1,05	
11	20	680	707,80	772,84	2,23	
12			Suma:	1343,79		
13			$Średnia_{(11-20)}$	134,38		

Źródło: Opracowanie własne w Excelu.

Rys. 23.3. Dane modelowe kwadraty błędów oraz współczynniki Janusowe z przedziału prognoz

Przyjrzyjmy się bliżej formule obliczania współczynnika Janusowego np. dla okresu prognozowanego 12: (=((SUMA($D$2:D3)/(A3-$F$2))/$F$4)). Rozpoczynamy obliczenia J^2 od okresu (*n* +2), przy czym *n* = 10. Współczynnik Janusowy powinien być mniejszy od 1. Dla prognozy 9, czyli okresu 19. przekracza już 1. Zatem zachodzi potrzeba ponownego określenia modelu ekonometrycznego liniowego na podstawie okresów (10-19):

=REGLINP(B2:B11;A2:A11;PRAWDA;PRAWDA).

a odpowiedz programu jest następująca:

a_1	a_0	Cechy
19,79	475,13	parametry modelu liniowego
0,62	3,85	błąd standardowy oszacowania parametrów
0,99	5,64	R^2, S_e
1016,08	8,00	F, iss
32323,50	254,50	ESS, RSS

Mamy zatem nowy model regresji liniowej: $\hat{y} = 475{,}13 + 19{,}79 \cdot t$. Możemy teraz obliczyć jeszcze wygasłe prognozy oraz kwadraty błędów prognoz (zob. rysunek 23.4) dla okresów (1-10).

	A	B	C	D
1	t	y	y^	$(y-y^)^2$
2	1	500	494,92	25,81
3	2	507	514,71	59,44
4	3	533	534,50	2,25
5	4	550	554,29	18,40
6	5	580	574,08	35,05
7	6	600	593,87	37,58
8	7	610	613,66	13,40
9	8	640	633,45	42,90
10	9	650	653,24	10,50
11	10	670	673,03	9,18
12			Suma:	254,50
13			Średnia:	25,45

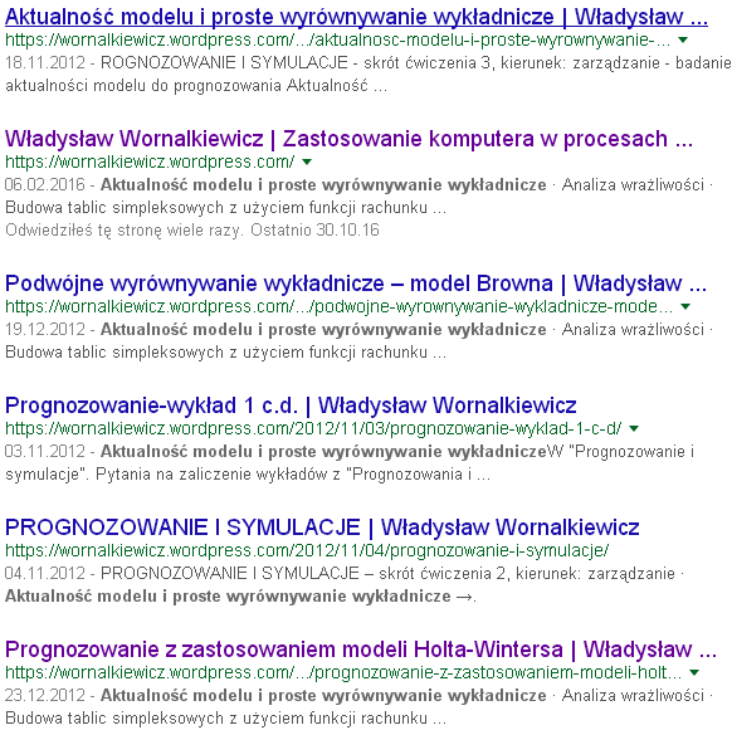
Źródło: Opracowanie własne w *Excelu*.

Rys. 23.4. Dane zmiennej i kwadraty błędów prognoz dla nowego modelu

Zatem dla nowego modelu średni błąd kwadratów wygasłych prognoz (25,45) jest mniejszy niż dla modelu wcześniejszego (60,182). Tak więc nowy model wyraźnie zmniejszył błąd prognozy. Pozostaje nam jednak pilnowanie dla kolejnych prognoz aktualności modelu współczynnikiem Janusowym.

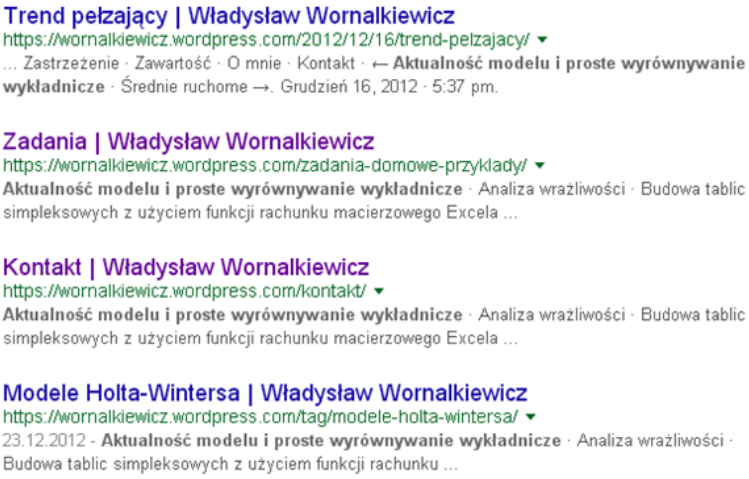
23.2. Odniesienia frazy *Aktualność modelu i proste wyrównywanie wykładnicze* we wpisach wyszukiwarki Google

Po zadaniu tej frazy w cudzysłowie wyszukiwarka Google w dniu badania udostępniła 90 wyników na liście w czasie 0,55 sekundy. Zgodnie z założeniem początkowym interesują nas tylko pozycje występujące na pierwszej stronie i to formatu PDF (zob. rysunki 23.5 oraz 23.6).



Źródło: Opracowanie własne w *Google*.

Rys. 23.5. Część 1. pierwszej strony raportu wyszukiwania



Źródło: Opracowanie własne w *Google*.

Rys. 23.6. Część 2. pierwszej strony raportu wyszukiwania

Dla tak rozbudowanej frazy na stronie pierwszej znalazły się tylko odniesienia do publikacji autora, co cieszy gdyż *Google* to wyszukiwarka światowa i cel popularyzatorski blogu rozpowszechniania metod ilościowych jest coraz bardziej realizowany.

24. Średnie ruchome

24.1. Tekst wpisu

Prognozowanie krótkoterminowe w oparciu o średnie ruchome stanowiło początek procesu formalizacji wygładzania wartości obserwacji. Wpis ten wprowadziłem do omawianego blogu w grudniu 2012 roku (zob. rysunek 24.1). Zainteresowanie tym wpisem wystąpiło dopiero od czerwca 2013 roku i obserwowany jest powolny wzrost.

Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Średnie ruchome	wornalkiewicz	Prognozowanie i symulacje	Model prostych średnich ruchomych, Model ważonych średnich ruchomych, Wagi, Średnie ruchome	—	0	Opublikowano 16.12.2012

Źródło: Opracowanie własne w *Wordpress*.

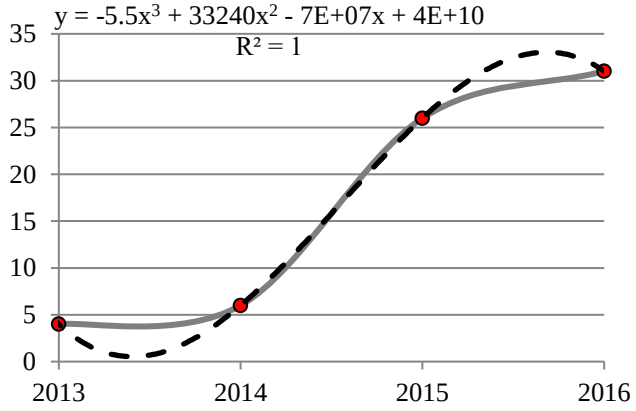
Rys. 24.1. Efekt wyszukiwania frazy „*Średnie ruchome*" w blogu „*Zastosowanie komputera w procesach decyzyjnych*"

Statystykę wejść internautów według miesięcy w latach 2013-2016 zaprezentowano na rysunku 24.2, przy czym na zakończenie roku 2016 było 31 wejść. Zobrazowanie tendencji trendem wielomianowym pokazano na rysunku 24.3.

	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2013						1	0	2	0	0	0	1	4
2014	1	0	0	0	1	0	1	0	1	1	1	0	6
2015	10	0	2	1	4	1	1	0	0	2	1	4	26
2016	2	1	2	1	3	18	3	0	0	1			31

Źródło: Opracowanie własne w *Wordpress*.

Rys. 24.2. Statystyka miesięczna wejść do wpisu „*Średnie ruchome*"



Źródło: Opracowanie własne w *Excelu*.

Rys. 24.3. Trend wielomianowy wejść do wpisu „*Średnie ruchome*" (linia przerywana)

Tekst wpisu „*Średnie ruchome*" stanowi skrót wykładu autora niniejszej pracy pt. „*Prognozowanie w oparciu o szeregi czasowe*". Gdy dane zjawisko jest złożone i trudno jednoznacznie wyznaczyć zmienne objaśniające to stosowane jest prognozowanie bazujące na obserwacjach w kolejnych okresach czasowych. W tym względzie używane są różne modele, począwszy od prostych średnich ruchomych, aż po modele Holta-Wintersa w których występuje

wyrównywanie wykładnicze: poziomu wartości obserwacji, zmieniającego się trendu, uwzględniona jest sezonowość. Prognozowanie bazujące na szeregach czasowych stosowane jest do ustalania prognoz krótkoterminowych, w tym bazując na modelach:

- prostych średnich ruchomych dwu i czterookresowych,
- ważonych średnich ruchomych czterookresowych.

Dla lepszego przedstawienia tych modeli posłużę się przykładem sprzedaży określonego produktu w kolejnych 12. miesiącach roku:

t	1	2	3	4	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**
y	120	124	122	123	125	128	129	127	129	128	130	132

W modelu prostych średnich ruchomych prognoza na kolejny okres stanowi średnią arytmetyczną kilku (np. dwóch lub czterech) wcześniejszych obserwacji. Mając określone wygasłe prognozy w oparciu o dwa lub cztery wcześniejsze okresy możemy porównać metody stosując określoną miarę błędów. Zastosowano tu miarę RMSE wyznaczona zamieszczonym wzorem, czyli pierwiastek ze średniej kwadratów odchyleń wartości obliczonej na podstawie modelu $\hat{y}_t$ od wartości obserwacji y_t:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}$$

Dla naszego przykładu bierzemy obserwacje 5-12, dla których wyznaczamy prognozy dwu i czterookresowe. Dla prostych średnich ruchomych dwuokresowych RMSE = 2,32, a dla czterookresowych 2,89. Zatem lepsza jest prognoza na kolejny okres bazująca na dwóch okresach.

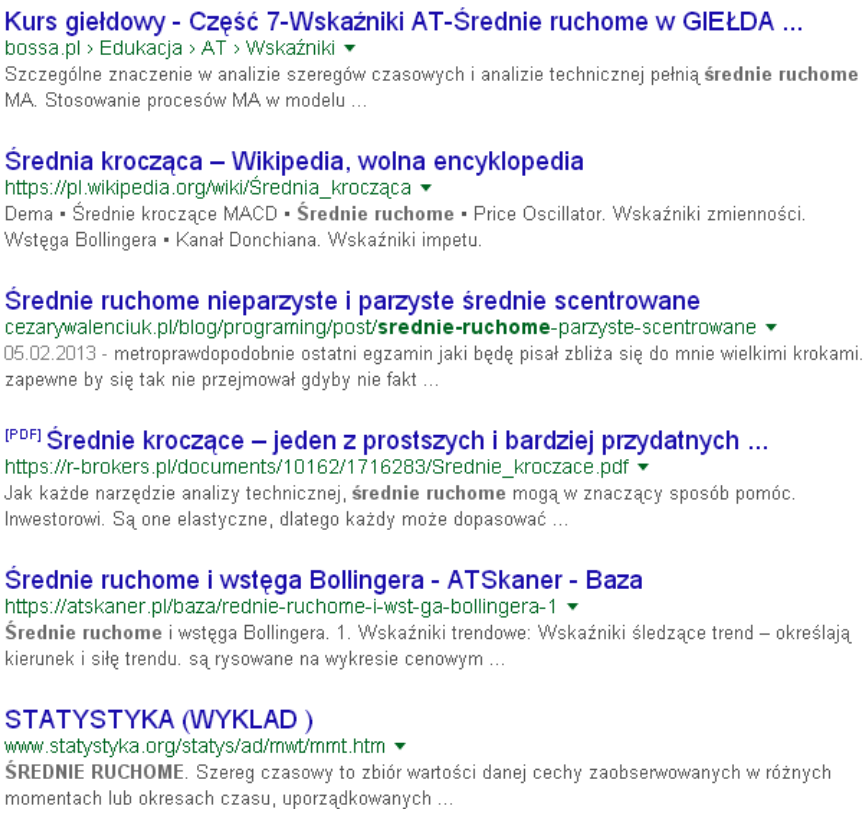
W modelach ważonych średnich ruchomych czterookresowych wprowadzono wagi dla kolejnych czterech okresów, przy czym suma tych wag równa się 1. Dla naszego przykładu prognoza na okres piąty przy umownych wagach (0,1; 0,2; 0,3; 0,4) wynosi:

$$F_5 = (120 \cdot 0{,}1) + (124 \cdot 0{,}2) + (122 \cdot 0{,}3) + (123 \cdot 0{,}4) = 122{,}6.$$

W obliczeniu wprowadzono „*umniejszanie*" znaczenia danych wygasłych przez wprowadzenie rosnącego szeregu wag (0,1; 0,2; 0,3; 0,4). Zachęcam Czytelnika do obliczenia w *Excelu* miary błędów RMSE, która dla naszego przykładu czterookresowego wynosi 2,54.

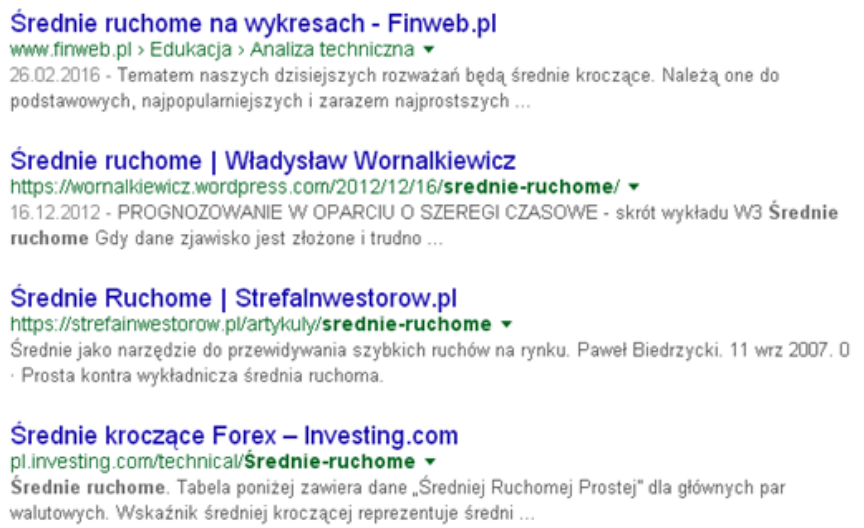
24.2. Odniesienia frazy *Średnie ruchome* we wpisach wyszukiwarki *Google*

Okazuje się, że fraza „*Średnie ruchome*" podana w cudzysłowie występuje w bardzo wielu pozycjach (wpisach) zgromadzonych w zasobach - dyskach wyszukiwarki *Google*. Udostępnione zostało aż 9710 pozycji skrótów do wpisów w czasie zaledwie 0,56 sekundy. Pozycje listy występujące na pierwszej stronie zamieszczono na rysunkach 24.4 oraz 24.5. Zaskakuje nas tu taka ogromna sprawność systemu wcześniejszego indeksowania oraz szybkiego udostępniania informacji na zapytanie w postaci frazy.



Źródło: Opracowanie własne w *Google*.

Rys. 24.4. Część 1. pierwszej strony raportu wyszukiwania na zapytanie „*Średnie ruchome*”



Źródło: Opracowanie własne w *Google*.

Rys. 24.5. Część 2. pierwszej strony raportu wyszukiwania na zapytanie „*Średnie ruchome*”

Mimo tak licznej listy udostępnionych pozycji system *Google* proponuje nam jeszcze skorzystanie z wyszukiwania do zadanej frazy „*Średnie ruchome*” podobnego - pisanego małymi literami- a mianowicie:

- średnie ruchome statystyka,
- średnie kroczące interpretacja,
- średnia krocząca excel,
- średnia ruchoma wzór,
- średnia krocząca giełda,
- średnia ruchoma prognozowanie,
- średnia ruchoma forex.

Pojawiła się tu fraza „*średnia ruchoma forex*", której rodzaj nie spotkałem jeszcze w literaturze. Zachęcam więc Czytelnika do zapoznania się z wpisami na ten temat. Cieszy mnie pojawienie się wpisu „*Średnie ruchome*" prowadzonego na *Wordpress* w ramach blogu „*Zastosowanie komputera w procesach decyzyjnych*".

Na stronie pierwszej występuje jedna publikacja formatu PDF „*Średnie kroczące - jeden z najprostszych i bardziej przydatnych wskaźników*" autorstwa Domu Maklerskiego Raiffeisen Bank Polska S.A. , a na stronie 18/18 jest rozwinięcie tekstu podanego w skrócie na liście wyszukiwania cytuję:

Jak każde narzędzie analizy technicznej, średnie ruchome mogą w znaczący sposób pomóc Inwestorowi. Są one elastyczne, dlatego każdy może dopasować ich parametry w celu wykorzystania średnich do swojej strategii. Jak każdy wskaźnik, prawidłowo używane są przydatne, należy jednak pamiętać, że nie jest to narzędzie idealne - średnia ruchoma także może być błędnie zinterpretowana i generować fałszywe sygnały.

Zainteresujmy się jeszcze fragmentem wpisu „Średnie ruchome i wstęga Bollingera …", gdzie spotykamy potraktowanie średnich ruchomych jako średnich kroczących. Tu pojawiło się rzadko stosowane pojęcie „*wstęga Bollingera*", które proponuję Czytelnikowi bliżej rozpoznać w Internecie[205]. W tej publikacji średnie ruchome (kroczące) to wskaźniki oparte na średniej notowań (cen zamknięcia) z określonej liczby sesji.

Zaglądnijmy jeszcze do publikacji „*Średnie kroczące Forex Investing.com*"[206]. Dziesięciodniowa średnia krocząca MA (*Moving Average*) to średnia z pewnej grupy danych. Obliczając następną wartość średniej, usuwa się pierwszy i dodaje jedenasty składnik sumy. Zadajemy sobie jednak pytanie, czy powinna to być prosta średnia krocząca (MA), ważona (WMA), wygładzona (SMA) czy też wykładnicza (EMA). Na rynkach terminowych wykorzystuje się połączenie średnich np. (4 i 9). Na rynku akcyjnym popularna jest średnia 50-dniowa, natomiast w analizie długoterminowej 200-dniowa. Do wyliczenia średniej kroczącej najczęściej używamy cen zamknięcia. Są firmy specjalizujące się w opracowaniach tego typu. Według wcześniej podanej publikacji popularne jest użycie pary średnich kroczących. Jeżeli szybsza z nich, ta o mniejszej wartości okresów, przecina wolniejszą, generowany jest sygnał kupna lub sprzedaży. W tym wypadku cały czas możemy być w rynku. Ponowne przecięcie średnich w kierunku odwrotnym generuje sygnał zamykania pozycji i otwarcia w kierunku przeciwnym. Tutaj przykładem są średnie 8 (MA) i 21 (MA)[207].

[205] https://atskaner.pl/baza/rednie -ruchome-i-wst-ga-bollingera-1.
[206] https://comparic.pl/systemy-inwestycyjne-oparte-o-srednie-kroczace-5-min-momentum-trade/.
[207] Ibidem.

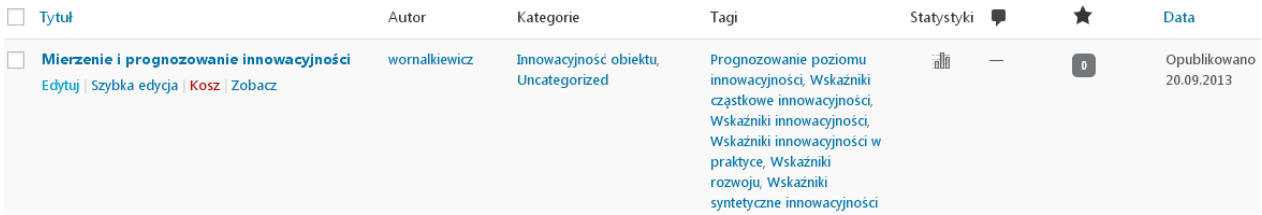
25. Mierzenie i prognozowanie innowacyjności

25. Tekst wpisu

Tematyka innowacyjności powraca od czasu do czasu w rywalizujących ze sobą krajach czy też regionach. Jako tagi (słowa kluczowe) w omawianym wpisie występują:

- prognozowanie poziomu innowacyjności,
- wskaźniki cząstkowe innowacyjności,
- wskaźniki innowacyjności,
- wskaźniki innowacyjności w praktyce,
- wskaźniki syntetyczne innowacyjności .

Wpis „*Mierzenie i prognozowanie innowacyjności*" opublikowałem na swoim blogu we wrześniu 2013 roku (zob. rysunek 25.1).

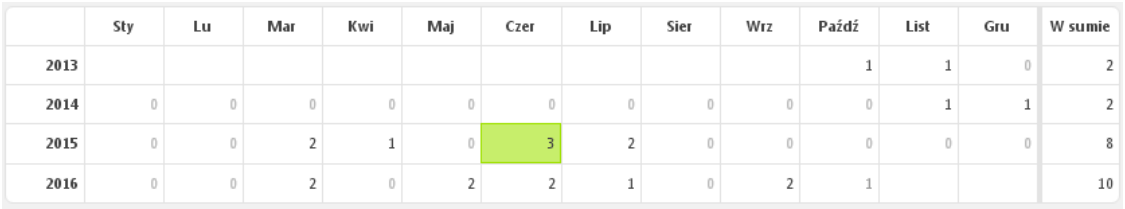


Tytuł	Autor	Kategorie	Tagi	Statystyki		Data
Mierzenie i prognozowanie innowacyjności Edytuj | Szybka edycja | Kosz | Zobacz	wornalkiewicz	Innowacyjność obiektu, Uncategorized	Prognozowanie poziomu innowacyjności, Wskaźniki cząstkowe innowacyjności, Wskaźniki innowacyjności, Wskaźniki innowacyjności w praktyce, Wskaźniki rozwoju, Wskaźniki syntetyczne innowacyjności	—	0	Opublikowano 20.09.2013

Źródło: Opracowanie własne w *Wordpress*.

Rys. 25.1. Komunikat o występowaniu wpisu wywołanego frazą „*Mierzenie i prognozowanie innowacyjności*"

Zainteresowanie internautów tym tematem w latach 2013-2016, według statystyki miesięcznej prowadzonej przez *Wordpress*, okazało się niewielkie. Obserwowany jest jednak powolny wzrost, a szkoda, gdyż wpis ten zawiera wiele praktycznych informacji o zastosowaniu wskaźników mierzenia innowacyjności. Na zakończenie roku 2016 było 14 odwiedzin (zob. rysunek 25.2).



	Sty	Lu	Mar	Kwi	Maj	Czer	Lip	Sier	Wrz	Paźdź	List	Gru	W sumie
2013										1	1	0	2
2014	0	0	0	0	0	0	0	0	0	0	1	1	2
2015	0	0	2	1	0	3	2	0	0	0	0	0	8
2016	0	0	2	0	2	2	1	0	2	1			10

Źródło: Opracowanie własne w *Wordpress*.

Rys. 25.2. Statystyka miesięczna wejść do wpisu „*Mierzenie i prognozowanie innowacyjności*"

Coraz częściej słyszymy o wydawaniu dużych środków finansowych, w tym dotacji z Unii Europejskiej na różne przedsięwzięcia. Oczekuje się od tego poprawy innowacyjności naszej gospodarki oraz usług. W tym względzie interesujące może być mierzenie i przewidywanie rozwoju w obszarze szeroko rozumianej kreatywności i innowacyjności wytworów pracy grup społecznych i produkcyjnych. Z tej inspiracji wyłoniłem na podstawie dostępnych źródeł, w tym głownie

internetowych propozycje wskaźników zarówno cząstkowych jak i budowanych na nich wskaźników syntetycznych, a celami głównymi w tym zakresie były[208]:

Potrzeba: wyłonienie wskaźników cząstkowych i miary syntetycznej.

Odniesienie: podmiotowo - nawiązanie do cech ekonomicznych obiektu.

Trudności: statystyka GUS-u ma odniesienie przedmiotowe.

Podejście - tłoczenie nowych rozwiązań do użytkowników.

Grupowanie - finansowanie badań.

Spojrzenie - kreatywne.

Porównywalność wskaźników z innymi regionami.

Formułowanie rankingu na podstawie wielu zmiennych.

Obszary kreatywne: komputeryzacja, teleinformatyka oraz biotechnologia.

Ocena oryginalności pomysłu - sprawne poruszanie się po mediach informacyjnych.

Określenie związków korelacyjnych między zmiennymi.

Istotność analizy stanu istniejącego oraz wskazanie zamierzeń strategicznych.

Umiejętność budowania modeli ekonometrycznych.

Wskazanie rekomendacji rozwoju badanego obiektu.

Wskaźniki rozwoju i innowacyjności powinny uwzględniać:

Badania jakościowe przy porównywaniach międzynarodowych.

Modelowanie ekonometryczne wymagane przede wszystkim badań ilościowych.

Zestaw standardowych wskaźników pomiaru rozwoju określonej branży.

Dominujące znaczenie cech produktów lub usług oferowanych odbiorcom.

Podejścia do innowacyjności, w tym popytowe, podażowe, popytowo-podażowe i modelowe.

Źródło, którym jest międzynarodowy program badań statystycznych (CIS).

Metodologie, z uwzględnieniem opracowań Amerykańskiej Izby Handlowej, Unii Europejskiej.

Bazowanie na uzyskaniu określonych efektów przy przyjętych nakładach.

Ocenę regionu według istniejącego poziomu rozwoju gospodarczego, wydajność pracy i sprawność emitowania nowych produktów.

Sposoby pomiaru, tj. w formie zestawu wskaźników cząstkowych, wskaźnika syntetycznego, modelu ekonometrycznego na wyselekcjonowanych wskaźnikach cząstkowych oraz prognozowania w oparciu o modele.

W formułowaniu wskaźników rozwoju innowacyjności trzeba skorzystać ze wskaźników ze sprawozdawczości GUS-u do pomiaru rozwoju sektorów, a w tym: wielkości zatrudnienia, wysokości obrotów, zysku brutto, nakładów inwestycyjnych, udziału sektorów kreatywnych w PKB. Wskaźnikami cząstkowymi do badania innowacyjności są:

Nakłady na innowacje jako udział w obrotach (ogółem).

Udział firm prowadzących działalność innowacyjną.

Firmy wprowadzające innowacje nietechnologiczne (organizacyjne, marketingowe), tj. procent wszystkich firm wybranych grup.

Eksport produktów.

Sprzedaż produktów nowych dla rynku (% obrotów ogółem).

Liczba zarejestrowanych wzorów użytkowych i znaków towarowych.

Wartość dodana brutto na zatrudnionego.

Wskaźniki cząstkowe są bazą do formułowania różnych wskaźników syntetycznych z wyodrębnieniem kreatywności i innowacyjności. Trzeba pamiętać jednak o modyfikacji zmiennych w czasie, a także rozważyć tzw. zmienne społeczne. Jako zmienne z obszaru regulacji i rezultatów proponuje się:

Regulacje:

Ulgi podatkowe dla osób pracujących w sektorach kreatywnych.

Stawka VAT na książki, prasę, nagrania oraz video.

Regulacje podatkowe zachęcające do sponsoringu i dotacji.

[208] Niniejsza część opracowania bazuje na: Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania,* Wydawnictwo Instytut Śląski, Opole 2015, rozdział 2.20.

Bezpośrednie wydatki publiczne na kulturę.

Wydatki publiczne na sztukę filmową.

Wydatki publiczne na telewizję publiczną.

Prawa autorskie w przemyśle muzycznym na mieszkańca.

Rezultaty kreatywności:

Wartość dodana w przemysłach kreatywnych jako % PKB.

Obroty w przemyśle muzycznym na mieszkańca.

Obroty w przemyśle wydawniczym na mieszkańca.

Obroty kin na mieszkańca.

Liczba wyprodukowanych filmów na mieszkańca.

Liczba wyprodukowanych nagrań na mieszkańca.

Liczba wydanych książek na mieszkańca.

Liczba złożonych wniosków na wzory użytkowe na milion mieszkańców.

Stopień innowacyjności stanowi uproszczony wskaźnik syntetyczny (SIL) jako średnia ważona (5-10) wskaźników składowych. Miarami łącznej oceny rozwoju i innowacyjności są:

Wartość dodana wybranych grup na zatrudnionego.

Udział wybranych grup w zatrudnieniu ogółem.

Udział eksportu dla wybranych grup w eksporcie ogółem.

Inwestycje w przedsiębiorstwach wybranych grup na zatrudnionego.

Nakłady na innowacje.

Udział dochodów ze sprzedaży produktów, które zostały wprowadzone lub ulepszone w ciągu ostatnich trzech lat, w dochodach ze sprzedaży w danym roku.

Bardzo istotna rolę odgrywa właściwy dobór zestawu zmiennych i następnie estymacja parametrów modelu ekonometrycznego. Badanie zmian parametru sumarycznego ma na celu ocenę łączną m.in. wydajności czynników produkcji. Odniesieniem do wyznaczenia wskaźników innowacyjności w praktyce była opublikowana w Internecie „*Analiza stanu innowacyjności województwa opolskiego*", której miała następujące cele[209]:

- wyłonienie zmiennych ⇒ wskaźniki cząstkowe ⇒ wskaźniki syntetyczne (średnie),
- model ekonometryczny ⇒ prognozy strategiczne,
- wyłonienie wskaźników cząstkowych,
- sformułowanie wskaźników średnich,
- zabazowanie na danych statystycznych Eurostat-u.

Umożliwiło to przedstawienie graficzne pozycji danego regionu na tle kraju, regionów UE oraz grupy regionów porównywalnych.

Kroki postępowania w opracowaniu analizy innowacyjności sprowadzają się do wyznaczenie następujących wskaźników cząstkowych:

Całkowite wydatki wewnętrzne na badania i rozwój jako % PKB.

Liczba wniosków patentowych złożonych do Europejskiego Urzędu Patentowego w przeliczeniu na 1 mln mieszkańców.

Zatrudnienie w usługach opartych na wiedzy jako % zatrudnienia ogółem.

Zatrudnienie w sektorze produkcji *high-tech* oraz *medium high-tech* jako % zatrudnienia ogółem.

Ponadto zachodzi potrzeba analizy korelacji pomiędzy poziomem innowacyjności regionów wyrażoną 50. zmiennymi dotyczącymi innowacyjności a zamożnością (*PKB per capita*). Zalecany jest podział pięćdziesięciu zmiennych według współczynnika korelacji na 4 grupy korelacji tj. bardzo wysokiej, wysokiej, przeciętnej oraz słabej. W kolejnych krokach określane są:

Ranking 16. województw w grupach korelacji, w tym pozycja danego województwa np. opolskiego.

Określenie zestawu zmiennych dla których dane województwo np. opolskie jest w czołówce (1-4) oraz w dalszych przedziałach (5-16).

---

[209] *Analiza stanu innowacyjności województwa opolskiego*, GEOPROFIT, Warszawa 2010, https://www.efs.2007-2013.gov.pl/analizyraportypodsumowania/baza_projektow_badawczych_efs/documents/analiza_stanu_innowacyjnosci.pdf .

Graficzna prezentacja wybranych dla dobrego wizerunku zmiennych na tle innych województw.

W opracowaniu analizy innowacyjności wyznaczane są też średnie wskaźniki, a w tym wskaźnik syntetyczny. W wymienionej wcześniej „*Analizie stanu innowacyjności województwa opolskiego*" opracowanie wskaźnika syntetycznego poprzedzone jest wyborem 22 zmiennych. Poddano je ponownej wnikliwej analizie i następnie przyjęto 8 następujących zmiennych traktowanych jako wskaźniki cząstkowe:

1. Jednostki B+R w przedsiębiorstwach na 1000 mieszkańców.
2. Nakłady na B+R na zatrudnionego w B+R.
3. Pracownicy naukowo-badawczy ogółem w przeliczeniu na 1000 mieszkańców.
4. Nakłady na działalność innowacyjną w przemyśle na mieszkańca.
5. Procent firm korzystających poprzez Internet z usług administracji.
6. Procent gospodarstw domowych z komputerem i dostępem do Internetu.
7. Przedsiębiorstwa, które prowadziły działalność innowacyjną jako % przedsiębiorstw przemysłowych.
8. Liczba komputerów z dostępem do Internetu na 100 uczniów we wszystkich szkołach oprócz uczelni wyższych.

Wyodrębnione zostały również podwskaźniki (średnie) przybliżające obszary wpływu, obejmujące wymienione wcześniej wskaźniki cząstkowe: *gospodarki* (1, 2, 4, 5, 7); *społeczeństwa* (3, 5); *sieci* (1, 5, 6, 8). Następnie opracowuje się prognozy poziomu innowacyjności na dłuższy horyzont lat oraz sporządza się analizę SWOT. Wskazuje się w niej słabe i mocne strony danego obiektu struktury terytorialnej w aspekcie rozwoju, kreatywności i innowacyjności.

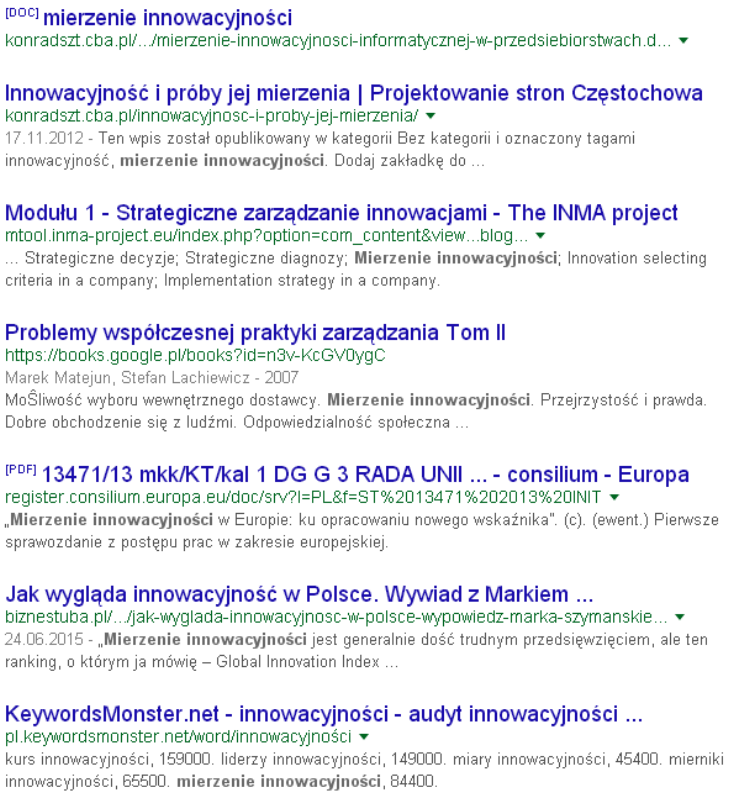
25.2. Rozpoznanie źródeł dostępnych w *Google* według frazy *Mierzenie i prognozowanie innowacyjności*

Dla tak sformułowanej szerszej frazy podanej w cudzysłowie wyszukiwarka *Google* udostępniła tylko trzy pozycje w czasie 0,46 sekundy. Dwie z nich to teksty zamieszczone w ramach blogu „*Zastosowanie komputera w procesach decyzyjnych*" a jedna to reklama. System sygnalizuje nam możliwość dodatkowego wyszukiwania z uwzględnieniem pominiętych wyników uznanych przez niego jako mniej trafne. Nie skorzystamy z tego, lecz spróbujmy zmienić frazę na krótszą „*Mierzenie innowacyjności*".

Źródło: Opracowanie własne w *Google*.

Rys. 25.3. Odpowiedź na zapytanie " *Mierzenie i prognozowanie innowacyjności*"

Dla tej złożonej frazy wystąpiły tylko dwa wpisy i to autora tej publikacji. Spróbujmy zmienić frazę na krótszą „*Mierzenie innowacyjności*". Teraz pojawiło się 100 odpowiedzi w czasie 0,31 sekundy, a na stronie pierwszej są pozycje wyszczególnione na rysunkach 25.4 i 25.5.



Źródło: Opracowanie własne w *Google*.

Rys. 25.4. Część 1. odpowiedzi na zapytanie ”*Mierzenie innowacyjności*”

Źródło: Opracowanie własne w *Google*.

Rys. 25.5. Część 2. odpowiedzi na zapytanie ” *Mierzenie innowacyjności*”

Czytając wygenerowaną listę rezultatów wyszukiwania dla frazy „*Mierzenie innowacyjności*” moim zdaniem na uwagę zasługują dwa wpisy: *Mierzenie innowacyjności* w formacie edytora tekstów Word oraz wpis oznaczony jako „13471 …”. Zajrzyjmy zatem do pierwszego z wymienionych, czyli *Mierzenie innowacyjności*, z którego fragment tekstu przytoczę:

Innowacyjność jest środkiem do celu jakim jest osiągnięcie przewagi konkurencyjnej na rynku. Dla polskich przedsiębiorstw bardziej opłacalne są innowacje naśladowcze, niż przełomowe. Biorąc pod uwagę nakłady finansowe, które przedsiębiorstwa są w stanie przeznaczyć na badania nad nowymi produktami wspomagającymi czy nowatorskie wdrożenia wychodzi na to, iż ryzyko niepowodzenia i strat finansowych skutecznie odstrasza ich przed tymi działaniami. Działania naśladowcze nie są aż tak kapitałochłonne, ale też nie powodują objęcia zdecydowanego przodownictwa w tym „wyścigu zbrojeń”. Mierzyć poziom innowacyjności w informatyzacji można mierząc pośrednio za pomocą poprawy produktywności powstałej na skutek wdrożenia innowacyjnych rozwiązań.

W opracowaniu wpisu autor jako podstawę tekstu przyjął trzy cztery pozycje literaturowe, a mianowicie:

1. Markowska M., *Problematyka pomiaru innowacyjności regionalnej w statystyce unijnej*, W: *Statystyka wczoraj, dziś i jutro*, Biblioteka Wiadomości Statystycznych, tom 56, GUS - PTS, Warszawa 2008.

2. Markowska M., Strahl D., *Propozycja pomiaru innowacyjności regionalnej typu Input - Output*. W: Statystyka w praktyce społeczno gospodarczej, pod red. W. Ostasiewicza, Akademia Ekonomiczna, Wrocław 2007, PN nr 1163, Wrocław 2007.

3. Markowska M., Strahl D., *Przegląd koncepcji pomiaru regionalnej innowacyjności w unijnej statystyce*, Prace Naukowe Akademii Ekonomicznej, Wrocław nr 1142, Wrocław, 2006.

Zainteresujmy się jeszcze wcześniej wymienionym drugim wpisem internetowym w formacie PDF o skrócie „13471…", w tym tekstem *Informacje o wskaźniku poziomu innowacyjności* spotykanym stronie 4/6. Zacytuje moim zdaniem warte uwagi dwa fragmenty tekstu:

W strategii 'Europa 2020" na rzecz inteligentnego, trwałego wzrostu gospodarczego sprzyjającego włączeniu społecznemu określono pięć głównych wskaźników. Jednym z nich jest poprawa warunków w przypadku badań i rozwoju, w celu zwiększenia poziomów połączonych inwestycji publicznych i prywatnych w badania i rozwój do 3% PKB. By uzupełnić ten wskaźnik intensywność inwestycji w badania i rozwój, Rada Europejska udzieliła Komisji mandatu do przygotowania pojedynczego wskaźnika innowacyjności.

oraz

Proponowany wskaźnik koncentruje się na czterech składnikach związanych z polityką:

1. Innowacjach technologicznych mierzonych liczbą patentów, co pokazuje zdolność gospodarki do przekładania wynalazków na skuteczne technologie, które mogą zostać wprowadzone na rynek.

