

**Władysław Wornalkiewicz
Roman Szarawara**

**WSTĘP DO MODELOWANIA
EKONOMETRYCZNEGO**

Przykłady zastosowania



**Zalecane do publikacji przez radę akademicką
Poltawski Instytut Ekonomii i Prawa
(protokół nr 5 z dnia 6 Luty 2025 r.)**

Recenzenci:

*Prof. Akademii Śląskiej w Katowicach dr Tetyana Nestorenko (Poland);
Doktor nauk ekonomicznych, profesor, kierownik katedry finanse, bankowość i ubezpieczenia
Państwo Poltawskie Uniwersytet Rolniczy Oleksija Zorya (Ukraina).*

Wornalkiewicz W., Szarawara R. Wstęp do modelowania ekonometrycznego. Przykłady zastosowania. Poltawa: Poltawski Instytut Ekonomii i Prawa Uniwersytetu „Ukraina”, 2025. 264 s.

DOI: [HTTPS://DOI.ORG/10.36994/978-966-388-702-9-2025-264](https://doi.org/10.36994/978-966-388-702-9-2025-264)

ISBN 978-966-388-702-9

Wraz z rozwojem komputeryzacji i jej implementacji do warunków życia, a w szczególności gospodarczego modna była ekonometria bazująca na statystyce. Rodziły się nowe aplikacje informatyczne, występowała potrzeba znajomości formułowania algorytmów realizacji funkcji w obszarach działalności użytkowników. Stopniowo jednak ten trend został zahamowany pojawieniem się w ofertach pakietów standardowych, w tym zintegrowanych do potrzeb zarządzania. Chętnych zdolnych studentów do poznawania ekonometrii było coraz mniej, również firmy nie zgłaszały potrzeby na tak wykształconych absolwentów studiów wyższych. Fala wdrażania techniki cyfrowej opanowała większość społeczeństwa, nauczyła go posługiwania się transakcjami online w świecie wirtualnym oraz komunikowania się przy pomocy urządzeń mobilnych. Nastąpiła teraz silna tendencja do automatyzacji komputerowej także prac umysłowych, w tym również w administracji państwowej, dydaktyce, sterowaniu gospodarką. Impulsem są dojrzałe już rozwiązania klasy „sztuczna inteligencja”, a także zastępowalność wielu obszarów pracy umysłowej robotami. Zauważalna jest większa ochota studiujących do poznania i budowy programów „inteligentnych” jako elementów modelowania cyfrowego zjawisk życia gospodarczego i społecznego. Z tej potrzeby opracowano niniejszy podręcznik, mający charakter pomocniczy do modelowania ekonometrycznego cech ekonomicznych. Poznanie i opanowanie przez analogię do zaprezentowanych przykładów, procedur analizy szeregów obserwacji, estymacji oraz budowy modeli jedno i wielorównaniowych, może być w przyszłości pomostem do zdobywania intratnych stanowisk w ogólnie określonej przestrzeni jaką jest AI.

УДК 330.43(075.8)
III25

ISBN 978-966-388-702-9

© Poltawski Instytut Ekonomii i Prawa, 2025
© Wornalkiewicz W., Szarawara R., 2025
@Uniwersytet „Ukraina”, 2025

SPIS TREŚCI



WPROWADZENIE	6
1. ROZPOZNANIE LITERATUROWE WYKORZYSTANIA PAKIETÓW PROGRAMOWYCH W STATYSTYCE OPISOWEJ	9
1.1. Miary klasyczne i pozycyjne	9
1.2. <i>Excel</i>	10
1.3. <i>GRET</i>	12
1.4. <i>EViews</i>	15
1.5. <i>STATISTICA</i>	18
1.6. <i>STATGRAPHICS</i>	18
1.7. <i>SPSS</i>	19
1.8. <i>SAS</i>	20
1.9. <i>Stata</i>	20
1.10. <i>R</i>	21
1.11. Dodatkowe nawiązanie do oprogramowania statystycznego	21
2. ŚRODOWISKO PAKIETU <i>R</i>	25
2.1. Instalowanie i uruchomienie <i>R</i>	25
2.2. Charakterystyka <i>R</i>	26
2.3. Elementy statystyki opisowej	26
2.4. Możliwości analizy wielowymiarowej w <i>R</i>	28
3. PRZYKŁADY RAPORTÓW STATYSTYKI OPISOWEJ	31
3.1. Wstęp	31
3.2. Arkusz kalkulacyjny <i>Excel</i>	31
3.3. <i>STATGRAPHICS</i>	41
3.4. Środowisko <i>R</i>	48
4. ZASTOSOWANIE KOMPUTERA W EKONOMETRII	54
4.1. Wstęp	54
4.2. Programy <i>GRET</i> , <i>EViews</i> , <i>WinQSB</i> , <i>Excel</i>	54
4.3. Informacja o innych programach	61
4.4. O oprogramowaniu symulacyjnym	62
5. WYBÓR KOMBINACJI ZMIENNYCH	67
5.1. Wstęp	67
5.2. Model ekonometryczny	67
5.3. Metoda wskaźnika wektora zmiennych objaśniających	67
5.4. Bazowanie na mierze błędów <i>ex ante</i>	69
5.5. Analiza korelacji	69
5.6. Metoda pojemności nośników informacji	70

5.7. Metoda grafowa i inne.....	70
5.8. Testy porównywania modeli	73
5.9. Metoda regresji krokowej oraz minimalizacji resztowej sumy kwadratów	74
5.10. Przykłady zastosowania metody Hellwiga.....	75
5.11. Przykład określenia zmiennych objaśniających	77
5.12. Procedura wyboru zmiennych dla modelu regresji z 4-ma zmiennymi objaśniającymi.....	83
6. DOBÓR MODELI TRENDU DLA ZMIENNYCH OBJAŚNIAJĄCYCH MODELU REGRESJI ZMIENNEJ OBJAŚNIANEJ	86
6.1. Wprowadzenie.....	86
6.2. Dobór stopnia wielomianu dla modeli trendów zmiennych objaśniających Wygenerowanego modelu zmiennej	87
6.3. Generowanie zmiennych o równaniach trendów	96
6.4. Określenie błędów prognozy	98
7. SYMULACJA DOBORU ZMIENNYCH OBJAŚNIAJĄCYCH DLA MODELU EKONOMETRYCZNEGO	101
7.1. Słowo wstępne.....	101
7.2. Modele zmiennych endogenicznych modelu Y (PKB).....	103
7.3. Propozycja wstępna równań wraz z parametrami modelu ekonometrycznego.....	110
7.4. Dalsza redukcja zmiennych.....	111
7.5. Generacja zmiennych endogenicznych	114
7.6. Określenie prognozy zmiennej Y na bazie wygenerowanych zmiennych.....	123
8. DOBÓR METOD I PARAMETRÓW	127
8.1. Wprowadzenie.....	127
8.2. Procedura pracy z modulem FC	129
8.3. Metody prognozowania	130
9. MODEL EKONOMETRYCZNY ZMIENNEJ „Liczba bezrobotnych”	137
9.1. Dane wejściowe.....	137
9.2. Formatowanie modelu ekonometrycznego	140
9.3. Weryfikacja modelu.....	142
10. WSPOMAGANIE KOMPUTEROWE BUDOWY MODELU PROCESU EKONOMICZNEGO	148
10.1. Wielomianowe modele trendu	148
10.2. Modele ze zmiennymi sezonowymi	152
10.3. Modele autoregresyjne	156
11. ESTYMACJA MODELI WIELORÓWNANIOWYCH W GRETL.....	159
11.1. Wstęp	159
11.2. Model wielorównaniowy	159
11.3. Model prosty	160
11.4. Model rekurencyjny.....	160

11.5. Przykład zastosowania <i>GRET</i> L dla modelu rekurencyjnego.....	160
11.6. Model o równaniach łącznie współzależnych	164
11.7. Przykład podwójnej metody najmniejszych kwadratów (2MKN)	166
11.8. Przykład zastosowania 2MKN w <i>GRET</i> L	169
12. PRZYKŁADY WPŁYWU NAKŁADÓW INWESTYCYJNYCH.....	174
12.1. Wybór zmiennych do określenia regresji jednoczynnikowych	174
12.2. Przykłady prezentacji graficznej wpływu nakładów inwestycyjnych na wybrane zmienne	178
12.3. Propozycja modelu ekonometrycznego produktu krajowego brutto.....	185
13. WERYFIKACJA MODELI EKONOMETRYCZNYCH WSPOMAGANA KOM- PUTEROWO	191
13.1. Wstęp.....	191
13.2. Modele jednorównaniowe	191
13.3. Zastosowanie funkcji <i>Excel</i> a.....	194
13.4. Dalsze badania modelu ekonometrycznego	201
13.5. Zastosowanie <i>GRET</i> L.....	209
14. TABLICE STATYSTYCZNE.....	217
15. DOSTĘP DO WYBRANYCH PUBLIKACJI W LATACH 2008-2025.....	234
16. BIBLIOGRAFIA	262

WPROWADZENIE



Obecne czasy znamienuje projektowanie i wdrażanie oprogramowania zwanego sztuczną inteligencją. Po okresie usprawnienia komunikacji, dzięki zbudowaniu infrastruktury telefonii komórkowej, postępu nauki w zakresie poznania funkcjonalności mózgu ludzkiego, nastała pora na robotyzację także czynności umysłowych. Coraz bardziej obserwujemy to w zakresie obsługi wirtualnej w czasie podróży, w operacjach prowadzenia biznesu, a także w sferze bezpieczeństwa, gdzie w czasie niepokoju tradycyjne metody walki są zastępowalne atakami uzbrojonych dronów na obiekty infrastruktury przeciwnika.

W procesie kształcenia coraz bardziej zauważalne jest głębokie zainteresowanie nowymi technikami, dążącymi do przyzwyczajania się społeczeństwa do współpracy z systemami informatycznymi, gdzie elementy sztucznej inteligencji są możliwie szybko implementowane do praktyki pracy na wielu stanowiskach.

Niniejsze opracowanie pomyślane jest głównie jako podręcznik pomocniczy dla studentów studiów niestacjonarnych, którzy są świadomi tego, że nadchodząca dekada to wezwanie do głębszego poznania statystyki opisowej i matematycznej, nabrania umiejętności modelowania jedno i wielorównaniowego zjawisk, przede wszystkim gospodarczych i ekonomicznych w przedsiębiorstwach produkcyjnych i usługowych.

Na wielu kierunkach studiów, w tym na *Zarządzaniu, Logistyce, Administracji* przez jakiś okres czasu występowała tendencja odchodzenia od przybliżenia sobie arkanów ekonometrii teoretycznej. Jest teraz jednak skłonność do poznania ekonometrii stosowanej, zwłaszcza na uczelniach typu akademie nauk stosowanych. Jednak tradycyjne podręczniki w tym zakresie są dość trudne i w małym stopniu stosuje się w nich prezentacje z wykorzystaniem techniki komputerowej. Z tego względu autorzy tej pracy mają nadzieję, że to opracowanie będzie ułatwieniem we wstępnym poznaniu modelowania ekonometrycznego. Stąd już na początku podjęto się przypomnienia elementów statystyki opisowej, sposobu korzystania z dostępnych programów komputerowych do formułowania modeli ekonometrycznych. Ponadto do wskazywania na ich podstawie prognoz, a także przeprowadzania weryfikacji poprawności skonstruowanego rozwiązania modelowego. Opanowanie tego, zwłaszcza przez studentów studiów zaocznych, umożliwi im lepsze przygotowanie się do okresu pracy z robotami komputerowymi. Również do współpracy z wysoce dopracowanymi „inteligentnymi” pakietami programowymi, które będą podstawą w różnych obszarach działalności ludzkiej.