2. Zatrudnienie w przypadku branż wymagających specjalistycznej wiedzy, co pokazuje, w jaki sposób wysoko wykwalifikowana siła robocza przyczynia się do produkcji innowacyjnych towarów i świadczenia innowacyjnych usług.

3. Konkurencyjność towarów i usług, które wymagają specjalistycznej wiedzy, co pokazuje zdolność państwa do wprowadzenia innowacji na rynki światowe.

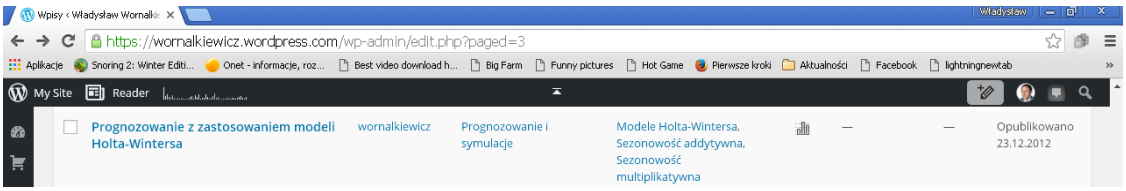
4. Zatrudnienie w szybko rosnących przedsiębiorstwach w sektorach innowacyjnych, wskazujące na zdolność państwa do szybkiego przekształcenia gospodarki, by wychodziła naprzeciw nowym potrzebom i by korzystała z pojawiającego się popytu.

Trzeba dodać, że zasugerowany wskaźnik ma na celu porównywanie w Unii Europejskiej osiągnięć w zakresie innowacyjności.

26. Modelowanie ekonometryczne ruchu internetowego

26.1. Zastosowanie modelu wielomianowego

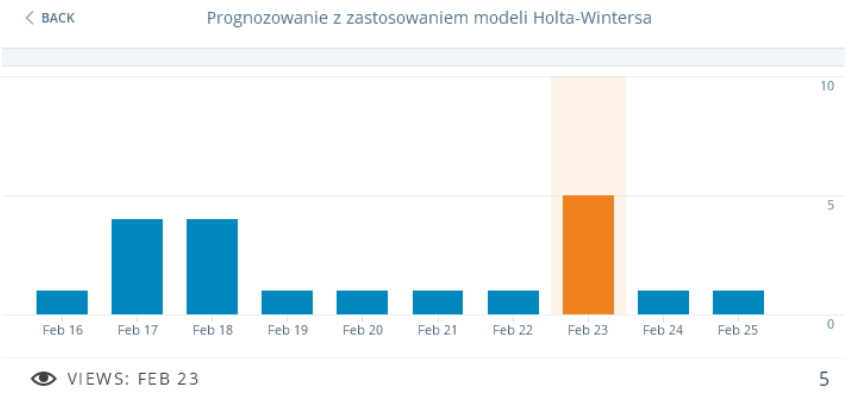
Niniejszy rozdział skupia się na statystyce ruchu internetowego obserwowanej na blogu autora pt. „*Zastosowanie komputera w procesach decyzyjnych*”. Blog ten na stronie głównej zawiera wiele wpisów (*postów*), ale najczęściej na przestrzeni okresu obserwowanego w tej pracy życia blogu wywoływany przez internautów był wpis pt. „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*”. Jego zaznaczenie w ramach szerszego zestawu wpisów pokazano na rysunku 26.1.



Źródło: Opracowanie własne według *WordPress*.

Rys. 26.1. Fragment zestawu wpisów dotyczący wpisu „Prognozowanie z zastosowaniem modeli Holta-Wintersa”

Statystykę wejść w przykładowych dniach lutego 2016 roku pokazano na rysunku 26.2.



Źródło: Opracowanie własne wedlug statystyki *WordPress*.

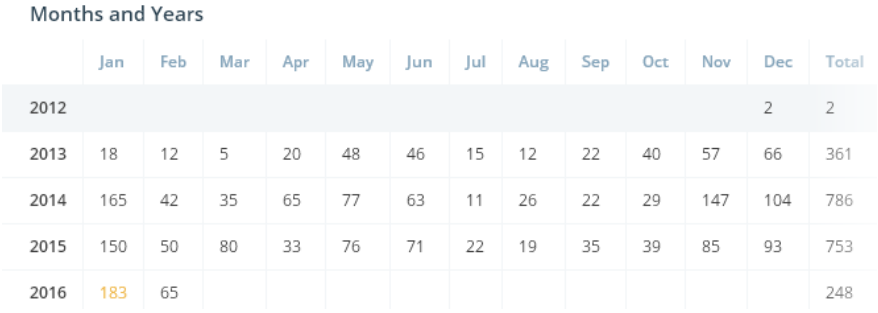
Rys. 26.2. Dzienna liczność wejść do wpisu „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*”

Jeśli pobierzemy z omawianego blogu statystykę z okresu przyjętego w tym opracowaniu poprzez adres:

https://wornalkiewicz.wordpress.com/wp-admin/index.php?page=stats&view=post&post=768)

w układzie lat i miesięcy podanych w skrótach angielskich (Jan - Dec), to zauważamy jako interesujące do modelowania ekonometrycznego kompletne dane z lat 2013-2015.

Na zamieszczonym rysunku 26.3 dotyczącej statystyki omawianego wpisu pokazano również podsumowania liczb wejść miesięcznych dla poszczególnych lat (kolumna *Total*).



Months and Years

	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Total
2012												2	2
2013	18	12	5	20	48	46	15	12	22	40	57	66	361
2014	165	42	35	65	77	63	11	26	22	29	147	104	786
2015	150	50	80	33	76	71	22	19	35	39	85	93	753
2016	183	65											248

Źródło: Opracowanie własne wedlug statystyki *WordPress.*

Rys. 26.3. Pełna statystyka wejść w układzie lat i miesiecy wpisu „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*”

Widzimy wyraźny skok zainteresowania problematyka metod ilościowych w roku 2014 w stosunku do roku poprzedniego (361 do 786). Zapiszmy dane lat 2013, 2014 oraz 2015 w formie tabeli Worda (zob. tabela 26.1). Zwróćmy uwage na zaznaczone na czerwono dwa okresy roku akademickiego tj. 09.2013/06.2014 i 09.2014/06.2015. Największe zainteresowanie przykładowym tematem występuje w miesiącach przygotowania się do egzaminów z wykładów czy też zaliczeń z ćwiczeń, czyli orientacyjnie w styczniu oraz maju.

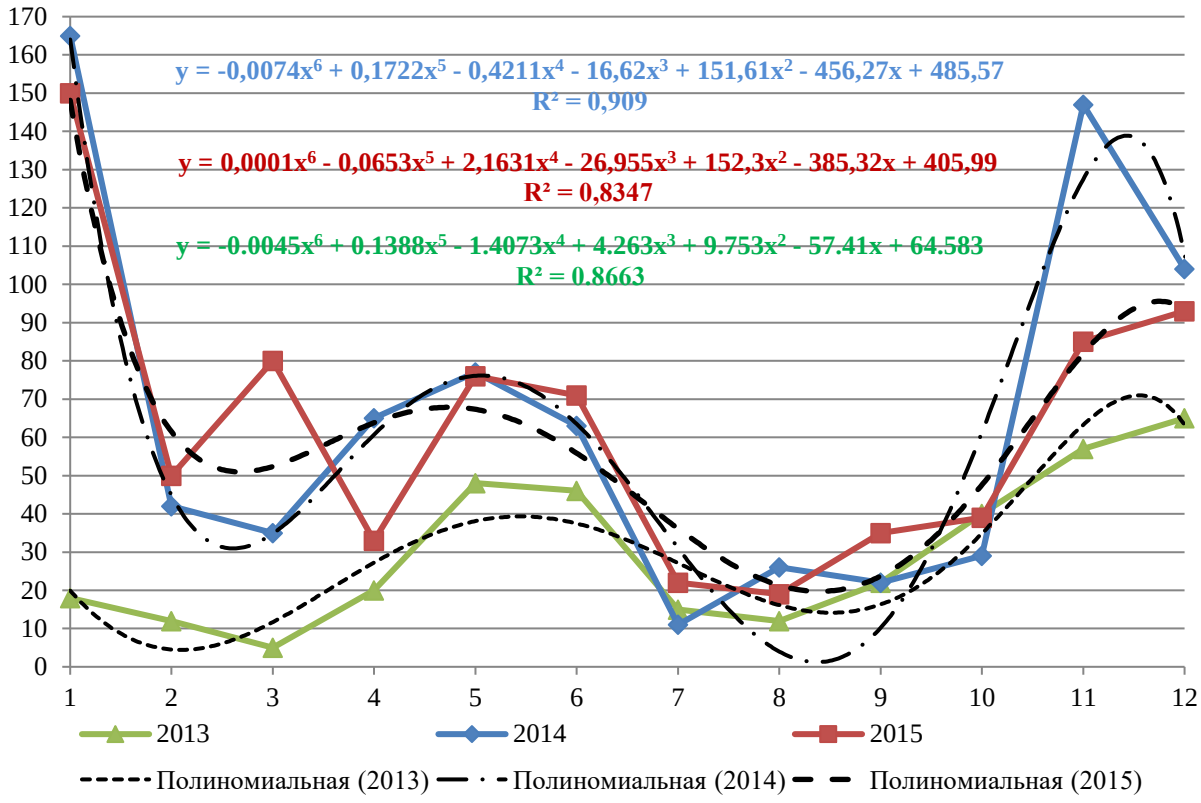
Tab. 26.1.

Dane statystyczne dotyczące przykładowo jednego z najpopularniejszych wpisów

Rok/M-c	O1	02	03	04	05	06	07	08	09	10	11	12
2013	18	12	5	20	48	46	15	12	22	40	57	65
2014	165	42	35	65	77	63	11	26	22	29	147	104
2015	150	50	80	33	76	71	22	19	35	39	85	93

Źródło: Opracowanie własne.

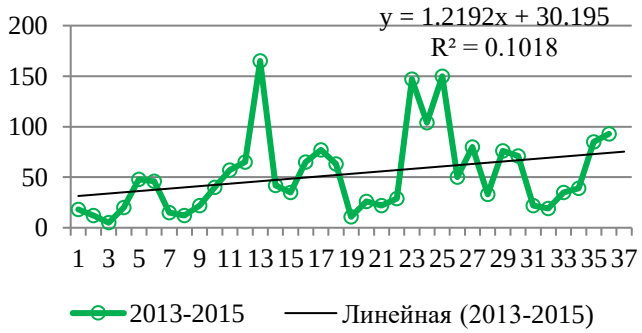
Na początek jednak zobrazujmy dane lat 2013, 2014, 2015 w formie modelu wielomianowego szóstego stopnia wykonanego przy zastosowaniu *Excela*. Na rysunku 26.4 widzimy naniesione dane statystyczne oraz wygenerowane modele wielomianowe wykonane oddzielnie dla roku 2013, 2014, 2015. Zaskakują nas wysokie współczynniki determinacji R^2, świadczące o dobroci dopasowania określonych modeli do empirii.



Źródło: Opracowanie własne według statystyki *WordPress*.

Rys. 26.4. Dane statystyczne oraz trendy wielomianowe liczby wejść do omawianego wpisu

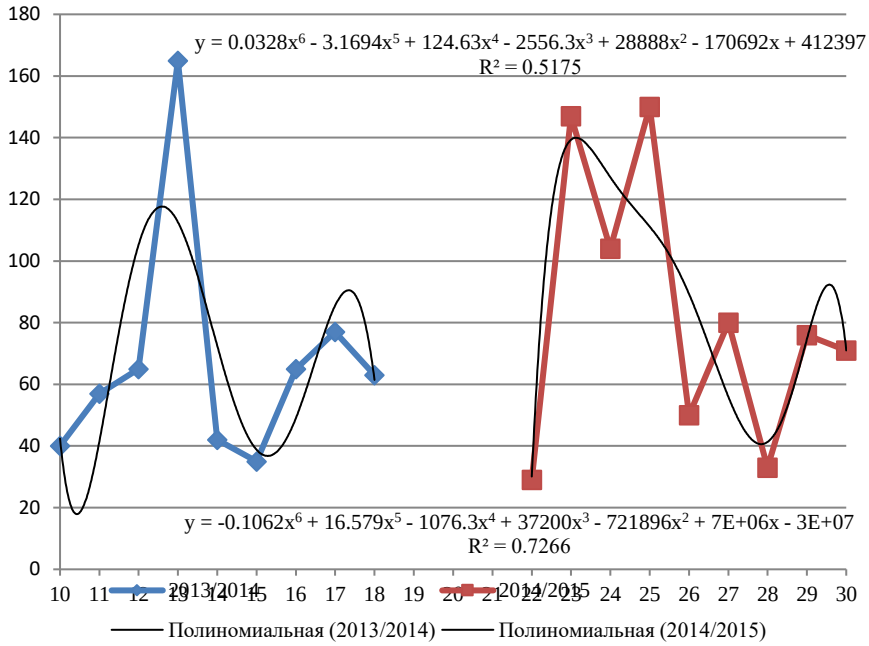
W celu zastosowania modelu Holta-Wintersa warto wcześniej poznać wystepowanie trendu liniowego w całym przedziale badawczym tj. w 36 miesiacach lat 2013-2015. Na wykonanym w *Excelu* wykresie (zob. rysunek 26.5) wystepuje nieznaczny trend liniowy określony współczynnikiem 1,2192, przy niskim wspólczynniku determinacji R^2 = 0,1018.



Źródło: Opracowanie własne w *Excelu*.

Rys. 26.5. Trend liniowy liczby wejść do omawianego wpisu w okresie 36. miesiecy

Określmy jeszcze trendy wielomianowe szóstego stopnia, przy których występuje najlepsze dopasowanie modelu do danych empirycznych w przedziałach czasowych roku aktywności akademickiej tj. miesiąca października do miesiąca czerwca włącznie. Rezultat z wykonania dwóch wykresów na jednej odciętej (miesiące 10.2013 do 06.2014 oraz 10.2014 do 6.2015) zaprezentowano na rysunku 26.6. Możemy zauważyć możliwie dobry R^2 = 0,7266 dla okresu 10.2014 do 6.2015.



Źródło: Opracowanie własne w *Excelu*.

Rys. 26.6. Trend wielomianowe liczmy wejść do wybranego wpisu dla dwóch okresów aktywności akademickiej

26.2. Modele Holta-Wintersa

Zasadniczo możemy wyróżnić trzy rodzaje modeli Holta-Wintersa: bez sezonowości, z sezonowością multiplikatywną, z sezonowością addytywną[210]. Wyrównywana jest wartość trendu z poprzedniego okresu.

Model Holta-Wintersa bez sezonowości Model bez sezonowości ma zastosowanie, gdy występuje trend i wahania okresowe. W modelu bez sezonowości oprócz parametru α występuje drugi parametr β odpowiednio do wygładzania wartości poziomu i trendu. Równania tego modelu są następujące:

$$S_n = \alpha\, Y_n + (1- \alpha)\, (S_{n-1} + b_{n-1})$$
$$b_n = \beta\, (S_n - S_{n-1}) + (1 - \beta)\, b_{n-1}$$
$$F_{n+m} = S_n + b_n\, m.$$

Gdzie Y_n odnosi się do zmiennej objaśnianej i są to obserwacje w kolejnych okresach szeregu czasowego. W modelu tym przyjmujemy następujące wartości początkowe:

$$S_1 = Y_1; \quad b_1 = [(Y_2 - Y_1) + (Y_4 - Y_3)]/2.$$

Działania optymalizacyjne sprowadzają się do takiego dobrania parametru α, β, aby uzyskać minimalna wartość pierwiastka kwadratu średniego błędu *RMSE* prognozy *ex post*.

Model Holta-Wintersa z sezonowościa multiplikatywną W modelu tym występuje trend, wahania okresowe oraz zmienna sezonowość zależna od trendu. Wyrównywanie sezonowości następuje poprzez zastosowanie parametru γ stosowanego do obliczania indeksów sezonowości. Postać modelu z sezonowością multiplikatywną jest następująca[211]:

$$Y_{n+1} = (S_n + b_n)\, I_{n-L+1} + z_{n+1}$$

gdzie:

I_{n-L+1} - wyrównana wartość indeksu sezonowości na okres $n + 1$,

[210] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania* cz. II, Wydawnictwo Instytut Śląski, Opole 2015, rozdz. 2.12. *Wyrównywanie trendu z zastosowaniem modelu Holta-Wintersa*.
[211] Ibidem, s. 504.

z_{n+1} – błąd w okresie n +1,

L – długość cyklu sezonowości (12 dla danych miesięcznych, 4 dla kwartalnych).

Wyrównana wykładniczo wartość Y_n po usunięciu sezonowości:

$$S_n = \alpha\ (Y_n/I_{n-L}) + (1 - \alpha)\ (S_{n-1} + b_{n-1}).$$

Wyrównana wykładniczo wartość trendu w okresie n:

$$b_n = \beta\ (S_n - S_{n-1}) + (1 - \beta)\ b_{n-1}.$$

Wyrównany wykładniczo indeks sezonowości w okresie n:

$$I_n = \gamma\ (Y_n/S_n) + (1 - \gamma)\ I_{n-L}.$$

Określenie wcześniej podanych wartości wyrównanych pozwala na określenie prognozy na okres $n + m$, przy czym m to horyzont prognozy:

$$F_{n+\mathrm{m}} = (S_n + b_n\ m)\ I_{n-L+m}.$$

Jako wartości początkowe możemy przyjąć: $\alpha = 0{,}2$; $\beta = 0{,}2$; $\gamma = 0{,}1$. Jednak dla doboru najlepszej wartości tych parametrów możemy zastosować moduł *Forecasting and Lineal Regression* programu WinQSB przy warunku minimum RMSE bez podawania wartości początkowych.

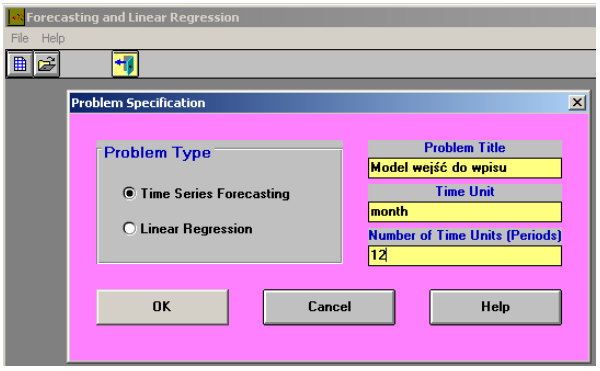
Model Holta-Wintersa z sezonowościąi addytywną. Model ten określony jest równaniami:

$$S_n = \alpha\ (Y_n - I_{n-L}) + (1 - \alpha)\ (S_n\text{-}1 + b_{n-1})$$
$$b_n = \beta\ (S_n - S_{n-1}) + (1 - \beta)\ b_{n-1}$$
$$I_n = \gamma\ (Y_n - S_n) + (1 - \gamma)\ I_{n-L}$$
$$F_{n+m} = S_n + b_n\ m + I_{\mathrm{n}-L+m}$$

Widzimy, że model Holta-Wintersa z sezonowością addytywną stanowi analogię do modelu multiplikatywnego. Zwróćmy jednak uwagę na odejmowanie indeksu sezonowości I_{n-L}.

26.3. Sformułowanie modelu multiplikatywnego

Powróćmy do naszych modeli wielomianowych rocznych ilości wejść do wpisu „*Prognozowanie z zastosowaniem modeli Holta-Wintersa*", lecz zastosujmy model z sezonowością multiplikatywną oraz moduł *Forecasting and Lineal Regression* (FC) programu WinQSB (zob. rysunek 26.7). Wykorzystamy również opcję *Holt-Winters Multiplicative Algorithm* (HWM) do automatycznego dobierania parametrów α, β, γ w celu osiągnięcia minimum błędu według miary RMSE[212]. Dobór przeprowadzamy ze skokiem co 0.01 (z kropką dziesiętną) bez podawania wartości początkowych.



Źródło: Opracowanie własne.

Rys. 26.7. Wstępne definiowanie modelu multiplikatywnego

[212] Zabazowano na rozdziale 2.12.2.2 *Model Holta-Wintersa z sezonowością multiplikatywną*, książki autora pt. „ *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*", s. 514.

Następną czynnościa jest wprowadzenie danych dotyczących obserwacji wejść do omawianego wpisu internetowego dla 12 miesiecy roku 2013 (zob. rysunek 26.8).

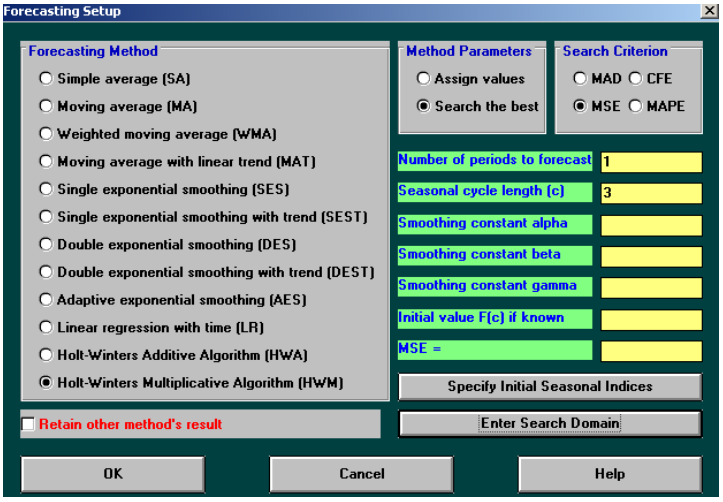
Month	Historical Data
1	18
2	12
3	5
4	20
5	48
6	46
7	15
8	12
9	22
10	40
11	57
12	65

Źródło: Opracowanie własne.

Rys. 26.8. Okno wprowadzania danych

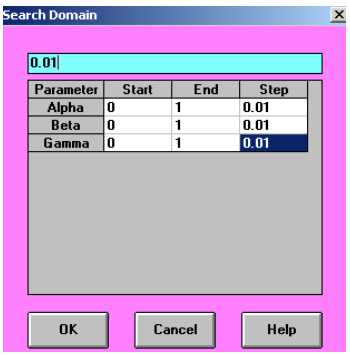
Pozostaje nam teraz wybór rodzaju modelu prognozowania *Holt-Winters Multiplicative Algorithm* (HWM) oraz opcji dotyczących procesu iteracyjnego optymalizacji parametrów α, β, γ. Wykorzystamy metodę szukania tych parametrów jako *Search the best*, czyli najlepszy wybór z punktu widzenia kryterium sumarycznego błędu *Search Criterion* jako MSE, z którego pierwiastek równa się RMSE. Przystąpienie do formułowania modelu prognozy na pierwszy miesiąc nowego roku wymaga podania opcji:

- *Number of periods to forecast* (liczby okresów prognozowania) równej 1,
- *Seasonal cycle length* (*c*) – ilość okresów cyklu sesonowości = 3,
- określenia skoku iteracyjnego (*Step*) doboru kolejnych parametrów α, β, γ co 0.01 z kropką dziesietną w podoknie *Enter Search Domain* (zob. rysunki 26.9 oraz 26.10).



Źródło: Opracowanie własne.

Rys. 26.9. Okno ustawienia opcji prognozowania



Źródło: Opracowanie własne.

Rys. 26.10. Podokno definiowania przedziałów iteracji parametrów modelu

Po kliknięciach na menu *Solve and Analyze/Perform Forecasting*, po pewnym czasie, gdyż proces optymalizacyjny jest przebiega na wielu wariantach trzech parametrów α, β, γ uzyskujemy rozwiązanie zaprezentowane na rysunku 26.11.

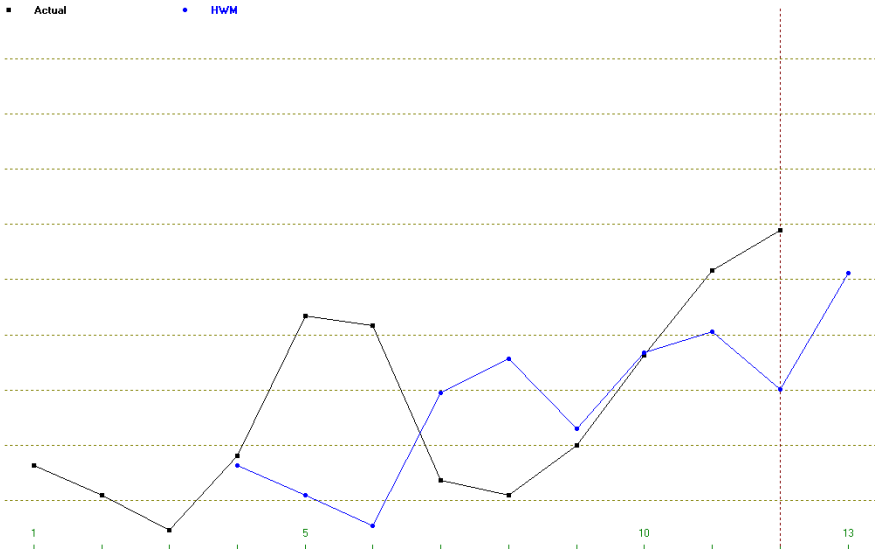
03-01-2016 Month	Actual Data	Forecast by HWM	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-square
1	18								
2	12								
3	5								
4	20	18	2	2	2	4	10	1	
5	48	12,08	35,92	37,92	18,96	647,1231	42,41666	2	
6	46	5,948	40,052	77,972	25,99067	966,1364	57,30096	3	
7	15	32,52462	-17,52462	60,44738	23,87416	801,3804	72,18343	2,531917	
8	12	39,38581	-27,38581	33,06157	24,57649	791,1008	103,3898	1,345252	0,8269053
9	22	25,4188	-3,418797	29,64277	21,0502	661,1987	88,74813	1,408194	0,7574807
10	40	40,53912	-0,5391235	29,10365	18,12005	566,7833	76,26237	1,606157	0,869369
11	57	44,76883	12,23117	41,33482	17,38394	514,6356	69,41184	2,377759	0,7938924
12	65	33,30332	31,69668	73,0315	18,97425	569,0849	67,11765	3,848981	0,6827483
13		56,52769							
CFE		73,0315							
MAD		18,97425							
MSE		569,0849							
MAPE		67,11765							
Trk.Signal		3,848981							
R-square		0,6827483							
		c=3							
		Alpha=0,03							
		Beta=1							
		Gamma=0,23							
		F(0)=11,6667							
		T(0)=0							
		S(1)=1,542857							
		S(2)=1,028571							
		S(3)=0,4285714							

Źródło: Opracowanie własne w programie *WinQSB*.

Rys. 26.11. Określenie prognozy na okres 13 (1’2014) oraz informacje dodatkowe generowane przez moduł FC

Określony przez moduł FC sumaryczny błąd prognozy RMSE jako pierwiastek z MSE równy 569,0849 wynosi 23,86. Optymalne parametry sa następujące: α = 0,03, β = 1, γ = 0.23. Wartość początkowa prognozy F(0) potrzebna w procedurze obliczania komputerowego jest średnią z trzech pierwszych obserwacji (18 + 12 + 5)/3 = 11,6667. Wygasła prognoza *Forecast by HWM* na okres 4. (04’2013) została przyjeta przez FC jako równa pierwszemu okresowi tj. 18. Mimo precyzyjnego doboru parametrów wartośc teoretyczna prognozy (około 57) wyraźnie odstaje od nagłego zainteresowania internautów omawianym wpisem na moim blogu w miesiącu 1’2014, gdzie liczba rzeczywista wejść wynosiła 165 (zob. tabela 26.1). Tak więc wobec akcyjności korzystających z wejścia - głównie studentów w okresie sesji egzaminacyjnej, zaprezentowany model może być stosowany dylko do badania zaległych prognoz lub powinniśmy przystąpić do możliwie najlepszego doboru cyklu obserwacji np. 5. okresów, co jednak wymaga dłuższego horyzontu badawczego danych

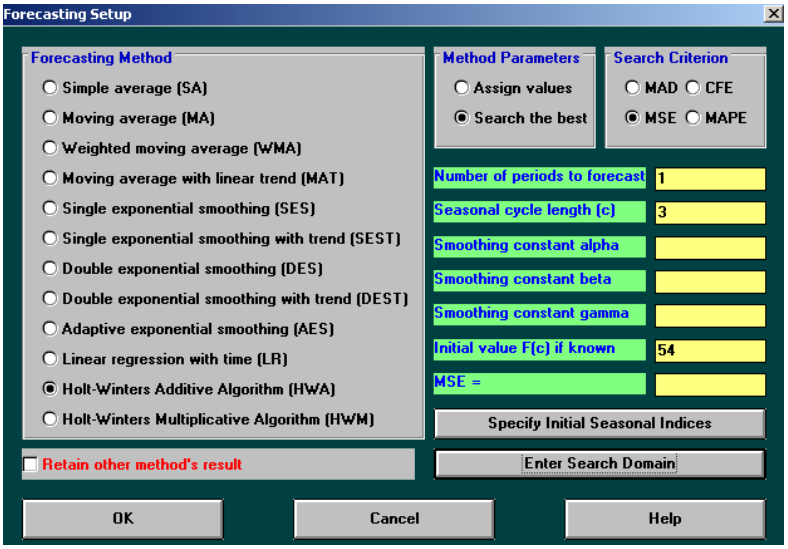
statystycznych. Zobaczmy jeszcze obraz graficzny kształtowania się danych empirycznych (*Actual*) i teoretycznych - modelowych (HWM) pokazany na rysunku 26.12.



Źródło: Opracowanie własne modułem FC programu *WinQSB*.

Rys. 26.12. Porównanie danych rzeczywistych i modelowych

Pouczeni wcześniejszym badaniem, przyjmijmy teraz do testowania „*sklejone*" szeregi danych dwóch okresów akademickich, tj miesiące 10'2013 do 06'2014 oraz 10'2014 do 06'2015. Zastosujmy model *Holt-WintersAdditive Algorithm* (HWA) i średnią jako *F*(0) z liczb wejść dla miesięcy 10'2013, 11'2013 i 12'2013: (40 + 57 + 65)/3 = 54. Pozostawiamy dobór opcji tego modelu taki jak dla HWM (zob. rysunek 26.13).



Źródło: Opracowanie własne modułem FC programu *WinQSB*.

Rys. 26.13. Opcje modelu HWA

Na rycinie 26.13 pokazano rezultat pracy modułu FC w modelu HWA. Zastosowanie modelu HWA przy cyklu trzech miesięcy okazało się bardzo niekorzystne, gdyż znacznie wzrosł MSE z którego pierwiastek, czyli współczynnik sumarycznego błędu prognoz RMSE = 49,6. Sugeruje to nam potrzebę przyjęcia innej długości cyklu sezonowości np. semestr, czyli 5 miesięcy.

03-01-2016 Month	Actual Data	Forecast by HWA	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal	R-square
1	40,0000								
2	57,0000								
3	65,0000								
4	165,0000	40,0000	125,0000	125,0000	125,0000	15625,0000	75,7576	1,0000	
5	42,0000	60,8750	-18,8750	106,1250	71,9375	7990,6330	60,3490	1,4752	0,7732
6	35,0000	69,6649	-34,6649	71,4601	59,5133	5727,6400	73,2468	1,2007	0,2028
7	65,0000	45,9826	19,0174	90,4775	49,3893	4386,1450	62,2495	1,8319	0,2390
8	77,0000	62,9483	14,0517	104,5292	42,3218	3548,4060	53,4494	2,4699	0,2575
9	63,0000	72,2244	-9,2244	95,3049	36,8056	2971,1870	46,9815	2,5894	0,2131
10	29,0000	49,8393	-20,8393	74,4655	34,5247	2608,7710	50,5355	2,1569	0,1324
11	147,0000	65,7830	81,2170	155,6825	40,3612	3107,2000	51,1248	3,8572	0,2187
12	104,0000	76,7370	27,2630	182,9455	38,9058	2844,5410	48,3570	4,7023	0,2639
13	150,0000	55,9319	94,0681	277,0135	44,4221	3444,9670	49,7925	6,2359	0,3862
14	50,0000	77,4020	-27,4020	249,6116	42,8748	3200,0480	50,2481	5,8219	0,2950
15	80,0000	86,6880	-6,6880	242,9235	39,8592	2937,1050	46,7574	6,0945	0,2879
16	33,0000	66,5184	-33,5184	209,4052	39,3715	2797,5960	50,9738	5,3187	0,2051
17	76,0000	83,5026	-7,5026	201,9026	37,0951	2601,7880	48,0380	5,4428	0,2011
18	71,0000	92,8648	-21,8648	180,0377	36,0798	2460,2070	46,8885	4,9900	0,1991
19		71,4370							
CFE		180,0377							
MAD		36,0798							
MSE		2460,2070							
MAPE		46,8885							
Trk.Signal		4,9900							
R-square		0,1991							
		c=3							
		Alpha=0,02							
		Beta=0,55							
		Gamma=0,01							
		F(0)=54							
		T(0)=0							
		S(1)=-14,0000							
		S(2)=3,0000							
		S(3)=11,0000							

Źródło: Opracowanie własne modułem FC programu *WinQSB*.

Rys. 26.14. Opcje modelu HWA dla horyzontu sklejonych dwóch okresów akademickich, tj. 18 miesięcy

26.4. Propozycja zastosowania modelu podwójnego wyrównywania wykładniczego z trendem

Podwójne wyrównywanie wykładnicze polega na zastosowaniu dwóch procedur wyrównywania przy tym samym parametrze wyrównujacym α. Oblicza się różnice między pojedynczo i podwójnie wygladzonymi wartościami zmiennej w celu estymacji trendu[213]. Szukając metody postępowania w modelowaniu rocznych szeregów obserwacji wejść do wpisu internetowego zawartego w konkretnym blogu weźmy pod uwagę osobno trzy szeregi czasowe lat 2013, 2014, 2015. Skorzystajmy teraz z metody DEST dla roku 2013 (zob. rysunek 26.14).

03-01-2016 Month	Actual Data	Forecast by DEST	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal
1	18,00							
2	12,00	18,00	-6,00	-6,00	6,00	36,00	50,00	-1,00
3	5,00	6,12	-1,12	-7,12	3,56	18,63	36,20	-2,00
4	20,00	-1,98	21,98	14,86	9,70	173,43	60,76	1,53
5	48,00	34,56	13,44	28,30	10,63	175,23	52,57	2,66
6	46,00	75,73	-29,73	-1,44	14,45	317,00	54,99	-0,10
7	15,00	44,60	-29,60	-31,03	16,98	410,15	78,71	-1,83
8	12,00	-15,41	27,41	-3,62	18,47	458,90	100,09	-0,20
9	22,00	8,45	13,55	9,93	17,85	424,49	95,28	0,56
10	40,00	31,73	8,27	18,20	16,79	384,92	86,99	1,08
11	57,00	57,84	-0,84	17,36	15,19	346,50	78,44	1,14
12	65,00	74,02	-9,02	8,35	14,63	322,39	72,57	0,57
13		73,18						
CFE		8,35						
MAD		14,63						
MSE		322,39						
MAPE		72,57						
Trk.Signal		0,57						
R-square		2,17						
		Alpha=0,99						
		F(0)=18						
		F'(0)=18						

Źródło: Opracowanie własne modułem FC programu *WinQSB*.

Rys. 26.14. Określenie prognoz wygasłych i prognozy na styczeń kolejnego roku 2014

[213] Szersze omówienie postaci zastosowanych równań występuje w mojej książce: *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, część II., Wydawnictwo Instytut Śląski, Opole 2015.

Widzimy dla wygasłych prognoz niski MSE = 322,29, którego pierwiastek, czyli współczynnik sumarycznego błedu wynosi 18. Nie kierujemy się na razie prognozą na styczeń 2014 roku, która wyraźnie odstaje od danej rzeczywistej (165). W kolejności określamy również modelem DEST wygasłe prognozy dla roku 2014 (zob. rysunek 26.15).

03-01-2016 Month	Actual Data	Forecast by DEST	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal
1	165,00							
2	42,00	165,00	-123,00	-123,00	123,00	15129,00	292,86	-1,00
3	35,00	64,14	-29,14	-152,14	76,07	7989,07	188,06	-2,00
4	65,00	19,57	45,43	-106,71	65,86	6014,04	148,67	-1,62
5	77,00	31,25	45,75	-60,96	60,83	5033,85	126,36	-1,00
6	63,00	50,83	12,17	-48,78	51,10	4056,72	104,95	-0,95
7	11,00	50,56	-39,56	-88,35	49,18	3641,46	147,40	-1,80
8	26,00	9,92	16,08	-72,27	44,45	3158,18	135,18	-1,63
9	22,00	8,25	13,75	-58,52	40,61	2787,03	126,09	-1,44
10	29,00	7,38	21,62	-36,90	38,50	2529,30	120,36	-0,96
11	147,00	15,27	131,73	94,83	47,82	4011,64	117,29	1,98
12	104,00	117,09	-13,09	81,74	44,67	3662,51	107,77	1,83
13		122,30						
CFE		81,74						
MAD		44,67						
MSE		3662,51						
MAPE		107,77						
Trk.Signal		1,83						
R-square		1,58						
		Alpha=0,41						
		F(0)=165						
		F'(0)=165						

Źródło: Opracowanie własne modułem FC programu *WinQSB*.

Rys. 26.15. Określenie prognoz wygasłych i prognozy na styczeń kolejnego roku 2015

Z powodu dużego skoku liczy wejść w styczniu 2014 roku obserwujemy dla roku 2014 znaczący wzrost MSE, zobaczmy jeszcze sytuację dla bardziej porównywalnego roku 2015 (rys. 26.16).

03-01-2016 Month	Actual Data	Forecast by DEST	Forecast Error	CFE	MAD	MSE	MAPE (%)	Tracking Signal
1	150,00							
2	50,00	150,00	-100,00	-100,00	100,00	10000,00	200,00	-1,00
3	80,00	66,00	14,00	-86,00	57,00	5098,00	108,75	-1,51
4	33,00	60,12	-27,12	-113,12	47,04	3643,83	99,89	-2,40
5	76,00	22,17	53,83	-59,29	48,74	3457,32	92,63	-1,22
6	71,00	47,43	23,57	-35,72	43,70	2876,94	80,74	-0,82
7	22,00	56,77	-34,77	-70,49	42,21	2598,95	93,63	-1,67
8	19,00	21,26	-2,26	-72,75	36,51	2228,40	81,95	-1,99
9	35,00	6,93	28,07	-44,68	35,45	2048,36	81,73	-1,26
10	39,00	17,67	21,33	-23,36	33,88	1871,30	78,73	-0,69
11	85,00	27,71	57,29	33,94	36,22	2012,43	77,60	0,94
12	93,00	71,71	21,29	55,23	34,87	1870,67	72,62	1,58
13		95,58						
CFE		55,23						
MAD		34,87						
MSE		1870,67						
MAPE		72,62						
Trk.Signal		1,58						
R-square		2,24						
		Alpha=0,42						
		F(0)=150						
		F'(0)=150						

Źródło: Opracowanie własne modułem FC programu *WinQSB*.

Rys. 26.16. Określenie prognoz wygasłych i prognozy na styczeń kolejnego roku 2016

26.5. Próba wyłonienia procedury wynikowej

Już z pobieżnie pokazanych testów zastosowania różnych modeli prognozowania (w oparciu o szeregi czasowe do modelowania ruchu internetowego na przykładzie liczby wejść do danego wpisu) zauważamy dużą trudność w formalizacji tego zjawiska. Wiele firm pozycjonowania stron WWW jest zainteresowanych opracowaniem lub pozyskaniem formuły przewidywania. Zaprezentowane przykłady niech będą przyczynkiem do rozwiązania modelowania ruchu internetowego. Wydaje się jednak za celowe stosowanie następującego postępowania bazujacego na:

- szeregach pełzajacych z trzech ostatnich lat,

- określaniu wstępnym prognoz na styczeń kolejnych trzech lat,

- określeniu prognozy wynikowej na podstawie tych prognoz cząstkowych metodą LR (*Linear regression with time*),

- uwzględnieniu współczynnika trendu z całego 36 miesiecznego okresu trzech lat, w naszym przypadu równego 1,2192 (zob. rysunek 26.5).

Po dopisaniu liczby wejść kolejnego miesiąca szeregi ulegają pełzaniu - przesunięciu o jeden okres. Obserwacja danych za styczeń 165, 150, 183 odpowiednio lat 2014, 2015, 2016 w miarę stabilnych danych statystycznych daje nadzieję na stopniowe doskonalenie modelowania ekonometrycznego. Rok 2013, który przyjęto do badania z racji braku szerszego horyzontu obserwacji wyraźnie jednak odstawał od obecnej tendencji liczby wejść do wpisu internetowego. Był to początek wejścia z tematem i dlatego w wyniku pełzania okresów po zakończeniu roku 2015 okres roku 2013 będzie pominięty. Spróbujmy teraz wykonać wymienione wcześniej czynności postepowania i porównać z daną rzeczywistą za miesiąc 01’2015, która wynosi 183 wejść.

F_1	F_2	F_3
73,18	122,30	95.58

Po skorzystaniu z komputerowej metody LR uzyskujemy prognozę wynikowa równą 119,42. Po uwzględnieniu współczynnika trendu liniowego 1,2192 dla całej zbiorowości 36. obserwacji skorygowana prognoza F = 1,2192 * 119,42 = 145,6, czyli około 146 wejść.

Gdy porównamy z daną rzeczywistą - 183 (zob. rysunek 26.3) to procentowy błąd oszacowania wynosi: [(183 - 146)/183] * 100 = 20,2%. Odchylenie to przekracza przyjmowane przeważnie 10% jako granica błędu. Zważajac jednak na rok 2013 oraz wyjatkowo nieprzewidywalną „*materię*” jaką jest prognozowanie liczby wejść do określonego wpisu przez internautów, oceniam podana procedure jako przydatną do celów praktycznych pozycjonowania stron WWW.

27. Określenie modelu PKB

27.1. Sformatowanie pliku wejściowego

Dokonano przeglądu środowisk programowych w których możliwe jest uzyskanie miar charakterystycznych statystyki opisowej. Jednak w tym miejscu ograniczono się do[214]:

- zastosowania funkcji elementarnych arkusza kalkulacyjnego *Excel*,
- narzędzi programowych *Excela* dla otrzymywania statystyki opisowej zbiorczej,
- programu *GRETL*, dostępnego w Internecie.

Dla prezentacji wykorzystane zostaną dane empiryczne, stanowiące zmienne modelu produktu brutto dla województw, uzyskanego na etapie obliczeń efektywności technicznej tych obiektów[215]. Dane są typu panelowego i zebrane zostały na podstawie roczników statystycznych województw z lat 2000-2008. Następujące zmienne dobrano intuicyjnie biorąc pod uwagę cechy charakterystyczne dla ochrony środowiska oraz ich rolę jako efektu czy też nakładu. Dla ułatwienia przytoczono tu procedurę przygotowania danych.

Efekty:

X_1 - produkt krajowy brutto (ceny bieżące) na 1 mieszkańca w zł,

oraz dalsze zmienne wchodzące do wzoru na efektywność techniczną w postaci $1/X_i$:

X_2 - emisja pyłowych przemysłowych zanieczyszczeń powietrza z zakładów szczególnie uciążliwych dla czystości powietrza w tys. t.,

X_3 - emisja gazowych (bez dwutlenku węgla) przemysłowych zanieczyszczeń powietrza z zakładów szczególnie uciążliwych dla czystości powietrza w tys. t.,

X_4 - grunty zdewastowane i zdegradowane wymagające rekultywacji i zagospodarowania (stan w dniu 31 XII) w ha,

X_5 - odpady wytworzone (w ciągu roku z wyłączeniem odpadów komunalnych) w tys. t.,

X_6 - ścieki przemysłowe i komunalne odprowadzone do wód lub do ziemi w hm^3,

X_7 - przestępstwa stwierdzone w zakończonych postępowaniach przygotowawczych w tys.,

X_8 - stopa bezrobocia rejestrowanego (stan w dniu 31 XII) w %.

Nakłady:

X_9 - pracujący w przemyśle (stan w dniu 31 XII) w tys.;

X_{10} - pracujący w budownictwie (stan w dniu 31 XII) w tys.,

X_{11} - pracujący w transporcie, gospodarce magazynowej i łączności (stan w dniu 31 XII) w tys.;

X_{12} - przeciętne miesięczne wynagrodzenie w przemyśle brutto w zł;

X_{13} - przeciętne miesięczne wynagrodzenie w budownictwie brutto w zł;

X_{14} - przeciętne miesięczne wynagrodzenie w transporcie, gospodarce magazynowej i łączności brutto w zł;

X_{15} - nakłady na działalność badawczą i rozwojową (ceny bieżące) w mln zł;

X_{16} - użytki rolne (stan w czerwcu) w tys. Ha;

X_{17} - pozyskanie drewna (grubizny) w dekametrach sześciennych;

X_{18} - produkcja sprzedana przemysłu (ceny bieżące) w mln zł;

X_{19} - produkcja budowlano-montażowa (ceny bieżące) w mln zł;

X_{20} - nakłady inwestycyjne w przemyśle (ceny bieżące) w mln zł;

X_{21} - nakłady inwestycyjne w budownictwie (ceny bieżące) w mln zł;

X_{22} - nakłady inwestycyjne w transporcie, gospodarce magazynowej i łączności (ceny bieżące) w mln zł;

X_{23} - nakłady inwestycyjne na ochronę środowiska w mln zł.

Program *GRETL* korzysta z zaimportowanych danych z *Excela* (z przecinkiem) ułożonych według 16. województw oraz 9 okresów czasowych (lata 1999-2007). Ze względu na obszerność danych panelowych (obiektowo-czasowych) wszystkie szeregi danych wczytano do kolumn arkusza

[214] Informacja dotyczy okresu opracowania niniejszej pracy przez autora.

[215] Duczmal M., *Efektywność działania obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski, Opole 2013, tab. 3.2.

kalkulacyjnego *Excel*, a dopiero później sporządzono odpowiednie pliki wejściowe do *GRETL'a*. Arkusz kalkulacyjny danych panelowych obejmuje macierz o 144 wierszach oraz 23 kolumnach (zmiennych). Zestawienie wartości zmiennych *X*1-*X*23 znajduje się w tabeli 3.2 cytowanej publikacji[216]. W tabeli 3.2 dla każdego z szesnastu województw obliczono wartość średnią zmiennej *X*1:

1 - łódzkie,
2 - mazowieckie,
3 - małopolskie,
4 - śląskie,
5 - lubelskie,
6 - podkarpackie,
7 - podlaskie,
8 - świętokrzyskie,
9 - lubuskie,
10 - wielkopolskie,
11 - zachodniopomorskie,
12 - dolnośląskie,
13 - opolskie,
14 - kujawsko-pomorskie,
15 - pomorskie,
16 - warmińsko-mazurskie.

Procedura postępowania w zakresie przeniesienia danych statystycznych z arkusza kalkulacyjnego *Excel* do bazy danych użytkownika w programie *GRETL* sprowadza się do etapów:

1. Skopiowanie wcześniej wprowadzonych manualnie danych na wyodrębniony skoroszyt o jednym arkuszu nadając nazwę przykładowo *GRETL.xls*. Rozpoczynamy od komórki A1 wpisując w pierwszym wierszu nazwy pól kolejno: *W* (symbol liczbowy województwa), *T* (symbol cyfrowy kolejnych okresów - lat 1999-2007), zmienne *X*1-*X*23, *n* (numer obserwacji od 1 -144). Cyfra 144 stanowi iloczyn ilości obiektów - 16 (województw) i 9 okresów.

2. Wywołanie programu *GRETL*.

3. Pobranie pliku *GRETL.xls* poprzez polecenie: *Plik/Otwórz dane/Import/Excel...*
Wskazanie pliku w podkatalogu np. *D:/MODELE-P-S*, a następnie polecenie: *Otwórz*

4. Podanie w oknie dialogowym początku zakresu importu: *Kolumna* 1 *wiersz* 1 *Arkusz* 1/*OK*
Pojawia się informacja:

Importowane dane są traktowane jako niedatowane (dane przekrojowe).
Czy chcesz interpretować dane jako szereg czasowy lub dane panelowe? Tak

5. Okno *Kreator struktury danych* Wybieramy *Dane panelowe/Do przodu*

6. Okno *Kreator struktury danych panelowych* Wybieramy *Stos szeregów czasowych/Do przodu*

7. Okno kreator struktury danych:
Liczba przekrojów (obiektów) w jednostce czasu - 16
Liczba jednostek czasu w przekroju - 9 *Do* przo*du*

7. Okno *Potwierdź strukturę danych*:
Dane panelowe (stos szeregów czasowych) 16 jednostek przekroju w 9 okresach/zastosuj.

Powstał zbiór *GRETL.xls** o strukturze:

0	const	1	*W*	2	*T*	3	*X*1	....	25	*X*23	26	*n*

[216] Ibidem.

Program przeniósł wyróżnienia województw *W*, okresy *T*, wartości zmiennych *X*1-*X*23 oraz *n* dodając numery obserwacji jako 1.1 - 16.9. Przykładowy fragment zaimportowanych danych przez *GRETL* dla województwa łódzkiego i trzech zmiennych podano w tabeli 27.1.

Tab. 27.1.

Dane dla województwa łódzkiego w zakresie zmiennych X1-X3 oraz obserwacji 1:1 -1:9

Obs.	W	*T*	*X*1	*X*2	*X*3
1:1	1	1	12682	14,5752	380,7771
1:2	1	2	14497	12,1000	327,9000
1:3	1	3	15723	10,3000	298,0000
1:4	1	4	17535	7,1000	264,3000
1:5	1	5	18492	7,7000	232,2000
1:6	1	6	19498	7,4000	230,2000
1:7	1	7	22274	7,9000	220,8000
1:8	1	8	23666	7,2000	201,1000
1:9	1	9	25521	7,8000	180,8000

Źródło: Opracowanie własne na podstawie danych *GUS*-u.

Następnie skopiowano zaimportowany plik danych *GRETL.xls** jako plik *GRETL*-a pod nazwą *GRETL.gdt*, który posłużył do estymacji parametrów modelu ekonometrycznego[217].

27.2. Charakterystyka zmiennych

GRETL umożliwia szybkie otrzymywanie statystyki opisowej w odniesieniu do zmiennych *X*1-*X*23 w ujęciu panelowym, a w ramach niej określenie:

- średniej,
- mediany,
- wartości minimalnej oraz maksymalnej (*Min*, *Max*),
- odchylenia standardowego (*s*),
- współczynnika zmienności (*V*),
- współczynnik skośności (*S*) i współczynnik kurtozy *K*.

Informacje te podano w tabeli 27.2. stanowiącej raport z *GRETL*-a dla 144 (1.1-16.9) obserwacji próby danych panelowych.

Tab. 27.2.

Statystyka opisowa dla zmiennych X1-X23 przy wykorzystaniu 1:1 - 16:9

X_i	*Średnia*	*Mediana*	*Min*	*Max*
*X*1	19144	18360	10383	44381
*X*2	8,6977	6,9960	1,4000	43,029
*X*3	126,90	64,800	9,5000	754,10
*X*4	4280,4	3641,0	1234,0	10852
*X*5	7712,1	3379,3	399,20	49038
*X*6	577,44	197,60	36,000	2745,6
*X*7	83,875	71,250	26,700	208,60
*X*8	17,578	16,750	7,8000	30,300
*X*9	188,85	159,95	58,600	600,04
*X*10	46,337	33,700	11,000	367,26
*X*11	46,366	36,912	16,200	152,00
*X*12	2079,5	2010,4	1490,0	3230,7
*X*13	1731,0	1663,6	1305,0	3042,3
*X*14	2182,6	2131,4	1570,2	3424,9
*X*15	323,48	145,05	12,700	2742,3

[217] Ibidem, s. 87.

*X*16	1061,2	1012,6	457,90	2419,3
*X*17	1775,8	1706,7	688,30	3899,7
*X*18	38943	24888	7463,3	1,8165E+05
*X*19	4927,1	3197,1	1159,0	23230
*X*20	2716,3	1905,0	558,40	11999
*X*21	687,52	486,03	76,568	3270,5
*X*22	872,13	432,16	134,00	7798,8
*X*23	399,44	318,64	86,545	1993,7
X_i	*s*	*V*	*S*	*K*
*X*1	5766,8	0,30124	1,3310	2,9203
*X*2	7,0371	0,80908	2,2147	6,2462
*X*3	159,48	1,2568	2,4567	5,9287
*X*4	2177,9	0,50882	1,5628	2,2389
*X*5	12180	1,5793	2,2257	3,4764
*X*6	713,65	1,2359	1,3893	0,71205
*X*7	48,629	0,57978	1,0088	0,13337
*X*8	4,8636	0,27669	0,49439	-0,19435
*X*9	121,56	0,64368	1,2253	0,92177
*X*10	41,091	0,88678	3,8767	24,570
*X*11	31,360	0,67637	1,6441	2,2576
*X*12	358,65	0,17247	0,89701	0,62085
*X*13	261,10	0,15084	1,8483	5,0775
*X*14	309,06	0,14160	1,0880	2,1168
*X*15	526,09	1,6263	3,1145	9,1345
*X*16	474,79	0,44739	1,0554	0,58792
*X*17	751,39	0,42314	0,61059	-0,40331
*X*18	36224	0,93019	1,9042	3,1951
*X*19	4333,9	0,87961	1,9968	3,9843
*X*20	2095,9	0,77162	1,4636	2,1660
*X*21	642,22	0,93412	1,9280	3,7895
*X*22	1425,7	1,6347	3,4583	11,378
*X*23	307,01	0,76860	2,1417	6,4796

Źródło: Opracowanie własne w *GRETL*.

Z tabeli 27.2. wynika, że dla 23 zmiennych *X*1-*X*23 współczynnik zmienności przekracza założoną wartość krytyczną *V** = 10%, czyli zmienne te wstępnie wchodzą do modelu ekonometrycznego.

27.3. Macierz korelacji

Kryterium wstępne według współczynnika zmienności nie pozwoliło nam na głębszą weryfikację zmiennych do budowy modelu ekonometrycznego. Posłużmy się więc analizą współczynników korelacji. W tym celu przyjmijmy zmienną *X*1 jako zmienną objaśnianą, a pozostałe zmienne *X*2-*X*23 jako objaśniające. Program *GRETL* umożliwia uzyskanie macierzy korelacji na podstawie zaimportowanych danych tj. pliku *GRETL.gdt*. Procedura w tym zakresie jest następująca:

Plik/Otwórz dane/Pliki użytkownika ..

lub bezpośrednio wskazanie pliku założonej bazy danych *GRETL.gdt* wywołując menu:

Dane/Zaznacz wszystko; Widok/Macierz korelacji.

Współczynniki korelacji liniowej uzyskujemy dla próby 1-114 obserwacji szeregu danych wg: *GRETL* podaje nam tą wartość krytyczną, która obliczana jest dla danych panelowych przy dwustronnym 5% obszarze krytycznym i dla n = 144. W podanym wcześnie wzorze: t_α = 1,9759 -

$$r^* = \sqrt{\frac{t_\alpha^{\,2}}{t_\alpha^{\,2} + n - 2}} \qquad r^* = \sqrt{\frac{(1{,}9759)^2}{(1{,}9759)^2 + 144 - 2}} \approx 0{,}1637$$

statystyka wyznaczona na podstawie rozkładu *t*-Studenta dla poziomu istotności α = 0,05 oraz ilości stopni swobody $k = iss = n - 2 = 144 - 2 = 142$. Dla wyznaczenia wartości krytycznej przyjęto wartość istniejącą w tablicy k = 144. Odpowiedzią programu *GRETL* jest raport zamieszczony w tabeli 27.3.

Tab. 27.3

Współczynniki korelacji liniowej między zmiennymi X1-X23

*X*1	*X*2	*X*3	*X*4	*X*5	
1,0000	0,1298	0,3336	0,2072	0,2435	*X*1
	1,0000	0,8315	0,4083	0,7959	*X*2
		1,0000	0,3127	0,7383	*X*3
			1,0000	0,3761	*X*4
				1,0000	*X5*
X6	*X7*	*X*8	*X*9	*X*10	
0,5130	0,5927	-0,2680	0,4790	0,4593	*X*1
0,2018	0,7155	-0,3236	0,8390	0,6100	*X*2
0,1903	0,7189	-0,3672	0,8664	0,5542	*X*3
0,3478	0,4427	-0,1644	0,5655	0,3709	*X*4
-0,0348	0,6298	-0,1681	0,6984	0,4112	*X*5
1,0000	0,5033	-0,2623	0,4081	0,5455	*X*6
	1,0000	-0,3008	0,8894	0,7878	*X*7
		1,0000	-0,4592	-0,4572	*X*8
			1,0000	0,7666	*X*9
				1,0000	*X*10
*X*11	*X*12	*X*13	*X*14	*X*15	
0,6408	0,8701	0,8989	0,8007	0,6779	*X*1
0,6484	0,2751	0,0566	0,0010	0,2954	*X*2
0,6409	0,4830	0,1986	0,1328	0,3160	*X*3
0,3861	0,1482	0,1597	0,1248	0,1800	*X*4
0,4828	0,5133	0,1515	0,0838	0,1188	*X*5
0,6403	0,3050	0,4966	0,5062	0,6837	*X*6
0,9333	0,5981	0,5250	0,5693	0,7319	*X*7
-0,4626	-0,3157	-0,4152	-0,0595	-0,4223	*X*8
0,8621	0,5226	0,4004	0,3194	0,5658	*X*9
0,8497	0,4431	0,4912	0,4604	0,7477	*X*10
1,0000	0,5835	0,6372	0,5905	0,8751	*X*11
	1,0000	0,8326	0,6852	0,5203	*X*12
		1,0000	0,8062	0,7435	*X*13
			1,0000	*0,6984*	*X14*
				1,0000	*X*15
*X*16	*X*17	*X*18	*X*19	*X*20	
0,2852	0,2382	0,7944	0,7785	0,6866	*X*1
0,0417	-0,2618	0,5628	0,4992	0,6631	*X*2
-0,0400	-0,2290	0,7100	0,5623	0,7693	*X*3
0,4949	0,2257	0,3944	0,3644	0,5177	*X*4
-0,1791	-0,0212	0,5168	0,4009	0,6840	*X*5
0,5974	0,1522	0,5588	0,6647	0,4311	*X*6
0,3872	0,0048	0,8777	0,8551	0,8433	*X*7
-0,2771	0,3938	-0,4504	-0,5000	-0,4838	*X*8
0,2917	-0,0607	0,8496	0,7843	0,9060	*X*9
0,4728	-0,0732	0,7406	0,7885	0,7198	*X*10

0,5301	-0,0176	0,9161	0,9486	0,8458 *X*11
0,0409	0,1121	0,7700	0,7003	0,7223 *X*12
0,4066	0,1623	0,7473	0,7888	0,6200 *X*13
0,4166	0,3287	0,6448	0,6933	0,4693 *X*14
0,6364	-0,0636	0,7968	0,9017	0,6316 *X*15
1,0000	0,1290	0,3880	0,5213	0,2905 *X*16
	1,0000	0,0332	0,0576	0,0555 *X*17
		1,0000	0,9438	0,9139 *X*18
			1,0000	0,8649 *X*19
				1,0000 *X*20

*X*21	*X*22	*X*23	
0,7361	0,6230	0,4635	X*1*
0,3074	0,2986	0,7179	X*2*
0,3896	0,2753	0,7008	X*3*
0,3802	0,1796	0,4367	X*4*
0,3009	0,1087	0,5981	X*5*
0,5551	0,6707	0,4448	X*6*
0,6916	0,6931	0,7746	X*7*
-0,5092	-0,3860	-0,4696	*X*8
0,6468	0,5411	0,8574	*X*9
0,7027	0,7885	0,7585	*X*10
0,7786	0,8566	0,8114	*X*11
0,6686	0,4660	0,4843	*X*12
0,7690	0,7003	0,4146	*X*13
0,6575	0,6635	0,3311	*X*14
0,7732	0,9601	0,5936	*X*15
0,4927	0,6485	0,2999	*X*16
0,1319	-0,0212	-0,0222	*X*17
0,7937	0,7438	0,7547	*X*18
0,8792	0,8662	0,7715	*X*19
0,7809	0,5899	0,8544	*X*20
1,0000	0,7575	0,6989	*X*21
	1,0000	0,5919	*X*22
		1,0000	*X*23

Źródło: Opracowanie własne w *GRETL*.

Postępowanie dla wyboru zmiennych objaśniających jest następujące:

a) sprawdzamy, czy $|r_j| \le r^*$ i eliminujemy zmienne X_j nieistotnie skorelowane z $Y = X1$:

Kierując się tym kryterium i przeglądając w powyższej macierzy korelacji pary relacji zmiennej objaśnianej *X*1 względem

$$|r_j| \le r^* = 0{,}1637$$

odpowiadających zmiennych objaśniających *X*2-*X*23 możemy wyeliminować tylko zmienną *X*2.

b) z pozostałych 21 potencjalnych zmiennych objaśniających X3-X23 wybieramy zmienną X_h wg kryterium:

$$r_h = \max_j (r_j)$$

Zmienną najsilniej skorelowaną ze zmienną *X*1 jest zmienna: *X*13 - przeciętne miesięczne wynagrodzenie w budownictwie brutto w zł dla której $r_{13} = 0{,}8989$.

c) badamy relacje powyższej zmiennej z pozostałymi zmiennymi i eliminujemy te dla których $|r_{hi}| > r^*$, gdyż są one zbyt silnie skorelowane ze zmienną X_h. Pozostały zmienne *X*4 oraz *X*5.

d) Wracamy do punktu b) i wybieramy zmienną *X*5 o współczynniku korelacji względem *X*1 równym 0,2435.

e) Realizując ponownie etap wg c) badamy relację zmiennej *X*5 ze zmienną *X*4. Ponieważ współczynnik korelacji równy 0,3761 > 0,1637 pomijamy zmienną *X*4 w modelu.

Zatem wstępnie wg analizy współczynników korelacji liniowy model ekonometryczny ma postać:

$$\hat{X}_1 = a_0 + a_1 X_{13} + a_2 X_5$$

A teraz zobaczymy jaki model zaproponuje *GRETL* stosując *eliminację a posteriori*, czyli stopniowe eliminowanie określonej zmiennej kierując się testem *t*-Studenta statystycznej istotności parametrów.

27.4. Określenie modelu ekonometrycznego

Parametry liniowego modelu ekonometrycznego możemy oszacować stosując klasyczną metodę najmniejszych kwadratów *KMNK* w *GRETL*. Skorzystajmy z pliku *GRETL.gdt* wywołując menu: *Dane/Zaznacz wszystko* ; *Model/Klasyczna metoda najmniejszych kwadratów*.

W oknie ze specyfikacją modelu definiujemy jako zmienną zależną (objaśnianą) *X*1 oraz określamy wstępnie zmienne niezależne (objaśniające) *X*2-*X*23. W rezultacie *eliminacji a posteriori* (16 kroków) otrzymujemy z *GRETL* raport z parametrami modelu regresji liniowej - model 16, zwane współczynnikami oraz statystykę opisową modelu (patrz tabela 27.4.). Do estymacji parametrów modelu zastosowano *KMNK* na bazie danych panelowych obejmujących 144 obserwacje. Do tego panelu włączono 16 jednostek danych przekrojowych (województw) oraz szeregi czasowe długości = 9. Zmienną zależną jest *X*1.

Tab. 27.4.

Parametry oraz statystyka opisowa modelu 16

Zmienna	Współczynnik	Błąd standardowy	*t*-Studenta	Wartość *p*	
const	-14175,2	1722,40	-8,230	1,41E-013	***
***X*5**	-0,124107	0,0246434	-5,036	1,49E-06	***
*X*6	0,685282	0,288771	2,373	0,0190	**
*X*8	188,507	36,3130	5,191	7,51E-07	***
*X*10	-14,1471	5,89481	-2,400	0,0178	**
*X*12	7,70218	1,11035	6,937	1,52E-010	***
***X*13**	6,94873	1,51385	4,590	1,00E-05	***
*X*18	0,0238954	0,0132364	**1,805**	0,0733	*
*X*20	0,827128	0,226411	3,653	0,0004	***

Średnia arytmetyczna zmiennej zależnej = 19143,6
Odchylenie standardowe zmiennej zależnej = 5766,83
Suma kwadratów reszt = 3,99299e+008
Błąd standardowy reszt = 1719,82
Współczynnik determinacji R^2 = 0,91604
Skorygowany współczynnik determinacji R^2 = 0,91106
Statystyka F (8, 135) = 184,106 (wartość p < 0,00001)
Logarytm wiarygodności = -1272,48
Kryterium informacyjne Akaike'a *AIC* = 2562,95

Źródło: Opracowanie własne programem *GRETL*.

Widzimy zachowanie w modelu ekonometrycznym zmiennych *X*5 oraz *X*13 określonych wstępnie w wyniku analizy współczynników korelacji. W raporcie występuje też określenie błędu standardowego $D(a_i)$oszacowania danego parametru a_i oraz obliczenie statystyki *t*-Studenta według wzoru:

$$t_{obl} = \frac{a_i}{D(a_i)}$$

Program sprawdza poniższy warunek:

$$|t_{obl}| > t_{\alpha}$$

przy czym t_{α} - wartość krytyczna wg rozkładu *t* -Studenta dla poziomu istotności α.
Jak już wiemy dla poziomu istotności α = 0,05 oraz ilości stopni swobody *iss* = 142 t_{α} = 1,9759.

Warto zauważyć, że raport *GRETL-a* wprowadza dogodne oznaczenia graficzne w postaci gwiazdek informujące o poziomie istotności. Program *GRETL* wskazuje na wysoki współczynnik determinacji R^2 = 0,91604. Wyeliminujmy jeszcze ze względu na poziom istotności oznaczony jedną gwiazdką odpowiadający kryterium *t*-Studenta zmienną *X*18. Po tej czynności uzyskujemy z *GRETL* kolejny *model* 17 (tabela 27.5).

Tab. 27.5

Parametry oraz statystyka opisowa modelu 17

Zmienna	Współczynnik	Błąd standardowy	*t*-Studenta	Wartość *p*	
const	-15763,3	1493,05	-10,56	2,11E-019	***
*X*5	-0,129212	0,0246830	-5,235	6,12E-07	***
*X*6	0,810494	0,282636	2,868	0,0048	***
*X*8	196,194	36,3607	5,396	2,94E-07	***
*X*10	-11,3945	5,74129	-1,985	0,0492	**
*X*12	8,05910	1,10165	7,316	2,00E-011	***
*X*13	7,37759	1,50746	4,894	2,75E-06	***
*X*20	1,09901	0,170458	6,447	1,83E-09	***

Średnia arytmetyczna zmiennej zależnej = 19143,6
Odchylenie standardowe zmiennej zależnej = 5766,83
Suma kwadratów reszt = 4,08938e+008
Błąd standardowy reszt = 1734,04
Współczynnik determinacji R^2 = 0,91401
Skorygowany współczynnik determinacji R^2 = 0,90958
Statystyka *F* (7, 136) = 206,511 (wartość p < 0,00001)
Logarytm wiarygodności = -1274,19
Kryterium informacyjne Akaike'a *AIC* = 2564,39

Źródło: Opracowanie własne programem *GRETL*.

Możemy zatem określić model jako:

$$\hat{X}_1 = -15763{,}3 - 0{,}1292X_5 + 0{,}8105X_6 + 196{,}194X_8 - 11{,}3945X_{10} + 8{,}0591X_{12} + 7{,}3776X_{13} + 1{,}099X_{20}$$

gdzie przypomnijmy sobie:

X_1 - produkt krajowy brutto (ceny bieżące) na 1 mieszkańca w zł,
X_5 - odpady wytworzone (w ciągu roku, z wyłączeniem odpadów komunalnych) w tys. t.,
X_6 - ścieki przemysłowe i komunalne odprowadzone do wód lub do ziemi w hm^3,
X_8 - stopa bezrobocia rejestrowanego (stan w dniu 31 XII) w %,
X_{10} - pracujący w budownictwie (stan w dniu 31 XII) w tys.,
X_{12} - przeciętne miesięczne wynagrodzenie w przemyśle brutto w zł,
X_{13} - przeciętne miesięczne wynagrodzenie w budownictwie brutto w zł,
X_{20} - nakłady inwestycyjne w przemyśle (ceny bieżące) w mln zł.

27.5. Dane panelowe modelu PKB województw

Zestawienie danych panelowych modelu PKB województw 16 województw oraz 9 okresów czasowych (lata 1999-2007) podano w tabeli 27.6. Przypomnijmy sobie dane modelu, którego postać ogólna jest następująca:

$$\hat{X}_1 = f\left(X_5, X_6, X_8, X_{10}, X_{12}, X_{13}, X_{20}\right)$$

Kształtowanie się wartości zmiennych modelu ekonometrycznego w 9. okresach czasowych (1999-2007) występuje w tabelach zamieszczonych w *Aneksie* A.2. publikacji *Efektywność działania obiektów struktury terytorialnej*, które obejmują równania trendów i współczynniki determinacji w układzie poszczególnych województw[218].

Tab. 27.6.

Zestawienie danych empirycznych zmiennych modelu

n	W	Rok	T	X1	X5	X6	X 8	X10	X12	X13	X20
1	łódzkie	1999	1	12682	5171,6	188,3	14,3	47,6	1504,67	1392,73	2434,6
2		2000	2	14497	5807,8	158,6	16,3	43,7	1664,10	1481,54	2670,4
3		2001	3	15723	5263,7	153,0	18,1	37,7	1757,34	1478,20	2687,7
4		2002	4	17535	5041,8	130,8	18,4	34,7	1829,93	1532,55	2290,1

[218] Duczmal M., *Efektywność działania obiektów struktury terytorialnej*, op. cit.

5		2003	5	18492	4881,6	134,5	20,7	32,8	1880,97	1566,82	2482,6
6		2004	6	19498	4884,3	124,0	19,5	35,5	1975,94	1609,64	3189,8
7		2005	7	22274	4449,5	127,7	17,9	37,3	2023,61	1663,61	3901,8
8		2006	8	23666	4120,7	127,4	14,7	41,0	2109,67	1759,55	3850,0
9	18876	2007	9	25521	3895,2	129,4	11,2	46,7	2262,39	1928,20	5479,1
10	mazowieckie	1999	1	20920	3692,3	2130,2	9,5	151,	2048,60	1923,69	6779,5
11		2000	2	23760	4383,5	2165,2	10,8	140,	2290,64	2152,64	5870,1
12		2001	3	26871	5059,9	2128,9	13,0	367,	2489,44	2244,38	5638,3
13		2002	4	30283	4916,2	2185,1	13,8	119,	2563,59	2281,75	5086,4
14		2003	5	31115	5917,7	2363,7	15,4	106,	2642,50	2337,03	5884,7
15		2004	6	32722	5842,3	2458,4	14,7	101,	2703,61	2404,13	6357,6
16		2005	7	36636	5855,9	2524,2	13,8	105,	2807,37	2483,79	6326,1
17		2006	8	40817	5773,3	2745,6	11,8	115,	2915,50	2727,02	6856,5
18	31945	2007	9	44381	6788,4	2700,6	9,0	129,	3128,25	3042,25	7602,7
19	małopolskie	1999	1	13029	10086,5	739,2	10,2	78,5	1680,79	1462,25	2997,4
20		2000	2	14231	10004,5	771,0	12,2	72,0	1823,86	1595,86	3076,1
21		2001	3	15826	9881,0	720,9	14,1	64,9	1925,97	1603,85	2297,5
22		2002	4	16788	9860,6	850,6	13,8	59,9	1998,86	1623,00	2231,3
23		2003	5	17659	10604,5	845,7	16,0	58,0	2033,68	1666,26	2345,9
24		2004	6	18522	10079,9	833,2	15,0	56,0	2138,40	1688,53	3049,8
25		2005	7	20671	9484,4	898,5	13,8	59,4	2230,23	1734,92	2765,4
26		2006	8	21989	8676,7	834,1	11,3	64,4	2371,77	1865,22	3810,8
27	18092	2007	9	24111	9026,9	764,8	8,7	71,9	2581,55	2069,37	4919,6
28	śląskie	1999	1	16014	49038,3	484,2	10,4	141,	2152,69	1543,59	7128,0
29		2000	2	17565	46888,7	445,2	12,9	122,	2303,87	1648,27	5907,2
30		2001	3	19509	45874,2	429,5	15,7	114,	2464,78	1675,03	5393,9
31		2002	4	21206	42715,6	389,0	16,5	103,	2563,66	1702,41	6123,1
32		2003	5	22627	42439,9	371,1	17,6	90,4	2635,48	1745,59	6482,3
33		2004	6	23718	43459,2	364,6	16,9	85,8	2765,91	1766,10	6130,4
34		2005	7	27177	42273,0	362,2	15,5	69,7	2884,10	1861,43	7025,7
35		2006	8	27792	41678,1	366,8	12,7	97,6	3041,70	1970,83	9516,4
36	22789	2007	9	29497	41408,2	372,3	9,2	106,	3230,68	2183,35	11999,2
37	lubelskie	1999	1	10383	4130,4	148,6	12,9	36,1	1552,43	1406,66	1194,9
38		2000	2	11112	4242,6	152,1	14,0	30,3	1695,42	1543,66	1282,5
39		2001	3	12146	3670,4	119,9	15,7	28,8	1802,41	1560,03	1249,2
40		2002	4	13614	3844,3	120,4	15,7	26,5	1852,95	1564,04	1055,5
41		2003	5	14300	4132,4	114,2	18,7	24,6	1897,73	1629,13	1146,4
42		2004	6	15019	4072,2	135,7	17,8	23,5	1979,69	1654,47	1176,5
43		2005	7	16777	5155,4	137,3	17,0	24,5	2052,16	1730,48	1299,8
44		2006	8	17591	4884,3	155,0	15,5	27,5	2146,18	1806,15	1347,5
45	14413	2007	9	18779	5122,5	157,6	13,0	31,2	2343,07	1927,93	1517,8
46	podkarpackie	1999	1	10866	1323,7	217,0	14,5	40,1	1549,66	1397,57	1794,5
47		2000	2	11685	1349,2	215,8	16,2	35,0	1694,11	1535,49	1567,0
48		2001	3	12608	1070,9	174,3	17,4	32,4	1789,19	1537,99	1382,1
49		2002	4	13870	1068,3	193,6	16,9	28,5	1826,93	1547,11	1537,2
50		2003	5	14569	1119,9	190,8	20,2	25,5	1906,32	1604,58	1700,3
51		2004	6	15436	1169,3	180,9	19,1	25,3	1960,44	1621,94	1716,4
52		2005	7	16886	1236,3	189,2	18,5	25,9	2016,94	1663,50	1986,2
53		2006	8	17789	1336,9	198,4	16,4	29,9	2122,80	1786,40	2026,1
54		2007	9	19024	1276,1	201,8	14,2	32,9	2300,37	1919,83	2607,8

55	podlaskie	1999	1	10928	699,1	44,4	12,5	18,8	1489,95	1430,05	591,5
56		2000	2	11580	717,8	40,8	13,8	16,8	1628,84	1546,03	581,9
57		2001	3	13174	744,6	39,1	15,1	15,5	1717,73	1555,95	558,4
58		2002	4	14727	807,1	38,8	15,1	14,4	1783,44	1545,31	587,1
59		2003	5	15719	927,5	40,2	16,9	13,8	1862,69	1636,04	700,8
60		2004	6	16105	887,4	40,4	16,1	12,9	1896,12	1729,32	796,0
61		2005	7	18056	927,9	39,3	15,6	13,2	1985,90	1859,26	980,1
62		2006	8	19075	957,4	39,7	13,3	14,8	2105,52	1959,59	1222,1
63	15529	2007	9	20396	1057,2	41,2	10,4	16,8	2339,81	2136,29	1493,5
64	świętokrzyski	1999	1	11056	1704,8	1155,9	15,1	28,6	1634,89	1539,65	1217,0
65		2000	2	12435	1913,9	1062,3	16,6	23,8	1856,01	1701,45	1163,6
66		2001	3	13868	1710,6	1075,4	18,4	22,6	1942,54	1655,97	956,5
67		2002	4	14843	1386,0	1034,9	18,5	20,6	1973,87	1667,55	932,3
68		2003	5	15977	1403,8	1073,1	22,0	19,6	1980,68	1668,68	801,0
69		2004	6	16911	1761,0	1005,3	22,0	17,6	2086,79	1739,36	1068,0
70		2005	7	18714	1598,6	825,9	20,6	18,1	2124,39	1575,16	1222,1
71		2006	8	19274	1579,6	1094,1	17,7	19,3	2221,71	1636,06	1190,0
72	16023	2007	9	21130	1637,7	1042,4	14,9	21,8	2415,57	1797,03	1652,2

n	W	Rok	T	X1	X5	X6	X 8	X10	X12	X13	X20
73	lubuskie	1999	1	13062	902,4	45,5	17,5	21,0	1497,13	1345,72	1391,4
74		2000	2	14444	864,5	41,8	21,3	17,2	1622,93	1551,56	893,5
75		2001	3	15900	802,4	39,8	24,4	15,5	1709,71	1590,86	1064,0
76		2002	4	17254	785,0	39,6	26,0	13,4	1768,99	1552,79	909,5
77		2003	5	17850	549,1	38,5	27,5	11,6	1809,44	1604,62	1205,5
78		2004	6	18396	623,1	38,2	25,6	11,0	1898,19	1602,33	1065,4
79		2005	7	21641	515,9	36,7	23,0	12,9	1972,10	1612,21	1154,3
80		2006	8	23241	708,6	36,9	19,0	14,3	2063,69	1682,31	1216,3
81	18502	2007	9	24733	727,0	36,0	14,0	16,7	2257,42	1784,03	1863,3
82	wielkopolskie	1999	1	15141	3608,0	1940,2	10,5	85,8	1609,25	1562,67	3946,2
83		2000	2	16747	3650,8	1769,3	12,5	77,0	1766,76	1738,90	4036,7
84		2001	3	18900	3705,2	1839,8	15,4	69,3	1868,85	1753,85	4081,2
85		2002	4	20634	4023,9	1762,8	15,9	66,5	1913,86	1730,96	3539,1
86		2003	5	21072	3731,8	1838,6	17,1	63,3	2018,03	1805,56	4388,8
87		2004	6	22256	4304,1	1745,2	15,9	59,9	2093,58	1806,35	5212,2
88		2005	7	26001	5196,3	1695,1	14,6	64,0	2132,75	1823,29	4943,5
89		2006	8	27553	4363,9	1700,6	11,7	70,7	2257,71	1890,97	5158,0
90	21954	2007	9	29279	4851,6	1694,4	7,8	81,4	2481,16	2116,09	6271,9
91	zachodnio-	1999	1	13980	5368,3	1640,4	18,1	38,3	1676,24	1388,22	1428,7
92	pomorskie	2000	2	15924	5298,6	1628,3	20,8	36,0	1808,40	1522,03	1578,3
93		2001	3	17489	4991,5	1531,2	24,7	32,7	1943,48	1547,57	1029,6
94		2002	4	19269	4319,3	1518,3	26,6	29,5	1928,68	1538,82	1049,0
95		2003	5	20196	5276,9	1509,5	28,4	25,5	1997,97	1564,42	1363,3
96		2004	6	20357	5181,8	1475,7	27,5	24,3	2054,47	1599,01	1900,3
97		2005	7	22494	5426,0	1419,8	25,6	26,7	2102,41	1647,29	1635,8
98		2006	8	23924	6974,7	1688,6	21,5	30,5	2198,32	1754,59	2481,1
99	19884	2007	9	25324	6841,0	1645,9	16,4	34,6	2407,72	1907,41	2533,5
100	dolnośląskie	1999	1	14290	32530,1	223,7	16,0	74,0	1782,18	1592,62	4426,1
101		2000	2	16273	32363,5	209,7	18,4	66,7	1997,14	1661,57	3528,2
102		2001	3	18324	33630,8	225,6	21,5	57,5	2157,94	1728,86	4393,8
103		2002	4	19758	31974,4	241,8	22,4	51,0	2241,66	1732,56	4084,8

104		2003	5	21193	32821,9	227,0	23,8	47,4	2331,82	1760,99	3944,8
105		2004	6	21986	34706,5	229,0	22,4	45,0	2470,79	1822,95	4582,6
106		2005	7	24632	34882,1	235,2	20,6	47,9	2610,14	1901,97	4922,8
107		2006	8	26620	35180,7	246,1	16,6	52,3	2766,50	2078,75	6353,3
108	21424	2007	9	29739	34299,7	231,7	11,4	59,8	3024,88	2293,91	7063,2
109	opolskie	1999	1	12635	2139,0	77,0	13,2	25,9	1670,60	1443,43	1050,9
110		2000	2	13320	2510,1	75,1	15,7	21,9	1813,17	1532,51	719,2
111		2001	3	15146	2328,3	75,4	18,2	19,4	1943,87	1521,78	682,9
112		2002	4	15780	1932,3	91,6	19,4	18,2	2014,13	1513,50	749,3
113		2003	5	16738	1824,9	82,5	21,4	15,3	2068,09	1544,76	780,6
114		2004	6	17173	1816,6	87,5	20,0	15,6	2147,48	1572,15	950,6
115		2005	7	20785	1643,0	82,0	18,7	16,0	2177,98	1625,33	1197,5
116		2006	8	21347	1846,8	93,3	16,2	17,8	2292,51	1761,95	1117,5
117	17252	2007	9	22347	1536,5	91,7	11,9	20,2	2549,77	1973,60	1638,1
118	kujawsko-	1999	1	13193	3054,6	171,0	16,9	42,7	1590,14	1404,04	1839,0
119	pomorskie	2000	2	14121	3096,9	143,8	19,2	40,4	1692,89	1542,43	1853,1
120		2001	3	15893	2749,5	134,2	21,9	35,2	1808,57	1587,89	2112,3
121		2002	4	17563	2786,1	131,2	22,5	32,3	1872,75	1614,15	2030,7
122		2003	5	18635	2295,1	125,6	24,6	29,3	1922,59	1604,59	1635,3
123		2004	6	19169	2550,3	116,3	23,6	28,2	1983,74	1653,08	1938,6
124		2005	7	21633	3150,6	116,3	22,3	29,6	2034,74	1717,97	2287,7
125		2006	8	22474	2088,6	117,4	19,2	34,1	2144,98	1813,52	2420,9
126	18554	2007	9	24301	2584,4	117,9	14,9	39,0	2281,94	1984,52	3048,2
127	pomorskie	1999	1	14129	2406,4	195,3	13,8	56,3	1635,20	1486,61	2336,8
128		2000	2	16120	1976,1	198,0	16,6	48,9	1862,08	1674,00	2113,9
129		2001	3	17824	1876,8	191,4	19,6	44,6	2006,64	1770,82	1643,1
130		2002	4	19301	1990,0	192,6	21,3	39,6	2054,64	1791,04	1515,9
131		2003	5	20346	2106,3	197,2	22,4	36,2	2123,74	1802,11	1527,6
132		2004	6	20929	2097,1	217,8	21,4	33,3	2244,00	1791,62	1800,9
133		2005	7	23616	2197,9	225,1	19,2	36,2	2376,48	1900,55	1991,7
134		2006	8	25308	2539,9	215,1	15,3	41,7	2509,87	2063,42	2212,6
135	20550	2007	9	27373	2466,9	220,9	10,7	48,4	2732,64	2271,08	2719,1
136	warmińsko-	1999	1	10986	399,2	91,1	22,4	28,2	1532,26	1305,01	880,7
137	mazurskie	2000	2	12341	459,6	83,6	25,8	23,2	1702,61	1486,12	848,5
138		2001	3	13210	450,2	69,9	28,9	21,1	1801,18	1523,85	712,0
139		2002	4	14065	443,3	68,6	28,9	18,4	1866,05	1533,07	874,8
140		2003	5	15258	518,1	67,6	30,3	16,7	1909,12	1603,11	980,2
141		2004	6	16504	594,4	67,4	29,2	15,9	1874,22	1628,17	1001,8
142		2005	7	18778	609,3	67,1	27,2	17,8	1879,30	1660,65	1554,6
143		2006	8	19709	753,2	66,3	23,6	19,9	1970,75	1781,30	1909,7
144	15762	2007	9	21005	894,8	66,2	18,7	24,1	2158,18	1946,59	1935,7
		średnia		**19144**							

Źródło: Opracowanie własne w *Excelu* na podstawie danych *GUS*-u.