Niniejszy podręcznik obejmuje 13 rozdziałów merytorycznych, zestaw niezbędnych tablic statystycznych oraz wskazanie dodatkowej bibliografii jednego z współautorów, opublikowanych wcześniej w formie książek, monografii indywidualnych, a także artykułów w publikacjach zbiorowych.

Aby nie szperać po różnych opisach pomocniczego oprogramowania w rozdziale pierwszym dokonano rozpoznania literaturowego w tym zakresie. Krótko przedstawiono funkcjonalności popularnych programów, stosowanych nie tylko do statystyki opisowej, z wykorzystaniem takich programów jak *GRETl*, *R*, *STATISTICA* i innych. Szczególną uwagę należy zwrócić w tym opracowaniu na wygodny w zastosowaniu pakiet *GRETl* oraz coraz popularniejszy zestaw programów *R*.

Rozdział drugi stanowi przybliżenie dogodności stosowania w statystyce i ekonometrii pakietu *R* wraz z jego licznymi modułami, w tym do analizy wielowymiarowej obserwowanego zjawiska ekonomicznego i nie tylko.

W kolejnym rozdziale trzecim pokazano przykłady tworzenia raportów z zakresu statystyki opisowej, w tym z wykorzystaniem cenionego przez użytkowników arkusza kalkulacyjnego *Excel* i programu *STATGRAPHICS*.

Rozdział czwarty poświęcono opisowi zastosowania standardowych programów do usprawnienia modelowania ekonometrycznego. Budując model określonej cechy objaśnianej np. gospodarczej należy wcześniej zgromadzić odpowiednie szeregi czasowe obserwacji cech objaśniających, co jest przedmiotem rozważań rozdziału piątego. Trudno jednak od razu wskazać liczbę zmiennych oraz ich rodzaje, posługując się więc własną intuicją. Kolejnym krokiem jest selekcja i wyeliminowanie zmiennych nieistotnie skorelowanych ze zmienną objaśnianą. Interesują nas bowiem zmienne objaśniające silnie skorelowane ze zmienną objaśnianą, a słabo wzajemnie zależne. Aby przybliżyć to zagadnienie pokazano przykłady zastosowania różnych metod. Mając kilka użytych metod do sformułowania modelu, trzeba wskazać rozwiązanie najbardziej adekwatne. Służą do tego odpowiednie testy porównywania modeli, biorące pod uwagę wielkość popełnionego błędu w porównaniu z danymi empirycznymi.

Warto w rozdziale piątym zwrócić uwagę na przykłady, a mianowicie zastosowanie metody Hellwiga, a także na wybór zmiennych do modelu regresji z czterema zmiennymi objaśniającymi.

Kontynuując studiowanie tego pomocniczego podręcznika przechodzimy do rozdziału szóstego, zajmującego się doбором rodzaju modeli trendów dla zmiennych objaśniających, wchodzących w skład modelu regresji zmiennej objaśnianej. Początkowym krokiem jest dobór i graficzne przedstawienie modeli trendów poszczególnych zmiennych objaśniających. Zmienne te, które same są wyrażeniami liniowymi lub nieliniowymi zależnymi od czasu, wchodzą w skład modelu wielorównaniowego. Taki model może służyć do prognozowania, w szczególności krótkookresowego, na podstawie wygenerowanych wartości trendów zmiennych objaśniających.

Zmienne objaśniające mogą też być określone innymi zmiennymi endogenicznymi oraz egzogenicznymi. Postępowanie w tym zakresie w celu doprowadzenia do określenia prognozy zmiennej objaśnianej Y na podstawie zmiennych objaśniających X_i stanowi temat dociekań w rozdziale 7. Wymaga to odpowiedniego wygenerowania wartości zmiennych endogenicznych.

Jesteśmy już w rozdziale ósmym, i tu posłużono się modulem FC (Prognozowanie) wchodzącym w skład pakietu *WinQSB*. Moduł ten pozwala na dobór i przeanalizowanie odpowiednich metod, a w tym z uwzględnieniem optymalnego doboru wartości parametrów budowanego modelu ekonometrycznego.

W rozdziale dziewiątym zaprezentowano przykład sformułowania modelu ekonometrycznego zmiennej „Liczba bezrobotnych”. Skorzystano z modułu FC wspomnianego już pakietu *WinQSB*. Podana procedura może być wykorzystana dydaktycznie, a nawet praktycznie do budowy modelu ekonometrycznego, także innych cech z życia gospodarczego. Warto zwrócić uwagę na podane postępowanie w zakresie weryfikacji utworzonego modelu.

Następny rozdział dziesiąty jest kontynuacją tworzenia modelu, a odnosi się do modelowania ekonometrycznego. Rozważania w tym rozdziale dotyczą wielomianowych modeli trendów, modeli ze zmiennymi sezonowymi, a także modeli autoregresyjnych.

Wspomniany już pakiet *GRET* posłużył do estymacji parametrów modeli wielomianowych w rozdziale 11. Rozpatrzono tu także model rekurencyjny i model o równaniach łącznie współzależnych. Na szczególną uwagę zasługuje tu zastosowanie podwójnej metody najmniejszych kwadratów, także ze wspomaganiem funkcjonalnością pakietu *GRET*.

Rozdział dwunasty jest przykładem wpływu nakładów inwestycyjnych, w szczególności na wielkość produktu krajowego brutto (PKB). Warto tu zwrócić uwagę na zaprezentowane graficzne przykłady wpływu nakładów inwestycyjnych od innych wybranych przykładowo zmiennych

objaśniających. Efektem końcowym rozważań w tym rozdziale jest propozycja modelu ekonometrycznego PKB.

Jesteśmy już przy rozdziale trzynastym, w którym uwaga skoncentrowana jest na procedurze weryfikacji utworzonego modelu ekonometrycznego ze wspomaganie komputerowym. Zastosowano tu funkcje *Excela*, a także wykorzystano możliwości pakietu *GRET*L w zakresie przeprowadzenia różnych testów badania poprawności zaproponowanego modelu, zwłaszcza z przeznaczeniem do prognozowania.

W procesie budowy modelu ekonometrycznego, a przede wszystkim w etapach weryfikacji słuszności przyjętych parametrów, potrzebne są tablice statystyczne. Przykłady niezbędnych tablic rozkładów stanowią zakończenie rozważań niniejszej pracy.

Autorzy



1. ROZPOZNANIE LITERATUROWE WYKORZYSTANIA PAKIETÓW PROGRAMOWYCH W STATYSTYCE OPISOWEJ



1.1. Miary klasyczne i pozycyjne

Zagadnienie zastosowań komputera do statystyki opisowej spotykamy w różnych publikacjach. Warto jednak na wstępie określić czym zajmuje się ta dyscyplina. Zdefiniowanie tego pojęcia spotykamy w pracy „Statystyka Elementy teorii i zadania”¹:

„Statystyka opisowa dostarcza metod i procedur gromadzenia, opracowania i prezentacji danych statystycznych. W statystyce opisowej nie ma odwołań do mechanizmów stochastycznych, które dały początek zaobserwowanym danym. Celem statystyki opisowej jest podanie zwięzłego opisu określonego materiału statystycznego. ...Uogólnienie wyników zaobserwowanych na podstawie badań cząstkowych jest możliwe tylko w przypadku, gdy te są odpowiednio przeprowadzone”. Analiza statystyczna powinna być przeprowadzona przy dużej liczbie parametrów, spośród których wyróżniamy miary klasyczne i pozycyjne. Miary klasyczne obliczane są według wszystkich obserwacji, przy czym wyróżniamy następujące:

- średnia (arytmetyczna),
- wariancja,
- odchylenie standardowe,
- współczynnik zmienności,
- współczynnik skośności,
- współczynnik asymetrii,
- współczynnik koncentracji (skupienia),
- współczynnik Lorenza.

Czasem w analizach oblicza się również średnią, po odrzuceniu 5% obserwacji skrajnych. Natomiast miary pozycyjne to: modalna, kwartył pierwszy, kwartył trzeci, odchylenie ćwiartkowe, współczynnik zmienności, rozstęp, współczynnik skośności.

Udogodnieniem jest raportowanie wymienionych miar korzystając z pakietów komputerowych, takich jak arkusz kalkulacyjny *Excel*, *GRET*L, *Eviews*, *STATISTICA*, *STATGRAPHICS*, *R*. Ostatni z wymienionych stanowi platformę wielu pakietów o specjalistycznych funkcjach. Jednak Maddala w swoim powszechnie stosowanym podręczniku ekonometrii stwierdza, że nie ma pakietu, który obejmował by rozwiązywanie wszystkich tematów objętych jego pracą². Jego zdaniem wiele jest programów o różnorodnych symbolach, które w swej funkcjonalności nieznacznie różnią się między sobą oraz dotyczą tych samych metod i testów podstawowych. Wymienia niektóre z programów komputerowych z lat 1990 przeznaczonych do ekonometrii, które zawierają również funkcje dotyczące analizy statystycznej³, a należą do nich *SHAZAM*, *MICROFIT*, *RATS*, *SAS*, *MICRO-TSP*.

Istnieją pakiety programowe zarówno ogólnie bezpłatnie dostępne jak i wymagające indywidualnych licencji. Jest to bardzo obszerna tematyka znajdująca swoje odbicia wycinkowe w Internecie w zakresie przykładów zastosowań. Wystarczy sięgnąć do strony www.feweb.vu.nl/econometriclinks/software/software.html,

¹ Praca zbiorowa, *Statystyka Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006.

² Maddala G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.

³ Ibidem, Dodatek D: *Programy komputerowe*.

aby się przekonać o obszerności dorobku w zakresie oprogramowania rozwijanego w miarę postępu w zakresie systemów operacyjnych, programów narzędziowych i sieciowych. Strona ta zawiera linki (połączenia) z dystrybutorami określonego oprogramowania. Lista programów komputerowych jest na bieżąco aktualizowana.

Omówienie obszarów zastosowania całej palety programów komputerowych przeznaczonych między innymi do statystyki opisowej jest zadaniem przekraczającym ramy niniejszej pracy. Z tego względu oparto się na przykładowych występujących w literaturze przedmiotu.

1.2. Excel

Pakiet ten jest powszechnym i uniwersalnym narzędziem wspomagającym analizy statystyczne, estymację oraz weryfikację modeli ekonometrycznych. Posługuje się arkuszem kalkulacyjnym tworzącym skoroszyt. Znając wzory na poszczególne statystyki, testy, czy równania modeli możemy przy pomocy wyrażeń napisać potrzebne formuły obliczeniowe. Ponadto *Excel* ma wbudowane funkcje statystyczne i matematyczne⁴. Do grupy statystycznych należą następujące funkcje elementarne:

- WSP.KORELACJI,
- ŚREDNIA,
- NORMALIZUJ,
- WARIANCJA
- WARIANCJA.POPUL
- KOWARIANCJA,
- NACHYLENIE
- ODCIĘTA,
- ODCH.STANDARDOWE
- ODCH.STANDARDOWE.POPUL.

a także funkcje:

- R.KWADRAT (współczynnik dopasowania),
- REGLINP (regresja liniowa),
- REGEXP (regresja wykładnicza),
- REGBLSTD (błąd standardowy reszt).

Istnieje również możliwość korzystania z opcji *Excel/Analiza danych/Histogram* w celu sporządzenia wykresów oraz tabel częstości występowania wartości danych w określonych przedziałach. Funkcje matematyczne, a w szczególności z zakresu rachunku macierzowego, umożliwiają estymację parametrów równań liniowych, potęgowych, wykładniczych, wielowymiarowych. Należą do nich: *TRANSPONUJ* (transpozycja macierzy), *MACIERZ.ILOCZYN*, *MACIERZ.ODW* (macierz odwrotna).