28. Podstawowe miary statystyki opisowej

28.1. Przedziały obserwacji[219]

Obrazem zmiennej jest rozkład, który można przedstawić jako szereg statystyczny. Jeśli mamy wszystkie obserwacje to taki szereg nazywamy szczegółowym *n* obserwacji. Zmienne oznaczamy dużymi literami np. *X*, a kolejne obserwacje tej zmiennej jako x_1, x_2, ...,x_n. Gdy jest duża liczba obserwacji to tworzy się szereg rozdzielczy o *K* przedziałach. Środek przedziału oznaczamy przez $\dot{x}_i$, a liczebność danego przedziału przez n_i, przy czym $i = 1, 2, \ldots, K$.

Przykład: Dla zobrazowania przedziałów podano dane przedstawiające wagi dwóch grup studentów studiów zaocznych magisterskich - kierunek *Zarządzanie* zmierzone razem na ćwiczeniach ze statystyki.

Tab. 28.1

Wagi studentów w kg

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Waga	68	65	43	92	77	70	75	65	62	69	59	50	51	64	48	56	66	89	98	65
Lp.	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
Waga	96	78	107	102	98	87	53	65	60	120	60	63	86	47	86	100	75	98	59	55
Lp.	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60
Waga	56	60	82	66	100	60	55	75	78	60	53	65	86	78	65	47	50	64	55	56
Lp.	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80
Waga	85	64	69	57	92	64	115	114	48	96	77	66	94	77	62	86	86	87	66	94

Źródło: Opracowanie własne.

Korzystamy teraz z funkcji Excela, przykładowo dla przedziału adresowego (B10-U13) wpisując w nawiasach wagi od pozycji 1-80:

=ŚREDNIA(B10:U13). Średnia ze wszystkich obserwacji: 72,96 kg.

Wartość minimalna określona funkcją MIN(B10:U13) wynosi 43, a wartość maksymalna przy zastosowaniu funkcji MAX(B10:U13) to 120. Liczba klas stanowi pierwiastek z ilości obserwacji. Zatem $K = \sqrt{80} = 8{,}9 \approx 9$ Rozpiętość przedziału wynosi:

$$h = \frac{x_{max} - x_{min}}{K} = \frac{120 - 43}{9} = 8{,}6 \approx 9$$

Przyjęto 9 klas danych pogrupowanych po 10 liczb przy zastosowaniu liczenia funkcją Excela (ILE.LICZB( )). Wymaga to jednak wcześniejszego ustawienia wszystkich liczb w jednej kolumnie, a następnie posortowania narastająco korzystając z menu (DANE-SORTUJ), przy czym n_i - liczność klasy.

Tab. 28.2

Jednakowe grupy przedziałowe poszczególnych klas

K	*h*	n_i
1	40-49	5
2	50-59	14
3	60-69	25
4	70-79	10

[219] Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, część I. Wydawnictwo Instytut Śląski, Opole 2014, rozdz. 1.5.1. *Miary rozkładu jednej zmiennej*.

5	80-89	10
6	90-99	9
7	100-109	4
8	110-119	2
9	120-129	1
	n	80

Źródło: Opracowanie własne w Excelu.

Dane po uporządkowaniu narastająco przedstawia tabela 28.3. Przykładowo przedział $K = 1$ to wagi: 43, 47, 47, 48, 48.

Tab. 28.3

Wagi studentów ze wskazaniem wartości minimalnej, maksymalnej i klas przedziałowych

43	47	47	48	***48***	50	50	51	53	53	55	55	55	56	56	56	57	59	***59***	60
60	60	60	60	62	62	63	64	64	64	64	65	65	65	65	65	65	66	66	66
66	68	69	***69***	70	75	75	75	77	77	77	78	78	***78***	82	85	86	86	86	86
86	87	87	***89***	92	92	94	94	96	96	98	98	***98***	100	100	102	***107***	114	**115**	**120**

Źródło: Opracowanie własne w Excelu.

Aby scharakteryzować rozkład zmiennej stosujemy miary położenia, zmienności oraz skośności (asymetrii).

28.2. Miary przeciętne (położenia)

Do miar przeciętnych oraz pozostałych zastosowano wzory zamieszczone w publikacji K. Jajugi[220]. Rozróżniamy średnie (arytmetyczna, harmoniczna, geometryczna) oraz przeciętne (wartość środkowa - mediana, kwartyle).

Średnie

Często stosowana średnia arytmetyczna np. zmiennej X jest sumą wartości wszystkich obserwacji podzielona przez jej liczebność.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Średnia arytmetyczna ma kilka własności, a mianowicie:

a) Suma odchyleń od wartości średniej równa się zero.

$$\sum_{i=1}^{n} (x_i - \bar{x}) = 0$$

b) Suma kwadratów odchyleń wartości zmiennej od średniej arytmetycznej jest najmniejsza.

c) Jeżeli do wszystkich wartości zmiennej X dodamy stałą c, to wówczas wartość średnia:

$$\bar{X} + c = \bar{X} + c$$

c) Iloczyn stałej i zmiennej

$$\bar{X} * c = \bar{X} + c$$

e) Jeżeli jedna zmienna jest sumą dwóch innych zmiennych (Y, Z) to średnia arytmetyczna jest sumą średnich arytmetycznych wspomnianych zmiennych.

$$\bar{Y} = \bar{X} + \bar{Z}$$

[220] Jajuga K. (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2002.

Przykłady: Przykład dotyczą $n = 14$ obserwacji zmiennej x_i *klasy drugiej* oraz własności c), d), e). Przyjęto $c = 4$.

Przykład 1. Pokazanie własności średniej.

Tab. 28.4

Obliczenia pomocnicze dla wykazania własności średniej

Lp.	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$x_i + c$	$x_i * c$
1	50	-4,64	21,56	54	200
2	50	-4,64	21,56	54	200
3	51	-3,64	13,27	55	204
4	53	-1,64	2,70	57	212
5	53	-1,64	2,70	57	212
6	55	0,36	0,13	59	220
7	55	0,36	0,13	59	220
8	55	0,36	0,13	59	220
9	56	1,36	1,84	60	224
10	56	1,36	1,84	60	224
11	56	1,36	1,84	60	224
12	57	2,36	5,56	61	228
13	59	4,36	18,98	63	236
14	59	4,36	18,98	63	236
	Średnia		Suma	Średnia	Średnia
	54,64		111,21	58,64	218,57

gdzie:

54,64 - średnia arytmetyczna zmiennej X np. =SUMA(B2:B15)/A15,
111,21 - suma kwadratów odchyleń $(x_i - \bar{x})^2$,
58,64 - średnia z (x_i + c),
218,57 - średnia iloczynu (x_i * c).
Źródło: Opracowanie własne w Excelu.

Przykład 2. Dla celu pokazania obliczania średniej dwóch zmiennych przyjmijmy, że 10-cio elementowe klasy czwarta oraz piąta stanowią dwa oddzielne zbiory X_4 oraz X_5.

Tab. 28.5

Obliczenia pomocnicze do obliczenia średniej stanowiącej sumę dwóch zmiennych

Lp.	X_{4i}	X_{5i}	$Y_i = X_{4i} + X_{5i}$
1	70	82	152
2	75	85	160
3	75	86	161
4	75	86	161
5	77	86	163
6	77	86	163
7	77	86	163
8	78	87	165
9	78	87	165
10	78	89	167
Średnia	76	86	162

gdzie:

76; 86;162 - średnie z danej kolumny obliczone funkcją Excela ŚREDNIA( ).
Źródło: Opracowanie własne.

Łatwo zauważyć, że średnia sumy dwóch zmiennych jest sumą średnich danych zmiennych.

Gdy wartości zmiennej pogrupowane są w szereg rozdzielczy (tab. 28.2) to średnia arytmetyczna - zwana ważoną obliczana jest wg wzoru:

$$\bar{x} = \frac{\sum_{i=1}^{K} n_i \dot{x}_i}{n}$$

gdzie: $\dot{x}_i$ - środek i-tego przedziału, n_i - liczność tego przedziału, K - liczba przedziałów (klas).

Przykład: n = 80 obserwacji - tabela 28.2.

Tab. 28.6

Obliczenia pomocnicze do średniej ważonej

K	h	n_i	$\dot{x}_i$	$n_i\dot{x}_i$
1	40-49	5	44,5	222,5
2	50-59	14	54,5	763,0
3	60-69	25	64,5	1612,5
4	70-79	10	74,5	745,0
5	80-89	10	84,5	845,0
6	90-99	9	94,5	850,5
7	100-109	4	104,5	418,0
8	110-119	2	114,5	229,0
9	120-129	1	124,5	124,5
		n		Suma
		80		5810,0

Źródło: Opracowanie własne w Excelu.

Po podstawiamy do wzoru na średnią ważoną: $\bar{x} = 5810/80 = 72{,}63$

Pamiętamy, że średnia arytmetyczna z danych szeregu szczegółowego nieuporządkowanego (tabela 28.1) wynosi 72,96. Tak więc różnica między tymi dwoma średnimi jest mała - 0,33.

Kolejną stosowaną miarą położenia jest *średnia harmoniczna* $\boldsymbol{H_x}$, stosowana wówczas, gdy wartości jednostek podawane są w formie odwrotności. Dla szeregu szczegółowego H_x obliczamy ze wzoru:

$$H_x = \frac{n}{\sum_{i=1}^{n} (1/x_i)}$$

gdzie:

x_i - różne od zera.

Natomiast średnia harmoniczna ważona obliczana jest według wzoru:

$$\boldsymbol{H_x} = \frac{\boldsymbol{n}}{\sum_{\boldsymbol{i=1}}^{\boldsymbol{K}} (\boldsymbol{n_i} / \dot{\boldsymbol{x}}_{\boldsymbol{i}})}$$

Przykład: Średnia harmoniczna

Tab. 28.7

Obliczenia pomocnicze do średniej harmonicznej ważonej

K	h	n_i	$\dot{x}_i$	$n_i/\dot{x}_i$
1	40-49	5	44,5	0,11235955
2	50-59	14	54,5	0,25688073
3	60-69	25	64,5	0,38759690
4	70-79	10	74,5	0,13422819
5	80-89	10	84,5	0,11834320

6	90-99	9	94,5	0,09523810
7	100-109	4	104,5	0,03827751
8	110-119	2	114,5	0,01746725
9	120-129	1	124,5	0,00803213
		n		Suma
		80		1,16842355

Źródło: Opracowanie własne w Excelu.

Po podstawieniu do wzoru otrzymamy: H_x = 80/1,16842355 = 68,47. Jeżeli porównamy średnią ważoną oraz harmoniczną obliczone dla szeregu rozdzielczego tych samych danych (tabela 28.1) to uzyskamy różnicę: 72,63 - 68,47 = 4,16.

Obliczymy teraz *średnią geometryczną* G_x bazując na danych z tabeli 28.1. Średnia ta stosowana jest między innymi do rozkładów, w których zachodzą znaczne różnice między wartościami. Dla szeregu szczegółowego jest ona pierwiastkiem *n*-tego stopnia z iloczynu *n* wyrazów - obserwacji zmiennej *X*, przy czym *i* = 1- *n*.

$$G_x = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

Natomiast dla szeregu rozdzielczego obliczamy ją z zależności:

$$\boldsymbol{G_x = \sqrt[n]{\dot{x}_1^{\,n_1}\dot{x}_2^{\,n_2}...\dot{x}_n^{\,n_n}}}$$

Mając dane n_i oraz $\dot{x}_i$ w tabeli 28.7 obliczamy G_x.

Tab. 28.8

Obliczenia pomocnicze do średniej geometrycznej

n_i	$\dot{x}_i$	$\dot{x}_i^{\,n_i}$
5	44,5	174501858
14	54,5	2,E+24
25	64,5	2,E+45
10	74,5	5,E+18
10	84,5	2,E+19
9	94,5	6,E+17
4	104,5	119251860
2	114,5	13110
1	124,5	125
n		Iloczyn
80		7,1E+147

gdzie:

E - podstawa logarytmu naturalnego.

Źródło: Opracowanie własne.

Suma n_i równa się 80. Obliczmy zatem pierwiastek 80-tego stopnia z otrzymanego iloczynu 7,1E+147 korzystając z wyrażenia *Excela* np.: (=E201^0,0125), gdzie: 1/80 = 0,0125, e = E ≈ 2,718282.

$$\boldsymbol{G_x = \sqrt[80]{7{,}1 * 2{,}718282^{\,147}} = 70{,}49}$$

Aby obliczyć pierwiastek możemy skorzystać z funkcji *Excela* (PIERWIASTEK( )) lub (POTĘGA( )). Porównajmy teraz wartości różnych średnich ważonych dla szeregu rozdzielczego:

$$H_x = 68{,}47 < G_x = 70{,}49 < \bar{x} = 72{,}63.$$

Zróbmy teraz przykład na średnią geometryczna będącą iloczynem dwóch innych zmiennych, korzystając z własności, że taka średnia geometryczna jest iloczynem średnich geometrycznych zmiennych: $G_{xy} = G_x\, G_y$.

Przykład: Iloczyn dwóch średnich geometrycznych

Tab. 28.9

Obliczenia pomocnicze do iloczynu dwóch średnich geometrycznych

Lp.	X_4	X_5	$Y_i = X_4 * X_5$
1	70	82	5740
2	75	85	6375
3	75	86	6450
4	75	86	6450
5	77	86	6622
6	77	86	6622
7	77	86	6622
8	78	87	6786
9	78	87	6786
10	78	89	6942
Średnia geometryczna	75,9633	85,9835	6531,5933

gdzie:

dane x_{4i} oraz x_{5i} wg tabeli 28.5; w wierszu dolnym obliczone średnie geometryczne.

Źródło: Opracowanie własne w Excelu.

Korzystając z wyrażenia *Excel*-a obliczamy średnie geometryczne dla zmiennej X_4 oraz X_5 jak również dla sumy iloczynów $y_i = x_{4i}*x_{5i}$ np. wyrażenie (=ŚREDNIA.GEOMETRYCZNA(E43:E52) do obliczenia średniej geometrycznej zmiennej X_4. Przybliżony wynik otrzymamy jeśli przemnożymy średnie geometryczne zmiennej X_4 oraz zmiennej X_5, przykładowo: 75,9633 * 85,9835 = 6531,5904

28.3. Przeciętne pozycyjne

Drugą grupę miar położenia stanowią przeciętne pozycyjne, do których wyznaczenia stosujemy szereg szczegółowy uporządkowany. Rozpocznijmy od mediany, dla której liczba

$$M_x = \frac{x_{n/2} + x_{n/2+1}}{2}$$

jednostek mających wartość nie większą od mediany jest równa liczbie jednostek mających wartość nie mniejszą od niej. W przypadku nieparzystej liczby obserwacji medianą jest obserwacja o numerze $(n + 1)/2$: $M_x = x_{(n+1)/2}$. Jeśli liczba obserwacji jest parzysta, mediana jest średnią arytmetyczną dwóch wartości leżących w środku uporządkowanego szeregu. Wyjaśnijmy to na przykładzie, gdzie zmienną X_4 występuje jako szereg uporządkowany o parzystej liczbie obserwacji.

Tab. 28.10

Szereg uporządkowany zmiennej X_4

n	1	2	3	4	5	6	7	8	9	10
x_{4i}	70	75	75	75	77	77	77	78	78	78

Źródło: Opracowanie własne; x_i według tabeli 28.9.

$$\boldsymbol{M}_x = \frac{77 + 77}{2} = 77$$

Dla szeregu rozdzielczego stosujemy wzór:

$$M_x = x_0 + h * \frac{(n+1)/2 - n_s}{n_0}$$

gdzie:

n - ogólna ilość obserwacji ,
n_s - liczba obserwacji w szeregu skumulowanym w przedziale poprzedzającym przedział zawierający medianę,
x_0 - dolna granica przedziału zawierającego medianę, ,
h - rozpiętość przedziału zawierającego medianę,
n_0 - liczba obserwacji w przedziale z medianą.

Przykład: Wykonano na podstawie danych szeregu rozdzielczego wagi studentów - tabela 28.6.

Tab. 28.11

Określenie liczebności skumulowanej

K	*h*	Ilość studentów	n_i	Liczebność skumulowana
1	40-49	< 50	5	5
2	50-59	< 60	14	**19**
3	60-69	< 70	**25**	44
4	70-79	< 80	10	54
5	80-89	< 90	10	64
6	90-99	< 100	9	73
7	100-109	< 110	4	77
8	110-119	< 120	2	79
9	120-129	< 130	1	80

gdzie:

$n_s = 19$, $x_0 = 60$, $h = 10$, $n_0 = 25$, $n = 80$ (ilość obserwacji)

Źródło: Opracowanie własne w *Excelu*.

$$\boldsymbol{M}_x = 60 + 10 * \frac{(80+1)/2 - 19}{25} = 68{,}6$$

Dla porównania w szeregu szczegółowym średnia ważona wynosiła 72,63.

Kolejne miary pozycyjne to *kwartyle*. Dzielą one szereg szczegółowy uporządkowany względem liczebności na cztery równe części. Przykładowo kwartyl pierwszy Q_1 to taka wartość zmiennej, która dzieli szereg uporządkowany tak, że 1/4 jednostek ma od niej nie większe, a 3/4 nie mniejsze wartości.

Przykład:

Tab. 28.12

Szereg wag studentów uporządkowany narastająco dla n = 80 obserwacji

1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*	*18*	*19*	*20*
43	47	47	48	*48*	50	50	51	53	53	55	55	55	56	56	56	57	59	*59*	60
21	*22*	*23*	*24*	*25*	*26*	*27*	*28*	*29*	*30*	*31*	*32*	*33*	*34*	*35*	*36*	*37*	*38*	*39*	*40*
60	60	60	60	62	62	63	64	64	64	64	65	65	65	65	65	65	66	66	66
41	*42*	*43*	*44*	*45*	*46*	*47*	*48*	*49*	*50*	*51*	*52*	*53*	*54*	*55*	*56*	*57*	*58*	*59*	*60*
66	68	69	*69*	70	75	75	75	77	77	77	78	78	*78*	82	85	86	86	86	86
61	*62*	*63*	*64*	*65*	*66*	*67*	*68*	*69*	*70*	*71*	*72*	*73*	*74*	*75*	*76*	*77*	*78*	*79*	*80*
86	87	87	*89*	92	92	94	94	96	96	98	98	*98*	100	100	102	*107*	114	115	120

Dla n = 20 Q_1 = 60, n = 40 Q_2 = 66, n = 60 Q_3 = 86. Najmniejsza wartość zmiennej - 43, a największa - 120.

Źródło: Opracowanie własne w *Excelu*.

Dla szeregu rozdzielczego kwartyl pierwszy obliczamy według wzoru:

$$Q_x = x_0 + h * \frac{n/4 - n_s}{n_0}$$

gdzie:

n - liczba obserwacji,
n_s - liczba obserwacji w szeregu skumulowanym - w przedziale poprzedzającym kwartyl 1,
x_0 - dolna granica przedziału zawierającego kwartał 1,
h - rozpiętość przedziału zawierającego kwartyl 1,
n_0 - liczba obserwacji w przedziale z kwartylem 1.

Dla obliczenia Q_3 zamiast $n/4$ wstawiamy $3n/4$.

Przykład: Obliczamy kwartyle szeregu rozdzielczego mając dane z tabeli 28.11.

h	n_i	Liczebność skumulowana
40-49	5	5
50-59	14	**19**
60-69	**25**	44
70-79	10	***54***
80-89	***10***	64
90-99	9	73
100-109	4	77
110-119	2	79
120-129	1	80

- Kwartyl 1: $n_s = 19$, $x_0 = 60$, $h = 10$, $n_0 = 25$,

$$\boldsymbol{Q}_1 = 60 + 10 * \frac{80/4 - 19}{25} = 60{,}4$$

- Kwartyl 2 - wg tabeli 28.12.

$$Q_2 = M_x = 68{,}6$$

- Kwartyl 3: $n_s = 54$ $x_0 = 80$, $h = 10$, $n_0 = 10$.

$$\boldsymbol{Q}_3 = 80 + 10 * \frac{(3 * 80)/4 - 54}{10} = 86$$

28.4. Wykres pudełkowy oraz histogram

Wykres pudełkowy składa się z prostokąta rozciągniętego na poziomie kwartyli i przedzielonego na poziomie mediany. Do niego dołączone są dwa odcinki. Sposób ich określenia omówiony zostanie na przykładzie danych z tabeli 28.12, gdzie:

$$x_{min} = 43,\ x_{max} = 120,\ Q_x = (Q_3 - Q_1)/2 = (86 - 60)/2 = 13.$$

Sprawdzamy warunki:

$$x_{min} = 43 > Q_1 - 3Q_x = 60 - 3 * 13 = 21 \text{ - warunek jest spełniony.}$$

Rysujemy odcinek od pudełka do x_{min}.

Podobnie rysuje się drugi odcinek łączący kwartyl trzeci z wartością największą x_{max}

$$x_{max} = 120 < Q_3 + 3Q_x = 86 + 3 * 13 = 125 \text{ - warunek jest spełniony.}$$

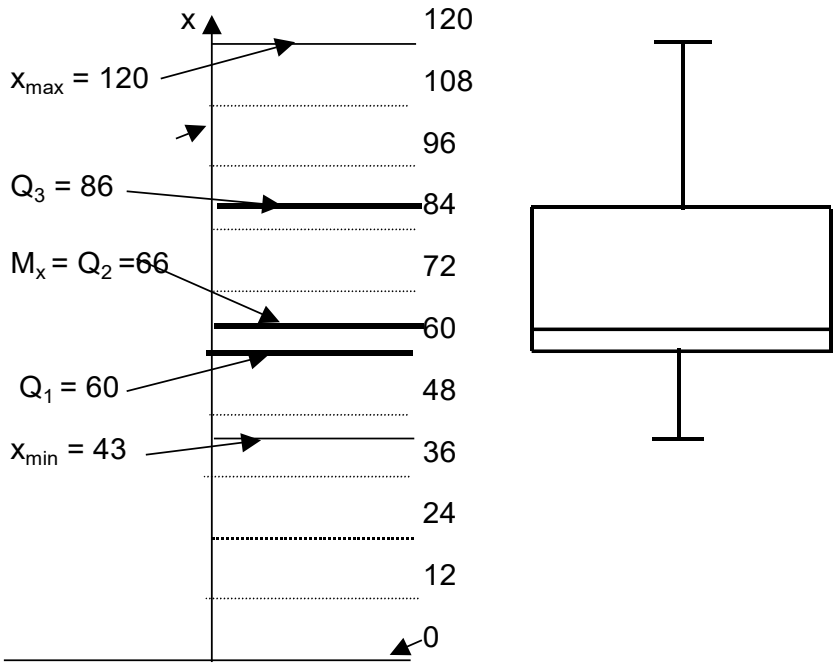
Ponadto zaznacza się wartości nietypowe jeśli występują. Obserwacja *i*-ta uznawana jest za nietypową jeśli zachodzi jeden z dwóch warunków:

$$x_i < Q_1 - 3Q_x = 60 - 3 * 13 = 21 \qquad x_i > Q_3 + 3Q_x = 125$$

Żaden z tych warunków nie występuje w naszym przykładzie, bo $x_1 = 43$, a $x_{80} = 10$, tak więc nie występują nietypowe wartości obserwacji.

Bazując na prezentacji graficznej (28.1) wyciągamy wnioski o:

- wartościach przeciętnych (środek pudełka - mediana, brzegi pudełka - kwartyle),
- zmienności (długość pudełka),
- skośności (położenie linii podziału pudełka w stosunku do brzegów pudełka - w rozkładach o asymetrii lewostronnej linia podziału jest bliżej wierzchołka pudełka).

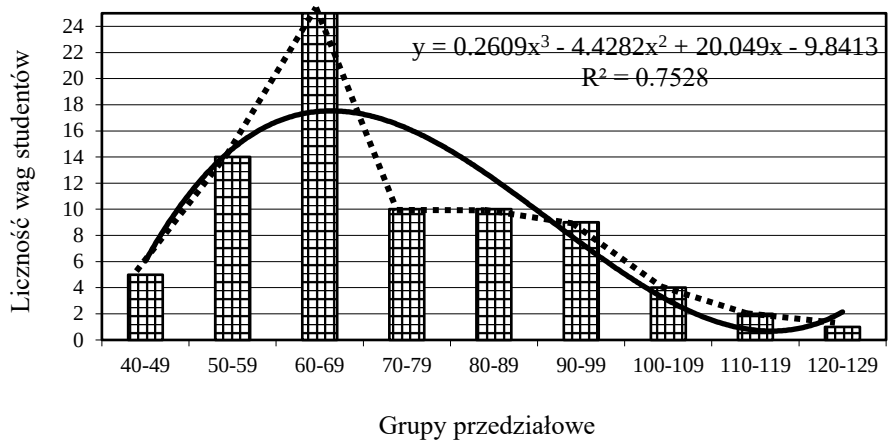


Źródło: Opracowanie własne w *Excelu*.

Rys. 28.1.Wykres pudełkowy zmiennej *X* dla szeregu szczegółowego

Przedstawmy teraz rozkład jednej zmiennej przy pomocy histogramu w układzie współrzędnych prostokątnych na przykładzie danych z tabeli 28.7. Pamiętamy z wcześniejszego przykładu, że średnia arytmetyczna z kolejnych 80-ciu obserwacji wynosi 72,96.

h	n_i
40-49	5
50-59	14
60-69	25
70-79	10
80-89	10
90-99	9
100-109	4
110-119	2
120-129	1



Źródło: Opracowanie własne wykonane w *Excelu*.

Rys. 28.2. Histogram na podstawie szeregu rozdzielczego zmiennej X

W histogramie na osi rzędnych zaznaczono liczebności wag studentów dla poszczególnych grup przedziałów klasowych w formie prostokątów. Ponadto narysowano diagram w postaci linii łamanej przerywanej oraz wyznaczono trend wielomianowy trzeciego stopnia oraz określono w *Excelu* współczynnik determinacji R^2. Gdy porównamy z wartością średnia $\bar{x} = 72{,}96$ to stwierdzamy, że otrzymany wykres ma asymetrię prawostronną.

29. Miary rozproszenia oraz miary asymetrii

29.1. Liczność występowania oraz obszar zmienności (rozstępu)

Miary rozproszenia, czyli zmienności, służą do scharakteryzowania stopnia zróżnicowania wartości badanej zmiennej. Miary zmienności dzielimy na absolutne i stosunkowe. Miarami absolutnymi są: rozstęp, odchylenie standardowe i wariancja oraz odchylenie przeciętne i ćwiartkowe. Zaprezentujmy zatem na danych umownych rozkład zmiennych X_1 oraz X_2 o podanej liczności n_i .

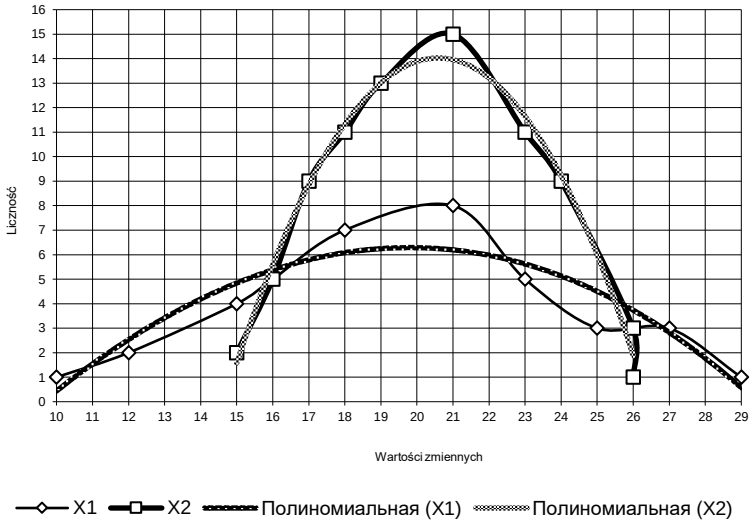
Tab. 29.1

Liczność występowania wartości zmiennych

X_1	n_1	X_2	n_2
10	1	12	2
12	2	14	5
15	4	17	9
16	5	18	11
18	7	19	13
21	8	21	15
23	5	23	11
25	3	24	9
27	3	26	4
30	1	28	2
18,3		20	średnia
6,53		4,08	odchylenie standardowe

Źródło: Opracowanie własne w Excelu.

W obliczeniu odchylenia standardowego wykorzystano funkcję ODCHYLENIE.STANDARDOWE.POPUL( ) traktując nasz szereg obserwacji jako populację.



Źródło: Opracowanie własne w *Excelu*.

Rys. 29.1. Wykres punktowy wartości empirycznych oraz trendy wielomianowe 3-go stopnia zmiennych X_1 i X_2.

Na rysunku 29.1 większemu skupieniu odpowiada krzywa o wyraźnie zaznaczonym maksimum dla liczności n_2 zmiennej X_2. Świadczy o tym również odchylenie standardowe równe 4,08.

Powróćmy do miar zmienności absolutnych. Rozstęp określony jest wzorem:

$$O_x = x_{\max} - x_{\min}$$

Przykład: Na podstawie danych z pomiaru wagi pewnej grupy studentów dla 80. obserwacji zmiennej X obliczamy rozstęp O_x. Wg tabeli 28.1 szeregu szczegółowego uporządkowanego obserwacji $x_{min} = 43$, $x_{max} = 120$, a $O_x = 120 - 43 = 77$.

29.2. Odchylenia, wariancja, odchylenie standardowe oraz współczynnik zmienności

Rozróżniamy odchylenia ćwiartkowe oraz przeciętne. Odchylenie ćwiartkowe, stanowi połowę różnicy między trzecim a pierwszym kwartylem.

$$Q_x = (Q_3 - Q_1)/2$$

Przykład: Tym razem przyjmijmy wcześniejsze obliczenia dla szeregu rozdzielczego wag studentów $Q_3 = 86$, $Q_1 = 60{,}4$ i określamy odchylenie ćwiartkowe:

$$\boldsymbol{Q_x = (86 - 60{,}4)/2 = 12{,}8}$$

Odchylenie ćwiartkowe jest średnią odległością mediany $M_x = Q_x = 68{,}6$ od kwartyli Q_1 oraz Q_3:

$$\boldsymbol{Q_x = [(86 - 68{,}6) + (68{,}6 - 60{,}4)]/2 = 12{,}8}$$

Pozycyjne miary zmienności (rozproszenia) opierają się na wartościach wybranych jednostek (zmiennych). Występują również klasyczne miary rozproszenia bazujące na wartościach wszystkich obserwacji, a należą do ich:

- przeciętne, które są średnią arytmetyczną bezwzględnych wartości odchyleń od średniej arytmetycznej z obserwacji danej zmiennej:

$$\boldsymbol{D_x = \frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}}$$

Obliczmy D_x dla szeregu szczegółowego wg danych x_i z pomiaru wag studentów, gdzie: $\bar{x} = 72{,}96$, $n = 80$.

Tab. 29.2

Obliczenia pomocnicze dla określenia przeciętnej zmiennej X

72,96	B	**C**	D	**E**	F	**G**	H	**I**	J	**K**	L	**M**	N	**O**	P	**R**
2	68	**4,96**	59	13,96	96	23,04	60	12,96	56	16,96	53	19,96	85	12,04	77	4,04
3	65	7,96	50	22,96	78	5,04	63	9,96	60	12,96	65	7,96	64	8,96	66	6,96
4	43	29,96	51	21,96	107	34,04	86	13,04	82	9,04	86	13,04	69	3,96	94	21,04
5	92	19,04	64	8,96	102	29,04	47	25,96	66	6,96	78	5,04	57	15,96	77	4,04
6	77	4,04	48	24,96	98	25,04	86	13,04	100	27,04	65	7,96	92	19,04	62	10,96
7	70	2,96	56	16,96	87	14,04	100	27,04	60	12,96	47	25,96	64	8,96	86	13,04
8	75	2,04	66	6,96	53	19,96	75	2,04	55	17,96	50	22,96	115	42,04	86	13,04
9	65	7,96	89	16,04	65	7,96	98	25,04	75	2,04	64	8,96	114	41,04	87	14,04
10	62	10,96	98	25,04	60	12,96	59	13,96	78	5,04	55	17,96	48	24,96	66	6,96
11	69	3,96	65	7,96	120	47,04	55	17,96	60	12,96	56	16,96	96	23,04	94	21,04
Suma		93,84		165,8		218,2		161		123,9		146,8		200		115,2
Razem	1224,6															

Źródło: Opracowanie własne w *Excelu*.

W tabeli obok kolumny wartości x_i podano kolumnę z wartością $Ix_i - \bar{x}I$, obliczoną przykładowo jako: I68 - 72,96I = 4,96. W komórce A1 znajduje się wartość średniej arytmetycznej - 72,96 obliczona wcześniej dla szeregu szczegółowego. Suma wszystkich sum kolumnowych wynosi 1224,6. Po podzieleniu przez 80 otrzymujemy $D_x \approx 15{,}31$. Dla obliczenia kolejnych wartości w kolumnach np. 4,96 wykorzystano wyrażenie (funkcję Excela) stosując adresowanie bezwzględne dla wartości średniej: (= MODUŁ.LICZBY(B2-$A$1)).

Dla szeregu rozdzielczego wzór na obliczenie odchylenia przeciętnego jest następujący:

$$D_x = \frac{\sum_{i=1}^{K} |\dot{x}_i - \bar{x}| n_i}{n}$$

Przykład:

Tab. 29.3

Obliczenia pomocnicze dla szeregu rozdzielczego według danych tabeli 5.7.

72,96	średnia						
K	*h*	n_i	$\dot{x}_i$	$\lvert\dot{x}_i - \bar{x}\rvert$	$(\dot{x}_i - \bar{x})^2$	$\lvert\dot{x}_i - \bar{x}\rvert n_i$	$(\dot{x}_i - \bar{x})^2 n_i$
1	40-49	**5**	**44,5**	**28,46**	809,9716	**142,3**	4049,858
2	50-59	14	54,5	18,46	340,7716	258,44	4770,802
3	60-69	25	64,5	8,46	71,5716	211,5	1789,29
4	70-79	10	74,5	1,54	2,3716	15,4	23,716
5	80-89	10	84,5	11,54	133,1716	115,4	1331,716
6	90-99	9	94,5	21,54	463,9716	193,86	4175,744
7	100-109	4	104,5	31,54	994,7716	126,16	3979,086
8	110-119	2	114,5	41,54	1725,572	83,08	3451,143
9	120-129	1	124,5	51,54	2656,372	51,54	2656,372
					SUMA	1197,68	**26227,73**
					D_x	**14,97**	

Źródło: Opracowanie własne w *Excelu*.

Suma wartości w kolumnie $|\dot{x}_i - \bar{x}|n_i$ wynosi 1197,68. Dzieląc przez n = 80 otrzymamy: $D_x \approx 14{,}97$.

Wariancję populacji dla szeregu szczegółowego obliczamy ze wzoru:

$$S_x^{\,2} = \frac{\sum_{1}^{n} (x_i - \bar{x})^2}{n}$$

Przykład: Dla danych z pomiaru wagi studentów obliczamy wariancję podnosząc do kwadratu wartości obliczeniowe z tabeli 29.2, np.: (=4,96^2) co daje 24,60.

Tab. 29.4

Obliczenia pomocnicze wariancji szeregu szczegółowego

24,60	**194,88**	**530,84**	**167,96**	**287,64**	**398,40**	**144,96**	**16,32**
63,36	527,16	25,40	99,20	167,96	63,36	80,28	48,44
897,60	482,24	1158,72	170,04	81,72	170,04	15,68	442,68
362,52	80,28	843,32	673,92	48,44	25,40	254,72	16,32
16,32	623,00	627,00	170,04	731,16	63,36	362,52	120,12
8,76	287,64	197,12	731,16	167,96	673,92	80,28	170,04
4,16	48,44	398,40	4,16	322,56	527,16	1767,36	170,04
63,36	257,28	63,36	627,00	4,16	80,28	1684,28	197,12
120,12	627,00	167,96	194,88	25,40	322,56	623,00	48,44
15,68	63,36	2212,76	322,56	167,96	287,64	530,84	442,68

1576,50	3191,30	6224,90	3160,94	2004,98	2612,14	5543,94	1672,22
Razem:	25986,89						
S_x^2	**324,84**						

Źródło: Opracowanie własne w *Excelu*.