Interesującym segmentem tego programu niezbędnym w statystyce opisowej oraz ekonometrii jest *Dodatki*, który instalowany jest poprzez wywołanie z menu opcji *Analysis ToolPak*. Między innymi dzięki temu segmentowi *Excel* znalazł szersze zastosowanie w dydaktyce przedmiotu statystyka opisowa, a fragmenty jego użycia, jak już wspomniano spotykamy w różnych pozycjach literaturowych do:

- określenia parametrów regresji liniowej modelu,
- raportowania podstawowej statystyki opisowej,

⁴ Microsoft: *Excel. Opis funkcji*, Microsoft Corporation, Ireland 1993.

- analizy reszt,
- przedstawienia histogramów poszczególnych zmiennych.

W zakresie funkcji elementarnych *Excelsa* ustawiamy się na aktywnej komórce i korzystamy z menu *Wstaw/Funkcje statystyczne*, a następnie wybieramy nazwę funkcji, przykładowo *Kwartyl*.

Możemy też wybrać ścieżkę dostępu „globalnego” *Narzędzia/Dodatki/Analiza danych/Statystyka opisowa*. Jeśli mamy wprowadzone dane do określonej kolumny to w oknie dialogowym wpisujemy lub wybieramy między innymi:

- zakres wejściowy komórek zawierających dane,
- grupowanie według (wierszy/kolumn),
- poziom ufności dla średniej,
- zakres wyjściowy (adres lewej górnej komórki dla tabeli wyników),
- statystyki podsumowujące dla: średniej, błędu standardowego, mediany, dominanty, odchylenia standardowego, wariancji, kurtozy, skośności, zakresu, minimum, maksimum.

Ponadto funkcje *Excelsa* można zastosować do obliczenia:

- współczynnika zmienności,
- współczynnika korelacji,
- współczynnika determinacji,
- oszacowania parametrów wielowymiarowego liniowego modelu regresji,
- statystyki regresji i analizy wariancji.

Wygodną funkcją *Excelsa* jest *SUMA.XMY.2*, która oblicza sumę kwadratów różnic pomiędzy wartościami w dwóch zakresach komórek.

Podjęcie tematyki analizy ekonometryczno-statystycznej w arkuszu kalkulacyjnym *Excel wersja 5.0* występuje w pracy pod redakcją Jajugi⁵. Uwaga skupiona jest na uaktywnieniu procedur programu *Excel* zawartych w znanym nam już zbiorze procedur *Dodatki Z* płaszczyzny statystyki opisowej interesujące są pokazane raporty możliwości realizacji poprzez menu *Narzędzia/Analiza danych* zestawień korelacji, kowariancji, statystyki opisowej, regresji i wykonanie histogramu. Cennym elementem *Excelsa* są jego funkcje matematyczne, zwłaszcza z zakresu rachunku macierzowego.

Publikacją, która zajmuje się wspomaganiem komputerowym procesu modelowania jest praca zbiorowa „Modelowanie ekonometryczne z Excelem” stanowiąca materiały pomocnicze do laboratorium z ekonometrii⁶. Spotykamy tu rozdział „Wstępna analiza danych” w którym przedstawiono funkcje elementarne *Excelsa* umożliwiające obliczenie:

- współczynnika zmienności stanowiącego stosunek odchylenia standardowego j - tej zmiennej do średniej j -tej zmiennej,
- współczynnika korelacji liniowej oraz otrzymanie macierzy korelacji zmiennych,
- współczynnika korelacji wielorakiej, gdy opisujemy przypadek w którym na określoną zmienną objaśnianą oddziałuje wiele zmiennych objaśniających.

Z fragmentarycznym zastosowaniem *Excelsa* do statystyki opisowej w kontekście budowy modelu ekonometrycznego spotykamy się w pracy pod redakcją Radzikowskiej⁷. Jednak z obszaru statystyki opisowej zastosowanie funkcji *Excelsa* ograniczono do:

- sprawdzenia wystarczającego zróżnicowania zmiennych,

⁵ Jajuga K. (red.), *Ekonometria Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003, Część 3.

⁶ Praca zbiorowa, *Modelowanie ekonometryczne z Excelem*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004, rozdz. 2.2. „Wstępna analiza danych”.

⁷ Radzikowska B. (red.), *Metody prognozowania. Zbiór zadań*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004, rozdz. 5.

- zbadania siły skorelowania zmiennych objaśniających ze zmienną objaśnianą oraz skorelowania wzajemnego,

- estymacji parametrów regresji liniowej poprzez zastosowanie już nam znanego menu: *Narzędzia/Analiza danych*,

- oceny dopasowania modelu do danych empirycznych.

Wychodząc naprzeciw oczekiwaniom studentów studiów zaocznych niezbędne elementy statystyki opisowej obliczone z zastosowaniem komputera pokazano w pracy „Wstęp do ekonometrii i badań operacyjnych”⁸. Z punktu widzenia statystyki opisowej należy zwrócić uwagę na następujące rozdziały i zawartą w nich tematykę:

Miary rozkładu jednej zmiennej (przykłady w Excelu):

- przedziały obserwacji,

- miary przeciętne,

- miary rozproszenia,

- miary asymetrii,

- rozkład wielu zmiennych, ze szczególnym określeniem normalizacji zmiennych.

W tomie „Statystyka”⁹ występuje *Aneks* w którym przedstawiono możliwości wykorzystania w analizach statystycznych pakietów *Excel* oraz *Eviews*.

1.3. GRETL

Oprogramowanie *GRETL* (*GNU Regression Econometrics Time-Series Library*) opracowane jest w wersji polskiej. Dostępne jest na stronie internetowej <http://www.gretl.sourceforge.net> lub kufel.torun.pl. Obejmuje podstawowe elementy z zakresu statystyki opisowej, zagadnienia ekonometrii oraz prognozowania. Oprogramowanie wydane jest przez Fundację Wolnego Oprogramowania (*Free Software Foundation*) i jest ogólnie bezpłatnie dostępne. Może być przez użytkowników modyfikowany jego kod źródłowy, jednak ryzyko w zakresie skuteczności działania funkcji ponosi sam użytkownik. Niniejszy opis dotyczy wersji *GRETL*'a o numerze 1.6.5¹⁰. 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 oraz 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, panelowych).

Oprogramowanie *GRETL* ma dostęp do serwera ricardo.ecn.wfu.edu poprzez menu:

Plik/Pliki baz danych/Instalowanie baz danych.

⁸ Duczmal M., Wornalkiewicz Wł., *Wstęp do ekonometrii i badań operacyjnych. Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski, Opole 2010.

⁹ Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, Tom I. *Statystyka*, Wydawnictwo C.H. Beck Warszawa 2003.

¹⁰ Procedurę instalacji pliku *gretl.install.exe* oraz plików z danymi pochodzącymi z podręczników ekonometrii podano w rozdziale 1.2. publikacji wymienionej w przypisie 3.

Aby pracować z *GRET*L'em pozyskany plik należy zapisać na dysku własnego komputera. W ramach tego programu występują przykłady modeli ekonometrycznych, a dla części z nich zestaw poleceń - komend. Wbudowano również tablice statystyczne rozkładów: normalnego, *t*-Studenta, chi-kwadrat, *F*-Fishera, Durbin-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 *GRET*L'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 funkcję wyznaczania macierzy korelacji. Oprócz *KMNK* w programie wykorzystuje się także inne metody estymacji będące modyfikacją metody najmniejszych kwadratów. Bardzo pomocne są informacje z procesu estymacji oraz specjalne funkcje *GRET*L'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 omawianego programu 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 Kufli¹¹ na uwagę zasługuje nawiązanie krótkiego wyjaśnienia teoretycznego podstawowych modeli procesów ekonomicznych z przedstawieniem na konkretnym pliku danych odpowiedniej funkcji *GRET*L'a. Daje to możliwość zrozumienia raportów programu z realizacji budowy modeli: wielomianowych trendu, wahań sezonowych oraz autoregresyjnych.

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 heteroskedastyczności wariancji (niestałości) to wówczas stosowana jest uogólniona metoda najmniejszych kwadratów (*UMNK*). W warunkach autokorelacji reszt program *GRET*L proponuje w ramach modeli szeregów czasowych metody: Cochran'a-Orcutta, Hirdertha-Lu, Prais-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 *TestTest heteroskedastyczności*. 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

¹¹ Procedurę instalacji pliku *gretl.install.exe* oraz plików z danymi pochodzącymi z podręczników ekonometrii podano w rozdziale 1.2. publikacji wymienionej w przypisie 5.

metoda najmniejszych kwadratów (2MНК) 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*.

Gdy szeregi czasowe są krótkie to w *GRET*L możemy zastosować dane panelowe, czyli przestrzenno-czasowe. Z takimi danymi mamy do czynienia, gdy obserwujemy kilka obiektów jednocześnie w tych samych okresach czasowych¹².

Poświęcona *GRET*L'owi publikacja Kufla obejmuje szereg tematów dotyczących możliwości jego menu na bazie plików danych oferowanych w Internecie przez tego autora¹³. Rozdział drugi „Dane statystyczne” podaje między innymi:

- podstawowe statystyki opisowe,
- rozkłady zmiennych,
- wykresy.

Natomiast w rozdziale trzecim podano tablice statystyczne zawarte w programie *GRET*L oraz umieszczony w nim kalkulator testów statystycznych. Do *GRET*L'a dane wprowadzamy przy użyciu funkcji *Plik/Nowy zbiór danych*. Możemy również dokonać importu danych – poszczególnych zmiennych przykładowo z arkusza kalkulacyjnego *Excel* poprzez menu: *Plik/Otwórz dane/Import pliku/Excel*.

*GRET*L umożliwia agregowanie szeregów czasowych przykładowo danych miesięcznych do kwartalnych korzystając z menu: *Dane/Agregowanie danych...* Podstawowe statystyki opisowe uzyskujemy wywołując menu: *Widok/Statystyki opisowe*. Efektem jest raport zbiorczy charakterystyk dla zmiennych modelu ekonometrycznego obejmujący średnią, medianę, minimum/maksimum, odchylenie standardowe, współczynnik zmienności oraz skośność oraz kurtozę.

Zaprezentowany w pracy Kufla program pozwala na uzyskanie rozkładu częstości dla danej zmiennej. Wykresy zmiennej otrzymamy wywołując menu: *Widok/Wykresy zmiennych*, w tym również wykres trójwymiarowy (3D). Jak już wspomniano udogodnieniem są tablice statystyczne rozkładów: normalnego, *t*-Studenta, *chi*-kwadrat, *F*-Fishera oraz Durбина-Watsona, do których dostęp uzyskujemy przez menu: *Narzędzia/Tablice statystyczne*. Wspomniany kalkulator testów statystycznych umożliwia wyznaczenie wartości statystyki dla testów istotności. Podając wyniki z próby oraz zakładaną wartość określonego parametru w populacji możemy wyznaczyć statystykę testową i empiryczną, co umożliwia podjęcie decyzji weryfikacyjnej.

Zastosowanie programu *GRET*L do wyznaczenia charakterystyk miar zmiennych losowych statystyki opisowej spotykamy w publikacji pod redakcją naukową Marii Osińskiej¹⁴. W tej pracy *GRET*L umożliwił między innymi:

- wyznaczenie macierzy współczynników korelacji liniowej Pearsona między poszczególnymi zmiennymi modelu ekonometrycznego,
- estymację parametrów modelu,

¹² Paweł Kufla jako przykład podaje „...badanie ekonomicznej efektywności ludności ...”, stopy bezrobocia w powiatach w Polsce w latach 1998-2002” oraz rozwiązuje ten przykład w *GRET*L (rozdział 2. publikacji podanej w przypisie 5).

¹³ Kufla T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu Gretl*. Wydawnictwo Naukowe PWN. Warszawa 2007.

¹⁴ Osińska M. (red.), *Ekonometria współczesna*, Towarzystwo Naukowe Organizacji i Kierownictwa, Toruń 2007, rozdz. 2. „Specyfikacja modelu ekonometrycznego”.

- określenie cech charakterystycznych poszczególnych zmiennych.
Natomiast to wykresów punktowych obserwacji zmiennych zastosowano funkcje *Excels*.

1.4. EViews

Statystyczny pakiet komputerowy *Eviews* posłużył Lipiec-Zajchowskiej do przeprowadzenia analiz, których wyniki są w formacie tego pakietu¹⁵. W aneksie publikacji zawarto opisy komputerowych programów *Excel* oraz *Eviews*. Wyniki analiz dyskryminacyjnych otrzymano stosując pakiet *Statistica*. Pakiet *Eviews* zastosowano do budowy liniowych modeli regresji z jedną zmienną objaśniającą (niezależną). W ramach tego dokonano oszacowania współczynników korelacji, estymacji parametrów modelu oraz otrzymano podstawowe statystyki dotyczące dopasowania modelu do danych empirycznych, a mianowicie uzyskano między innymi:

- współczynniki przy zmiennych objaśniających,
- błędy standardowe oszacowania parametrów,
- wartości testu *t*-Studenta,
- współczynnik determinacji,
- skorygowany współczynnik determinacji,
- odchylenie standardowe składnika resztowego,
- wartość testu Durбина-Watsona,
- średnia zmiennej zależnej,
- odchylenie standardowe zmiennej zależnej,
- wartość testu Fishera.

Menu pakietu *Eviews* w wersji 3.11 obejmuje opcję *File* (zbiór), a w ramach niej polecenia do zainicjowania nowego zbioru roboczego (*New*), przy czym możemy podać częstotliwość - okresy czasowe danych, otwarcie istniejącego (*Open*) oraz zapisanie zbioru korzystając z polecenia *Save As*. Wartości danych wprowadzamy do poszczególnych kolumn zmiennych w arkuszu tego programu, a zamiast przecinka używamy kropki dziesiętnej, dla odznaczenia części ułamkowej.

Ponadto menu główne zawiera opcje: *Edit*, *Objects*, *View*, *Procs*, *Quick*, *Options*, *Windows*, *Help*. Program *EViews* wersja 3.11 umożliwia również uzyskanie wykresów danych po wyborze opcji *Quick*, a następnie polecenia *Graph* i zaznaczeniu serii wartości danych danej zmiennej. Analogicznie jak w arkuszu kalkulacyjnym *Excel* możliwe jest uzyskanie po wyborze odpowiedniego polecenia (komendy) następujących wykresów:

- wykres liniowy (*Line Graph*);
- wykres liniowy warstwowy dla danych przekrojowych, przykładowo zmiennych X_1 , X_2 , X_3 (*Stacked Graph*);
- wykres słupkowy (*Bar Graph*);
- liniowo-słupkowy (*Mixed Bar & Line*);
- diagram korelacyjny (*Scatter Diagram*);
- wykres kołowy (*Pie Chart*).

Wykorzystując opcję *Procs* możemy poleceniem *Generate Series* wygenerować nową zmienną $Y1$ według notacji komputerowej bazującą na zmiennej już istniejącej w bazie pakietu *EViews*. W tym celu w oknie dialogowym wpisujemy przykładowo: $Y1 = 2 * Y$. Polecenie wygenerowania nowej

¹⁵ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom II. *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.

zmiennej możemy też wpisać bezpośrednio w pasku komend znajdującym się pod menu głównym, np.: $Y1=2*Y^{16}$.

Możemy użyć oprócz typowych operatorów działań arytmetycznych; +, -, * (mnożenie), / (dzielenie), ^ (potęgowanie) także:

d (y, n) - n-ta różnica wartości zmiennej Y,

log (y) - logarytm naturalny wartości zmiennej.

Oprócz wymienionych operatorów *EViews* zawiera funkcje pozwalające napisanie równań i następnie na realizację określonych działań¹⁷.

Współczynnik korelacji liniowej *Pearsona* uzyskujemy po wywołaniu ścieżki menu: *Quick/Group Statistics/Correlations*, a następnie wskazaniu zmiennych, dla których chcemy określić wzajemne skorelowanie.

Z punktu widzenia celu tej pracy istotna jest możliwość uzyskania w programie *EViews* przedziałów wartości dla szeregu rozdzielczego, histogramu oraz statystyki opisowej danej zmiennej i dlatego w tym celu korzystamy ze ścieżki menu:

Quick/Series Statistics/Histogram and Stats.

Istnieje możliwość przeniesienia fragmentu danych między programem *EViews* a programami *Excel*, *Word*. W tym celu począwszy od menu głównego wybieramy ścieżkę *Edit/Copy* a następnie polecenie *Wklej*. Natomiast do importu/eksportu arkusza danych służy opcja *Procs* a ramach niej polecenia *Import danych* lub *Eksport danych*.

EViews użyto również do oszacowania parametrów liniowego dwuwymiarowego modelu ekonometrycznego uzyskując także statystykę dopasowania modelu. Program zastosowano też do wielowymiarowych modeli regresji, gdy liczba zmiennych objaśniających jest większa od dwóch. W rozdziale 1.14.2 wymienionej w przypisie 15. publikacji spotykamy przykład oszacowania programem *EViews* nieliniowego wielowymiarowego modelu regresji z opóźnioną zmienną endogeniczną Y_{t-1} . Oprócz zastosowania pakietu do modeli przyczynowo-skutkowych wykorzystano *EViews* także do modeli rozwoju zjawiska w czasie. Interesujący jest przykład doboru funkcji trendu spośród liniowej, potęgowej oraz wykładniczej.

Gdy występuje sezonowość to dla kwartałów możemy wprowadzić zmienne zero-jedynkowe i wówczas estymujemy parametry modelu obejmującego czas oraz zmienne. Zastosowanie to dotyczy zarówno modeli multiplikatywnych jak i addytywnych. Zjawiska w czasie opisywane są dla celów prognozowania za pomocą modeli:

- prostych średnich ruchomych, przykładowo dwu lub cztero-okresowych;
- ważonych średnich ruchomych;
- prostego wyrównywania wykładniczego z parametrem wyrównującym α ;
- adaptacyjnego wyrównywania wykładniczego z sygnałem adaptacyjnym $TST\alpha$;
- podwójnych średnich ruchomych;
- podwójnego wyrównywania wykładniczego - model Browna;
- Holta-Wintersa (bez sezonowości, z sezonowością multiplikatywną lub addytywną).

Z publikacji pod redakcją naukową Lipiec-Zajchowskiej możemy się dowiedzieć jak w programie *Eviews* wprowadzać dane do modelu regresji, wykonać odpowiednie wykresy, przeprowadzić operacje na szeregach danych, dokonać estymacji parametrów, sporządzić wykres rozrzutu reszt oraz otrzymać miary statystyki opisowej.

¹⁶ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom I. *Statystyka* op. cit, Aneks s. 103.

¹⁷ Ibidem, s.: 118.

EViews jest też pomocny w modelach trendu. Możemy prognozować bazując na modelu autoregresyjnym, czy też wyrównywania wykładniczego. Opcje programu: *Histogram, statystyki opisowe, szereg rozdzielczy* dają możliwość sporządzenia podstawowej statystyki opisowej (ilościowej oraz graficznej). Warto jeszcze zaznaczyć, że pracując w systemie operacyjnym *Windows* mamy możliwość dokonania importu/eksportu plików między *Eviews* a arkuszem kalkulacyjnym *Excel* lub edytorem tekstów *Word*. Zastosowanie pakietu *Eviews* spotykamy również w tomie I. *Statystyka*¹⁸. Korzystanie z menu dla pozyskania statystyki opisowej oraz niżej wymienionych typów wykresów podano w rozdziale I. aneksu do w/w tomu *Statystyka*:

- liniowy, liniowy warstwowy;
- słupkowy, słupkowy warstwowy, liniowo-słupkowy;
- korelacyjny;
- kołowy.

W ramach ilościowej statystyki opisowej otrzymujemy określenie: średniej wartości zmiennej, mediany, wartości maksymalnej oraz minimalnej, odchylenia standardowego, współczynników skośności oraz kurtozy.

Ponadto zaprezentowano wykorzystanie w analizie statystycznej arkusza kalkulacyjnego *Excel* na przykładzie szeregu 10-cio elementowego. Istnieje możliwość skorzystania z określonej funkcji statystycznej poprzez menu: *Wstaw/Funkcje statystyczne*. Dla celów statystyki opisowej możemy poprzez menu *Excels: Narzędzia/Dodatki* wygenerować opcję *Analiza danych*, a w ramach niej funkcję *Statystyka opisowa*. Po akceptacji tej opcji raport wygenerowany przez komputer obejmuje między innymi charakterystyki zmiennej:

- średnia;
- błąd standardowy (oszacowania średniej) na określonym przez użytkownika poziomie ufności, przykładowo 95%;
- mediana;
- tryb (dominanta zwana również modalną);
- wariancja i odchylenie standardowe próbki;
- współczynniki skośności oraz kurtozy

oraz zakres, minimum oraz maksimum.

Biorąc pod uwagę wspomaganie komputerowe analizy statystycznej, z zacytowanej w tomie „*Statystyka*”¹⁹, na uwagę zasługują pozycje:

- Dobosz M., *Wspomagana komputerowo statystyczna analiza badań*, Warszawa 2001.
- Kościacz P. (red.), *Nowoczesne pakiety komputerowe w podejmowaniu decyzji*, UW,

Warszawa 1998.

- Lipiec-Zajchowska M, Sikora E., Zajchowski J., *Modele podejmowania decyzji w pakietach komputerowych*, UW, Warszawa 1995.

Dalsze rozpoznanie literaturowe przeprowadzono na podstawie informacji z Internetu. Dostęp uzyskano poprzez hasło: książka/pakiet programowy/nazwa pakietu.

¹⁸ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom I. *Statystyka*. Wydawnictwo C.H. Beck, Warszawa 2003.

¹⁹ Ibidem, *Bibliografia*, s.: 127.

1.5. STATISTICA

Wydanie trzecie podręcznika „Statystyka z pakietem komputerowym STATISTICA PL Teoria i zadania” wykorzystuje wersję STATISTICA 8²⁰. Z zakresu statystyki opisowej książka zawiera:

- tablice statystyczne najważniejszych rozkładów,
- skorowidz podstawowych pojęć i algorytmów statystycznych,
- alfabet grecki w nazewnictwie statystycznym,
- przedstawienie metod analizy statystycznej łącząc teorię podstaw statystyki z przykładami zastosowań ekonomicznych.

W książce znajdujemy też zagadnienie weryfikacji hipotez statystycznych dotyczące standardowych błędów losowych drugiego rodzaju oraz wskaźników mocy. Z obszaru statystyki opisowej na uwagę zasługują podrozdziały:

- podstawowe pojęcia statystyczne,
- szeregi rozdzielcze,
- zakres analiz opisowych i indukcyjnych,
- opisowa analiza porównawcza struktur,
- opis statystycznego pakietu komputerowego STATISTICA PL.

Ponadto są podrozdziały dotyczące charakterystyk opisowych zmiennych:

- indywidualne dane statystyczne,
- miary położenia,
- miary różnicowania,
- standaryzacja,
- miary asymetrii i kurtioza.

Warto zapoznać się w cytowanej publikacji z rozkładami wielu zmiennych, estymacją parametrów, analizą regresji oraz analizą korelacji, a w tym z miarami korelacji krzywoliniowej.