Ostatni wiersz stanowią sumy kolumnowe, a suma z tych sum (25986,89) podzielona przez n = 80 daje $S_x^2 \approx 324{,}84$. Dla szeregu rozdzielczego obliczenie wariancji populacji następuje według danych z tabeli 29.3 korzystając ze wzoru:

$$S_x^2 = \frac{\sum_{i=1}^{K} (\dot{x}_i - \bar{x}) n_i}{n}$$

Czasami w obliczeniach stosowana jest wariancja próby dla szeregu rozdzielczego.

$$S_x^2 = \frac{\sum_{i=1}^{K} (\dot{x}_i - \bar{x})^2 n_i}{n-1}$$

Przykład: Wykorzystamy dane w kolumnie $(\dot{x}_i - \bar{x})^2 n_i$ w tabeli 29.3. Suma wyników w tej kolumnie daje 26227,73, co po podzieleniu przez 80 wynosi $S_x^2 = 327{,}85$. Gdy podzielimy przez $(n - 1) = 79$, to wówczas $S_x^2 = 332$.

Do określenia miary rozproszenia używa się często odchylenia standardowego S_x:

$$S_x = \sqrt{S_x^2}$$

Zatem dla powyższego przykładu szeregu szczegółowego oraz rozdzielczego przykładowo dla populacji:

$$S_x = \sqrt{324{,}84} = 18{,}02 \qquad S_x = \sqrt{327{,}85} = 18{,}11$$

Natomiast przyjmując, że obserwacje z stanowią próbę to:

$$S_x = \sqrt{332} = 18{,}22$$

Współczynnik zmienności V_x , który stanowi iloraz odchylenia standardowego do wartości średniej zmiennej obliczmy według wzoru:

$$V_x = \frac{S_x}{\bar{x}}$$

Przykład:

$$V_x = \frac{18{,}22}{72{,}96} = 0{,}25$$

Natomiast względem miar pozycyjnych, przykładowo dla szeregu szczegółowego danych z pomiaru wagi studentów:

$$V_x = \frac{Q_x}{M_x} \qquad V_x = \frac{13}{66} = 0{,}2$$

29.3. Zastosowanie funkcji Excela do wyznaczania miar statystyki

Zastosowanie menu *Narzędzia-Analiza danych-Statystyka opisowa.* Zamiast obliczać "ręcznie" miary rozkładu jednej zmiennej można posłużyć się programem *Excel* i skorzystać z opcji: *Narzędzia-Analiza danych-Statystyka opisowa*. W opcji tej podajemy:

- zakres wejściowy stosu danych źródłowych uporządkowany narastająco i zakres wyjściowy jako nowy arkusz,
- zaznaczamy statystyki podsumowujące,
- *k*-tą największą oraz *k*-tą najmniejszą.

Opcja ta umożliwia obliczenie szeregu miar w tym wcześniej wymienionych. Niektóre miary obliczane są dla próby tj. (*n* -1) = (80 - 1) = 79 obserwacji, co zaznaczono w tabeli 29.4.

Tab. 29.4

Statystyka opisowa zmiennej X

Nazwa cechy	Wartość
Średnia (dla n = 80)	72,96
Mediana	66
Odchylenie standardowe (n-1)	18,14
Wariancja próbki (n-1)	328,95
Kurtoza (n-1)	-0,44
Skośność	0,59
Zakres (rozstęp)	77
Minimum	43
Maksimum	120
Suma (wartości obserwacji)	5837
Licznik n)	80
Największa wartość obserwacji	120
Najmniejsza wartość obserwacji	43

Źródło: Opracowanie własne w *Excelu*.

Sformułowanie składni funkcji elementarnych Excela z zakresu statystyki opisowej

Tab. 29.5

Funkcji Excela dla regresji i opisu statystyki

Nazwa	Składnia	Przykład
Parametry linii *y = mx + b*	REGLINP(znane_*y*;znane_*x*;stała;stats)	REGLINP(A38:A43;B38:B43;PRAWDA;PRAWDA) Y = 65142,86 * x - 729428,57
Wyznaczenie kolejnej wartości wg linii	REGLINW(znane_*y*;znane_*x*;nowe_*x*;;const)	REGLINW(A38:A43;B38:B43;B39:B40;PRAWDA) np. 2-wa m-ce 17 i 18
Wyznaczenie prognozy wg linii	REGLINX(*x*;znane_*y*;znane_*x*)	REGLINX(B39;A38:A43;B38:B43) np. B39 = 20
Parametry krzywej wykładniczej *y = b * mx*	REGEXPP(znane_*y*;znane_*x*;stała;stats) m:INDEKS(REGEXPP(znane_*y*;znane_*x*);1) b:INDEKS(REGEXPP(znane_*y*;znane_*x*);2)	REGEXPP(A38:A43;B38:B43;PRAWDA;PRAWDA) Y= 118,456 * 1,655^{x}
Wyznaczenie kolejnej wartości wg krzywej	REGEXPW(znane_*y*;znane_*x*;nowe_*x*)	REGEXPW(A38:A43;B38:B43;B39:B40:{17;18})) np. 2-wa m-ce 17 i 18
Odcięta "b" linii	ODCIĘTA(znane_*y*;znane_*x*)	ODCIĘTA({4;2;4;5;6};{13;19;12;15;17})

$Y = ax + b$		
Nachylenie "*a*" linii	NACHYLENIE(znane_*y*;znane_*x*)	NACHYLENIE({4;2;4;5;6};{13;19;12;15;17})
Wariancja próby dla *n* -1	WARIANCJA(liczba1;liczba2;...)	WARIANCJA(3;4;5;8;91;2;4;5) = 6,78
Wariancja populacji *n*	WARIANCJA.POPUL(liczba1;liczba2;...)	WARIANCJA.POPUL(3;4;5;8;91;2;4;5) = 6,0247
Odchylenie standardowe próby dla (*n* – 1)	ODCH.STANDARDOWE(liczba1;liczba2;...)	ODCH.STANDARDOWE(4;2;4;5;6) = 1,4832397
Odchylenie standardowe populacji *n*	ODCH.STAND.POPUL(liczba1;liczba2;...)	ODCH.STAND.POPUL(4;2;4;5;6) = 1,3266499
Średnia arytmetyczna	ŚREDNIA(liczba1;liczba2;...) lub SUMA(A1:E1)/ILE.LICZB(A1:E1)	ŚREDNIA(8;6;9;5;2) = 6
Standaryzacja wartości "*x*"	NORMALIZUJ(x;średnia;standard_odch)	NORMALIZUJ(4;3;1,2)
Odchylenie standardowe reszt dla "*n*"	REGBŁSTD(znane_y;znane_x)	REGBŁSTD(B20:B35;C20:C35)
Test F dwóch wariancji	TEST.F(tablica1;tablica2)	TEST.F({3;4;5;8;9;1;2;4;5}; {6;19;3;2;1;4;5;17;1}) = 0,013749
Test *t* dwóch średnich	TEST.T(tablica1;tablica2;strony; typ) strony = 2 - dwustronny obszar krytyczny typ - 1 znane wariancje dwóch populacji	TEST.T({3;4;5;8;9;1;2;4;5};{6;19;3;2;1;4;5;17;1};2;1) = 0,4975584 WARIANCJA(3;4;5;8;9;1;2;4;5) = 6,78 WARIANCJA(6;19;3;2;1;4;5;17;1) = 46,03
Kowariancja	KOWARIANCJA(tabela1;tabela2)	KOWARIANCJA({4;2;4;5;6};{13;19;12;15;17}) = - 0,84
Współczynnik korelacji *R*	WSP.KORELACJI(tablica1;tablica2) lub PEARSON(tablica1;tablica2)	WSP.KORELACJI(A24:A39;B24:B39) PEARSON(A24:A39;B24:B39)
Kwadrat współczynnika korelacji R^2	R.KWADRAT(znane_*y*;znane_*x*) Współczynnik determinacji	R.KWADRAT({4;2;4;5;6};{13;19;12;15;17}) = 0,0611142 - 6,1 %
Dopasowany R kwadrat R^{-2}	Funkcje elementarne Excel-a dla 1-nej zmiennej obliczenia wg wzoru: $R^{-2} = 1 - [(n - 1)/(n - 2) * (1- R^2)$	1-(16-1)/(16-2)*(1-0,339667) = 0,2925

Źródło: Opracowanie własne na podstawie informacji z *Pomocy* programu *Excel*.

29.4. Miary asymetrii

Klasyczny współczynnik asymetrii. Miary asymetrii (skośności) rozkładu jednej zmiennej służą do badania czy obserwacje są równomiernie rozłożone wokół wartości przeciętnej. Określony jest dla szeregu szczegółowego według:

$$A_s = \frac{(1/n) * \sum_{i=1}^{n} (x_i - \bar{x})^3}{s_x^{\,3}}$$

Dla rozkładów prawostronnie asymetrycznych A_s przyjmuje wartości dodatnie, natomiast dla lewostronnie jest wielkością ujemną.

Pozycyjny współczynnik asymetrii. Bazuje na kwartylach Q_1 i Q_3 oraz medianie M_x:

$$A_Q = \frac{Q_3 - 2 * M_x + Q_1}{Q_3 - Q_1}$$

A_Q przyjmuje wartości w przedziale <-1, 1>, natomiast A_s z przedziału <-∞, ∞).

Przykład: Obliczmy klasyczny współczynnik asymetrii dla szeregu szczegółowego n = 80. obserwacji wagi studentów, gdzie $\bar{x}$ = 72,96, wyznaczając najpierw:

$$\sum_{i=1}^{n} (x_i - \bar{x})^3$$

które wynosi 271585,43. Dla n = 80 i wcześniej obliczonego S_x = 18,02. Zatem jest to rozkład prawostronny. A teraz określmy pozycyjny współczynnik asymetrii na podstawie kwartyli z pomiaru wagi studentów na podstawie szeregu szczegółowego oraz rozdzielczego.

$$A_s = \frac{(1/80) * 271585,43}{18,02^{\,3}} = 0,58$$

Szereg szczegółowy: Q_1 = 60; M_x = 66; Q_3 = 86.

$$A_Q = \frac{86 - 2 * 66 + 60}{86 - 60} = 0,54$$

Szereg rozdzielczy bazuje na wcześniejszych obliczeniach: Q_1 = 60,4 ; M_x = 68,6; Q_3 = 86.

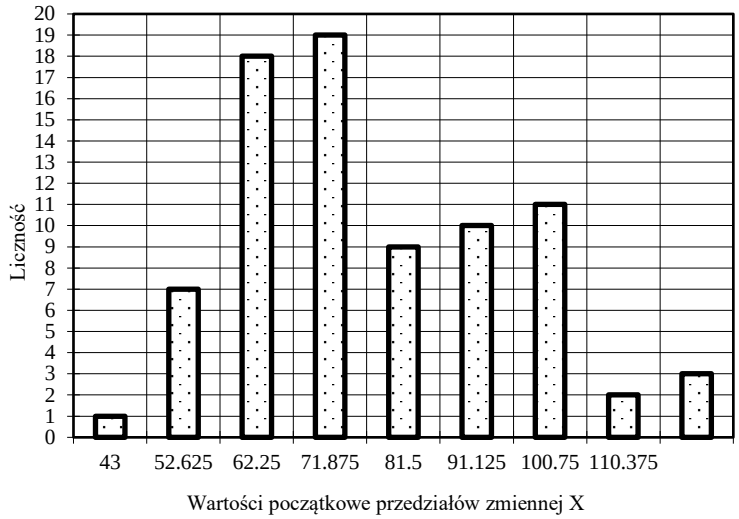
$$A_Q = \frac{86 - 2 * 68,6 + 60,4}{86 - 60,4} = 0,36$$

Histogram. Posługując się opcją *Narzędzia-Excel-Analiza* dla danych przykładu otrzymujemy tabelę 29.6 z zakresem przedziałów co 9,6 oraz histogram - zbiór prostokątów (rysunek 29.6).

Tab. 29.6

Liczność występowania wartości obserwacji - określona w *Excelu*

Rozpoczęcie przedziału	Liczność
43,0	1
52,6	7
62,3	18
71,9	19
81,5	9
91,1	10
100,8	11
110,4	2
Więcej	3



Źródło: Opracowanie własne w *Excelu*.

Rys. 29.3. Histogram zmiennej *X*

Gdy porównamy ze średnią o wartości 72,96 to widzimy wyraźnie rozkład prawostronnie asymetryczny. Gdy połączymy linią łamaną środki wierzchołków prostokątów uzyskujemy linię łamana zwaną diagramem.

29.5. Wybrane tabele statystyczne

29.6. Dystrybuanta rozkładu normalnego[221]
29.7. Wartości krytyczne $t(P, n)$ rozkładu *t*- Studenta
29.8. Wartości krytyczne $F(\alpha, r_1, r_2)$ rozkładu *F*
29.9.Wartości krytyczne χ^2 (α, r) rozkładu χ^2
29.10.Wartości krytyczne rozkładu Durbina-Watsona dla $\alpha = 0{,}05$[222]

[221] Tabele: 29.6-29.9, Zieliński R., *Tablice statystyczne*, PWN, Warszawa 1972.
[222] Tabela 29.10, Lipiec-Zajchowska (red.) *Wspomaganie procesów decyzyjnych*, tom II. *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003, s. 219.

Tab. 29.6

Dystrybuanta rozkładu normalnego

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{u} e^{-u^2/2} du$$

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,500 000	0,503 989	0,507 978	0,511 966	0,515 953	0,519 939	0,523 922	0,527 903	0,531 881	0,535 856
0,1	0,539 828	0,543 795	0,547 758	0,551 717	0,555 670	0,559 618	0,563 559	0,567 495	0,571 424	0,575 345
0,2	0,579 260	0,583 166	0,587 064	0,590 954	0,594 835	0,598 706	0,602 568	0,606 420	0,610 261	0,614 092
0,3	0,617 911	0,621 720	0,625 516	0,629 300	0,633 072	0,636 831	0,640 576	0,644 309	0,648 027	0,651 732
0,4	0,655 422	0,659 097	0,662 757	0,666 402	0,670 031	0,673 645	0,677 242	0,680 822	0,684 386	0,687 933
0,5	0,691 462	0,694 974	0,698 468	0,701 944	0,705 401	0,708 840	0,712 260	0,715 661	0,719 043	0,722 405
0,6	0,725 747	0,729 069	0,732 371	0,735 653	0,738 914	0,742 154	0,745 373	0,748 571	0,751 748	0,754 903
0,7	0,758 036	0,761 148	0,764 238	0,767 305	0,770 350	0,773 373	0,776 373	0,779 350	0,782 305	0,785 236
0,8	0,788 145	0,791 030	0,793 892	0,796 731	0,799 546	0,802 337	0,805 105	0,807 850	0,810 570	0,813 267
0,9	0,815 940	0,818 589	0,821 214	0,823 814	0,826 391	0,828 944	0,831 472	0,833 977	0,836 457	0,838 913
1,0	0,841 345	0,843 752	0,846 136	0,848 495	0,850 830	0,853 141	0,855 428	0,857 690	0,859 929	0,862 143
1,1	0,864 334	0,866 500	0,868 643	0,870 762	0,872 857	0,874 928	0,876 976	0,879 000	0,881 000	0,882 977
1,2	0,884 930	0,886 861	0,888 768	0,890 651	0,892 512	0,894 350	0,896 165	0,897 958	0,899 727	0,901 475
1,3	0,903 200	0,904 902	0,906 582	0,908 241	0,908 877	0,911 492	0,913 085	0,914 657	0,916 207	0,917 736
1,4	0,919 243	0,920 730	0,922 196	0,923 641	0,925 066	0,926 471	0,927 855	0,929 219	0,930 563	0,931 888
1,5	0,933 193	0,934 478	0,935 745	0,936 992	0,938 220	0,939 429	0,940 620	0,941 792	0,942 947	0,944 083
1,6	0,945 201	0,946 301	0,947 384	0,948 449	0,949 497	0,950 529	0,951 543	0,952 540	0,953 521	0,954 486
1,7	0,955 435	0,956 367	0,957 284	0,958 185	0,957 070	0,959 941	0,960 796	0,961 636	0,962 462	0,963 273
1,8	0,964 070	0,964 852	0,965 620	0,966 375	0,967 116	0,967 843	0,968 557	0,969 258	0,969 946	0,970 621
1,9	0,971 283	0,971 933	0,972 571	0,973 197	0,973 810	0,974 412	0,975 002	0,975 581	0,976 148	0,976 705
2,0	0,977 250	0,977 784	0,978 308	0,978 822	0,979 325	0,979 818	0,980 301	0,980 774	0,981 237	0,981 691
2,1	0,982 136	0,982 571	0,982 997	0,983 414	0,983 823	0,984 222	0,984 614	0,984 997	0,985 371	0,985 738
2,2	0,986 097	0,986 447	0,986 791	0,987 126	0,987 455	0,987 776	0,988 089	0,988 396	0,988 696	0,988 989
2,3	0,989 276	0,989 556	0,989 830	0,990 097	0,990 358	0,990 613	0,990 863	0,991 106	0,991 344	0,991 576
2,4	0,991 802	0,992 024	0,992 240	0,992 451	0,992 656	0,992 857	0,993 053	0,993 244	0,993 431	0,993 613
2,5	0,993 790	0,993 963	0,994 132	0,994 297	0,994 457	0,994 614	0,994 766	0,994 915	0,995 060	0,995 201
2,6	0,995 339	0,995 473	0,995 604	0,995 731	0,995 855	0,995 875	0,996 093	0,996 207	0,996 319	0,996 427
2,7	0,996 533	0,996 636	0,996 736	0,996 833	0,996 926	0,997 020	0,997 110	0,997 197	0,997 282	0,997 365
2,8	0,997 445	0,997 523	0,997 599	0,997 673	0,997 744	0,997 814	0,997 882	0,997 948	0,998 012	0,998 074
2,9	0,998 134	0,998 193	0,998 250	0,998 305	0,998 359	0,998 411	0,998 462	0,998 511	0,998 559	0,998 605
3,0	0,998 650	0,998 694	0,998 736	0,998 777	0,998 817	0,998 856	0,998 893	0,998 930	0,998 965	0,998 999
3,1	$0,9^3 0\,324$	$0,9^3 0\,646$	$0,9^3 0\,957$	$0,9^3 1\,260$	$0,9^3 1\,553$	$0,9^3 1\,836$	$0,9^3 2\,112$	$0,9^3 2\,378$	$0,9^3 2\,636$	$0,9^3 2\,886$
3,2	$0,9^3 3\,129$	$0,9^3 3\,363$	$0,9^3 3\,590$	$0,9^3 3\,810$	$0,9^3 4\,024$	$0,9^3 4\,230$	$0,9^3 4\,429$	$0,9^3 4\,623$	$0,9^3 4\,810$	$0,9^3 4\,991$
3,3	$0,9^3 5\,166$	$0,9^3 5\,335$	$0,9^3 5\,499$	$0,9^3 5\,658$	$0,9^3 5\,811$	$0,9^3 5\,959$	$0,9^3 6\,103$	$0,9^3 6\,242$	$0,9^3 6\,376$	$0,9^3 6\,505$
3,4	$0,9^3 6\,631$	$0,9^3 6\,752$	$0,9^3 6\,869$	$0,9^3 6\,982$	$0,9^3 7\,091$	$0,9^3 7\,197$	$0,9^3 7\,299$	$0,9^3 7\,398$	$0,9^3 7\,493$	$0,9^3 7\,585$
3,5	$0,9^3 7\,674$	$0,9^3 7\,759$	$0,9^3 7\,842$	$0,9^3 7\,922$	$0,9^3 7\,999$	$0,9^3 8\,074$	$0,9^3 8\,146$	$0,9^3 8\,215$	$0,9^3 8\,282$	$0,9^3 8\,347$
3,6	$0,9^3 8\,409$	$0,9^3 8\,469$	$0,9^3 8\,527$	$0,9^3 8\,583$	$0,9^3 8\,637$	$0,9^3 8\,689$	$0,9^3 8\,739$	$0,9^3 8\,787$	$0,9^3 8\,834$	$0,9^3 8\,879$
3,7	$0,9^3 8\,922$	$0,9^3 8\,964$	$0,9^4 0\,039$	$0,9^4 0\,426$	$0,9^4 0\,799$	$0,9^4 1\,158$	$0,9^4 1\,504$	$0,9^4 1\,858$	$0,9^4 2\,159$	$0,9^4 2\,468$
3,8	$0,9^4 2\,765$	$0,9^4 3\,052$	$0,9^4 3\,327$	$0,9^4 3\,593$	$0,9^4 3\,848$	$0,9^4 4\,094$	$0,9^4 4\,331$	$0,9^4 4\,558$	$0,9^4 4\,777$	$0,9^4 4\,988$
3,9	$0,9^4 5\,190$	$0,9^4 5\,385$	$0,9^4 5\,573$	$0,9^4 5\,753$	$0,9^4 5\,926$	$0,9^4 6\,092$	$0,9^4 6\,253$	$0,9^4 6\,406$	$0,9^4 6\,554$	$0,9^4 6\,696$
4,0	$0,9^4 6\,833$	$0,9^4 6\,964$	$0,9^4 7\,090$	$0,9^4 7\,211$	$0,9^4 7\,327$	$0,9^4 7\,439$	$0,9^4 7\,546$	$0,9^4 7\,649$	$0,9^4 7\,748$	$0,9^4 7\,843$
4,1	$0,9^4 7\,934$	$0,9^4 8\,022$	$0,9^4 8\,106$	$0,9^4 8\,186$	$0,9^4 8\,263$	$0,9^4 8\,338$	$0,9^4 8\,409$	$0,9^4 8\,477$	$0,9^4 8\,542$	$0,9^4 8\,605$
4,2	$0,9^4 8\,665$	$0,9^4 8\,723$	$0,9^4 8\,778$	$0,9^4 8\,832$	$0,9^4 8\,882$	$0,9^4 8\,931$	$0,9^4 8\,978$	$0,9^5 0\,226$	$0,9^5 0\,655$	$0,9^5 1\,066$
4,3	$0,9^5 1\,460$	$0,9^5 1\,837$	$0,9^5 2\,199$	$0,9^5 2\,545$	$0,9^5 2\,876$	$0,9^5 3\,193$	$0,9^5 3\,497$	$0,9^5 3\,788$	$0,9^5 3\,066$	$0,9^5 4\,332$
4,4	$0,9^5 4\,587$	$0,9^5 4\,831$	$0,9^5 5\,065$	$0,9^5 5\,288$	$0,9^5 5\,502$	$0,9^5 5\,706$	$0,9^5 5\,902$	$0,9^5 6\,089$	$0,9^5 6\,268$	$0,9^5 6\,439$
4,5	$0,9^5 6\,602$	$0,9^5 6\,759$	$0,9^5 6\,908$	$0,9^5 7\,051$	$0,9^5 7\,187$	$0,9^5 7\,318$	$0,9^5 7\,442$	$0,9^5 7\,561$	$0,9^5 7\,675$	$0,9^5 7\,784$
4,6	$0,9^5 7\,888$	$0,9^5 7\,987$	$0,9^5 8\,081$	$0,9^5 8\,172$	$0,9^5 8\,258$	$0,9^5 8\,340$	$0,9^5 8\,419$	$0,9^5 8\,494$	$0,9^5 8\,566$	$0,9^5 8\,634$
4,7	$0,9^5 8\,699$	$0,9^5 8\,761$	$0,9^5 8\,821$	$0,9^5 8\,877$	$0,9^5 8\,931$	$0,9^5 8\,983$	$0,9^5 0\,320$	$0,9^5 0\,789$	$0,9^5 1\,235$	$0,9^5 1\,661$
4,8	$0,9^6 2\,067$	$0,9^6 2\,453$	$0,9^6 2\,822$	$0,9^6 3\,173$	$0,9^6 3\,508$	$0,9^6 3\,827$	$0,9^6 4\,131$	$0,9^6 4\,420$	$0,9^6 4\,696$	$0,9^6 4\,958$
4,9	$0,9^6 5\,208$	$0,9^6 5\,446$	$0,9^6 5\,673$	$0,9^6 5\,889$	$0,9^6 6\,094$	$0,9^6 6\,289$	$0,9^6 6\,475$	$0,9^6 6\,652$	$0,9^6 6\,821$	$0,9^6 6\,981$

Wartości krytyczne $t(P, n)$ rozkładu t-Studenta; $P\{|t_n| > t(P, n)\} = P$

n / P	0,5	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	318,3088	636,6192
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	22,3271	31,5991
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	10,2145	12,9240
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688
6	0,7176	1,5398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	4,7853	5,4079
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2969	4,7809
10	0,6998	1,3732	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7884	4,1405
15	0,6912	1,3406	1,7530	2,1314	2,6025	2,9457	3,2860	3,7328	4,0728
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651
18	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216
19	0,6876	1,3277	1,7291	2,0930	2,5395	2,9609	3,1737	3,5794	3,8834
20	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495
21	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,4668	3,7455
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066
27	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896
28	0,6834	1,3125	1,7911	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739
29	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594
30	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
32	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,3653	3,6218
34	0,6818	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,3479	3,6007
36	0,6814	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,3326	3,5821
38	0,6810	1,3042	1,6890	2,0244	2,4286	2,7116	2,9803	3,3190	3,5657
40	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510
42	0,6804	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,2960	3,5377
44	0,6801	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,2861	3,5258
46	0,6799	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,2771	3,5150
48	0,6796	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,2689	3,5051
50	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,2614	3,4960
55	0,6790	1,2971	1,6730	2,0040	2,3961	2,6682	2,9247	3,2561	3,4764
60	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,2317	3,4602
65	0,6783	1,2947	1,6686	1,9971	2,3851	2,6535	2,9060	3,2204	3,4466
70	0,6780	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,2108	3,4350
80	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163
90	0,6772	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,1833	3,4019
100	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,1737	3,3905
120	0,6765	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,1595	3,3735
150	0,6761	1,2872	1,6551	1,9759	2,3515	2,6090	2,8492	3,1455	3,3566
200	0,6757	1,2858	1,6525	1,9719	2,3451	2,6006	2,8385	3,1315	3,3398
250	0,6755	1,2849	1,6510	1,9695	2,3414	2,5956	2,8322	3,1232	3,3299
300	0,6753	1,2844	1,6499	1,9679	2,3388	2,5923	2,8279	3,1176	3,3233
400	0,6751	1,2837	1,6487	1,9659	2,3357	2,5882	2,8227	3,1107	3,3150
500	0,6750	1,2832	1,6479	1,9647	2,3338	2,5887	2,8195	3,1066	3,3101
∞	0,6745	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905

Tab. 29.8

Wartości krytyczne $F(\alpha, r_1, r_2)$ rozkładu F dla $\alpha = 0{,}05$

r_1 / r_2	1	2	3	4	5	6	7	8	9	10	12	14	16	18	20
1	161	200	216	225	230	234	237	239	241	242	244	245	246	247	248
2	18,5	19,0	19,2	19,2	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,4
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,71	8,69	8,67	8,66
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,87	5,84	5,82	5,80
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,64	4,60	4,58	4,56
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,96	3,92	3,90	3,87
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,53	3,49	3,47	3,44
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,24	3,20	3,17	3,15
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,03	2,99	2,96	2,94
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,86	2,83	2,80	2,77
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,74	2,70	2,67	2,65
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,64	2,60	2,57	2,54
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,55	2,51	2,48	2,46
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,48	2,44	2,41	2,39
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,42	2,38	2,35	2,33
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,37	2,33	2,30	2,28
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,33	2,29	2,26	2,23
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,29	2,25	2,22	2,19
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,26	2,21	2,18	2,16
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,22	2,18	2,15	2,12
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,20	2,16	2,12	2,10
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,17	2,13	2,10	2,07
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,15	2,11	2,07	2,05
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,13	2,09	2,05	2,03
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,11	2,07	2,04	2,01
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,09	2,05	2,02	1,99
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,08	2,04	2,00	1,97
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,06	2,02	1,99	1,96
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,05	2,01	1,97	1,94
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,04	1,99	1,96	1,93
32	4,15	3,29	2,90	2,67	2,51	2,40	2,31	2,24	2,19	2,14	2,07	2,01	1,97	1,94	1,91
34	4,13	3,28	2,88	2,65	2,49	2,38	2,29	2,23	2,17	2,12	2,05	1,99	1,95	1,92	1,89
36	4,11	3,26	2,87	2,63	2,48	2,36	2,28	2,21	2,15	2,11	2,03	1,98	1,93	1,90	1,87
38	4,10	3,24	2,85	2,62	2,46	2,35	2,26	2,19	2,14	2,09	2,02	1,96	1,92	1,88	1,85
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,95	1,90	1,87	1,84
42	4,07	3,22	2,83	2,59	2,44	2,32	2,24	2,17	2,11	2,06	1,99	1,93	1,89	1,86	1,83
44	4,06	3,21	2,82	2,58	2,43	2,31	2,23	2,16	2,10	2,05	1,98	1,92	1,88	1,84	1,81
46	4,05	3,20	2,81	2,57	2,42	2,30	2,22	2,15	2,09	2,04	1,97	1,91	1,87	1,83	1,80
48	4,04	3,19	2,80	2,57	2,41	2,29	2,21	2,14	2,08	2,03	1,96	1,90	1,86	1,82	1,79
50	4,03	3,18	2,79	2,56	2,40	2,28	2,20	2,13	2,07	2,03	1,95	1,89	1,85	1,81	1,78
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,86	1,82	1,78	1,75
80	3,96	3,11	2,72	2,49	2,33	2,21	2,13	2,06	2,00	1,95	1,88	1,82	1,77	1,73	1,70
100	3,94	3,09	2,70	2,46	2,31	2,19	2,10	2,03	1,97	1,93	1,85	1,79	1,75	1,71	1,68
125	3,92	3,07	2,68	2,44	2,29	2,17	2,08	2,01	1,96	1,91	1,83	1,77	1,72	1,69	1,65
150	3,90	3,06	2,66	2,43	2,27	2,16	2,07	2,00	1,94	1,89	1,82	1,76	1,71	1,67	1.64
200	3,89	3,04	2,65	2,42	2,26	2,14	2,06	1,98	1,93	1,88	1,80	1,74	1,69	1,66	1,62
300	3,87	3,03	2,63	2,40	2,24	2,13	2,04	1,97	1,91	1,86	1,78	1,72	1,68	1,64	1,61
500	3,86	3,01	2,62	2,39	2,23	2,12	2,03	1,96	1,90	1,85	1,77	1,71	1,66	1,62	1,59
1000	3,85	3,00	2,61	2,38	2,22	2,11	2,02	1,95	1,89	1,84	1,76	1,70	1,65	1,61	1,58
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,69	1,64	1,60	1,57

Wartości krytyczne $F(\alpha, r_1, r_2)$ rozkładu F dla $\alpha = 0{,}05$

r_1 / r_2	22	24	26	28	30	35	40	45	50	60	80	100	200	500	∞
1	249	249	249	250	250	251	251	251	252	252	252	253	254	254	254
2	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5	19,5
3	8,65	8,64	8,63	8,62	8,62	8,60	8,59	8,59	8,58	8,57	8,56	8,55	8,54	8,53	8,53
4	5,79	5,77	5,76	5,75	5,75	5,73	5,72	5,71	5,70	5,69	5,67	5,66	5,65	5,64	5,63
5	4,54	4,53	4,52	4,50	4,50	4,48	4,46	4,45	4,44	4,43	4,41	4,41	4,39	4,37	4,37
6	3,86	3,84	3,83	3,82	3,81	3,79	3,77	3,76	3,75	3,74	3,72	3,71	3,69	3,68	3,67
7	3,43	3,41	3,40	3,39	3,38	3,36	3,34	3,33	3,32	3,30	3,29	3,27	3,25	3,24	3,23
8	3,13	3,12	3,10	3,09	3,08	3,06	3,04	3,03	3,02	3,01	2,99	2,97	2,95	2,94	2,93
9	2,92	2,90	2,89	2,87	2,86	2,84	2,83	2,81	2,80	2,79	2,77	2,76	2,73	2,72	2,71
10	2,75	2,74	2,72	2,71	2,70	2,68	2,66	2,65	2,64	2,62	2,60	2,59	2,56	2,55	2,54
11	2,63	2,61	2,59	2,58	2,57	2,55	2,53	2,52	2,51	2,49	2,47	2,46	2,43	2,42	2,40
12	2,52	2,51	2,49	2,48	2,47	2,44	2,43	2,41	2,40	2,38	2,36	2,35	2,32	2,31	2,30
13	2,44	2,42	2,41	2,39	2,38	2,36	2,34	2,33	2,31	2,30	2,27	2,26	2,23	2,22	2,21
14	2,37	2,35	2,33	2,32	2,31	2,28	2,27	2,25	2,24	2,22	2,20	2,19	2,16	2,14	2,13
15	2,31	2,29	2,27	2,26	2,25	2,22	2,20	2,19	2,18	2,16	2,14	2,12	2,10	2,08	2,07
16	2,25	2,24	2,22	2,21	2,19	2,17	2,15	2,14	2,12	2,11	2,08	2,07	2,04	2,02	2,01
17	2,21	2,19	2,17	2,16	2,15	2,12	2,10	2,09	2,08	2,06	2,03	2,02	1,99	1,97	1,96
18	2,17	2,15	2,13	2,12	2,11	2,08	2,06	2,05	2,04	2,02	1,99	1,98	1,95	1,93	1,92
19	2,13	2,11	2,10	2,08	2,07	2,05	2,03	2,01	2,00	1,98	1,96	1,94	1,91	1,89	1,88
20	2,10	2,08	2,07	2,05	2,04	2,01	1,99	1,98	1,97	1,95	1,92	1,91	1,88	1,86	1,84
21	2,07	2,05	2,04	2,02	2,01	1,98	1,96	1,95	1,94	1,92	1,89	1,88	1,84	1,82	1,81
22	2,05	2,03	2,01	2,00	1,98	1,96	1,94	1,92	1,91	1,89	1,86	1,85	1,82	1,80	1,78
23	2,02	2,00	1,99	1,97	1,96	1,93	1,91	1,90	1,88	1,86	1,84	1,82	1,79	1,77	1,76
24	2,00	1,98	1,97	1,95	1,94	1,91	1,89	1,88	1,86	1,84	1,82	1,80	1,77	1,75	1,73
25	1,98	1,96	1,95	1,93	1,92	1,89	1,87	1,86	1,84	1,82	1,80	1,78	1,75	1,73	1,71
26	1,97	1,95	1,93	1,91	1,90	1,87	1,85	1,84	1,82	1,80	1,78	1,76	1,73	1,71	1,69
27	1,95	1,93	1,91	1,90	1,88	1,86	1,84	1,82	1,81	1,79	1,76	1,74	1,71	1,69	1,67
28	1,93	1,91	1,90	1,88	1,87	1,84	1,82	1,80	1,79	1,77	1,74	1,73	1,69	1,67	1,65
29	1,92	1,90	1,88	1,87	1,85	1,83	1,81	1,79	1,77	1,75	1,73	1,71	1,67	1,65	1,64
30	1,91	1,89	1,87	1,85	1,84	1,81	1,79	1,77	1,76	1,74	1,71	1,70	1,66	1,64	1,62
32	1,88	1,86	1,85	1,83	1,82	1,79	1,77	1,75	1,74	1,71	1,69	1,67	1,63	1,61	1,59
34	1,86	1,84	1,82	1,80	1,80	1,77	1,75	1,73	1,71	1,69	1,66	1,65	1,61	1,59	1,57
36	1,85	1,82	1,81	1,79	1,78	1,75	1,73	1,71	1,69	1,67	1,64	1,62	1,59	1,56	1,55
38	1,83	1,81	1,79	1,77	1,76	1,73	1,71	1,69	1,68	1,65	1,62	1,61	1,57	1,54	1,53
40	1,81	1,79	1,77	1,76	1,74	1,72	1,69	1,67	1,66	1,64	1,61	1,59	1,55	1,53	1,51
42	1,80	1,78	1,76	1,74	1,73	1,70	1,68	1,66	1,65	1,62	1,59	1,57	1,53	1,51	1,49
44	1,79	1,77	1,75	1,73	1,72	1,69	1,67	1,65	1,63	1,61	1,58	1,56	1,52	1,49	1,48
46	1,78	1,76	1,74	1,72	1,71	1,68	1,65	1,64	1,62	1,60	1,57	1,55	1,51	1,48	1,46
48	1,77	1,75	1,73	1,71	1,70	1,67	1,64	1,62	1,61	1,59	1,56	1,54	1,49	1,47	1,45
50	1,76	1,74	1,72	1,70	1,69	1,66	1,63	1,61	1,60	1,58	1,54	1,52	1,48	1,46	1,44
60	1,72	1,70	1,68	1,66	1,65	1,62	1,59	1,57	1,56	1,53	1,50	1,48	1,44	1,41	1,39
80	1,68	1,65	1,63	1,62	1,60	1,57	1,54	1,52	1,51	1,48	1,45	1,43	1,38	1,35	1,32
100	1,65	1,63	1,61	1,59	1,57	1,54	1,52	1,49	1,48	1,45	1,41	1,39	1,34	1,31	1,28
125	1,63	1,60	1,58	1,57	1,55	1,52	1,49	1,47	1,45	1,42	1,39	1,36	1,31	1,27	1,25
150	1,61	1,59	1,57	1,55	1,53	1,50	1,48	1,45	1,44	1,41	1,37	1,34	1,29	1,25	1,22
200	1,60	1,57	1,55	1,53	1,52	1,48	1,46	1,43	1,41	1,39	1,35	1,32	1,26	1,22	1,19
300	1,58	1,55	1,53	1,51	1,50	1,46	1,43	1,41	1,39	1,36	1,32	1,30	1,23	1,19	1,15
500	1,56	1,54	1,52	1,50	1,48	1,45	1,42	1,40	1,38	1,34	1,30	1,28	1,21	1,16	1,11
1000	1,55	1,53	1,51	1,49	1,47	1,44	1,41	1,38	1,36	1,33	1,29	1,26	1,19	1,13	1,08
∞	1,54	1,52	1,50	1,48	1,46	1,42	1,39	1,37	1,35	1,32	1,27	1,24	1,17	1,11	1,00

Tab. 29.9

Wartości krytyczne χ^2 rozkładu χ^2 ; $P\{\chi^2_r \geq \chi^2(\alpha, r)\} = \alpha$

r /α	0,9995	0,999	0,995	0,99	0,975	0,95	0,90	0,80	0,70	0,60
1	$0{,}0^6393$	$0{,}0^5157$	$0{,}0^4393$	$0{,}0^3157$	$0{,}0^3982$	$0{,}0^2393$	0,0158	0,0642	0,148	0,275
2	$0{,}0^2100$	$0{,}0^2200$	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	1,022
3	0,0153	0,0243	0,0717	0,115	0,216	0,352	0,584	1,005	1,424	1,869
4	0,0639	0,0908	0,207	0,297	0,484	0,711	1,064	1,649	2,195	2,753
5	0,158	0,210	0,412	0,554	0,831	1,145	1,610	2,343	3,000	3,655
6	0,299	0,381	0,676	0,872	1,237	1,635	2,204	3,070	3,828	4,570
7	0,485	0,598	0,989	1,239	1,690	2,167	2,833	3,822	4,671	5,493
8	0,710	0,857	1,344	1,645	2,180	2,773	3,490	4,594	5,527	6,423
9	0,972	1,153	1,735	2,088	2,700	3,325	4,168	5,380	6,393	7,357
10	1,265	1,479	2,156	2,558	3,247	3,940	4,865	6,179	7,267	8,295
11	1,587	1,834	2,603	3,053	3,816	4,575	5,578	6,989	8,148	9,237
12	1,934	2,214	3,074	3,571	4,404	5,226	6,304	7,807	9,034	10,182
13	2,305	2,617	3,565	4,107	5,009	5,892	7,042	8,634	9,926	11,129
14	2,697	3,041	4,075	4,660	5,629	6,571	7,790	9,467	10,821	12,079
15	3,108	3,483	4,601	5,229	6,262	7,261	8,547	10,307	11,721	13,030
16	3,536	3,942	5,142	5,812	6,908	7,962	9,312	11,152	12,624	13,983
17	3,980	4,416	5,697	6,408	7,564	8,672	10,085	12,002	13,531	14,937
18	4,439	4,905	6,265	7,015	8,231	9,390	10,865	12,857	14,440	15,893
19	4,912	5,407	6,844	7,633	8,907	10,117	11,651	13,716	15,352	16,850
20	5,398	5,921	7,434	8,260	9,591	10,851	12,443	14,578	16,266	17,809
21	5,896	6,447	8,034	8,897	10,283	11,591	13,240	15,445	17,182	18,768
22	6,404	6,983	8,643	9,542	10,982	12,338	14,041	16,314	18,101	19,729
23	6,924	7,529	9,260	10,196	11,688	13,091	14,848	17,187	19,021	20,690
24	7,453	8,085	9,886	10,856	12,401	13,845	15,659	18,062	19,943	21,652
25	7,991	9,649	10,520	11,524	13,120	14,611	16,473	18,940	20,867	22,616
26	8,538	9,222	11,160	12,198	13,844	15,379	17,292	19,820	21,792	23,579
27	9,093	9,803	11,808	12,879	14,573	16,151	18,114	20,703	22,719	24,544
28	9,656	10,391	12,461	13,565	15,308	16,928	18,939	21,588	23,647	25,509
29	10,227	10,986	13,121	14,256	16,047	17,708	19,768	22,475	24,577	26,475
30	10,804	11,588	13,787	14,953	16,971	18,493	20,599	23,364	25,508	27,442
31	11,389	12,196	14,458	15,655	17,539	19,281	21,434	24,255	26,440	28,409
32	11,979	12,811	15,134	16,362	18,291	20,072	22,271	25,148	27,373	39,376
33	12,576	13,431	15,815	17,073	19,047	20,867	23,110	26,042	28,307	30,344
34	13,176	14,057	16,501	17,789	19,806	21,664	23,952	26,938	29,242	31,313
35	13,788	14,688	17,192	18,509	20,569	22,465	24,797	27,836	30,178	32,282
36	14,401	15,324	17,887	19,233	21,336	23,269	25,643	28,735	31,115	33,252
37	15,020	15,965	18,586	19,960	22,106	24,075	26,492	29,635	32,053	34,222
38	15,644	16,611	19,289	20,691	22,878	24,884	27,343	30,537	32,992	35,192
39	16,273	17,262	19,996	21,426	23,654	25,695	28,196	31,441	33,932	36,163
40	16,906	17,916	20,707	22,164	24,433	26,509	29,051	32,345	34,872	37,134
41	17,544	18,575	21,421	22,906	25,215	27,326	29,907	33,251	35,813	38,105
42	18,186	19,238	22,138	23,650	25,909	28,144	30,765	34,157	36,775	39,077
43	18,832	19,905	22,859	24,398	26,785	28,965	31,625	35,065	37,698	40,050
44	19,482	20,576	23,584	25,148	27,575	29,787	32,487	35,974	38,641	41,022
45	20,137	21,251	24,311	25,901	28,366	30,612	33,350	36,884	39,585	41,955
46	20,794	21,929	25,041	26,657	29,160	31,439	34,215	37,795	40,529	42,968
47	21,456	22,610	25,775	27,416	29,959	32,268	35,081	38,708	41,474	43,942
48	22,121	23,295	26,511	28,177	30,755	33,098	35,949	39,621	42,420	44,915
49	22,789	23,983	27,249	28,491	31,555	33,930	36,818	40,534	43,366	45,889
50	23,461	24,674	27,991	29,707	32,357	34,764	37,689	41,449	44,313	46,864
51	24,136	25,368	28,735	30,475	33,162	35,600	38,560	42,365	45,261	47,838
52	24,814	26,065	29,481	31,246	33,968	36,437	39,433	43,281	46,209	48,813
53	25,495	26,765	30,230	32,018	34,776	37,276	40,308	44,199	47,157	49,788
54	26,179	27,468	30,981	32,793	35,586	38,116	41,183	45,117	48,106	50,764
55	26,866	28,173	31,735	33,570	36,398	38,958	42,060	46,036	49,054	51,739

Tab. 29.9 c.d.