1.6. STATGRAPHICS

Podręcznik „Statystyka stosowana”²¹ zawiera wybrane metody statystyczne, potrzebne do analiz zjawisk gospodarczych. Zaprezentowane zostały przy zastosowaniu pakietu komputerowego STATGRAPHICS. W opracowaniu występuje podział metod statystycznych na metody opisu - analizy danych oraz metody wnioskowania. Podkreślono rolę analizy danych i prognozowania w podejmowaniu decyzji. Celem skryptu „Statystyczna analiza danych za pomocą komputerowych metod statystycznych” jest przedstawienie tych metod stosowanych w statystyce²². Wykorzystany został plik procedur statystycznych znany pod nazwą „Recepty numeryczne”, a także pakiet statystyczno-graficzny STATGRAPHICS nazywany również jako STATGRAF.

Na uwagę zasługuje rozdział trzeci w którym podane są ogólne zasady korzystania z STATGRAF-u. W skrypcie warto zwrócić uwagę na rozdziały obejmujące statystykę opisową, podające sposób instalowania, czy też korzystania z oprogramowania, a mianowicie:

Wybrane informacje wprowadzające do obsługi pakietu STATGRAF, a w ramach niego:

- instalacja i uruchomienie,

²⁰ Luszniewicz A., Słaby T., *Statystyka z pakietem komputerowym STATISTICA PL. Teoria i zastosowania*, Wydawnictwo C.H. Beck, Warszawa 2008.

²¹ Luszniewicz A., Słaby T., *Statystyka stosowana*, PWE.

²² Drabska E., *Statystyczna analiza danych za pomocą komputerowych metod statystycznych*, Politechnika Białostocka, 1992 (wg strony www w Internecie: //brun.fuw.edu.pe/edu/STAT).

- struktura,
- wprowadzanie i wyprowadzanie danych,
- środowisko systemowe,
- sporządzanie raportów,
- atrybuty graficzne i wykresy funkcji.
- komputerowe wspomaganie rozwiązywania wybranych problemów statystycznych.

1.7. SPSS

„Prognozowanie szeregów czasowych za pomocą pakietu SPSS” to kolejna pozycja literatury w której dla ilustracji zagadnienia stosuje się komputer²³. Stanowi rodzaj przewodnika korzystania z modułu *SPSS Trends*. Praktyczne podejście do zagadnień pozwala na samodzielne analizowanie danych, a także wiarygodne prognozowanie wielości cech czy też wskaźników. „Pakiety statystyczne dostępne na WNE UW” jest publikacją internetową²⁴ i zawierającą opisy programów *SPSS*, *EViews*, *SAS* oraz *STATA*.

Pakiet *SPSS (Statistical Package for the Social Sciences)* wersja 10. dla *Windows* początkowo przeznaczony był do opracowywania ankiet. Obecnie posiada procedury analizy danych przekrojowych oraz szeregów czasowych, możliwości wizualizacji danych i wyników. Dostępna jest też wersja polska (10 PL) tego pakietu. *SPSS* ma budowę modułową, a z zakresu statystyki opisowej interesuje nas *Base System* (moduł obowiązkowy), zawierający podstawowe procedury statystyczne, regresję liniową, zarządzanie danymi oraz edycję raportów oraz następujące moduły opcjonalne:

- *Professional Statistics*, procedury statystyczne w zakresie regresji logistycznej, nieliniowej i analizy probitowej;
- *Advanced Statistics*, zaawansowane metody analizy;
- *Tables*, narzędzie do budowy tablic wyników;
- *Conjoint*, techniki analityczne do pomiaru łącznego oddziaływania wielu cech;
- *Missing Values Analysis*, narzędzie do analizy zbiorów danych zawierających braki;
- *Answer Tree*, modelowanie i klasyfikacja za pomocą drzew decyzyjnych.

SPSS zawiera funkcje umożliwiające obliczanie statystyk opisowych, testowanie parametrycznych i nieparametrycznych hipotez statystycznych, analizę danych w podgrupach. Przetwarza dane o strukturze tablicowej. Wiersze stanowią obserwacje traktowane jako obiekty, a kolumny dotyczą zmiennych. Obserwacjom mogą być przypisane wagi. Zaletą omawianego pakietu jest możliwość pozyskania danych z plików zapisanych w formie tekstu, formacie pakietu *SAS*, pochodzących z *dBase* oraz *Excela*. Autor informacji o pakiecie *SPSS*²⁵ zaleca dla początkującego użytkownika podręcznik wydany przez SPSS Polska²⁶.

Kolejnym pakietem zaprezentowanym w ramach publikacji internetowej „Pakiety statystyczne dostępne na WNE UW” jest *EViews*, który już wstępnie przedstawiono. Zdaniem autorki tej części artykułu internetowego był on rozwijany przez ekonomistów przede wszystkim do analizy danych finansowych, przewidywań makroekonomicznych i symulacji²⁷. Program *EViews* wersja 4.0 pracuje w środowisku *Windows* i przeczytać może dane zapisane w formacie tekstowym, *DBF*, *Excel*

²³ Żądło T., Wywiał J., *Prognozowanie szeregów czasowych za pomocą pakietu SPSS*.

²⁴ Praca zbiorowa, Pakiety statystyczne dostępne na WNE UW (wg strony www.wne.uw.edu.pl/panel/materialy/opis-programow.pdf).

²⁵ Pęczkowski M., *Pakiet statystyczny SPSS*.

²⁶ Górniak J., Wachnicki J., *Pierwsze kroki w analizie danych SPSS PL for Windows*, SPSS Polska, Kraków 2000.

²⁷ Ciecieląg J., *Pakiet statystyczny EViews*.

i *Lotus*. Istnieje też możliwość przeniesienia danych za pomocą komend *Kopiuj, Wklej*. Program dostarcza między innymi niezbędnych narzędzi do analizy danych, regresji oraz prognozowania. Dokonać możemy estymacji jedno i wielorównaniowych. W *Eview* występują procedury analizy danych panelowych, przy czym wygodnym trybem tego programu jest praca interaktywna.

W programie występuje pojęcie „obiektu” którym może być zmienna, grupa zmiennych, wektor współczynników, równanie, model, złożone polecenie. Przykładowo obiekt „równanie” zawiera:

- oszacowania współczynników modelu,
- zestaw statystyk,
- macierz wariancji,
- macierz kowariancji.

Obiekty obejmujące również dane, wyniki analizy stanowią plik roboczy i przechowywane są w pamięci komputera. Zorganizowany jest on według numerów obserwacji lub skali czasowej.

1.8. SAS

Według Karola Kuhla, autora fragmentu „Analiza danych panelowych za pomocą pakietu SAS” wymieniony pakiet umożliwia²⁸:

- dostęp do zbiorów w różnych środowiskach programowych oraz zarządzanie danymi,
- prowadzenie statystycznej i ekonometrycznej analizy danych,
- prezentację ilościową oraz graficzną wyników analiz.

SAS cechuje znaczna ilość procedur statystycznych, a ponadto dostęp między innymi do zbiorów danych typu *ASCII*, *SPSS*, *Excel*, *Access*.

1.9. Stata

Stata jest pakietem statystycznym pracującym w środowiskach *Windows*, *Unix*, *Macintosh*²⁹. Pozwala na zarządzanie, analizę i wykonanie wykresów danych. Edytor danych jest w postaci prostego arkusza, a polecenia trzeba wpisywać po jednej linijce w oknie *Command Line*. Wyniki prezentowane są w oknie *Results*. W *Stata* są dodatkowe okna w które można wpisywać nazwy zmiennych i poprzednio używane komendy. Alternatywnym rozwiązaniem jest tworzenie programów i procedur w plikach typu *ASCII*. Pliki te mogą zawierać lokalne zmienne, podprogramy i odwołania do innych programów oraz procedur. Dane mogą być wpisywane ręcznie w edytorze danych programu, importowane w formacie *ASCII*, kopiowane z arkusza danych w innym programie.

Zaprezentowane rozpoznanie literaturowe wraz z omówieniem podstawowej funkcjonalności pakietów są tylko wąskim obrazem obszernej literatury zastosowania techniki komputerowej. Dotyczy to zarówno wspomagania procesów obliczeń z zakresu statystyki opisowej, matematycznej jak również modelowania ekonometrycznego. Wielu jest bowiem twórców wycinkowych modułów tworzących pakiety. Rozwijają się i powstają nowe platformy sprzętowe oraz systemy operacyjne i sieciowe. Powoduje to, że starsze oprogramowanie nie nadąża za falą doskonalenia arsenału metod statystycznych oraz ekonometrycznych. Z tego względu godne uwagi jest każde wydanie książkowe, czy też publikacja internetowa dotycząca zwłaszcza możliwości samodzielnego instalowania

²⁸ Praca zbiorowa, *Pakiety statystyczne dostępne na WNE UW*, fragment: Kuhl K., „Analiza danych panelowych za pomocą SAS”.

²⁹ Praca zbiorowa, *Pakiety statystyczne dostępne na WNE UW*, fragment: Tomaszewski A. „Możliwości przeprowadzenia badań statystycznych na danych panelowych przy pomocy programu *Stata*”.

oprogramowania bezpłatnego ogólnie dostępnego. Warto jeszcze wspomnieć o „Wielowymiarowej analizie statystycznej ...”³⁰ oraz o „Statystyce w biznesie...”³¹

1.10. R

Pakiet komputerowy *R* jest narzędziem do analizy danych i pozwala na: wykonanie analizy statystycznej, przedstawienie otrzymanych wyników w formie raportu oraz wykresów, przekazanie raportu drogą e-mailową. Istotną zaletą tego pakietu jest bezpłatna dostępność w Internecie do różnych zastosowań. Interesujące jest przedstawienie statystycznych, zarówno ilościowych jak i graficznych, możliwości w „Przewodniku po pakiecie *R*”³². W książce tej znajduje się opis bibliotek oraz funkcji użytkowych, a z zakresu statystyki opisowej warto uwagi są rozdzielają:

- typy zmiennych i operacje na nich, w tym wektory i macierze;
- operacje wejścia/wyjścia, obejmujące zapisywanie i odczytywanie danych;
- liczbowe statystyki opisowe;
- popularne rozkłady zmiennych losowych;
- funkcje graficzne z paletą różnorodnych wykresów.

Nazwa *R* obejmuje zarówno język o tej samej nazwie jak i środowisko programowe, a w tym zbiór bibliotek. W *R* dostępnych jest wiele funkcji statystycznych. Jest projektem otwartym popularnym na wielu uczelniach. Obecnie jego rozwojem kieruje fundacja *The R Foundation for Statistical Computing*. Język *R* jest językiem interpretowanym, a nie kompilowanym. Podajemy ciąg komend do wykonania. Wygodne jest jednak korzystanie z nakładek i *GUI*, czyli graficznego interfejsu użytkownika. *R* to platforma wielu pakietów programowych, z pośród których z zakresu statystyki opisowej należy wymienić *lima* - analiza danych mikromacierzowych oraz *rgl* - grafika trójwymiarowa.

1.11. Dodatkowe nawiązanie do oprogramowania statystycznego

Ważnym elementem przygotowania zgromadzonych danych jest analiza statystyczna³³. Według artykułu podanego w przypisie obejmuje ona: analizę opisową, graficzną reprezentację danych i wyników, dopasowanie postaci rozkładu, metody wnioskowania statystycznego, test *t*-Studenta, analizę regresji, testy nieparametryczne. Wymienione funkcje są podstawą funkcjonalności statystycznych programów komputerowych. Metody statystyczne są także częścią pakietów matematycznych. Teraz przedstawione zostaną wybrane programy analizy statystycznej.

MINITAB firmy *Statistical Software*. Pierwsza wersja tego programu powstała jeszcze w latach siedemdziesiątych. Jego 13-ta wersja jest programem prostym w obsłudze.

S-PLUS opracowany przez firmę *MathSoft* cechuje również łatwość obsługi. Umożliwia wykonanie analizy statystycznej danych i prezentację graficzną. Zastosowany język *S* umożliwia kompleksową analizę statystyczną oraz wizualizację danych. Ponadto pozwala na kontrolę procesu analizy danych w trybie *on-line*. Jest kompatybilny z innymi produktami firmy *MathSoft*.

STATGRAPHICS Plus firmy *Statistical Graphisc Corporation* może być narzędziem do celów dydaktycznych oraz do analizy danych o niedużych rozmiarach. Program ma budowę modułową,

³⁰ Frątczak E (red.), *Wielowymiarowa analiza statystyczna. Teoria - przykłady zastosowań z systemem SAS*, SGH, Warszawa 2009.

³¹ Praca zbiorowa, *Statystyka w biznesie. Zastosowanie narzędzi Excela*, Wydawca Elipsa 2005.

³² Biecek P., *Przewodnik po pakiecie R*, Oficyna Wydawnicza GiS, Wrocław 2008.

³³ Niniejszy rozdział opracowano na podstawie publikacji w Internecie - strona: <http://mk.po.opole.pl/treść/programy.htm#statystyczne>

w tym moduł bazowy. W *STATGRAPHICS* oprócz klasycznych miar statystyki opisowej danej zmiennej w zestawieniu sumarycznym występują także:

- średnia geometryczna (*Geometric mean*);
- średnia po odcięciu 5% wartości skrajnych (5% *Trimmed mean*);
- współczynnik zmienności (*Coeff. of variation*);
- błąd standardowy oszacowania średniej (*Standard error*);
- kwartył pierwszy Q_1 (*Lower quartile*);
- kwartył trzeci Q_3 (*Upper quartile*);
- przedział między kwartylami, czyli $2Q = Q_3 - Q_1$ (*Interquartile range*) stanowiący dwa odchylenia ćwiartkowe;
- standaryzowany współczynnik asymetrii (*Std. Skewness*)³⁴;
- standaryzowany współczynnik kurtozy (eksesu, spłaszczenia).

Wyjaśnienie standaryzacji zarówno współczynnika asymetrii jak i kurtozy występuje w Internecie³⁵:

Współczynnik skośności (*Skewness*) zwany również współczynnikiem asymetrii w pakiecie *STATGRAPHICS* wyrażony jest wzorem:

$$g_1 = \frac{n^2}{(n-1)(n-2)} \frac{M_3}{s^3}$$

Współczynnik ten charakteryzuje skośność rozkładu badanej cechy. W praktyce przyjmuje się, że dla $|g_1| < 0,5$ rozkład jest symetryczny, natomiast dla $|g_1| > 1$ mocno skośny. Znak współczynnika wskazuje, w którą stronę jest skośna cecha. Dla wartości dodatniej mamy skośność w prawo, a dla ujemnej w lewo.

Współczynnik skośności standaryzowany (*Standardized Skewness*) określony jest jako:

$$\sqrt{\frac{n}{6}} g_1$$

Autor publikacji internetowej³⁶ wyróżnia jeszcze indeks skośności Pearsona (*Pearson Index of Skewness*), który wyrażony jest wzorem, przy czym w liczniku (średnia minus wartość mediany), a w mianowniku odchylenie standardowe.

$$I = \frac{3(\bar{X} - X_{mod})}{s}$$

Wartość indeksu I interpretuje się podobnie jak współczynnika skośności g_1 .

Współczynnik spłaszczenia – kurtaza (*Kurtosis*) w pakiecie *STATGRAPHICS* określony wzorem:

$$g_2 = \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} \frac{M_4}{s^4} - 3 \frac{(n-1)^2}{(n-2)(n-3)}$$

Dla cechy o rozkładzie normalnym lub zbliżonym do rozkładu normalnego kurtaza jest równa zero. Wartości mniejsze od zera świadczą o spłaszczeniu rozkładu w porównaniu do rozkładu normalnego, natomiast wartości dodatnie o tym, że rozkład charakteryzuje względna szczytowość.

Standaryzowany współczynnik kurtozy tzw. kurtaza standardowa (*Standard Kurtosis*): $\sqrt{\frac{n}{24}} g_2$

W powyższych wzorach występują momenty I -tego rzędu i tak w odniesieniu do średniej, wariancji, występujących w obliczeniach współczynnika skośności, czy też współczynnika kurtozy wyrażone są wzorami:

³⁴ Internet <http://www.google.pl>, hasło: „standardized skewness”.

³⁵ Internet: http://krawczyk.hostil.pl/images/S_opisowa_part2_marco.pdf.

³⁶ Ibidem.

$$m_l = \frac{1}{n} \sum_{i=1}^n x_i^l$$

lub dla danych pogrupowanych:

$$m_l = \frac{1}{n} \sum_{i=1}^k \bar{x}_i^l n_i$$

Moment centralny rzędu l :

$$M_l = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^l$$

Dla danych pogrupowanych: $M_l = \frac{1}{n} \sum_{i=1}^k (x_i - \bar{x})^l n_i$

przy czym:

$$\bar{x} = m_1 \quad \text{oraz} \quad s^2 = \frac{n}{n-1} M_2$$

Wersja *Statgraphics Plus Professional* należy do najpopularniejszych pakietów statystycznych. Oprogramowanie to składające się z kilku modułów (które grupują jego różne funkcje) przeznaczony jest do badania i analizy statystycznej. Przykładowo *StatReport* generuje raporty na podstawie danych empirycznych.

STATISTICA PL jest produktem firmy *Statsoft*. Występuje spolszczona wersja tego pakietu, którą cechuje obszerny zestaw metod statystycznych i operacji na danych, łatwość w obsłudze oraz dobra grafika, a występujące funkcje obejmują:

- podstawowe statystyki,
- regresję wielokrotną,
- statystyki nieparametryczne i rozkłady,
- analizę wariancji,
- estymację nieliniową,
- szeregi czasowe i prognozowanie,
- analizę skupień,
- inne nowe metody, w tym ze statystyki matematycznej.

SPSS, jest pakietem opracowanym przez firmę *SPSS Inc.* Zasadniczym elementem tego pakietu jest moduł *SPSS Base*. Zawiera statystyki wielowymiarowe, tablice i wykresy. Moduł *Advanced Statistics* zawiera funkcje do analizy wariancji, modelowania liniowego i inne. Kolejny moduł *Categories* analizuje profile produktu na podstawie preferencji konsumenta, natomiast *CHAID* zajmuje się modelowaniem segmentowym. *Exact Test* przeznaczony jest do pracy z małymi zbiorami danych, a *Missing Value Analysis* wykrywa brakujące dane i zastępuje je wartościami estymowanymi. *Professional Statistics* udostępnia wyszukiwane funkcje regresji. Moduł *Tables* umożliwia sporządzanie raportów tabelarycznych. Narzędziem do prognozowania jest moduł *Trend*.

SYSTAT 7.0 to kolejny pakiet firmy *SPSS Inc.* Posiada funkcje obróbki danych i prezentacji wyników. Z zakresu statystyki opisowej zawiera klasyczne procedury obliczania statystyk oraz testy nieparametryczne. Ponadto obejmuje procedury dotyczące korelacji, szeregów czasowych, modele liniowe i analizę skupień.

Program *nQuery Advisor* jest produktem firmy *Statistical Solution Ltd.* Ułatwia planowanie eksperymentu, określenie liczebności próby i jej cech oraz dobór grupy statystycznej. Wielkość próby określana jest testem Wilcoxa/Manna-Whitneya dla ciągłych lub dyskretnych wartości.

STATLES, firmy *NWP Associates* umożliwia analizę statystyczną własnych danych poprzez Internet lub w obrębie użytkowników tworzących własną sieć lokalną typu *Intranet*. Zawiera cechy

programu *STATGRAPHICS* oraz *Execustat*. Pakiet *STATLES* jest zestawem dostępnych funkcji umieszczonych na serwerze i przetwarzanych za pomocą przeglądarki stron WWW. Analiza statystyczna obejmuje: metody opisowe zarówno numeryczne jak i graficzne, testowanie hipotez, analizę wariancji, analizę regresji, kontrole jakości, prognozowanie szeregów czasowych, planowanie eksperymentów. Stosując narzędzia tego pakietu uzyskujemy wykresy rozrzutu, histogramy, wykresy pudełkowe, wykresy reszt w skali prawdopodobieństw rozkładu normalnego. Dodatek do tego pakietu - *Statistical Interpreter* szacuje istotność uzyskanych wyników.

2. ŚRODOWISKO PAKIETU R



2.1. Instalowanie i uruchomienie R

Pierwszym etapem po zapoznaniu się z pakietem *R* jest instalowanie środowiska z podstawowymi bibliotekami³⁷. Otrzymany zestaw funkcji jest zasadniczo wystarczający dla zadań ze statystyki opisowej. Służy bowiem do analizy danych i tworzenia wykresów. W kolejnym etapie możemy doinstalować dodatkowe pakiety, wzmacniając w ten sposób wersję podstawową. Można zainstalować *R*, przykładowo wersję 2.7.1 ze skompilowanego pliku instalacyjnego, a procedura w tym zakresie jest następująca:

- wybór lokalnego serwera, przykładowo wrocławskiego zwanego *mirror*, na którym umieszczony jest plik instalacyjny³⁸,
- ściągnięcie pliku *R-2-7.1-win32.exe*,
- uruchomienie i postępowanie dalej według podanych podpowiedzi programu.

Wynikiem jest utworzenie w katalogu *C:\Program Files\R\R-2-7.1* struktury podkatalogów z niezbędnymi plikami użytkowymi³⁹. Uruchomienie programu *R.exe* jest w trybie tekstowym, a *Rgui.exe* w graficznym zwanym okienkowym. Oba te pliki znajdują się w podkatalogu *bin* zawierającym pliki wykonywalne. Gdy program sygnalizuje brak określonego pakietu to można go doinstalować funkcją *install.packages(utils)* wpisując w miejsce *utils* nazwę pakietu. Przykładowo dodanie do podstawowego zbioru bibliotek platformy *R* pakietu *Rcmdr* wraz z wszystkimi podpakietami niezbędnymi do jego działania posiadającego dogodne dla użytkownika rozwijalne menu główne jest następujące: *install.packages („Rcmdr”, dependencies=TRUE)*. Alternatywnym rozwiązaniem jest wybranie z menu trybu okienkowego opcji:

packages/install packages(s).

Pakiety pobierane są ze wskazanego przez użytkownika lokalnego serwera tzw. *mirrors* i umieszczone w katalogu *library*.

Uruchomienie platformy *R* polega na ładowaniu pakietów podstawowych: *base*, *graphics*, *stats*. Jeśli określony jest pakiet użytkowy, przykładowo *Rcmdr*, wspomagający interfejs graficzny, został już załadowany i znajduje się na dysku komputera użytkownika, to jego uaktywnienie w pamięci (dodatkowych funkcji, zbiorów danych) następuje poprzez polecenie: *library(Rcmdr)*.

Jak już nadmieniono, pakiet *Rcmdr* umożliwia pracę w trybie z interfejsem graficznym, co jest niezbędne przy początkowym kontakcie użytkownika z platformą *R*. Godnym uwagi jest zamieszczenie w „Przewodniku po pakiecie *R*” skorowidza funkcji z zaznaczeniem w nawiasie nazwy pakietu, przykładowo funkcja: *wilcox.test(stats)*. Gdy mamy sytuację odwrotną - znamy funkcję, a nie znamy nazwy pakietu to korzystamy z pomocy: *help.search(utils)*. Natomiast dla wskazania z którego pakietu chcemy skorzystać dla wybrania funkcji używamy operatora ::

Powróćmy jeszcze do uruchomienia środowiska programowego *R*. Jak już wiemy przy pierwszym uruchomieniu korzystamy z pliku *Rgui.exe* umieszczonym w katalogu *bin*. Na ekranie pojawia się znak zachęty > i system oczekuje na wprowadzenie polecenia. Menu ekranu pakietu *Rgui* zawiera między innymi opcje, a w ramach nich funkcje:

³⁷ Ibidem

³⁸ Adresy mirrorów znajdują się pod linkiem: <http://cran.r-project.org/mirrors.html>.