Wartości krytyczne χ^2 rozkładu χ^2 ; $P\{\chi^2_r \geq \chi^2(\alpha, r)\} = \alpha$

r /α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
1	0,455	0,708	1,074	1,642	2,706	3,841	5,024	6,636	7,879	10,828	12,116
2	1,386	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816	15,202
3	2,366	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266	17,730
4	3,357	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467	19,997
5	4,351	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515	22,105
6	5,345	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458	24,103
7	6,346	7,283	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322	26,018
8	7,344	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,125	27,868
9	8,343	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877	29,666
10	9,342	10,473	11,781	13,442	15,987	18,307	20,483	23,209	25,188	29,588	31,420
11	10,341	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264	33,136
12	11,340	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909	34,821
13	12,340	13,636	15,119	16,985	19,812	22,362	24,735	27,688	29,819	34,528	36,478
14	13,339	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123	38,109
15	14,339	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697	39,719
16	15,338	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252	41,308
17	16,338	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790	42,879
18	17,338	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312	44,434
19	18,338	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820	45,973
20	19,337	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315	47,498
21	20,337	21,991	23,858	26,171	29,615	32,671	35,479	38,932	41,401	46,797	49,010
22	21,337	23,031	24,939	27,301	30,813	33,924	36,781	40,289	42,796	48,268	50,511
23	22,337	24,069	26,018	28,429	32,007	35,172	38,086	41,638	44,181	49,728	52,000
24	23,337	25,106	27,096	29,553	33,196	36,415	39,364	42,980	45,558	51,179	53,479
25	24,337	26,143	28,172	30,675	34,382	37,652	40,646	44,314	46,928	52,618	54,947
26	25,336	27,179	29,246	31,795	35,563	38,885	41,923	45,642	48,290	54,052	56,407
27	26,336	28,214	30,319	32,912	36,741	40,113	43,194	46,963	49,645	55,476	57,858
28	27,336	29,249	31,391	34,027	37,916	41,337	44,461	48,278	50,993	56,892	59,300
29	28,336	30,283	32,461	35,139	39,087	42,557	45,722	49,588	52,336	58,301	60,735
30	29,336	31,316	33,530	36,250	40,256	43,773	46,979	50,892	53,672	59,703	62,162
31	30,336	32,349	34,598	37,359	41,422	44,985	48,232	52,191	55,003	61,098	63,582
32	31,336	33,381	35,665	38,466	42,585	46,194	49,480	53,486	56,328	62,487	64,995
33	32,336	34,413	36,731	39,572	43,745	47,400	50,725	54,776	57,648	63,870	66,402
34	33,336	35,444	37,795	40,676	44,903	48,602	51,966	56,061	58,964	65,247	67,803
35	34,336	36,475	38,859	41,778	46,059	49,802	53,203	57,342	60,275	66,619	69,199
36	35,336	37,505	39,922	42,879	47,212	50,998	54,437	58,619	61,581	67,985	70,588
37	36,336	38,535	40,984	43,978	48,363	52,192	55,668	59,892	62,882	69,346	71,972
38	37,335	39,564	42,045	45,076	49,513	53,384	56,895	61,162	64,181	70,703	73,351
39	38,335	40,593	43,105	46,173	50,660	54,572	58,120	62,428	65,476	72,055	74,725
40	39,335	41,622	44,165	47,269	51,805	55,768	59,342	63,691	66,766	73,402	76,095
41	40,335	42,651	45,224	48,363	52,949	56,942	60,561	64,950	68,053	74,745	77,459
42	41,335	43,679	46,282	49,456	54,090	58,124	61,777	66,206	69,336	76,084	78,820
43	42,335	44,706	47,339	50,548	55,230	59,304	62,990	67,459	70,616	77,419	80,176
44	43,335	45,734	48,396	51,639	56,369	60,481	64,201	68,709	71,893	78,749	81,528
45	44,335	46,761	49,452	52,729	57,505	61,656	65,410	69,957	73,166	80,077	82,876
46	45,335	47,787	50,507	53,818	58,641	62,830	66,617	71,201	74,437	81,400	84,220
47	46,335	48,814	51,562	54,906	59,774	64,001	67,821	72,443	75,704	82,720	85,560
48	47,335	49,840	52,616	55,993	60,907	65,171	69,023	73,683	76,969	84,037	86,897
49	48,335	50,866	53,670	57,079	62,038	66,339	70,222	74,919	78,231	85,351	88,231
50	49,335	51,892	54,723	58,164	63,167	67,505	71,420	76,154	79,490	86,661	89,561
51	50,335	52,917	55,775	59,248	64,295	68,669	72,616	77,386	80,747	87,968	90,881
52	51,335	53,942	56,827	60,332	65,422	69,832	73,810	78,616	82,001	89,272	92,211
53	52,335	54,967	57,879	61,414	66,548	70,993	75,002	79,843	83,253	90,573	93,531
54	53,335	55,992	58,930	62,496	67,673	72,153	76,192	81,069	84,502	91,872	94,849
55	54,335	57,016	59,980	63,577	68,796	73,311	77,380	82,292	85,749	93,167	96,163

Wartości krytyczne χ^2 rozkładu χ^2 ; $P\{\chi^2_r \geq \chi^2(\alpha, r)\} = \alpha$

r /α	0,9995	0,999	0,995	0,99	0,975	0,95	0,90	0,80	0,70	0,60
56	27,555	28,881	32,490	34,350	37,212	39,801	42,937	46,955	50,005	52,715
57	28,248	29,591	33,248	35,131	38,027	40,646	43,816	47,876	50,956	53,691
58	28,943	30,305	34,008	35,913	38,844	41,492	44,696	48,797	51,909	54,667
59	29,640	31,020	34,770	36,698	39,662	42,339	45,577	49,718	52,857	55,643
60	30,340	31,738	35,535	37,485	40,482	43,188	46,459	50,641	53,809	56,620
61	31,043	32,459	36,301	38,273	41,303	44,038	47,342	51,564	54,761	57,597
62	31,748	33,181	37,068	39,063	42,126	44,889	48,226	52,487	55,714	58,574
63	32,455	33,906	37,838	39,855	42,950	45,741	49,111	53,412	56,666	59,551
64	33,165	34,633	38,610	40,649	43,776	46,595	49,996	54,336	57,619	60,528
65	33,877	35,362	39,383	41,444	44,603	47,450	50,883	55,262	58,573	61,506
66	34,591	36,093	40,158	42,240	45,431	48,305	51,770	56,188	59,527	62,484
67	35,307	36,826	40,935	43,038	46,261	49,162	52,659	57,115	60,481	63,461
68	36,025	37,561	41,713	43,838	47,092	50,020	53,548	58,042	61,436	64,440
69	36,745	38,298	42,494	44,639	47,924	50,879	54,438	58,970	62,391	65,418
70	37,467	39,036	43,275	45,442	48,758	51,739	55,329	59,898	63,346	66,396
71	38,192	39,777	44,058	46,246	49,592	52,600	56,221	60,827	64,302	67,375
72	38,918	40,520	44,843	47,051	50,428	53,462	57,113	61,756	65,258	68,353
73	39,646	41,262	45,629	47,858	51,265	54,325	58,006	62,686	66,214	69,332
74	40,376	42,010	46,417	48,666	52,103	55,189	58,900	63,616	67,170	70,311
75	41,107	42,757	47,206	49,475	52,942	56,054	59,795	64,547	68,127	71,290
76	41,841	43,507	47,997	50,286	53,782	56,920	60,690	65,478	69,084	72,270
77	42,576	44,258	48,788	51,097	54,623	57,786	61,586	66,409	70,042	73,249
78	43,313	45,010	49,582	51,910	55,466	58,654	62,483	67,341	70,999	74,229
79	44,051	45,764	50,376	52,725	56,309	59,522	63,308	68,274	71,957	75,208
80	44,791	46,520	51,172	53,540	57,153	60,391	64,278	69,207	72,915	76,188
81	45,533	47,277	51,969	54,357	57,998	61,261	65,176	70,140	73,874	77,168
82	46,276	48,036	52,767	55,174	58,845	62,132	66,076	71,074	74,833	78,148
83	47,021	48,796	53,567	55,993	59,692	63,004	66,976	72,008	75,792	79,128
84	47,767	49,557	54,368	56,813	60,540	63,876	67,876	72,943	76,751	80,108
85	48,515	50,320	55,170	57,634	61,389	64,749	68,777	73,878	77,710	81,089
86	49,264	51,085	55,973	58,456	62,239	65,623	69,679	74,813	78,670	82,069
87	50,015	51,850	56,777	59,279	63,089	66,498	70,581	75,749	79,630	83,050
88	50,767	52,617	57,582	60,103	63,941	67,373	71,484	76,685	80,590	84,031
89	51,521	53,386	58,389	60,928	64,793	68,249	72,387	77,622	81,550	85,012
90	52,276	54,155	59,196	61,754	65,647	69,126	73,291	78,558	82,511	85,993
91	53,032	54,926	60,005	62,581	66,501	70,003	74,196	79,496	83,472	86,974
92	53,790	55,698	60,815	63,409	67,356	70,882	75,101	80,433	84,433	87,955
93	54,549	56,472	61,625	64,238	68,211	71,760	76,006	81,371	85,394	88,936
94	55,309	57,246	62,437	65,068	69,068	72,640	76,912	82,309	86,356	89,917
95	56,070	58,022	63,250	65,898	69,925	73,520	77,818	83,248	87,317	90,899
96	56,833	58,799	64,063	66,730	70,783	74,400	78,725	84,187	88,279	91,881
97	57,597	59,577	64,878	67,562	71,642	75,282	79,633	85,126	89,241	92,862
98	58,362	60,356	65,694	68,396	72,501	76,164	80,541	86,065	90,204	93,844
99	59,128	61,136	66,510	69,230	73,361	77,046	81,449	87,005	91,166	94,826
100	59,896	61,918	67,328	70,075	74,222	77,929	82,358	87,945	92,129	95,808

Funkcje pomocnicze $p_1(\alpha)$ i $p_2(\alpha, \nu)$ $\left(\nu = 1/\sqrt{2r}\right)$

ν/α	0,9995	0,999	0,995	0,99	0,975	0,950	0,90	0,80	0,70	0,60
0,00	6,652	5,700	3,757	2,941	1,894	1,137	0,428	-0,194	-0,483	-0,624
0,01	6,537	5,691	3,757	2,945	1,901	1,145	0,436	-0,189	-0,487	-0,622
0,02	6,522	5,681	3,758	2,949	1,908	1,153	0,443	-0,183	-0,491	-0,620
0,03	6,506	5,670	3,758	2,953	1,914	1,160	0,451	-0,177	-0,495	-0,618
0,04	6,489	5,659	3,758	2,956	1,921	1,168	0,459	-0,171	-0,499	-0,616
0,05	6,471	5,647	3,758	2,959	1,927	1,176	0,466	-0,165	-0,503	-0,614
0,06	6,452	5,635	3,757	2,962	1,933	1,183	0,474	-0,159	-0,507	-0,612
0,07	6,432	5,621	3,755	2,964	1,940	1,191	0,481	-0,153	-0,511	-0,609
0,08	6,411	5,607	3,754	2,967	1,946	1,198	0,489	-0,147	-0,516	-0,607
$p_1(\alpha)$	-2673	-3,0902 3231	-2,5758 2930	-2,3263 4787	-1,9599 6398	-1,6448 5363	-1,2815 5157	-0,8416 2123	-0,5244 0051	-0,2533 4710

Wartości krytyczne χ^2 rozkładu χ^2 ; $P\{\chi^2_r \geq \chi^2(\alpha, r)\} = \alpha$

r /α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
56	55,335	58,040	61,031	64,658	69,918	74,468	78,567	83,513	86,994	94,460	97,475
57	56,335	59,064	62,080	65,737	71,040	75,624	79,752	84,733	88,236	95,751	98,784
58	57,335	60,088	63,129	66,816	72,160	76,778	80,936	85,950	89,477	97,039	100,090
59	58,335	61,111	64,178	67,894	73,279	77,931	82,117	87,166	90,715	98,324	101,394
60	59,335	62,135	65,226	68,972	74,397	79,082	83,298	88,379	91,952	99,607	102,695
61	60,335	63,158	66,274	70,049	75,514	80,232	84,476	89,591	93,186	100,888	103,993
62	61,335	64,181	67,322	71,125	76,630	81,381	85,654	90,802	94,419	102,166	105,289
63	62,335	65,204	68,369	72,201	77,745	82,529	86,830	92,010	95,649	103,442	106,583
64	63,335	66,226	69,416	73,276	78,860	83,675	88,004	93,217	96,878	104,716	107,875
65	64,335	67,249	70,462	74,351	79,971	84,821	89,177	94,422	98,105	105,988	109,164
66	65,335	68,271	71,508	75,425	81,086	85,965	90,349	95,626	99,330	107,258	110,451
67	66,335	69,293	72,554	76,498	82,197	87,108	91,519	96,828	100,554	108,526	111,736
68	67,334	70,315	73,600	77,571	83,308	88,250	92,688	98,028	101,776	109,791	113,018
69	68,334	71,337	74,645	78,643	84,418	89,391	93,856	99,227	102,996	111,055	114,299
70	69,334	72,358	75,689	79,715	85,527	90,531	95,023	100,425	104,215	112,317	115,578
71	70,334	73,380	76,734	80,786	86,635	91,670	96,189	101,621	105,432	113,577	116,854
72	71,334	74,401	77,778	81,857	87,743	92,808	97,353	102,816	106,648	114,835	118,129
73	72,334	75,422	78,822	82,927	88,850	93,945	98,516	104,010	107,862	116,092	119,402
74	73,334	76,443	79,865	83,997	89,956	95,081	99,678	105,202	109,074	117,346	120,673
75	74,334	77,464	80,908	85,066	91,061	96,217	100,839	106,393	110,286	118,599	121,942
76	75,334	78,485	81,951	86,135	92,166	97,351	101,999	107,582	111,495	119,850	123,209
77	76,334	79,505	82,994	87,203	93,270	98,484	103,158	108,771	112,704	121,100	124,475
78	77,334	80,526	84,036	88,271	94,374	99,617	104,316	109,958	113,911	122,348	125,739
79	78,334	81,546	85,078	89,338	95,476	100,749	105,473	111,144	115,117	123,594	127,001
80	79,334	82,566	86,120	90,405	96,578	101,879	106,629	112,329	116,321	124,839	128,261
81	80,334	83,586	87,161	91,472	97,608	103,010	107,783	113,512	117,524	126,082	129,520
82	81,334	84,606	88,202	92,538	98,780	104,139	108,937	114,695	118,726	127,324	130,778
83	82,334	85,626	89,243	93,604	99,880	105,267	110,090	115,876	119,927	128,565	132,033
84	83,334	86,646	90,284	94,669	100,980	106,395	111,242	117,057	121,126	129,804	133,288
85	84,334	87,665	91,325	95,734	102,079	107,522	112,393	118,236	122,325	131,041	134,540
86	85,334	88,685	92,365	96,799	103,177	108,648	113,544	119,414	123,522	132,277	135,792
87	86,334	89,704	93,405	97,863	104,275	109,773	114,693	120,591	124,718	133,512	137,041
88	87,334	90,723	94,445	98,927	105,372	110,898	115,841	121,767	125,913	134,745	138,290
89	88,334	91,742	95,484	99,991	106,469	112,022	116,989	122,942	127,106	135,978	139,537
90	89,334	92,761	96,524	101,054	107,565	113,145	118,136	124,116	128,299	137,208	140,782
91	90,334	93,780	97,563	102,117	108,661	114,268	119,282	125,289	129,491	138,438	142,027
92	91,334	94,799	98,602	103,179	109,756	115,390	120,427	126,462	130,681	139,666	143,269
93	92,334	95,818	99,641	104,241	110,850	116,511	121,571	127,633	131,871	140,893	144,511
94	93,334	96,836	100,679	105,303	111,944	117,632	122,715	128,803	133,059	142,119	145,751
95	94,334	97,855	101,717	106,364	113,038	118,752	123,858	129,973	134,247	143,344	146,990
96	95,334	98,873	102,755	107,425	114,131	119,871	125,000	131,141	135,433	144,567	148,228
97	96,334	99,892	103,793	108,486	115,223	120,990	126,141	132,309	136,619	145,789	149,465
98	97,334	100,910	104,831	109,547	116,315	122,108	127,282	133,476	137,803	147,010	150,700
99	98,334	101,928	105,868	110,607	117,407	123,225	128,422	134,642	138,987	148,230	151,934
100	99,334	102,946	106,906	111,667	118,498	124,342	129,561	135,807	140,169	149,449	153,167

Funkcje pomocnicze $p_1(\alpha)$ i $p_2(\alpha, \nu)$ $\left(\nu = 1/\sqrt{2r}\right)$

ν/α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005
0,00	-0,667	-0,624	-0,483	-0,194	0,428	1,137	1,894	2,941	3,757	5,700	6,552
0,01	-0,667	-0,626	-0,479	-0,200	0,421	1,129	1,887	2,937	3,755	5,708	6,565
0,02	-0,667	-0,628	-0,476	-0,206	0,413	1,121	1,880	2,933	3,754	5,716	6,578
0,03	-0,667	-0,630	-0,472	-0,212	0,405	1,113	1,873	2,928	3,752	5,723	6,590
0,04	-0,666	-0,631	-0,468	-0,218	0,398	1,105	1,866	2,923	3,750	5,730	6,601
0,05	-0,666	-0,633	-0,464	-0,224	0,390	1,097	1,859	2,918	3,748	5,736	6,612
0,06	-0,666	-0,635	-0,460	-0,229	0,382	1,089	1,851	2,913	3,745	5,741	6,622
0,07	-0,666	-0,637	-0,457	-0,235	0,375	1,081	1,844	2,908	3,742	5,747	6,631
0,08	-0,666	-0,639	-0,453	-0,241	0,367	1,073	1,836	2,902	3,739	5,741	6,640
$p_1(\alpha)$	0,0000 0000	0,2533 4710	0,5244 0051	0,8416 2123	1,2815 5157	1,6448 5363	1,9599 6398	2,3263 4787	2,5758 2930	3,0902 3231	3,2905 2673

Tab. 29.10

Wartości krytyczne rozkładu Durbina-Watsona dla $\alpha = 0{,}05$

T	$k = 1$		$k = 2$		$k = 3$	
	dL	dU	dL	dU	dL	dU
6	0,610	1,400				
7	0,700	1,356	0,467	1,896		
8	0,763	1,332	0,559	1,777	0,368	2,287
9	0,824	1,320	0,629	1,699	0,455	2,128
10	0,879	1,320	0,697	1,641	0,525	2,016
11	0,927	1,324	0,758	1,604	0,595	1,928
12	0,971	1,331	0,812	1,579	0,658	1,864
13	1,010	1,340	0,861	1,562	0,715	1,816
14	1,045	1,350	0,905	1,551	0,767	1,779
15	1,077	1,361	0,946	1,543	0,814	1,750
16	1,106	1,371	0,982	1,539	0,857	1,728
17	1,133	1,381	1,015	1,536	0,897	1,710
18	1,158	1,391	1,046	1,535	0,933	1,696
19	1,180	1,401	1,074	1,536	0,967	1,685
20	1,201	1,411	1,100	1,537	0,998	1,676
21	1,221	1,420	1,125	1,538	1,026	1,667
22	1,239	1,429	1,147	1,541	1,053	1,664
23	1,257	1,437	1,168	1,543	1,078	1,660
24	1,273	1,446	1,188	1,546	1,101	1,656
25	1,288	1,454	1,206	1,550	1,123	1,654
26	1,302	1,461	1,224	1,553	1,143	1,652
27	1,316	1,469	1,240	1,556	1,162	1,651
28	1,328	1,476	1,255	1,560	1,181	1,650
29	1,341	1,483	1,270	1,563	1,198	1,650
30	1,352	1,489	1,284	1,567	1,214	1,650
31	1,363	1,496	1,297	1,570	1,229	1,650
32	1,373	1,502	1,309	1,574	1,244	1,650
33	1,383	1,508	1,321	1,577	1,258	1,651
34	1,393	1,514	1,333	1,580	1,271	1,652
35	1,402	1,519	1,343	1,584	1,283	1,653
36	1,411	1,525	1,354	1,587	1,295	1,654
37	1,419	1,530	1,364	1,590	1,307	1,655

30. Słowniczek wybranych pojęć z zakresu modelowania prognoz

W opracowaniu powtarzają się niektóre określenia dotyczące metod lub modeli prognozowania. Odniesione są one do różnych horyzontów ustalania prognoz, a mianowicie dla krótkich, średnich czy też dużych okresów czasowych. Słowniczek ten jest pomyślany przede wszystkim jako rozwinięcie opisów częściowo zamieszczonych w poszczególnych rozdziałach. Może też stanowić materiał pomocniczy dla studentów w czasie zaliczania takich przedmiotów jak statystyka opisowa, ekonometria, przewidywanie cech ekonomicznych.

***A**lgorytmy uczenia maszynowego. https://azure.microsoft.com/pl-pl/resources/cloud-computing-dictionary/what-are-machine-learning-algorithms/#overview.* *Stanowią* kod ułatwiający eksplorowanie, analizowanie i rozumienie złożonych zestawów danych. Każdy algorytm to skończony zestaw jednoznacznych instrukcji krok po kroku, które może wykonać komputer, aby osiągnąć określony cel. Celem jest ustanowienie lub odnalezienie wzorców, których ludzie mogą używać do przewidywania lub kategoryzowania informacji

Analiza dyskryminacyjna. https://pl.wikipedia.org/wiki/Analiza_dyskryminacyjna. Zespół metod wielowymiarowej analizy danych. Zawiera metody, które czynią z tej techniki niezwykle efektywne narzędzie do zagadnień klasyfikacyjnych i eksploracji danych. Jej zadaniem jest rozstrzyganie, które zmienne niezależne (predyktory) w najlepszy sposób dzielą dany zbiór przypadków na występujące w naturalny sposób grupy, opisane jakościową zmienną zależną.

*Analiza harmoniczna. https://pl.wikipedia.org/wiki/Analiza_harmoniczna.*Analiza fourierowska - dział matematyki obejmujący teorię i zastosowania szeregu Fouriera i transformaty Fouriera. Analiza ta prowadzi do utworzenia modelu stanowiącego sumę składowych harmonicznych (harmonik), tj. funkcji sinusoidalnych i cosinusoidalnych w określonym przedziale czasowym.

Analiza regresji. https://mfiles.pl/pl/index.php/Analiza_regresji. Statystyczna metoda służąca modelowaniu związków pomiędzy dwiema lub więcej zmiennymi. Model regresji stosuje się do analitycznego wyjaśnienia związku między zmienną objaśnianą - zależną, a zmienną (zmiennymi) objaśniającymi.

Analiza widma osobliwego (SSA). https://www.gamingdeputy.com/pl/android/narzedzie-do-przewidywania-przyszlosci-aktualnosci-mit/. SSA imputuje i prognozuje pojedyncze szeregi czasowe. Imputacja to proces zastępowania brakujących wartości lub korygowania przeszłych wartości dzięki odpowiedniemu algorytmowi.

***B**adanie panelowe. https://pl.wikipedia.org/wiki/Badanie_panelowe.* Rodzaj analizy trendu w badaniach socjologicznych polegający na kilkukrotnych pomiarach tych samych jednostek analizy (osób, organizacji) w celu określenia dynamiki zmian badanych cech oraz ich wyjaśnienia.

Burza mózgów. https://pl.wikipedia.org/wiki/Burza_m%C3%B3zg%C3%B3w. Technika wywodząca się z psychologii społecznej, która ma na celu doskonalenie decyzji grupowych. Burza mózgów jest również formą dyskusji dydaktycznej, wykorzystywaną jako jedna z metod nauczania. Zalicza się ją wówczas do metod aktywizujących, która stanowi podgrupę metod problemowych. Jest to jedna z tak zwanych metod heurystycznych.

***E**konometryczny model grawitacji. file:///C:/Users/wlode/Downloads/Pietrzak_Wilk_Odleglosc_ekonomiczna_w_modelowaniu_zjawisk.pdf.* Koncepcja modelu grawitacji w modelowaniu zjawisk ekonomicznych nawiązuje do prawa powszechnego ciążenia Newtona, które głosi, że „*każdy obiekt we wszechświecie przyciąga każdy inny obiekt z siłą, która jest wprost proporcjonalna do iloczynu ich mas i odwrotnie proporcjonalna do kwadratu odległości między ich środkami*". Zatem istotną zmienną objaśniającą w modelu grawitacji jest odległość. W pierwszych pracach z zakresu modelu grawitacji odległość rozumiano

w sensie fizycznym. Jednak konstrukcja modelu pozwala uwzględniać także inne rodzaje odległości, w tym tzw. odległość ekonomiczną.

M*etoda analogii historycznych. https://pl.wikipedia.org/wiki/Prognozowanie_analogowe.* Przenoszenie zaobserwowanych analogii zmian z jednych zjawisk na inne zachodzące później w tym samym obiekcie np. na tempo rozwoju komputeryzacji może być miarą rozwoju Internetu.

Metoda analogii przestrzenno-czasowych.
https://pl.wikipedia.org/wiki/Prognozowanie_analogowe. Polega na przenoszeniu z jednych zjawisk do drugich, prawidłowości zmian w czasie, np. wzrost zainteresowania firmą po wprowadzeniu nowego modelu w jednym kraju, może być podstawą do prognozowania wzrostu zainteresowania w innym kraju po wprowadzeniu w nim nowego produktu.

Metoda analogii przestrzennych. *https://pl.wikipedia.org/wiki/Prognozowanie_analogowe.* Przewidywanie wystąpienia zjawiska na danym obszarze po zaobserwowaniu go na innych obszarach.

Metoda ankietowa. https://pl.wikipedia.org/wiki/Metoda_ankietowa. Metoda badań naukowych polegające na stawianiu pytań, a następnie – na opracowywaniu odpowiedzi. Stanowi ona trzon metodologiczny nauk społecznych. Metoda ankietowa odgrywa znaczną rolę w naukach psychologicznych. Ankieta jest zbiorem pytań stawianych pisemnie, celem uzyskania odpowiedzi mających służyć do rozwiązania pewnego problemu. Odpowiedzi te są następnie opracowywane statystycznie, a wyniki objaśniane - interpretowane.

Metoda buzz session. https://witalni.pl/baza_wiedzy/techniki-kreatywnego-myslenia/. Zwana również Philips 66 to odmiana burzy mózgów polegająca na podzieleniu zespołu ludzi na grupy po 6 osób, które będą pracować nad pomysłami i rozwiązaniami przez 6 minut. Po tym czasie grupy przedstawiają swoje wyniki i proces opracowywania pomysłów jest powtarzany jeszcze kilka razy, z uwzględnieniem i rozwinięciem pomysłów opracowanych wcześniej przez inne grupy. Pozwala to na stopniowe wyszukiwanie najlepszych i najbardziej trafnych rozwiązań.

Metoda delficka. https://pl.wikipedia.org/wiki/Metoda_delficka. Należy do grupy metod heurystycznych, w których do podejmowania decyzji wykorzystuje się wiedzę, doświadczenie i opinie ekspertów z danej dziedziny. Wykorzystywana jest do określenia prawdopodobieństwa lub czasu zajścia przyszłych zdarzeń. Postawioną prognozę uzyskuje się poprzez przeprowadzenie serii ankiet wśród ekspertów.

Metoda ekonometryczna. https://mfiles.pl/pl/index.php/Model_ekonometryczny. Model ekonometryczny za pomocą równania (układu równań) pomaga wyjaśnić mechanizm zmian zachodzących w badanym obszarze. Opisuje powiązania między danymi wielkościami ekonomicznymi. Jest to formalny matematyczny zapis istniejących prawidłowości ekonomicznych. Zbudowanie modelu ekonometrycznego wymaga nie tylko dobrej znajomości teorii ekonomii oraz wiedzy matematyczno- ekonomicznej, lecz również znajomości praktyki ekonomicznej. Model ekonometryczny powinien posiadać nie tylko wartość poznawczą z punktu widzenia teorii ekonomii, ale również wartość praktyczną, czyli aby mógł służyć jako narzędzie wnioskowania w przyszłości.W szerszym pojęciu model ekonometryczny obejmuje nie tylko te modele, które zawierają matematyczny opis istniejących prawidłowości ekonomicznych, ale również i te, które obok opisu rzeczywistości pozwalają na wybór rozwiązań optymalnych. W takich modelach, obok zależności funkcyjnych wyznaczają obszar dopuszczalnych decyzji, występuje funkcja kryterium, zwana również funkcją celu, umożliwiająca wybór wariantu optymalnego spośród dopuszczalnych. Natomiast w węższym sensie model ekonometryczny odnosi się jedynie do modeli opisujących prawidłowości istniejące w gospodarce, a więc modelami, którymi zajmuje się klasyczna teoria ekonomii.

Metoda kolejnych przybliżeń. 3434-6612-1-PB.pdf. Jest jedną z podstawowych metod dowodzenia twierdzeń o istnieniu rozwiązań równań różniczkowych zwyczajnych, a także równań cząstkowych pewnych typów. Znanych jest kilka wariantów tej metody, przy czym różnice między nimi są dwojakiego typu: są to mianowicie różnice w konstrukcji ciągu kolejnych przybliżeń oraz różnice w metodach dowodów zbieżności.

Metoda macierzy powiązań. https://iung.pl/wp-content/uploads/2009/10/zesz14.pdf . Metoda macierzy powiązań jest podobna do delfickiej. Macierz taką konstruuje się i przedstawia w niej oszacowane dwa typy danych ważnych dla przyszłego rozwoju. Pierwsze mówią o prawdopodobieństwie, że dane zjawisko wystąpi w określonym czasie, drugie o sile wzajemnych powiązań i ich prawdopodobieństwie.

Metoda naiwna. https://pl.wikipedia.org/wiki/Metoda_naiwna. Metoda prognostyczna dotycząca analizy szeregów czasowych bez tendencji. Metoda ta stosowana jest przy stałym poziomie zjawiska i niewielkich wahaniach przypadkowych (niski współczynnik zmienności V < 5-10%) i przy założeniu, że nie wystąpią istotne zmiany najważniejszych czynników. Jest to prognoza typu: "jutro będzie tak jak dziś". Metoda ta polega na tym, iż wartość z ubiegłego okresu (t-1) przypisywana jest do okresu bezpośrednio następującego po nim jako wartość prognozowana.

Metoda najmniejszych kwadratów. https://pl.wikipedia.org/wiki/Metoda_najmniejszych_kwadrat%C3%B3w. Standardowa metoda przybliżania rozwiązań układów nadokreślonych, tzn. zestawu równań, w którym jest ich więcej niż zmiennych. Nazwa „*najmniejsze kwadraty*" oznacza, że końcowe rozwiązanie tą metodą minimalizuje sumę kwadratów błędów przy rozwiązywaniu każdego z równań. W statystyce wykorzystuje się ją do estymacji i wyznaczania linii trendu na podstawie zbioru danych w postaci par liczb. Najczęściej jest stosowana przy regresji liniowej, ale może też być stosowana do statystycznego wyznaczania parametrów nieliniowych linii trendu.

Metoda regionalizacyjna. https://dspace.uni.lodz.pl/bitstream/handle/11089/38077/Majdzinska_Regionalizacja%20demograficzna.pdf?sequence=1&isAllowed=y. Wielowymiarowa analiza porównawcza, ze szczególnym uwzględnieniem metod taksonomicznych dla jednostek terytorialnych o różnej skali wielkości. W zbiorze zmiennych diagnostycznych uwzględnione są np. informacje z zakresu demograficznych struktur ludności, ruchu naturalnego oraz migracyjnego.

Metoda średniej ruchomej ważonej. http://www.ekonometria.4me.pl/metoda-sredniej-ruchomej.htm. Obserwacjom nowszym nadawane są większe wagi, które są świeższe i mają większy wpływ na prognozę. Wnioskowanie takie, określamy terminem postarzania informacji.

Metoda wywiadu. https://pl.wikipedia.org/wiki/Metoda_ankietowa. Równie często stosowana metoda w naukach humanistycznych, polegająca na stawianiu pytań dla określonych celów.

Metoda prognozowania. *https://mfiles.pl/pl/index.php/Prognozowanie*. Sposób przetworzenia danych z przeszłości wraz ze sposobem przejścia od przetworzonych danych do prognozy.

Metody ekspertyz. https://iung.pl/sir/zeszyt14_2.pdf. Wśród intuicyjnych metod prognozowania należy wyróżnić: metodę indywidualnych ekspertyz (ocen rzeczoznawców) oraz metodę ekspertyz zespołowych (równoległych lub kolejnych). Metody ekspertyz polegają na zasięgnięciu opinii w formie ekspertyzy dotyczącej określonej problematyki, która ma ukazać najbardziej prawdopodobny przebieg rozwoju danej dziedziny lub grupy zjawisk (zjawiska) w określonym odcinku czasu. Rozróżnia się metodę równoległych i kolejnych ekspertyz.

Model ARIMA. https://support.predictivesolutions.pl/ekspress/download/EB54/mat/model_ARIMA_istotnosc_parametrow.pdf. Jest to autoregresyjny zintegrowany model średniej ruchomej składający się z: procesu autoregresyjnego (AR), procesu średniej ruchomej, stopnia integracji).

Model ARMA. http://www.ekonometria.wne.uw.edu.pl/uploads/Main/wyklad18_ps.pdf. Składa się z dwóch części: modelu autoregresyjnego, modelu średniej ruchomej.

Model bayesowski (BMA). https://web.sgh.waw.pl/~atoroj/ekonometria_bayesowska/12_BMA.pdf. Bayesowskie łączenie wiedzy służy do rozwiązywania problemu niepewności odpowiedniego doboru specyfikacji modelu (doboru zmiennych). BMA czasami określane jest także mianem bayesowskiego uśredniania modeli.

Model Holta-Wintersa (HW). https://excelraport.pl/index.php/2018/03/05/prognozowanie-w-excelu-modelem-holta-wintersa/. Technika prognozowania wykorzystujących tzw. wygładzenie

wykładnicze. Wygładzenie polega na stworzeniu ważonej średniej ruchomej, której wagi określa się według schematu - im starsza informacja o badanym zjawisku, tym mniejszą wartość stanowi ona dla aktualnej prognozy. W modelu oblicza się prognozy wygasłe, czyli takie które dotyczą okresu, w których rzeczywista wartość już się zrealizowała oraz prognozy rzeczywiste stawiane na okres, który jeszcze nie zaistniał.

Model Lee-Cartera. https://en-m-wikipedia-org.translate.goog/wiki/Lee%E2%80%93Carter_model?_x_tr_sl=en&_x_tr_tl=pl&_x_tr_hl=pl&_x_tr_pto=sc. Algorytm numeryczny stosowany w prognozowaniu śmiertelności i przewidywanej długości życia. Dane wejściowe do modelu to macierz współczynników umieralności w zależności od wieku uporządkowana monotonicznie według czasu, zwykle z wiekiem w kolumnach i latami w wierszach. Dane wyjściowe to prognozowana macierz współczynników śmiertelności w tym samym formacie, co dane wejściowe.

Model logit. https://pl.economy-pedia.com/11040708-logit-model. Jest to binarny model wyboru, który opiera się na standardowym logistycznym rozkładzie skumulowanym. Mówiąc dokładniej, w tym modelu logit jest funkcją polegającą na obliczeniu logarytmu naturalnego ilorazu szans obliczanego jako $p/(1-p)$, gdzie p to prawdopodobieństwo, a uzyskany wynik jest zmienną zależną L_i.

$$L_i = \ln\left(\frac{P_i}{1-P_i}\right) = a + bX$$

W powyższym wzorze a i b są współczynnikami modelu ekonometrycznego, które można określić metodą najmniejszych kwadratów lub metodą największej wiarygodności, a X jest wektorem zmiennych niezależnych.

Model panelowy. http://www.ekonometria.wne.uw.edu.pl/uploads/Main/wyklad26_ps.pdf. Dane panelowe to dane, które jednocześnie posiadają cechy danych przekrojowych i szeregów czasowych, czyli obserwujemy te same obiekty w kolejnych momentach czasu.

Model probit. https://en-m-wikipedia-org.translate.goog/wiki/Probit_model?_x_tr_sl=en&_x_tr_tl=pl&_x_tr_hl=pl&_x_tr_pto=sc.
W statystyce model probitowy to rodzaj regresji, w której zmienna zależna może mieć tylko dwie wartości, przy czym celem modelu jest oszacowanie prawdopodobieństwa.

Model rekurencyjny. https://enauczanie.pg.edu.pl/moodle/pluginfile.php/316115/mod_resource/content/1/PiS7-modele%20wielor%C3%B3wnaniowe.pdf. Charakteryzuje się występowaniem powiązań bezpośrednich między zmiennymi łącznie współzależnymi.

Model statystyczny. https://pl.wikipedia.org/wiki/Model_statystyczny. Hipoteza lub układ hipotez sformułowanych w sposób matematyczny (odpowiednio w postaci równania lub układu równań), który przedstawia zasadnicze powiązania występujące pomiędzy rozpatrywanymi zjawiskami rzeczywistymi. Modele statystyczne używane w ekonometrii noszą nazwę modeli ekonometrycznych.