³⁹ Biecek P., *Przewodnik po pakiecie R*, op. cit., rozdz. 1.3.1.

- *File* (wskazanie pliku tekstowego z listy komend w języku *R*, otwarcie wbudowanego edytora tekstów do edycji nowego pliku, wyświetlenie zawartości wskazanego pliku, odczytanie obszaru roboczego z informacjami o wszystkich obiektach znajdujących się w pamięci komputera);

- *Edit* (standardowe operacje do kopiowania i wklejania informacji do i ze schowka);

- *View* (włączanie/wyłączanie pasków stanu wraz z narzędziami);

- *Misc* (przerwanie aktualnie wykonywanego polecenia);

- *Packages* (załadowanie wybranego pakietu *R*);

- *Windows* (określenie sposobu określania okien);

- *Help* (lista skrótów klawiszowych przyspieszających realizację zadania).

2.2. Charakterystyka *R*

W środowisku *R* możliwe jest wykonywanie operacji oraz odwoływanie się do obiektów (list, wektorów, macierzy i ramek danych). Ramka danych jest strukturą przechowywania danych tabelarycznych. Zależności między zmiennymi opisywane są za pomocą formuł. Obiekty z formułami stanowią argumenty funkcji statystycznych i statystyk opisowych. Stosowane są operatory arytmetyczne i logiczne. Nazwy wszystkich zmiennych tworzą przestrzeń nazw. Język *R* zawiera funkcje - komendy do zarządzania przestrzenią nazw. Przykładowo *ls()* wyświetla nazwy obiektów.

Najpopularniejszym sposobem tworzenia wykresów w *R* jest zastosowanie funkcji *plot()*. Występują również funkcje umożliwiające operacje na plikach i katalogach. Krokiem początkowym jest wczytanie danych do analizy statystycznej. Pakiet *R* podobnie jak inne pakiety umożliwia zapisywanie i odczytywanie danych z: plików tekstowych, plików utworzonych w innych formatach, schowka systemowego, bazy danych.

Zalecane jest, ze względu na stosowanie kropki dziesiętnej w *R*, zapisywanie pliku importowanego z *Excelsa* w formacie *csv*. Ponadto pliki graficzne w *R* zapisywać można w formatach *bmp*, *pdf* i innych, gdy zbiory są duże to możemy korzystać z baz danych. W tym względzie istnieje możliwość połączenia z *Accessową* bazą danych, czy też korzystania z plików *Excelsa* jako bazy danych⁴⁰. Środowisko programowe *R*, jak już nadmieniono, to bardzo obszerny zbiór funkcji zarządzania obiektami, funkcji matematycznych oraz procedur statystycznych. Skupmy się jednak na obszarze nas interesującym, tj. statystyce opisowej, która stanowi zobrazowanie danych.

Przewiduje się, że oprogramowanie *R* stanowić będzie w przyszłości podstawowy pakiet do nauki statystyki w szkołach wyższych⁴¹.

2.3. Elementy statystyki opisowej

Podstawowe statystyki opisowe dla *n*-elementowego wektora służą do oceny parametrów rozkładu na podstawie próby. Realizowane są odpowiednimi funkcjami, które wyszczególniono w tabeli 3.1. zamieszczonej w rozdziale „Wybrane procedury statystyczne”⁴². Dla ich zaprezentowania wymienimy chociaż do obliczenia jakich cech służą:

- wartość minimalna/maksymalna w próbie;
- średnia arytmetyczna, ważona, geometryczna, harmoniczna;
- mediana;
- kwantyl wybranego rzędu;

⁴⁰ Biecek P., *Przewodnik po pakiecie R*, op. cit., rozdz. 2.3.4.

⁴¹ Ibidem, rozdz. 3. *Wybrane procedury statystyczne*.

⁴² Ibidem, s.: 126.

- przedział zmienności próby;
- rozstęp międzykwartylowy, czyli różnica między górnym a dolnym kwartylem;
- wariancja próby;
- odchylenie standardowe (próby);
- macierz korelacji i kowariancji;
- współczynnik zmienności określany jako iloraz odchylenia standardowego do średniej z

próby;

- kurtoza jako miara względnego spłaszczenia/szczytowości rozkładu empirycznego w porównaniu z rozkładem normalnym, który ma kurtozę równą zero;
- skośność jako miara asymetryczności;
- dominanta, czyli wartość występująca najczęściej w próbie.

Jak wiemy, graficzną reprezentacją zmiennych są wykresy i z tego obszaru statystyka graficzna obejmuje:

- histogram, przedstawiający liczebność w poszczególnych przedziałach danej zmiennej;
- wygładzona wersja histogramu zwana jądrowym estymatorem gęstości;
- wykres pudełkowy, przedstawiający cechy pozycyjne zmienności pojedynczej zmiennej;
- wykres skrzypcowy, stanowiący odbite symetrycznie jądrowe oceny gęstości;
- wykres rozrzutu, zwany też wykresem rozproszenia, który stosowany jest do przedstawienia zależności między parą zmiennych;
- wykres mozaikowy, stosowany dla 3-ch zmiennych, gdzie liczebności odpowiednich kombinacji zmiennych reprezentowane są przez pola;
- wykres balonowy, w którym liczebnościom odpowiadają odpowiednie powierzchnie kół, zamiast prostokątów jak w wykresie mozaikowym.

Na uwagę zasługują podstawowe rozkłady zmiennych losowych, a w tym rozkładów normalnego, jednostajnego, wykładniczego i t -Studenta. Listę funkcji obsługujących wybrane rozkłady prawdopodobieństwa w R , zarówno ciągłych jak i dyskretnych wyszczególniono w tabeli 3.2⁴³. Środowisko R obejmuje także dalsze procedury statystyczne, które jednak wykraczają poza zasięg statystyki opisowej, a są elementami ekonometrii, czy też statystyki matematycznej. Należy tu wymienić między innymi estymację parametrów metodą maksimum funkcji wiarygodności. Proces przygotowania danych statystycznych jest pracochłonny i czasem trwa dłużej niż sam etap analizy. Środowisko programowe R daje pewne udogodnienia dla tzw. przetwarzania wstępnego i należy tu wymienić:

- procedurę radzenia sobie z brakującymi obserwacjami;
- standaryzację, która polega na odjęciu od każdej obserwacji wartości średniej i podzieleniu tej różnicy przez odchylenie standardowe;
- transformacje nieliniowe, jeśli dane są „skośne”.

Powróćmy jeszcze do możliwości tworzenia, zarządzania i modyfikowania grafiki danych statystycznych w środowisku R . Zainteresują nas sposoby wizualizacji danych pakietem *ggplot*, który daje większość kolorowych graficznych statystyk opisowych, a mianowicie wykresy:

- paskowy tzw. jednowymiarowy wykres rozrzutu;
- róża wiatrów, odpowiadający histogramowi, lecz dla danych o charakterze kątowym;
- słonecznikowy, na którym dla każdego punktu zaznaczone są również liczebności jego wystąpienia;
- rozrzutu (3D) dla trzech zmiennych;
- kołowy;

⁴³ Biecek P., *Przewodnik po pakiecie R*, op. cit., s. 144.

- słupkowy;
- kropkowy, dla lepszego zaznaczenia elementów wykresu;
- otoczkowy, z znaczeniem otoczki obszaru obserwacji;
- torbowy, z otoczką obrysowującą 50% punktów i zewnętrzną obrysowującą prawie wszystkie punkty.

Ponadto *R* umożliwia przedstawienie graficzne macierzy wykresów rozrzutu dla każdej pary podanych zmiennych i wiele innych specjalistycznych reprezentacji danych z doбором kolorów i linii.

2.4. Możliwości analizy wielowymiarowej w *R*

Zaawansowana statystyka opisowa to również analiza wielowymiarowa. Publikację temu tematowi opracowano z wykorzystaniem środowiska *R*⁴⁴. Obejmuje ona oprócz analizy skupień, klasyfikacji, rozpoznawania obrazów z obszaru statystyki i ekonometrii również wiele innych funkcji. W ramach statystycznej analizy wielowymiarowej występują zagadnienia:

- typy skal pomiarowych i ich charakterystyka,
- normalizacja zmiennych,
- miary odległości,
- charakterystyki rozkładu wielu zmiennych.

Istotnym elementem są szerokie możliwości graficznej prezentacji rozkładów zmiennych, w tym również przestrzeni trójwymiarowej. Oprogramowanie *R* w wersji 2.7.x ma około 1500 pakietów rozszerzających jego podstawowe możliwości⁴⁵. Jest więc obszernym systemem z większością pakietów umieszczonych na serwerach CRAN (*Comprehensive R Archive Network*), chociaż zalecane jest korzystanie z kopii na serwerze lokalnym w pobliżu miejsca zamieszkania użytkownika.

W obliczeniach statystycznych macierze są jako tablice dwuwymiarowe. Niezbędne są więc - podobnie jak w *Excelu*, tablice umożliwiające następujące działania na macierzach:

- mnożenie macierzy;
- obliczanie wyznacznika macierzy kwadratowej;
- określenie macierzy odwrotnej;
- wyznaczenie macierzy składowej, gdy znamy iloczyn macierzy i jedną z macierzy będącą składnikiem tego iloczynu;
- określenie macierzy transponowanej.

W statystycznej analizie wielowymiarowej różniamy również pojęcia *obiekt* oraz *zmienna*, przy czym obiektem jest określona rzecz, osoba lub zdarzenie. Zbiór obiektów oznaczamy jako:

$$A = \{A_i\}_1^n = \{A_1, \dots, A_n\}$$

Zmienną jest charakterystyka opisująca zbiorowość obiektów. Znając zbiór obiektów i zbiór zmiennych możemy zapisać macierz danych, w której występuje x_{ij} czyli obserwacja j -ta zmiennej w i -tym obiekcie. Gdy dodamy kolejną zmienną - okres czasu, to uzyskamy ujęcie trójwymiarowe tzw. kostkę danych w której x_{ijt} jest to wartość j -tej zmiennej w i -tym obiekcie w okresie t ⁴⁶.

⁴⁴ Walesiak M., Gatnar E. (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009.

⁴⁵ Ibidem, rozdz. 1.4.

⁴⁶ Ibidem, s. 63.

Celem wspomnianej normalizacji, którą bywa standaryzacja jest pozbawienie mian wartości zmiennych i ujednolicenie ich rzędów wielkości. Istnieją różne formuły normalizacyjne dla danych metrycznych, a przykładowo znana nam już standaryzacja wyrażona jest wzorem⁴⁷:

$$z_{ij} = (x_{ij} - \bar{x}_j) / s_j$$

Gdzie:

x_{ij} - obserwacja j -ta w i -tym obiekcie,

$\bar{x}_j$ - średnia j -tej zmiennej,

s_j - odchylenie standardowe j -tej zmiennej.

W odniesieniu do rozkładu wielu zmiennych ciągłych rozróżniamy następujące podstawowe charakterystyki:

- wektor wartości przeciętnych, który w odniesieniu do próby dotyczy zestawu średnich arytmetycznych z wartości j -tej zmiennej:

$$\bar{x}^T = [\bar{x}_1, \dots, \bar{x}_m]$$

Gdzie: m - liczba zmiennych.