Model trendu. http://alina.gleska.pracownik.put.poznan.pl/st_wyklad11.pdf. Konstrukcja formalna za pomocą której przedstawia się przebieg badanego zjawiska w czasie.

Model wektorowej autoregresji (VAR). https://pl.wikipedia.org/wiki/Model_wektorowej_autoregresji.Vector autoregression to model ekonometryczny. Podstawowe cechy modeli wektorowej autoregresji w kanonicznej formie to: wszystkie zmienne modelu są endogeniczne, nie występują sztuczne ograniczenia dotyczące liczby zmiennych występujących w pojedynczym równaniu (warunek wymiaru i rzędu), łatwość prognozowania.

Model wektorowej autoregresji. https://pl.wikipedia.org/wiki/Modele_VAR. Model *Vector autoregression model* (VAR) podobnie jak w przypadku modeli AR nie posiada podziału na zmienne egzogeniczne oraz endogeniczne. Dzieje się tak, gdyż każda zmienna uwzględniona w modelu ma wpływ na pozostałe zmienne, a proces modelowania dotyczy każdej zmiennej.

Modele adaptacyjne. http://www.ekonometria.4me.pl/trend-pelzajacy.htm. Modele przy konstrukcji których konstrukcji odrzucamy dokuczliwe założenie, przyjmowane przy budowie modeli analitycznych tendencji rozwojowej, mówiące o niezmienności mechanizmu rozwojowego badanych zjawisk. Ze względu na sporą elastyczność modeli adaptacyjnych, mamy możliwość ujęcia nieregularnych zmian szeregu czasowym, co czyni z nich przydatne narzędzie budowy prognoz w krótkim okresie. Modele adaptacyjne mają szeroki wachlarz typów np. model trendu pełzającego.

Modele demograficzne ekonomiczne. Bazują na zjawisku, ze ludzie dostosowują się do zmian ekonomicznych, podążają więc za możliwością uzyskania pracy oraz wyższych wynagrodzeń.

Modele demograficzne grawitacyjne (wielkość migracji jest funkcją liczby mieszkańców miejsc przeznaczenia i pochodzenia oraz odległości między nimi).

Modele wielowymiarowe. http://manuals.pqstat.pl/statpqpl:wielowympl. Wielowymiarowe modele regresji dają możliwość badania wpływu wielu zmiennych niezależnych (wielu czynników) oraz ich interakcji na jedną zmienną zależną.

***P**rognozowanie analogowe. https://pl.wikipedia.org/wiki/Prognozowanie_analogowe.* Metoda prognozowania przyszłości danej zmiennej poprzez użycie informacji o zachowaniu innej zmiennej, której zmiany w czasie są podobne ale nierównoczesne. Jeśli zmienna prognozowana i zmienne wykorzystane do predykcji są tego samego typu i mają tego samego typu miana, to mamy do czynienia z prognozowaniem na podstawie zmiennych jednoimiennych, w przeciwnym wypadku jest to prognozowanie na podstawie zmiennych różnoimiennych.

Prognozowanie. https://mfiles.pl/pl/index.php/Prognozowanie. Racjonalne, naukowe, przewidywanie przyszłych zdarzeń, czyli wnioskowanie o zdarzeniach nieznanych na podstawie zdarzeń znanych.

***R**egesja kwantylowa. https://www.ibm.com/docs/pl/spss-statistics/28.0.0?topic=regression-quantile.* Modeluje relację między zbiorem predyktorów (zmiennych niezależnych) i konkretnych centyli (lub kwantyli) zmiennej przewidywanej (zależnej), najczęściej medianę. Kwantyl rzędu 1/2 to inaczej mediana. Kwantyle rzędu 1/4, 2/4, 3/4 są inaczej nazywane kwartylami.
Kwantyle rzędu 1/5, 2/5, 3/5, 4/5 to inaczej kwintyle. Kwantyle rzędu 1/10, 2/10,..., 9/10 to inaczej decyle. Kwantyle rzędu 1/100, 2/100,..., 99/100 to inaczej percentyle.

Rodzaje modele podejmowania decyzji. https://mfiles.pl/pl/index.php/Modele_podejmowania_decyzji. Wyróżniamy modele, gdzie racjonalne problematyka racjonalności w procesie podejmowania decyzji jest rozpatrywana w nawiązaniu do warunków i założeń, które należy spełnić, zaleceń praktycznych (jak powinien postępować racjonalny menedżer), modelowych ujęć procesów decyzyjnych (w jaki sposób powinny przebiegać procesy decyzyjne). Ponadto modele behawioralne ograniczonej racjonalności, w których decydent postępuje zgodnie z regułami pełnej racjonalności, ale nie jest „doskonały" w swoich zadaniach. Występują również modele heurystyczne zakładające całkowite odejście od pełnej racjonalności, które przedstawiają rzeczywiste procesy decyzyjne. Modele „*podwójnego procesu*" integrują podejście zorientowane na racjonalną analizę z podejściem zorientowanym na intuicję.

***S**ieć neuronowa. https://pl.wikipedia.org/wiki/Sie%C4%87_neuronowa.* System przeznaczony do przetwarzania informacji, którego budowa i zasada działania są w pewnym stopniu wzorowane na funkcjonowaniu fragmentów rzeczywistego (biologicznego) systemu nerwowego. Na przesłankach biologicznych oparte są schematy sztucznych neuronów wchodzących w skład sieci oraz (w pewnym stopniu) jej struktura. Jednak schematy połączeń neuronów w sieci neuronowej są wybierane arbitralnie, a nie stanowią modelu rzeczywistych struktur nerwowych. Wyróżniającą cechą sieci neuronowej jako narzędzia informatycznego jest możliwość komputerowego rozwiązywania przy jej pomocy praktycznych problemów bez ich uprzedniej matematycznej formalizacji. Dalszą zaletą jest brak konieczności odwoływania się przy stosowaniu sieci do jakichkolwiek teoretycznych założeń na temat rozwiązywanego problemu. Najbardziej znamienną

cechą sieci neuronowej jest jej zdolność uczenia się na podstawie przykładów i możliwość automatycznego uogólniania zdobytej wiedzy (generalizacja).

Synektyka. https://witalni.pl/baza_wiedzy/techniki-kreatywnego-myslenia/. Metoda rozwiązywania problemów w sposób twórczy, która polega na szukaniu powiązań między elementami z pozoru nie mającymi ze sobą związku, co pozwala na stworzenie nowych, wcześniej nie branych pod uwagę rozwiązań. Wykorzystana jest w grupowych dyskusjach, skupiających przedstawicieli różnych grup naukowych/zawodowych. Każda z osób podczas rozmowy przedstawia swój punkt widzenia w odniesieniu do tematów, w których się specjalizuje, a wszyscy wspólnie szukają powiązań w różnych dziedzinach, które mogą wykreować rozwiązanie problemu.

System GFS. https://pl.wikipedia.org/wiki/Global_Forecast_System. GFS to globalny system prognozowania pogody krótko i średnioterminowej. Najważniejszym składnikiem systemu jest numeryczny model prognozy pogody obsługiwany przez amerykańską National Oceanic and Atmospheric Administration (NOAA). Model uruchamiany jest cztery razy w ciągu doby i generuje maksymalnie 16-dniowe prognozy pogody, przy czym im dłuższy okres prognozy tym mniejsza jej rozdzielczość powierzchniowa i czasowa. Ogólnie przyjmowane jest, że sprawdzalność generowanej przez model GFS prognozy pogody powyżej siódmego (7) dnia jest mało dokładna. Model GFS, wraz z europejskim modelem ECMWF (*European Center for Medium-range Weather Forecast*), kanadyjskim GEM (*Global Environmental Multiscale Model*) i modelem NOGAPS, należy do czterech najpowszechniej wykorzystywanym modeli numerycznych na świecie.

System prognoz klimatycznych - CFS). https://www.ncei.noaa.gov/products/weather-climate-models/climate-forecast-system. CFS modeluje interakcje między oceanami, lądem i atmosferą Ziemi w skali globalnej. Model jest tworzony przez kilkudziesięciu naukowców pod kierunkiem National Centres for Environmental Prediction (NCEP).

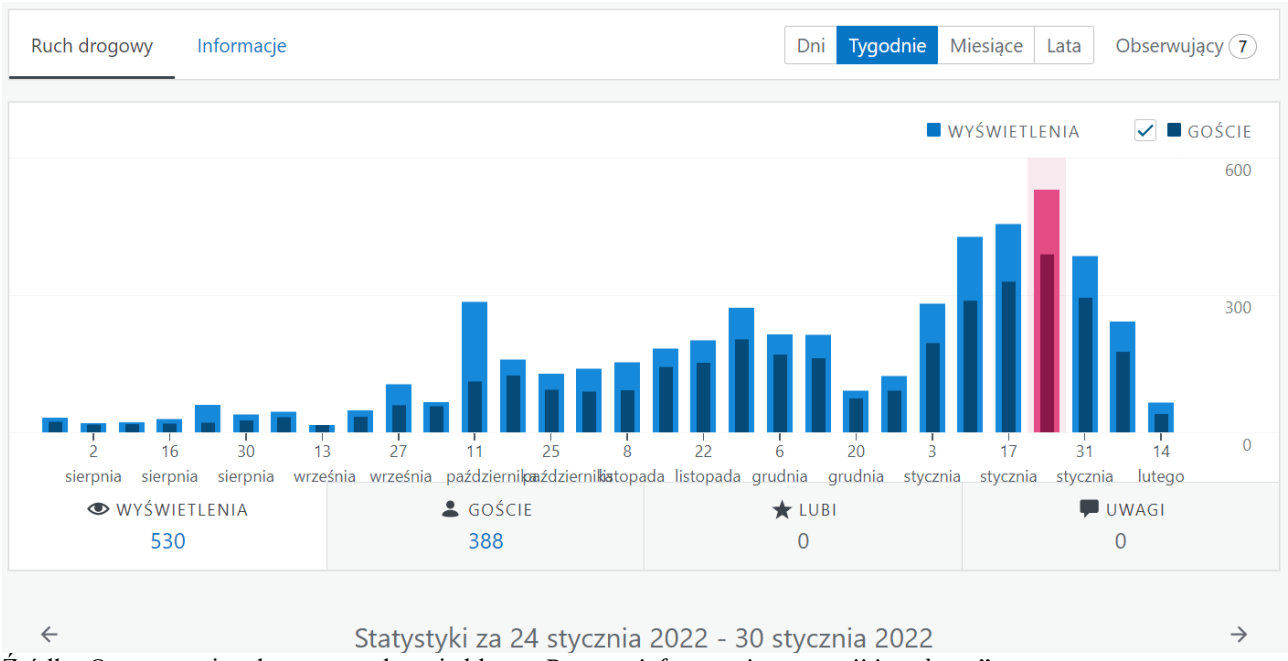
***Ś**rednia krocząca. https://www.xtb.com/pl/edukacja/srednia-kroczaca*. Jest to np. wskaźnik, który podąża za trendem bazując na przeszłych cenach i obliczana jest poprzez wybranie określonego okresu i podzielenie go przez liczbę okresów. Pomaga ona na wygładzanie zachowania ceny.

Średnia ruchoma prosta. http://www.ekonometria.4me.pl/metoda-sredniej-ruchomej.htm. Modele średniej ruchomej służyć mogą zarówno do wygładzania szeregu czasowego jak i do prognozowania. Kolejne wartości prognoz wygasłych powstają poprzez obliczanie średniej arytmetycznej dla wybranej liczby elementów - stałej wygładzania k. Aby określić jaka średnia będzie najlepsza obliczamy prognozy wygasłe dla każdej średniej ruchomej $^{(k)}$ *a następnie obliczamy średni kwadratowy błędu prognozy ex post. Spośród badanych średnich wygrywa ta, która posiada błąd najmniejszy czyli można powiedzieć, iż jest najlepiej dopasowana do danych rzeczywistych szeregu. Średnia ruchoma prosta ma wadę polegającą na tym, iż każda z przyjętych do wygładzania obserwacji ma jednakową równą 1 wagę, czyli taki sam udział w szacunku prognozy.*

***Z**mienne latentne. http://www.ptde.org/pluginfile.php/1262/mod_page/content/16/47.pdf.* W diagnostyce edukacyjnej problem pomiaru jest szczególnie złożony, gdyż zmienne nas interesujące nie są bezpośrednio obserwowalne. Takie konstrukty psychologiczne jak inteligencja, kompetencja lub wiedza są zmiennymi latentnymi, które posiadają znacznie empiryczne jedynie poprzez obserwowalne konsekwencje tych zmiennych.

31. Moje publikacje w latach 2008-2022

Coraz częściej spotykam się z dużym zainteresowaniem internautów moimi publikacjami i wpisami, a świadczy o tym tygodniowa statystyka w *WordPress* na blogu: https://wornalkiewicz.wordpress.com/ (zob. rysunek 31.1).



Źródło: Opracowanie własne na podstawie blogu „*Procesy informacyjne w teorii i praktyce*".

Rys. 31.1. Statystyka w układzie tygodni na blogu moim blogu

Oprócz tego forum, współpracy z sympatykami metod ilościowych w domenie „*Google*" w Internecie, znajdują się pozycje ze zmiankami o moim dorobku naukowym na stronie www WSZiA Opole.

Dalej zostaną zaprezentowane wybrane z tej strony moje pozycje publikacji zarówno jako książki, monografie indywidualne oraz artykuły w monografiach zbiorowych. Zaprezentowane zostaną publikacje sygnalizowane w domenie Google po wywołaniu autora „*Władysław Wornalkiewicz*". Całość opracowano w układzie kolejności narastającej lat (2008-2022).

2008

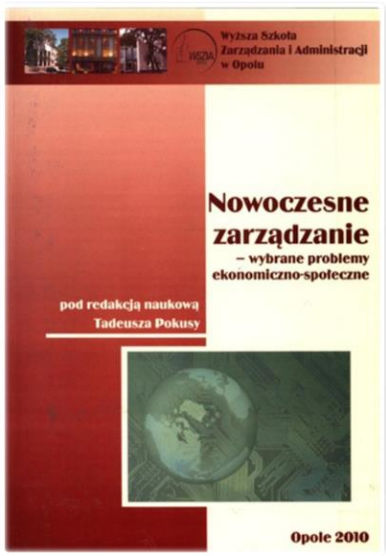
1. „*Ekonomiczno-społeczne problemy współczesnego zarządzania i komunikacji*", monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2008, 255 s., artykuł „*Dobór metod i parametrów w budowaniu modeli ekonometrycznych*", strony: 190-204, ISBN 978-83-88980-69-5, 978-83-7511-098-2.
Link:
https://wordpress.com/media/wornalkiewicz.wordpress.com.

2010

1. „*Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*”, monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Marian Duczmal, Opole: Wydawnictwo Instytut Śląski, 2010, 597 s., ISBN 978-8388980-94-7, 978-83-62105-41-0.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Wstep_do_ekonometrii_i_badan_operacyjnych.pdf.

2. „*Zarządzanie i polityka społeczna - wybrane problemy*”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2010, 443 s., artykuł „*Estymacja modeli wielorównaniowych w GRETL*” strony: 255-276, ISBN 978-83-88980-90-9, 978-83-62105-23-6.
Link: Link: https://wordpress.com/media/wornalkiewicz.wordpress.com.

3. „*Nowoczesne zarządzanie - wybrane problemy ekonomiczno-społeczne*”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2010, 242 s., artykuł: „*Propozycja określenia efektywności województw*”, strony: 189-208, ISBN 978-83-88980-98-5, 978-83-62105-53-3.
Link: https://integro.bg.polsl.pl/172200374055/ksiazka/nowoczesne-zarzadzanie-wybrane-problemy-ekonomiczno-spoleczne?bibFilter=17.

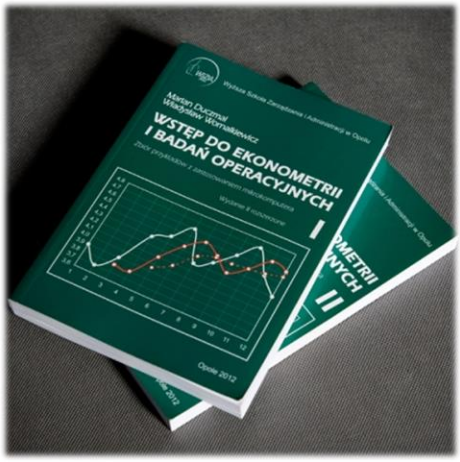


2011

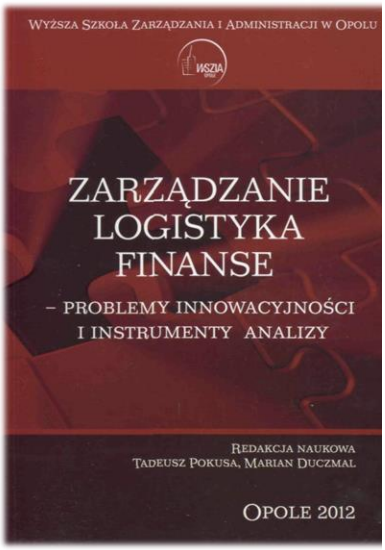
1. „*Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne*”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2011, 394 s., artykuł: „*Model ekonometryczny zmiennej „liczba bezrobotnych*””, strony: 172-186, ISBN 978-83-62683-20-8, 978 -83-7511-135-4.
Link:
https://wordpress.com/media/wornalkiewicz.wordpress.com.

2012

1. „*Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*”, wydanie II rozszerzone, część I, monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Marian Duczmal, Opole: Wydawnictwo Instytut Śląski, 2012, 411 s., ISBN 978-83-62683-30-7, 978-83-7511-259-0.
2. „*Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*”, wydanie II rozszerzone, część II, monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Marian Duczmal, Opole: Wydawnictwo Instytut Śląski, 2012, 223 s., ISBN 978-83-62683-30-7, 978-83-7511-259-0.



3. „*Zarządzanie Logistyka Finanse - Problemy innowacyjności i instrumenty analizy*”, monografia zbiorowa recenzowana, Opole: Wyższa Szkoła Zarządzania i Administracji w Opolu, 2012, … s., artykuły: „*Model wielorównaniowy PKB*”, strony: 223-249, „Etapy i procedury budowy modelu produktu brutto podregionów”, strony: 261-279, ISBN
Link:
https://w.bibliotece.pl/1798765/Zarz%C4%85dzanie+logistyka+finanse.

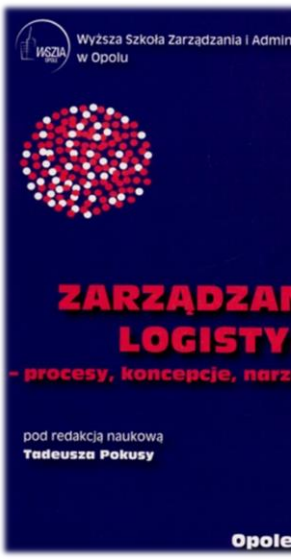


2013

1. „*Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*”, skrypt nr 1/2013 - pomocniczy do przedmiotu Ekonometria, Opole: Wyższa Szkoła Zarządzania i Administracji w Opolu, 2013, 252 s., ISBN 978-83-62683-44-4, 978-83-7511-187-3.
Link: https://wordpress.com/media/wornalkiewicz.wordpress.com; https://books.google.pl/books/about/Metoda_badania_przyczynowo_skutkowego_zw.html?id=0d0DoQEACAAJ&redir_esc=y.

2. „*Modele ekonometryczne PKB obiektów struktury terytorialnej*”, monografia indywidualna recenzowana, Opole: Wydawnictwo Instytut Śląski, 2013, 343 s., ISBN 978-83-62683-36-9, 978-83-7511-170-5.
Link: https://integro.bs.katowice.pl/32403087968/wornalkiewicz-wladyslaw/modele-ekonometryczne-pkb-obiektow-struktury-terytorialnej.

3. „*Zarządzanie Logistyka - procesy, koncepcje, narzędzia*”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2013, 304 s., , artykuł „*Programowanie wielokryterialne z zastosowaniem WinQSB*”, strony: 168-184, ISBN 978-83-62683-49-9, 978-83-7511-192-7.
Link: https://integro.ciniba.edu.pl/integro/192504222555/ksika/zarzdzanielogistykakoncepcjeprocesynarzdzia.



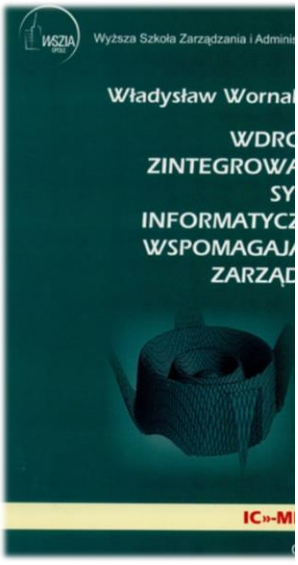
4. „*Społeczno-ekonomiczne uwarunkowania zarządzania i administracji - innowacyjność, komunikacja*”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2013, 251 s., artykuł „*Wybór lokalizacji obiektu z zastosowaniem Expert Choice*”, strony: 124-140, ISBN 978-83-62683-53-6, 978-83-7511-103-4.
Link: https://wordpress.com/media/wornalkiewicz.wordpress.com.

2014

1. „*Formułowanie modeli ekonometrycznych na potrzeby zarządzania”* cz. I, monografia indywidualna recenzowana (podręcznik akademicki), Opole: Wydawnictwo Instytut Śląski, 2014, 665 s., ISBN 978-83-62683-64-2, 978-83-7511-210-8.
2. „*Formułowanie modeli ekonometrycznych na potrzeby zarządzania”* cz. II, monografia indywidualna recenzowana (podręcznik akademicki), Opole: Wydawnictwo Instytut Śląski, **2015**, 665 s., ISBN 978-83-62683-64-2, 978-83-7511-210-8.
Link: https://integro.bg.polsl.pl/172600881896/wornalkiewicz-wladyslaw/formulowanie-modeli-ekonometrycznych-na-potrzeby-zarzadzania.

2015

1. „*Wdrożenie zintegrowanego systemu informatycznego wspomagającego zarządzanie*”, monografia indywidualna recenzowana (podręcznik akademicki), Opole: Wydawnictwo Instytut Śląski, 2015, 370 s., ISBN 978-83-62683-67-3, 978-83-7511-226-9.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Wornalkiewicz_Wdrozenie_zintegrowanego_systemu_informatycznego.pdf.



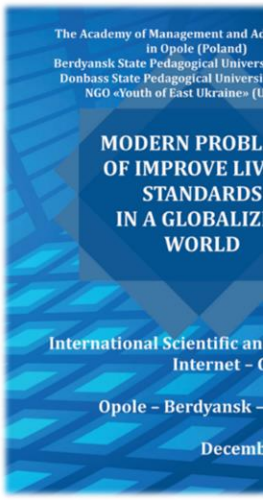
2. „*Ekonomiczno-społeczne uwarunkowania rozwoju gospodarczego - zarządzanie informacją i nowymi technologiami*”, monografia zbiorowa recenzowana, Opole: *Wydawnictwo „Instytut Śląski”,* 2015, 507 s., artykuły i strony: *„Przejawy wdrożenia systemów informatycznych*” (85-108), „*Modelowanie procesów zarządzania*” (124-146), współautorstwo: Ryszard Broszkiewicz, „*EDI w procesie logistycznym*” (278-301), „*Analityka biznesowa*” (356-375).
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/2_2021.pdf.

3. *„Koncepcje i narzędzia współczesnej logistyki - monografia upowszechniająca projekt „WSZiA w Opolu w perspektywie realizacji Strategii Europa 2020 Wzmocnienie potencjału dydaktycznego Uczelni w obszarze kształcenia kadr dla sektora logistyki poprzez wdrożenie programu rozwojowego””,* praca zbiorowa, Opole: Wydawnictwo Instytut Śląski, 102 s., artykuł „*Implementacja systemów klasy ERP w logistyce*”, ISBN 978-83-62683-75-8, 978-83-7511-235-1.
Link:
https://wordpress.com/media/wornalkiewicz.wordpress.com.

2016

1. „*Wprowadzenie do projektowanie systemów informatycznych zarządzania*” Część 1, monografia indywidualna recenzowana (podręcznik), Opole: Wydawnictwo Wyższej Szkoły Zarządzania i Administracji w Opolu, 2016, 328 s., ISBN 978-83-62683-97-0, 978-83-7511-243-6.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Projektowanie_systemow_informatycznych_zarzadzania.pdf.
2. „*Wprowadzenie do projektowania systemów informatycznych zarządzania*” część 2, monografia indywidualna recenzowana (podręcznik), Opole: Wydawnictwo Wyższej Szkoły Zarządzania i Administracji w Opolu, 2016, 567 s., ISBN 978-83-62683-97-0, 978-83-7511-243-6.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Wprowadzenie_do_projektowania_Czesc2.pdf.

3. „*Conference Proceedings of the International Scientific Internet-Conference Modern Problems of Improve Living Standards in a Globalized World*”, materiały pokonferencyjne - monografia zbiorowa recenzowana, Opole - Berdyansk - Slavyansk), 2016, (electronic edition), 534 s., artykuł „*Rozwiązanie problemu transportowego metodą VAM*” (22-28), ISBN 978-83-62683-871.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2016_modern_problems_of_imrpve_living_standards_in_a_globalized_world_slavyansk.pdf.



4. „*Social and Economic Priorities in the Context of Sustainable Development*”, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2016, 444 s., artykuł „*Product promotion and company image in Internet*” (*Promocja produktu i wizerunku firmy w Internecie*), strony: 138-148, , ISBN 978 - 83 - 62683 -78 - 9.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2016_priorytet_spoleczno_gospodarcze_w_kontekscie_zrownowazonego_rozwoju.pdf.

5. „*20 lat Wyższa Szkoła Zarządzania i Administracji w Opolu - interdyscyplinarność badań*”, monografia zbiorowa recenzowana jubileuszowa, Opole: Wydawnictwo Instytut Śląski, 2016, 375 s., artykuł: „*Wyszukiwarki i media społecznościowe*”, strony: 320-333, ISBN 978-83-62683-86-4, 978-83-7511-247-4.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2016_wyzsza_szkola_zarzadzania_i_administracji_w_opolu_interdyscyplinarnosc_badan.pdf.

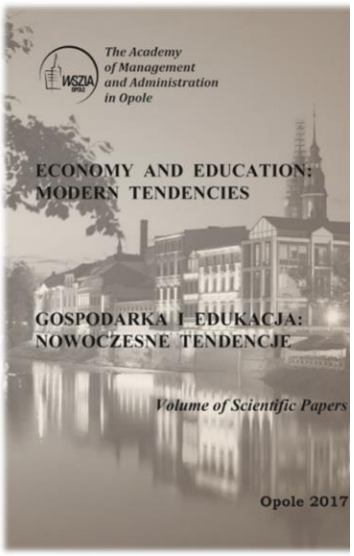
6. „СОЦІАЛЬНО-ЕКОНОМІЧНИЙ РОЗВИТОК: РОЛЬ ІНФОРМАЦІЇ ТА НОВИХ ТЕХНОЛОГІЙ”, monografia zbiorowa recenzowana, Бердянський державний педагогічний університет, 2016, 295 s., artykuł „*Symulacja biznesowa*”, strony: 197-208, ISBN 978-617-7291-44-1.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2016_%D0%9E%D0%A6%D0%86%D0%90%D0%9B%D0%AC%D0%9D%D0%9E_%D0%95%D0%9A%D0%9E%D0%9D%D0%9E%D0%9C%D0%86%D0%A7%D0%9D%D0%98%D0%99.pdf.

7. „*Освіта і суспільство. Міжнародний збірник наукових праць*”, Бердянський державний педагогічний університет, 2016, 418 s., artykuł *„ Ranking metod ilościowych w Internecie*”, strony: 275-285, ISBN 978-617-7291-80-9 (електронне видання).
Link: https://bdpu.org.ua/wp-content/uploads/2019/10/Papers_Berdyansk_2016.pdf.

8. „*Basic Trends in Public Sector*” („*Podstawowe tendencje w sektorze publicznym*”), monografia zbiorowa recenzowana, artykuł: *Designing of managerial consoles* (*Projektowanie pulpitów menadżerskich*), strony: 58-74.
Opole: The Academy of Management and Administration in Opole, 2016, 200 s., ISBN 978-83-62683-79-6.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2016_podstawowe_tendencje_w_sektorze_publicznym_duczmal_nestoreka-1.pdf.

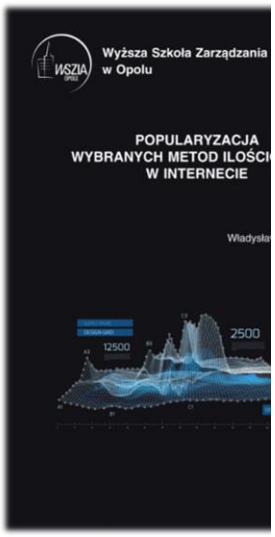
2017

1. „*ECONOMY AND EDUCATION: MODERN TENDENCIES*” - *GOSPODARKA I EDUKACJA: NOWOCZESNE TENDENCJE*, Volume of Scientific Papers, monografia zbiorowa recenzowana, Wyższa Szkoła Zarządzania i Administracji w Opolu, 2017, 362 s., artykuł „*Prognozowanie z wykorzystaniem zasady postarzania informacji*”, strony: 39-43, ISBN 978 - 83 - 62683 - 24 - 6.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2017_economy_andeducation_modern_tendencies.pdf.

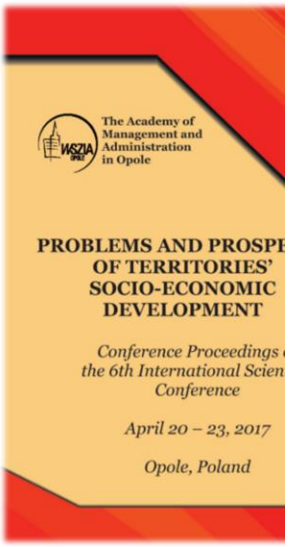


2. „*Modern Problems of Management: Economics, Education, Health Care and Pharmacy*”, Conference Proceedings of the 5 th International Scientific Conference, 232 s., Opole, The Academy of Management and Administration in Opole, 2017, artykuł „*Zastosowanie konwertera plików*”, strony: 25-27, 978-83-62683-23-9.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2017_modern_problems_of_management_econimics_education_health_care_and_pharmacy.pdf.

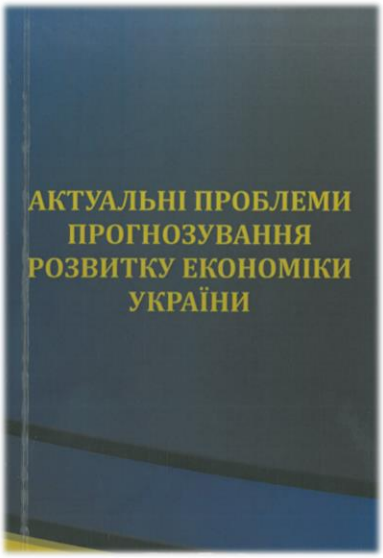
3. „*Popularyzacja wybranych metod ilościowych w Internecie*”, monografia indywidualna, Wyższa Szkoła Zarządzania i administracji w Opolu, Opole, 2017, 351 s.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2017_popularyzacja_wybranych_metod_ilosciowych_w_interneice_wornalkiewicz.pdf.

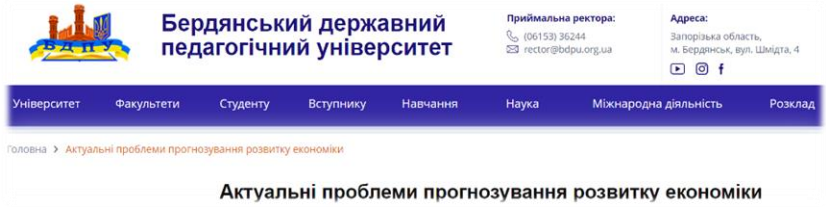


4. „*Problems and Prospects of Territories' Socio-Economic Development*”, Conference Proceedings of the 6 th International Scientific Conference, materiały pokonferencyjne, monografia zbiorowa recenzowana, The Academy of Management and Administration in Opole, 2017, 272 s., artykuł: „*Optymalizacja marszrutyzacji przewozów z zastosowaniem funkcji Excela*”, strony: 41-43, ISBN 978-83-62683-10-9.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2017_problems_and_prodprcts_pf_territories_socio_economic_development.pdf.



5. „*Актуальні проблеми прогнозування розвитку економіки…*”
(„*Bieżące problemy prognozowania ekonomiki Ukrainy*”), zbiór prac naukowych - monografia zbiorowa recenzowana, 370 s., КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА (Kijowski Narodowy Uniwersytet im. Tarasa Szewczenki) i inne, 2017, „*Optymalizacja przewozów z zastosowaniem funkcji Excela*” strony: 154-169, ISBN 978-617-7291-98-4.
Link: https://bdpu.org.ua/actual-problems-of-forecasting-economic-development/.

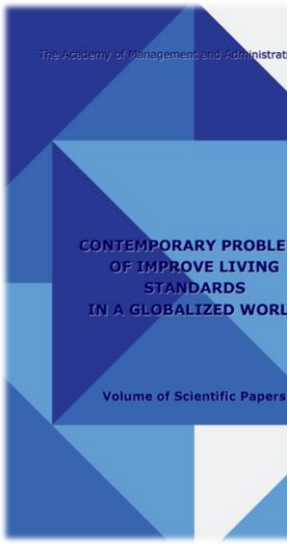




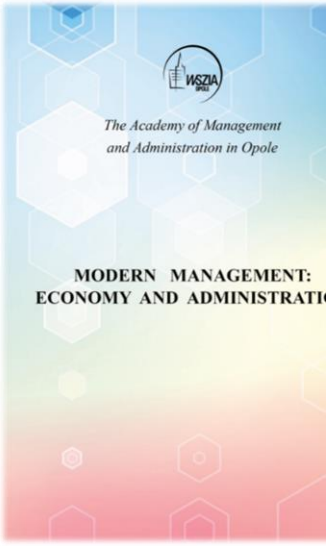
6. „*Transformations in contemporary society: economic aspects*” (*Przemiany w nowoczesnym społeczeństwie: aspekty ekonomiczne*”, monografia zbiorowa recenzowana, Wyższa Szkoła Zarządzania i Administracji w Opolu, 2017, 348 s., artykuł: „*Forecasting using the multiplicative model*” („*Prognozowanie z zastosowaniem modelu multiplikatywnego*”), strony: 205-212, ISBN 978-83-62683-96-3.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2017_przemiany_w_nowoczesym_spoleczenstwie_aspekty_ekonomiczne_nosferanko_pokusa.pdf.

2018

1. „*Contemporary Problems of Improve Living Standards in a Globalized World*”, Volume of Scientific Papers, monografia zbiorowa recenzowana, The Academy of Management and Administration in Opole, Opole, 2018, electronic edition, 770 s., artykuł „*Echa migracji w wybranej literaturze* (*Stosowane metody i modele*)”, strony: 516-527, ISBN 978-83-946765-1-3.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2018_contemporary_problems_of_improwve_living_in_a_globalized_world.pdf.

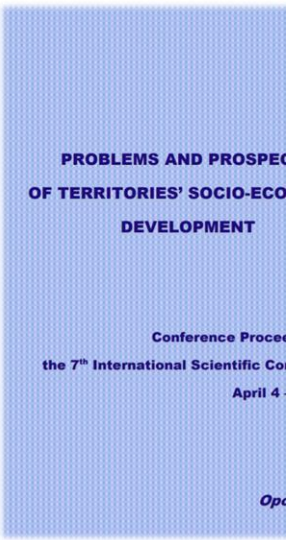


2. „*MODERN MANAGEMENT: ECONOMY AND ADMINISTRATION*” (*NOWOCZESNE ZARZĄDZANIE: EKONOMIA I ADMINISTRACJA*), monografia zbiorowa recenzowana, The Academy of Management and Administration in Opole, Opole, 2018, 218 s., artykuł ”*Applications used in designing websites*” (*Aplikacje stosowane w projektowaniu stron www*) , strony: 154-160, ISBN 978-83-62-683-27-7.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2018_modern_management_economy_and_administration.pdf.



3. „*Modern Problems of Management: Economics, Education, Health Care and Pharmacy*” (*Współczesne problemy zarządzania: Ekonomia, edukacja, opieka zdrowotna i farmacja*), Conference Proceedings of the 6th International Scientific Conference, Opole, The Academy of Management and Administration in Opole, 2018, 116 s., artykuł „*Potrzeba utworzenia piramidy wskaźników procesów ludnościowych*”, strony: 22-25, ISBN 978-83-946765-1-3.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2018_modern_problems_of_management_economics_education_health_nad_pharmcy.pdf.

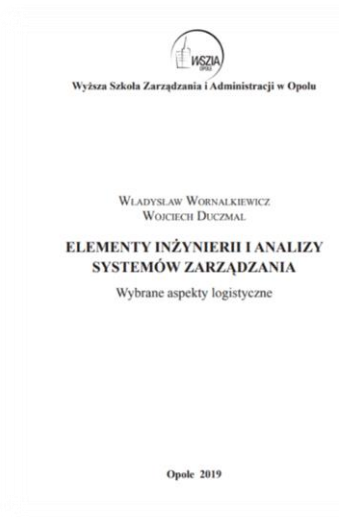
4. „*Problems and Prospects of Territories' Socio-Economic Development*" Conference Proceedings of the 7 th International Scientific Conference, Opole, materiały pokonferencyjne, monografia zbiorowa recenzowana, The Academy of Management and Administration in Opole, 2018, 345 s., artykuł „*Ocena nieefektywności gospodarowania zasobami ludności*", strony: 70-72, ISBN 978 - 83 - 62683 - 25 - 3.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2018_problems_andprospects_of_territories_socio_economic_development.pdf.



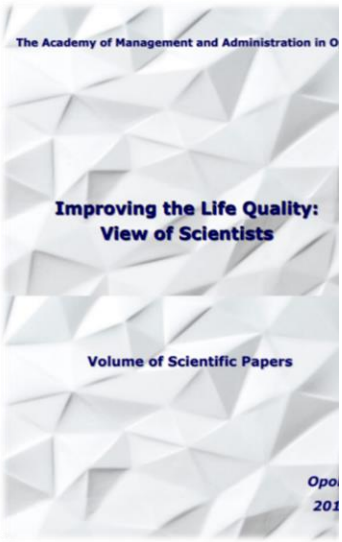
5. „*Uwarunkowania demograficzne rozwoju społecznego i ekonomicznego miasta Nysa i powiatu nyskiego*", Opole-Nysa, Wyższa Szkoła Zarzadzania i Administracji w Opolu, 2018, 260 s., artykuł: „*Prognozowanie migracji ludności z uwzględnieniem wag harmonicznych*", strony: 212-233,
ISBN 978-83-946765-0-6.

2019

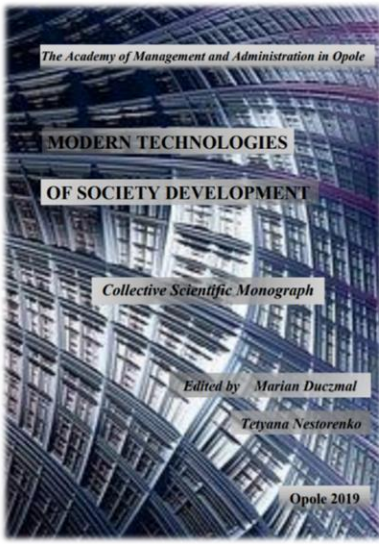
1. „*Elementy inżynierii i analizy danych systemów zarządzania Wybrane aspekty logistyczne*", monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Wojciech Duczmal, Opole, Wydawnictwo Wyższej Szkoły Zarządzania i Administracji w Opolu, 2019, 341 s., ISBN 978-83-946765-8-2.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Elementy_inzynierii_i_analizy_systemow_zarzadzania.pdf.



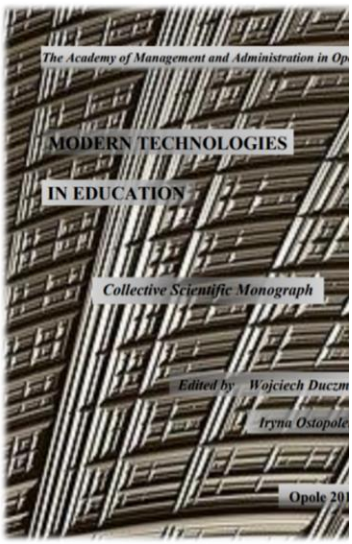
2. „*Improving the Life Quality: View of Scientists*”, Volume of Scientific Papers, monografia zbiorowa recenzowana, The Academy of Management and Administration in Opole, Opole, 2019, electronic edition, 660 s., artykuły: „*Pozyskiwanie danych o odległościach dla potrzeb zagadnienia transportowego*” strony: 118-136, „*Prognozowanie zewnętrznych usług transportowych*” strony: 137-156, „*Zastosowanie taksonomii wrocławskiej*” (157-188), ISBN 978-83- 946765-3-7.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_improving_the_life_quality_view_of_scientists.pdf.



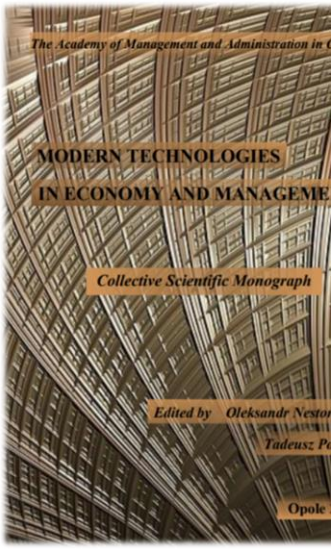
3 „*Modern Technologies of Society Development*”, Collective Scientific Monograph, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2019, 410 s., artykuł: „*Verification of the research when comparing in pairs*” (*Weryfikacja badania przy porównywaniu parami*), strony: 119-129, ISBN 978- 83-946765-6-8.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_modern_technologies_in_education.pdf.



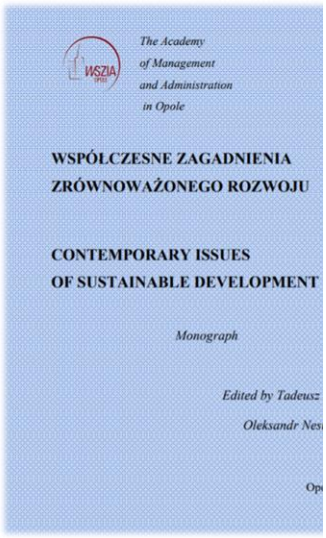
4. „*Modern Technologies in Education*”, Collective Scientific Monograph, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2019, 495 s., artykuł: „*Converting PDF to DOC*” (*Konwersja pliku formatu PDF na DOC*), strony: 184-194, ISBN 978-83- 946765-5-1.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_modern_technologies_in_education_dauczan.pdf.



5. „*Modern Technologies in Economy and Management*”, Collective Scientific Monograph, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2019, 493 s., artykuł: „*Mortality modeling*” (*Modelowanie umieralności*), strony: 148-160, ISBN 978-83-946765-4-4.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_moderntechnologies_in_ecenomy_and_managment.pdf.



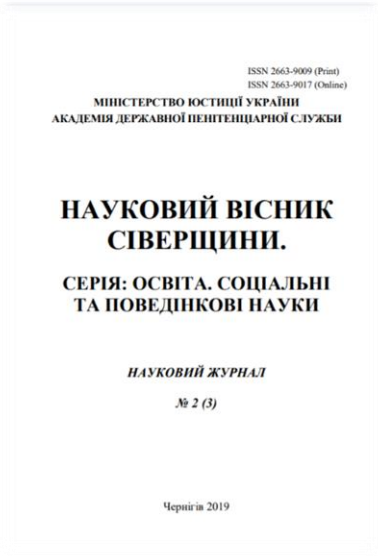
6. „*Contemporary issues of sustainable development*”, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2019, 452 s., artykuł: „*Technology of optimization solutions in decision-making task*” (*Techniki rozwiązań optymalizacyjnych zadania decyzyjnego*), strony: 63-82, ISBN 978 - 83 - 946765 - 7 - 5.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_wspolczesne_zagadnienia_zrownowazonego_rozwoju.pdf.



7. „*Наука ІІІ тисячоліття: пошуки, проблеми*”, перспективи розвитку, 2019, Збірник тез, БЕРДЯНСЬКИЙ ДЕРЖАВНИЙ ПЕДАГОГІЧНИЙ УНІВЕРСИТЕТ РАДА МОЛОДИХ УЧЕНИХ, 291 s., artykuł „*Conversion of Audio File to Text File*” (*Konwersja pliku audio na tekstowy*), strony: 199-200, УДК 378:001. (063) Н 34.

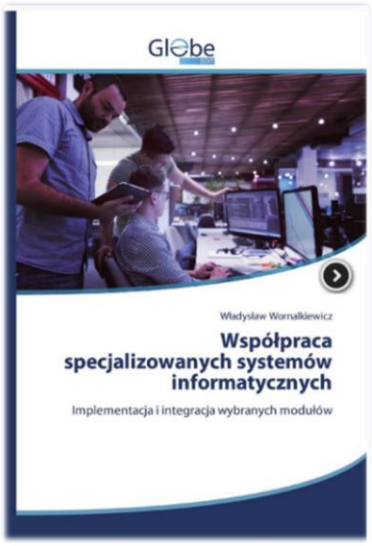
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_%D0%9D%D0%B0%D1%83%D0%BA%D0%B0_%D0%86%D0%86%D0%86_%D1%82%D0%B8%D1%81%D1%8F%D1%87%D0%BE%D0%BB%D1%96%D1%82%D1%82%D1%8F.pdf.

8. „*Науковий вісник Сіверщини. Серія: Освіта*”, Соціальні та поведінкові науки: науковий журнал / Академія Державної пенітенціарної служби”. Чернігів: Академія ДПтС, 2019. № 2 (3). 208 с. artykuł „*Przyszłość → magistrala drogowa Hamburg-Szanghaj*”, strony: 178-207, ISSN 2663-9009 (Print), ISSN 2663-9017 (Online).



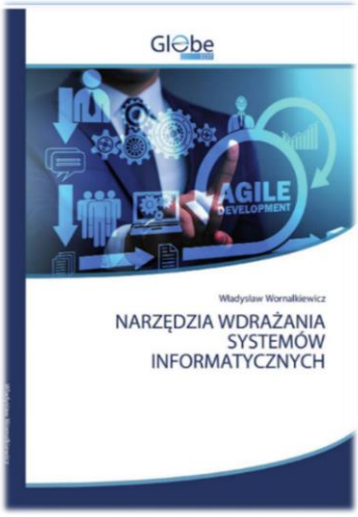
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_%D0%9D%D0%90%D0%A3%D0%9A%D0%9E%D0%92%D0%98%D0%99_%D0%92%D0%86%D0%A1%D0%9D%D0%98%D0%9A.pdf.

9. „*Współpraca specjalizowanych systemów informatycznych Implementacja i integracja wybranych modułów*”, monografia indywidualna, wydawca: GlobeEdit (International Book Market Service Ltd. , Member of OmniScriptum Publishing Group), Republic of Moldova, druk: Printforce - United Kingdom, 2019, 52 s., ISBN 978-613-42041-4.
Link: https://www.morebooks.de/store/gb/book/wsp%C3%B3%C5%82praca-specjalizowanych-system%C3%B3w-informatycznych/isbn/978-613-9-42041-4.

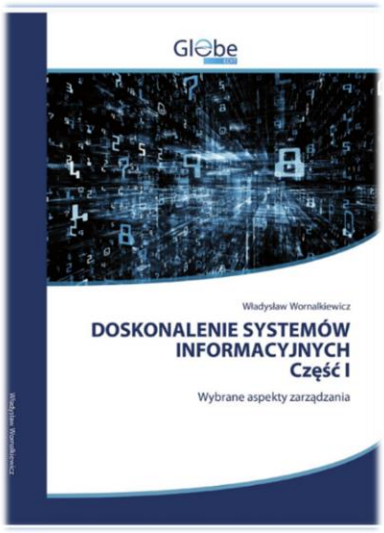


2020

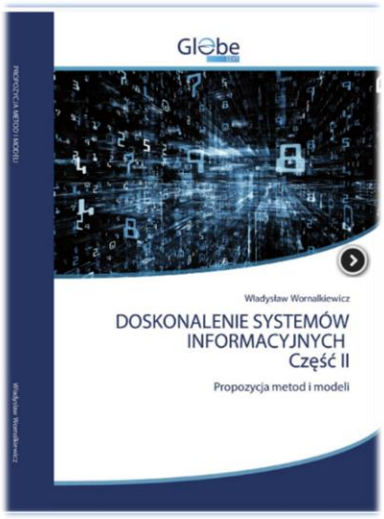
1. *„Narzędzia wdrażania systemów informatycznych”*, monografia indywidualna, wydawca: GlobeEdit (International Book Market Service Ltd. , Member of OmniScriptum Publishing Group), Republic of Moldova, druk: Printforce - United Kingdom, 2020, 140 s., ISBN 978-620-0-61025-6.
Link: https://www.morebooks.de/gb/p_978-620-0-61025-6; https://www.globeedit.com; https://www.morebooks.de/store/gb/book/narz%C4%99dzia-wdra%C5%BCania-system%C3%B3w-informatycznych/isbn/978-620-0-61025-6.



2. „*Doskonalenie systemów informacyjnych*” Część I „*Wybrane aspekty zarządzania*”, monografia indywidualna, wydawca: GlobeEdit (International Book Market Service Ltd. , Member of OmniScriptum Publishing Group), Republic of Moldova, druk: Printforce - United Kingdom, 2020, 388 s.,
ISBN 978-620-0-59233-0.
Link:
https://www.morebooks.de/store/gb/book/doskonalenie-system%C3%B3w-informacyjnych-cz%C4%99%C5%9B%C4%87-i/isbn/978-620-0-59233-0.



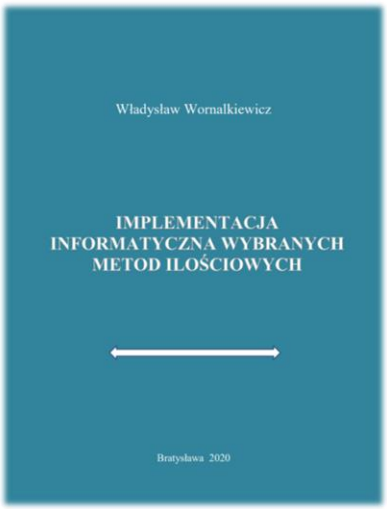
3. „*Doskonalenie systemów informacyjnych*” Część II „*Propozycja metod i modeli*”, monografia indywidualna, wydawca: GlobeEdit (International Book Market Service Ltd. , Member of OmniScriptum Publishing Group), Republic of Moldova, druk: Printforce - United Kingdom, 2020, 452 s., ISBN 978-620-0-59542-3.
Link:
https://www.morebooks.de/store/gb/book/doskonalenie-system%C3%B3w-informacyjnych-cz%C4%99%C5%9B%C4%87-ii/isbn/978-620-0-59542-3.



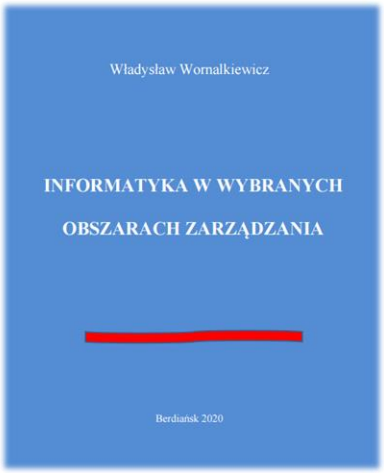
4. „*Освіта і суспільство V*: Міжнародний збірник наукових праць”, Бердянський державний педагогічний університет, Ополє: видавництво Вищої школи управління і адміністрації в Ополє, 422 с., 2020, artykuły: „*MOBILE APPLICATIONS IN LOGISTICS*” („*Aplikacje mobilne w logistyce*”) (strony: 277-287), „*METADATA EDITING PROGRAMS*”, „*Programy edycji metadanych*” (strony: 288-299), ISBN 978-83-66567-00-9.

Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2020_%D0%9E%D0%A1%D0%92%D0%86%D0%A2%D0%90_%D0%86_%D0%A1%D0%A3%D0%A1%D0%9F%D0%86%D0%9B%D0%AC%D0%A1%D0%A2%D0%92%D0%9E_V.pdf.

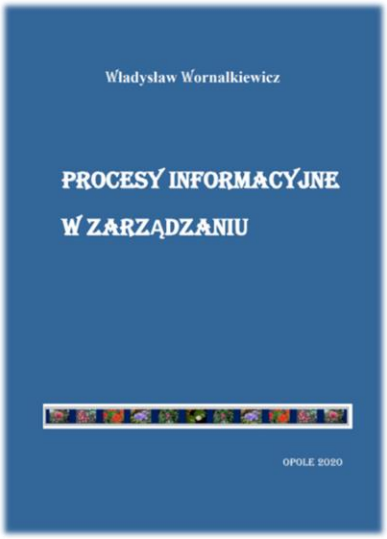
5. *„Implementacja informatyczna wybranych metod ilościowych*”, monografia indywidualna recenzowana, Bratysława: Wydawnictwo DENAKYR, s. r. o., 2020, 500 s., ISBN 978-80-973568-0-4.
Link:
https://wornalkiewicz.files.wordpress.com/2020/02/implementacja-informatyczna-wybranych-metod-iloc59bciowych.pdf.



6. „*Informatyka w wybranych obszarach zarządzania*”, monografia indywidualna recenzowana, Berdiańsk: Wydawca „Svidler A.L.”, 2020, 450 s., ISBN 978-617-627-145-1.
Link:
https://wornalkiewicz.files.wordpress.com/2020/10/informatyka-w-wybranych-obszarach-zarzadzania.pdf.

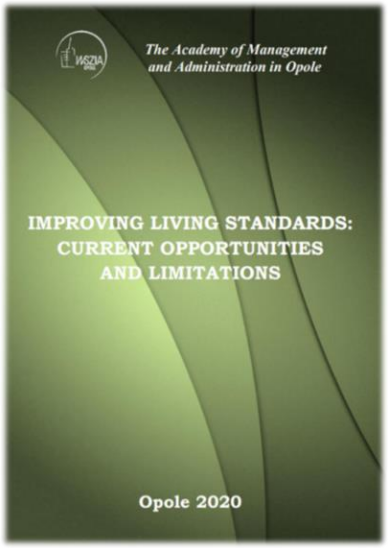


7. „*Procesy informacyjne w zarządzaniu*”, monografia indywidualna recenzowana, Opole: Wydawnictwo - Wyższa Szkoła Zarządzania i Administracji w Opolu, 2020, 373 s., ISBN 978-83-665-22-1.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Procesy_informacyjne_w_zarzadzaniu.pdf.

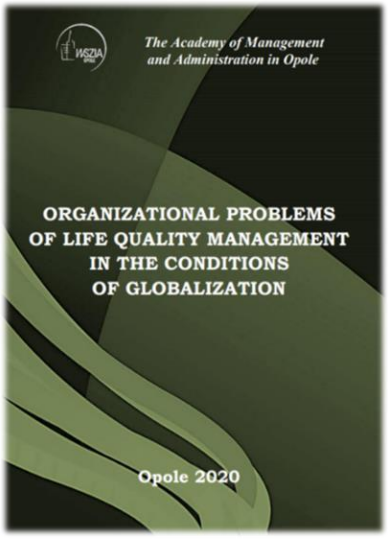


8. „*Journal of Modern Economic Research*”, Bratysława, monografia zbiorowa recenzowana, artykuł „*Innovative logistic solutions*” (*Innowacyjne rozwiązania logistyczne*), strony: 53-63, współautorstwo: Maksym Kutsenko, 2020, ISSN 2644-4380 nadruk; 2644-6332 online.
Link: https://denakyrpublishing.science/index.php/jmer/article/view/40.

9. „*Improving living standards: current opportunities and limitations*”, monografia zbiorowa recenzowana, 594 s., Opole: The Academy of Management and Administration in Opole, 2020, artykuł „*POS systems*” (*Systemy klasy POS*), strony: 139-167, ISBN 978 - 83 - 66567 - 21 - 4.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Zbirnyk_Osvita-i-suspilstvo-VI_new.pdf.

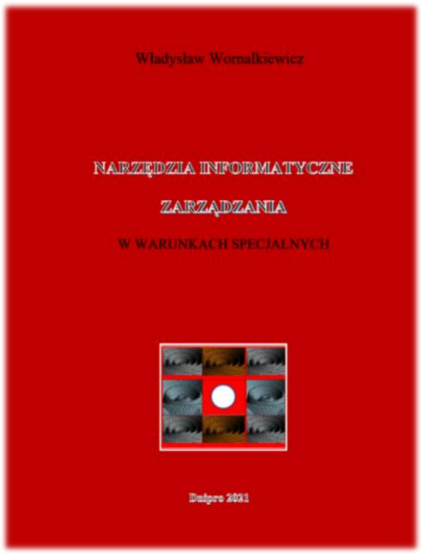


10. „*Organizational problems of life quality management in the conditions of globalization*”, Opole: The Academy of Management and Administration in Opole, 428 s., 2020, artykuł „*Existing and intended logistic projects*” (*Istniejące i zamierzone przedsięwzięcia logistyczne*), strony: 309-334, ISBN 978 - 83 - 66567 - 20 - 7.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/2_2021.pdf.

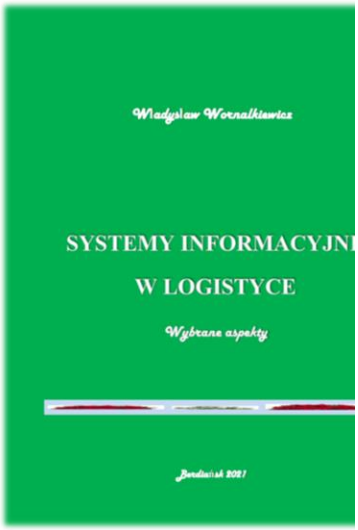


2021

1. *„Narzędzia informatyczne zarządzania w warunkach specjalnych*", monografia indywidualna recenzowana, Dnipro: Wydawca „Svidler A.L.", 2021, 288 s., ISBN 978-617-627-168-0.
Link: https://dspace.bdpu.org/handle/123456789/4082

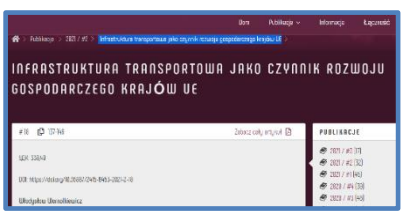


2. „*Systemy informacyjne w logistyce Wybrane aspekty*", monografia indywidualna recenzowana, Dnipro: Wydawca „Svidler A/L.", 2021, 376 s., ISBN 978-617-627-157-4.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Systemy_informacyjne_w_logistyce_Wornalkiewicz.pdf.



3. „*Економічні науки*", monografia zbiorowa recenzowana, Вісник Хмельницького національного університету, 2021, № 5, Том I, artykuł: *THE OVERSIZED CARGO FORWARDING: ECONOMIC AND LEGAL ASPECTS*, UDC 330, strony: 175-180, ISSN 2307-5740, współautorstwo: YURII KRAVCHYK, ALLA KAPLUNOVSKA.
Link: vknu-es-2021-n-5-298-175-180[1].pdf (khnu.km.ua). 3. „*Економічні науки*", monografia zbiorowa recenzowana, Вісник Хмельницького національного університету, 2021, № 5, Том I, artykuł: *THE OVERSIZED CARGO FORWARDING: ECONOMIC AND LEGAL ASPECTS*, UDC 330, strony: 175-180, ISSN 2307-5740, współautorstwo: YURII KRAVCHYK, ALLA KAPLUNOVSKA. Link: vknu-es-2021-n-5-298-175-180[1].pdf (khnu.km.ua).

4. „*Ukrainian Journal of Applied Economics*", 2021, Volume 6, Nr 2, artykuł „*Transport infrastructure as a factor of the EU countries' economic development*" („*Infrastruktura transportowa jako czynnik rozwoju gospodarczego krajów UE*"), strony: 137-146, współautorstwo:
Alla Kaplunovska, Olena Padchenko, ISSN 2415-8453.
Link: https://doi.org/10.36887/2415-8453-2021-2-18.



5. Huzhou University's Multicultural Center (Wielokulturowe Centrum Huzhou - Chiny), UDC 656.021.2, artykuł „*The perspective of increasing of road capacity*", („*Perspektywa rozwoju dróg szosowych*"), współautorstwo: Ievegen Medvediev, Seriy Soroka, 2021, strony: 12-24.
Link:
https://wornalkiewicz.files.wordpress.com/2021/08/uniwersytet-chiny.pdf; http://kwh.zjhu.edu.cn.

2021 年 第 3 期

目录

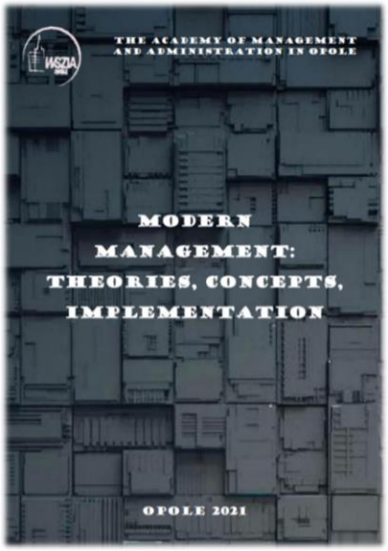
跨文化研究

3
12
25
39
54
70
85
92
101
109
121
143
154
174

内部资料 免费赠阅

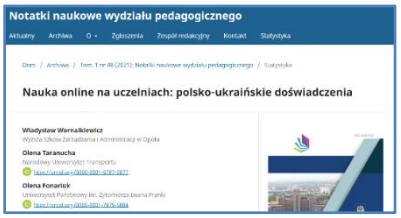
6. „*SCIENTIFIC NOTES OF THE PEDAGOGICAL DEPARTMEN*", artykuł „*ONLINE LEARNING AT UNIVERSITIES: POLISH-UKRAINIAN EXPERIENCE*" „*Nauczanie online na uniwersytetach: polsko-ukraińskie doświadczenia*", strony: 123-132, УДК 378.147.31, 2021, współautorstwo: Olena Taranukha, Olena Fonariuk.
Link: 17560-Текст статті-34489-1-10-20210721 (2).pdf; https://periodicals.karazin.ua.

7. „*Modern management: theories, concepts, implementation*", monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2021, 430 s., artykuł „*UNBLOCKING THE "ODRA - DANUBE" WATERWAY*" (*Udrożnienie drogi wodnej Odra - Dunaj*), strony: 121-141, ISBN 978-83-66567-24-2.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/7_2021.pdf.

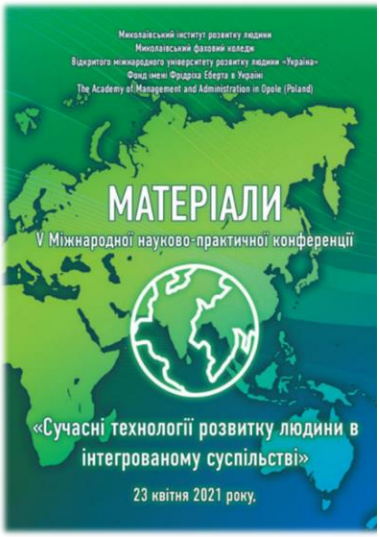


8. „*ОСВІТА І СУСПІЛЬСТВО VI*” Міжнародний збірник наукових праць, monografia zbiorowa recenzowana, Bierdiańsk: Wydawnictwo - Wyższa Szkoła Zarządzania i Administracji w Opolu, 2021, artykuł „*Opportunities to make milk reception logistics more modern*” („*Możliwości unowocześnienie logistyki odbioru mleka*”)„ strony: 328-343, ISBN 978-83-66567-26-9.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Zbirnyk_Osvita-i-suspilstvo-VI_new.pdf.

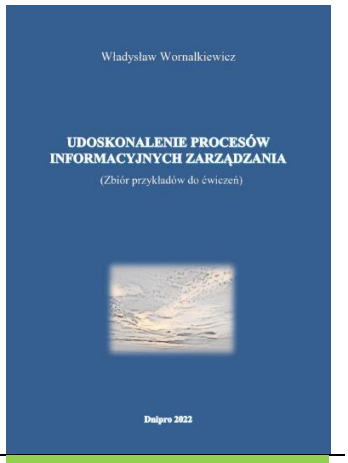
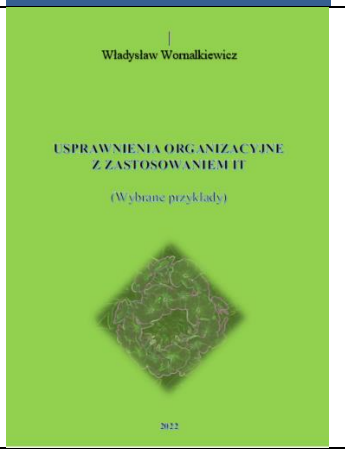
9. „*SCIENTIFIC NOTES OF THE PEDAGOGICAL DEPARTMEN*”, artykuł „*ONLINE LEARNING AT UNIVERSITIES: POLISH-UKRAINIAN EXPERIENCE*” „*Nauczanie online na uniwersytetach: polsko-ukraińskie doświadczenia*”, strony: 123-132, УДК 378.147.31, 2021, współautorstwo: Olena Taranukha, Olena Fonariuk.
Link: 17560-Текст статті-34489-1-10-20210721 (2).pdf, https://periodicals.karazin.ua; https://periodicals.karazin.ua/pedagogy/article/view/17560.



10. „*СУЧАСНІ ТЕХНОЛОГІЇ РОЗВИТКУ ЛЮДИНИ В ІНТЕГРОВАНОМУ СУСПІЛЬСТВІ*” Матеріали V Міжнародної науково-практичної конференції (*Materiały V Międzynarodowej naukowo-praktycznej konferencji*), 2021, Миколаївський інститут розвитку людини (Instytut w Mikolaiv - Ukraina), monografia zbiorowa recenzowana pokonferencyjna, 308 s., artykuł „*Socio-logistical aspects of the Vistula spit dug-through*”, („*Społeczno-logistyczne aspekty przekopu Mierzei Wiślanej*”), strony: 302-304, удк 371: 378.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/Spoleczno_logistyczne_aspekty_przekopu.pdf.



2022

1. „*Udoskonalenie procesów informacyjnych zarządzania (Zbiór przykładów do ćwiczeń)*, monografia indywidualna recenzowana, Dnipro: Wydawca „Svidler A.L.”, 241 s., ISBN 978-617-627-170-3.	
2. „*Usprawnienia organizacyjne z zastosowaniem IT (Wybrane przykłady)*, monografia indywidualna recenzowana, Kijów: Wydawca „Majster Knyg”, s. 340, ISBN 978-617-7652-59-4. Link do strony internetowej Narodowego Uniwersytetu Gospodarki Miejskiej im. O.M. Beketova w Charkowie.: http://eprints.kname.edu.ua/61802/1/Monografia_7.pdf	
3. „*Zasilanie alternatywne pojazdów samochodowych*”, Zeszyty naukowe WSZT Katowice, Nr 2, w druku, 2022, s. 28.	
4. „*Przedsięwzięcia w zakresie budowy sieci 5G*”, Zeszyty naukowe WSZT Katowice, w druku, s. 21.	
5. Artykuły w ramach monografii EDUKACJA I SPOŁECZEŃSTWO VII, zbiór artykułów naukowych, Opole 2022, strony: 278-320, ISBN 978-83-66567-41-2: „*Modelowanie biznesowe z zastosowaniem UML*”, stron: 16, „*Potrzeba scalania systemów klasy ERP*”, stron: 12, „*Wspomaganie dystrybucji systemem Dynamics NAV*”, stron: 15. Wydanie przez: ANS Opole, Wydział Pedagogiczny Uniwersytetu w Preszowie (Słowacja), Państwowy Uniwersytet Pedagogiczny w Bierdańsku (Ukraina).	

Bibliografia

Analiza stanu innowacyjności województwa opolskiego, GEOPROFIT, Warszawa 2010.

Anholcer M., Gaspars H., Owczarkowski A., *Przykłady i zadania z badań operacyjnych i ekonometrii*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, Poznań 2003.

Blaik P., *Logistyka*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2001.

Brzezińska M., *W cieniu europejskiej twierdzy Obrazy Zachodu wśród Afrykanów Gwinea Bissau*, Wydawnictwo Uniwersytetu Warszawskiego, 2017.

Budyta-Budzyńska M., *Polacy na Islandii Rekonstrukcja przestrzeni obecności*, Scholar, Warszawa 2017.

Castles S., Miller M.J., *Migracje we współczesnym świecie*, Wydawnictwo Naukowe PWN, Warszawa 2020.

Cebul K., Fehler W., Podgórzanka R., *Migracje jako wyzwanie dla Unii Europejskiej i wybranych państw członkowskich*, Difin, Warszawa 2017.

Cisło W., *Imigranci u bram. Kryzys uchodźcy i męczeństwo chrześcijan XXI w.*, Wydawca: Biały Kruk, 2017.

Coyle J.J., Bardi E.J., Langley C.J. Jr , *Zarządzanie Logistyczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2002.

Cymanow P., *Ekonomiczno-społeczne następstwa migracji na obszarach górskich na przykładzie Karpat Polskich*, Difin, Warszawa 2018.

Czerwiński Z., Guzik B., *Prognozowanie ekonometryczne*, Polskie Wydawnictwo Ekonomiczne, Warszawa 1980.

Dittman P., *Metody prognozowania sprzedaży w przedsiębiorstwie,* Akademia Ekonomiczna we Wrocławiu, Wrocław 2000.

Duczmal M., *Efektywność działania obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski, Opole 2013.

Duczmal M., *Metody optymalizacji w zarządzaniu*, Wydawnictwo Instytut Śląski, Opole 2014.

Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski w Opolu, Wydanie II rozszerzone, 2012.

Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski , Opole 2010.

Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania,* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003.

Fihel A., *Starzenie się społeczeństwa a polityka fiskalna i migracyjna*, Wydawnictwo Uniwersytetu Warszawskiego, 2017.

Gardocka T., Sobczak J.(red. nauk.) *Uchodźcy w Polsce i Europie. Stan prawny i rzeczywistość*, Adam Marszałek, Toruń 2010.

Gawon Yoon, *A Comparison of Long-term Forecasting Methods: A Comment*, http://www.kostat.go.kr/attach/journal/3-2-6.PDF.

Guth M., *Przewidywanie domyślnych prawdopodobieństw dla testów warunków skrajnych. Porównanie modeli*, artykuł internetowy, 02.2022.

Han C., *Pracownicy migracyjni w Chinach*, Adam Marszałek, Toruń 2017.

Jajuga K. (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2002.

Jaremczuk E.J., (red. nauk.), *Migracje i kryzys uchodźczy w Europie. Rzeczywistość i wyzwania, FNCE,* Poznań 2017.

Kisperska-Moroń D., Krzyżaniak S., *Logistyka,* Biblioteka Logistyka, Poznań 2009.

Klaus W., *Przemoc ze strony najbliższych w doświadczeniach życiowych. Analiza kryminologiczna,* Scholar, Warszawa 2017.

Kopczewska K., Kopczewski T., Wójcik P., *Metody ilościowe w R: aplikacje ekonomiczne i finansowe,* CeDeWu, Warszawa 2009.

Krasowicz S., *Niematematyczne metody prognozowania w rolnictwie, Studia i raporty IUNG - PIB* 19 ZESZYT 14 2009.

Krawczyk R., *Młodzi emigranci polscy pracujący w Wielkiej Brytanii. Adaptacja zawodowa a poczucie szczęścia,* Difin, Warszawa 2017.

Krywult-Albańska M., *Imigranci polscy w Kanadzie. Socjologiczna analiza procesów adaptacji*, Impuls, Kraków 2015.

Kufel T.: *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu Gretl,* Wydawnictwo Naukowe PWN, Warszawa 2007.

Kurkiewicz J. (red.), *Demograficzne uwarunkowania i wybrane społeczno-ekonomiczne konsekwencje starzenia się ludności w krajach europejskich,* Wydawnictwo Uniwersytetu Ekonomicznego w Krakowie, format: ibuk.

Lipiec-Zajchowska (red.): *Wspomaganie procesów decyzyjnych*. Tom III. *Badania operacyjne*. Wydawnictwo C.H. Beck Warszawa 2003.

Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych,* Tom II, *Ekonometria,* Wydawnictwo C.H. Beck, Warszawa 2003.

Majdzińska A., *Regionalizacja demograficzna Wybrane metody i ich aplikacje*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2016.

Mendyk E., *Ekonomika i organizacja transportu*, Wydawnictwo Wyższa Szkoła Logistyki, Poznań 2002.

Migacz J., *Prognozowanie zewnętrznych usług transportowych* (*na podstawie Fabryki Armatur „Głuchołazy"*), praca licencjacka, Wyższa Szkoła Zarządzania i Administracji, Opole 2017.

Nadolska J., Stawarz P., Wojtaszyk K.A., *Unia Europejska i wybrane państwa świata wobec kryzysu migracyjnego*, Aspra, 2017.

Nowak E., *Zaawansowana rachunkowość zarządcza*, Polskie wydawnictwo Ekonomiczne, Warszawa 2017.

Nowak E., *Zarys metod ekonometrii. Zbiór zadań,* Wydawnictwo Naukowe PWN, Warszawa 2002.

O*cena realizacji polityki EU Net-Zero w wybranych państwach członkowskich za pomocą algorytmów automatycznego prognozowania*, Międzynarodowy Dziennik. Informacje geograficzne (isprs), Wydział Biznesu Międzynarodowego i Ekonomii, Uniwersytet Ekonomiczny w Bukareszcie.

Okólski M., Fihel A., *Demografia. Współczesne zjawiska i teorie*, Wydawnictwo Scholar, 2017, *pdf*.

Olkiewicz A., *Jak żyć szczęśliwie w innym kraju*, tłumacz: Inga Sawicka, Wydawnictwo: Czarna Owca, Warszawa 2014.

Osińska M. (red. naukowa), *Ekonometria współczesna,* Towarzystwo Naukowe Organizacji i Kierownictwa. Toruń 2007.

Pędziwiatr K., Siewierska-Chmaj A., Matyja R., *Polska polityka migracyjna W poszukiwaniu nowego modelu,* Wydawnictwo Uniwersytetu Warszawskiego, 2015.
Pianko-Kluczyńska K., Konca-Kędzierska K., Jaczewski A., Mandal A., Droździoł R., Wdowikowski M., Bałdysz Z., Wrońska W., Kózka K., Nykiel G., Strzyżewski T., Marta Gruszczyńska M., *Polska na mapie historii prognozowania dalekiego zasięgu*, Magazyn Popularno-Naukowy „*Obserwator*", IMGW-PIB/Ośrodek Modelowania Meteorologicznego Analizy i Prognozy Dalekiego Zasięgu, lipiec 2022, wersja angielska.
Praca zbiorowa, *Poprawa dokładności w przewidywaniu wskaźników kosztów budowy na poziomie miasta dzięki połączeniu liniowej ARIMA i nieliniowej ANNS, Journal of Management in Engineering*, Uniwersytet Teksasu w Arlington, 11.2021.
Przytuła P., *Migracje międzynarodowe i ekspatriacja. Perspektywa indywidualna, organizacyjna, społeczno-kulturalna*, Difin, Warszawa 2017.
Radzikowska B. (red.), *Metody prognozowania. Zbiór zadań,* Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004.
Robin A., *Angela Merkel i kryzys migracyjny Dzień po dniu* (*Die Getriebenen*), tłumacz: Marcin Masny, Wydawca: Teologia Polityczna, Warszawa 2017.
Rocznik Statystyczny Rzeczypospolitej Polskiej 2007. Dział: *Rachunki narodowe*.
Rocznik Statystyczny Rzeczypospolitej Polskiej 2007. Dział: *Rynek pracy*.
Roczniki statystyczne GUS, lata 2011-2016.
Rossa A., Socha L., Szymański A., *Hybrydowe modelowanie procesów demograficznych z wykorzystaniem rozmytych przełączających układów dynamicznych*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015.
Sa'ad A., Nyoungue A., Hajej Z., *Analiza porównawcza produkcji energii wiatrowej. Planowanie oparte na sztucznej sieci neuronowej i podejście teoretyczne*, Materiały z Międzynarodowej Konferencji Inżynierii Przemysłowej i Zarządzania Operacyjnego, Stambuł, 03.2022.
Stawiński R., *Uchodźczy polscy w latach dziewięćdziesiątych jako nowy problem społeczny*, Difin, Warszawa 2017.
Stokowski F., *Demografia*, Polskie Wydawnictwo Ekonomiczne, Warszawa 2015.
Sugita K., *Prognozowanie z użyciem modeli autoregresyjnych wektorów bayesowych: porównanie bezpośrednich i iterowanych metod wieloetapowych Asian Journal of Economics and Banking*, 2022, tom. 6, nr 2.
Szostek R., *Uogólniony model Holta na przykładzie prognozowania liczby pasażerów w transporcie lotniczym w Polsce, Ekonometria*, 2(36) 2012.
Sztajat A., Moridpour S., Best B., Rumi S., *Przegląd modeli przewidywania degradacji nawierzchni, Journal of Advanced Transportation*, tom 2022, Wiley Hindawi, Uniwersytet RMIT, Melbourne, Australia.
Szukalski P. (red.), *Procesy demograficzne w województwie łódzkim w XXI wieku*, Wydawnictwo Uniwersytetu Łódzkiego, Łódź 2015.
Walesiak M., Gatnar E. (red. nauk.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009.
Wargentin Robin, *Long-term and Short-term Forecasting Techniques for Regional Aiport Planning*, Degree Project in Mathematics second cycle, 30 credits Stockholm, Sweden 2016.
Winiecka K., *Strategia i migracja - adaptacja Rodziny młodych migrantów polskich* w Londynie po przystąpieniu Polski do Unii Europejskiej, Wydawnictwo Naukowe Katedra, 2018.
Wojtaszyk K.A., Szymańska J. (red. nauk.), *Uchodźcy w Europie Uwarunkowania, istota, następstwa*, ASPRA.

Wornalkiewicz W., *Dobór metod i parametrów w budowaniu modeli ekonometrycznych*, Ekonomiczno-społeczne problemy współczesnego zarządzania i komunikacji, (red.) Duczmal. M, Pokusa T, Wyższa Szkoła Zarządzania i Administracji , Opole 2008.
Wornalkiewicz W., *Formułowanie modeli ekonometrycznych na potrzeby zarządzania*, Wydawnictwo Instytut Śląski w Opolu, 2014.
Wornalkiewicz W., *Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*, Skrypt Wyższej Szkoły Zarządzania i Administracji w Opolu, 2013.
Wornalkiewicz W., *Model ekonometryczny zmiennej „Liczba bezrobotnych*", monografia, *Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne*, red. nauk. T. Pokusa, Wydawnictwo Instytut Śląski, Opole 2011.
Wornalkiewicz W., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski, Opole 2013.
Wornalkiewicz W., *Popularyzacja wybranych metod ilościowych w Internecie, Wyższa Szkoła Zarzadzania i Administracji*, Opole 2017.
Wornalkiewicz W., *Przejawy wielowymiarowości współczesnego zarządzania - formy i instrumenty ekonomiczno-społeczne*, red. naukowa T. Pokusa, artykuł *Model ekonometryczny zmiennej „liczba bezrobotnych"*, Wydawnictwo Instytut Śląski, Opole 2011.
Współczesne polskie migracje Strategie - skutki społeczne - reakcja państwa, Wydawnictwo Uniwersytetu Warszawskiego.
Zeliaś A., Pawełek B., Wanat S., *Prognozowanie ekonomiczne Teoria Przykłady Zadania*, Wydawnictwo Naukowe PWN, Warszawa 2021.
Zieliński R., *Tablice statystyczne*, PWN, Warszawa 1972.

Wydanie naukowe

Władysław Wornalkiewicz

MODELOWANIE EKONOMETRYCZNE PODSTAWĄ PROGNOZOWANIA

Monografia

Підписано до друку 29.04.2025. Формат 60x84/8.
Папір офсетний. Гарнітура Таймс. Друк офсетний.
Ум. друк. арк. 41.2. Наклад 300. Зам. № 212

Надруковано в “Видавництво Людмила”.
Свідоцтво про внесення до Державного реєстру
суб’єктів видавничої справи серія ДК № 5303 від 02.03.2017.
“ Видавництво Людмила ”
03148, Київ, а/с 115.
Тел./факс: + 38 050 469 7485, 068 340 8332
E-mail: lesya3000@ukr.net