- macierz kowariancji, którą dla próby określamy jako⁴⁸:

$$S = \begin{bmatrix} s_{11} & \dots & s_{1m} \\ \dots & \dots & \dots \\ s_{m1} & \dots & s_{mm} \end{bmatrix}$$

Gdzie:

$$s_{jh} = \frac{1}{n-1} \sum_{i=1}^n (x_{ij} - \bar{x}_j) (x_{ih} - \bar{x}_h)$$

- macierz korelacji, która dla próby możemy ująć jako:

$$R = \begin{bmatrix} r_{11} & \dots & r_{1m} \\ \dots & \dots & \dots \\ r_{m1} & \dots & r_{mm} \end{bmatrix}$$

Przy czym:

$$r_{ih} = r_{hj}; \quad r_{jh} \in [-1; 1]$$

$$r_{jh} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j) (x_{ih} - \bar{x}_h)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 (x_{ih} - \bar{x}_h)^2}}$$

Kolejny raz przejdźmy do graficznej prezentacji danych w środowisku R , lecz teraz w ujęciu wielu zmiennych. Funkcje trybu graficznego pozwalają:

- wykonać wykresy punktowe wartości dwóch zmiennych w różnej formie edycyjnej,
- przedstawić empiryczny rozkład częstości zmiennych, co pozwala następnie na sporządzenie histogramu, który stanowi oszacowanie trójwymiarowej funkcji gęstości dla zmiennych X_1 i X_2 .

Zagadnieniem, które jest pomostem między statystyką opisową a ekonometrią jest analiza regresji wielorakiej. Bada się w niej zależność zmiennej objaśnianej Y od wielu (co najmniej dwóch) zmiennych objaśniających $(X_0, X_1, \dots, X_m)$. Liniowy model takiej regresji do której wykorzystuje się klasyczna metodą najmniejszych kwadratów ($KMNK$) jest następujący⁴⁹:

⁴⁷ Ibidem, tabela 2.2.

⁴⁸ Ibidem, s. 74.

⁴⁹ Ibidem, rozdz. 5.

$$Y_t = b_0 X_{0t} + b_1 X_{1t} + \dots + b_m X_{mt} + e_t = \sum_{j=0}^m b_j X_{jt} + e_t$$

Gdzie:

$b_0, b_1, \dots, b_n$ - parametry strukturalne,

e_t - składnik losowy,

$t = 1, \dots, T$ - numer obserwacji.

Równanie macierzowe estymacji parametrów b w modelu określone jest jako:

$$\hat{b} = (X^T X)^{-1} X^T y$$

Oszacowanie parametrów modelu pozwala na obliczenie regresyjnej sumy kwadratów odchyleń - reszt (*SSR*) oraz resztowej sumy kwadratów odchyleń (*SSE*), a wykonanie tego umożliwia określenie następujących miar dopasowania modelu do danych empirycznych:

- współczynnika determinacji R^2 ,
- współczynnika zbieżności, $\Phi^2 = 1 - R^2$
- skorygowanego o liczbę obserwacji współczynnika determinacji $\bar{R}^2$
- standardowy błąd oceny stanowiący pierwiastek z wariancji resztowej⁵⁰:

$$s = \hat{\sigma} = \sqrt{\frac{SEE}{T - (m + 1)}}$$

Ponadto można obliczyć statystykę F oceny istotności całości zestawu parametrów modelu ekonometrycznego:

$$F = \frac{MSR}{MSE} = \frac{SSR/m}{SSE/(T - m - 1)}$$

* * *

Pakiet *R* posiada jeszcze więcej funkcji oceny modelu, lecz to jest raczej domena ekonometrii, niż statystyki i dlatego odsyła się dociekliwego Czytelnika do rozdziału 5. „Analiza regresji wielorakiej”⁵¹. Warto jeszcze nadmienić, że podstawowe charakterystyki modelu uzyskujemy w środowisku *R* stosując pakiet *stats*.

W *R* do klasyfikacji obserwacji x_i stosowane są różne metody między innymi metoda K -najbliższych sąsiadów. Dana obserwację przydziela się do tej klasy do której należy najwięcej spośród tych K sąsiadów. K musi być znacznie mniejsze niż liczebność najmniejszej klasy N .

⁵⁰ Ibidem, tabela 5.1.

⁵¹ Ibidem, s. 135 i dalsze.

3. PRZYKŁADY RAPORTÓW STATYSTYKI OPISOWEJ



3.1. Wstęp

Zmienne modelu ekonometrycznego produktu brutto województw były bazą do przykładów zastosowania narzędzi programowych do statystyki opisowej zarówno ilościowej jak i graficznej. Dla możliwości sortowania danych według województw, a w ramach nich według okresów czasowych, nadano województwom oraz latom (1999-2007) numery porządkowe. Fragment tak przygotowanych danych dla województwa łódzkiego obejmującego tylko zmienne $X_1, X_5, X_6, X_8, X_{10}, X_{12}, X_{13}, X_{20}$ prezentuje tabela 3.1.

Tab. 3.1. Dane statystyczne modelu ekonometrycznego

W	T	X_1	X_5	X_6	X_8	X_{10}	X_{12}	X_{13}	X_{20}	n
1	1	12682	5171,6	188,3	14,3	47,6	1504,67	1392,73	2434,6	1
1	2	14497	5807,8	158,6	16,3	43,7	1664,10	1481,54	2670,4	2
1	3	15723	5263,7	153,0	18,1	37,7	1757,34	1478,20	2687,7	3
1	4	17535	5041,8	130,8	18,4	34,7	1829,93	1532,55	2290,1	4
1	5	18492	4881,6	134,5	20,7	32,8	1880,97	1566,82	2482,6	5
1	6	19498	4884,3	124,0	19,5	35,5	1975,94	1609,64	3189,8	6
1	7	22274	4449,5	127,7	17,9	37,3	2023,61	1663,61	3901,8	7
1	8	23666	4120,7	127,4	14,7	41,0	2109,67	1759,55	3850,0	8
1	9	25521	3895,2	129,4	11,2	46,7	2262,39	1928,20	5479,1	9

Źródło: Opracowanie własne w Excelu.

3.2. Arkusz kalkulacyjny Excel

Możliwości generowania raportów ilościowych i histogramów na przykładzie zmiennych z poziomu życia jest możliwe z zastosowaniem programu Excel. W tym miejscu przytoczono w pełni procedurę raportowania uzyskaną z menu głównego Excela na danych statystycznych, by móc porównać wyniki z innymi środowiskami programowymi. Skorzystano z menu: *Narzędzia/Analiza danych*, a w ramach niej z opcji:

- *statystyka opisowa* (podająca podstawowe miary rozkładu zmiennej),
- *korelacja* (dla uzyskania macierzy współczynników korelacji liniowej Pearsona),
- *kovariancja* (w celu przedstawienia macierzy kowariancji między zmiennymi modelu ekonometrycznego),
- *histogram* (dający możliwości podawania przedziałów klasowych oraz graficznego rozkładu danej zmiennej na podstawie jej próby).

Przed przystąpieniem do otrzymania raportów statystyki opisowej zasadniczo dokonujemy wprowadzenia danych do arkusza kalkulacyjnego programu Excel. W naszym przykładzie autor skorzystał z istniejącego zbioru wejściowego *GRETL.xls* obejmującego zmienne X_1 - X_{23} . Był on potrzebny do sformułowania modelu ekonometrycznego zmiennej X_1 - produkt krajowy brutto (ceny bieżące) na 1. mieszkańca określonego województwa. Pełne dane statystyczne obejmują 144 obserwacje, będące danymi panelowymi województw (1-16) i okresów czasowych (1-9).

Ponadto *Excel* daje możliwość określenia regresji wielu zmiennych stosując znaną nam już funkcję *REGLINP* dla siedmiu zmiennych objaśniających lub korzystając ze ścieżki menu: *Narzędzia/Analiza danych/Regresja*. W efekcie uzyskujemy raporty:

- statystyka regresji,
- wnioskowanie statystyczne (z określeniem parametrów modelu liniowego wielowymiarowego),

- reszty i ich standaryzacja,
- analiza wariancji.

Przyjmijmy jako zmienną objaśnianą $Y = X_1$, a następnie wypełniamy następujące argumenty w oknie dialogowym:

Zakres wejściowy X_1 : \$C\$1:\$C\$145.

Tytuły: v (z zaznaczeniem występowania nazw w pierwszym wierszu danej kolumny).

Poziom ufności: 95% (odnoszący się do oszacowania średniej zmiennej X_1 na poziomie istotności $\alpha = 0,05$).

Opcje wyjścia: Nowy arkusz.

Składniki resztowe, a w ramach tego zaznaczamy:

- reszty,
- standaryzowane składniki resztowe,
- rozkład reszt,
- rozkład linii dopasowanej,
- rozkład prawdopodobieństwa normalnego.

W wyniku działania programu *Excel* uzyskujemy raporty podsumowujące, a mianowicie:

Statystyki regresji

Tab. 3.2. Raport statystyki regresji

Wielokrotność R (R_w)	0,956039
R^2	0,91401
Dopasowany R^2 ($\bar{R}^2$)	0,909584
Błąd standardowy reszt s	1734,042
Obserwacje n	144

Źródło: Opracowanie własne w *Excelu*.

Gdzie:

R - macierz korelacji między zmiennymi objaśniającymi $X_5, X_6, X_8, X_{10}, X_{12}, X_{13}, X_{20}$,

R^* - macierz rozszerzona:

$$R^* = \begin{bmatrix} 1 & R_o^T \\ R_o & R \end{bmatrix}$$

R_o - wektor zmiennej objaśnianej i zmiennych objaśniających,

R_w - współczynnik korelacji wielorakiej:

$$R_w = \sqrt{1 - (\det R^* / \det R)}$$

R^2 - współczynnik determinacji:

$\bar{R}^2$ - skorygowany współczynnik determinacji:

$$R^2 = 1 - \phi^2 = 1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$\bar{R}^2 = 1 - \frac{n-1}{n-2} (1 - R^2)$$

s - stanowi pierwiastek z wariancji resztowej dla n -obserwacji i k ilości zmiennych objaśniających:

Analiza wariancji

$$s(n) = \sqrt{\frac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}}$$

Tab. 3.2. Raport z analizy wariancji

		df	SS	MS = SS/df	F	Istotność F
Regresja	E	$k = 7$	4346710632	620958662	206,5112	3,41639E-69
Resztkowy	U	$n - k - 1 = 136$	408938450,7	3006900,37		
Razem	T	$n - 1 = 143$	4755649083			

Źródło: Opracowanie własne w Excelu.

Gdzie:

Edf - liczba zmiennych objaśniających,

ESS - wyjaśniona suma kwadratów odchyłeń teoretycznych zmiennej objaśnianej:

$$ESS = \sum_{i=1}^n (y_i - \bar{y})^2$$

Udf - liczba obserwacji z uwzględnieniem parametrów modelu ekonometrycznego,

USS - niewyjaśniona suma kwadratów odchyłeń (reszt) przez model:

$$USS = \sum_{i=1}^n (y_i - \hat{y})^2 \quad TSS = \sum_{i=1}^n (y_i - \bar{y})^2$$

TSS - całkowita (łączna) suma kwadratów odchyłeń zmiennej objaśnianej od jej wartości średniej,

F - statystyka Fishera.

Wnioskowanie statystyczne

Tab. 3.4. Raport z wnioskowania statystycznego

Zmienna	Współczyn- niki (parametry)	Błąd standardowy (oszacowania para- metru)	t Stat (statystyka t-Stu- denta)	Wartość- p	Dolne 95%	Górne 95%
Przecię- cie	-15763,31	1493,05	-10,56	2,11E-19	-18715,9069	-12810,7
X5	-0,129212	0,024683	-5,235	6,12E-07	-0,17802395	-0,0804
X6	0,810494	0,282636	2,868	0,0048	0,251564986	1,369423
X8	196,194	36,3607	5,396	2,94E-07	124,2880807	268,0992
X10	-11,3945	5,74129	-1,985	0,0492	-22,7482785	-0,04078
X12	8,0591	1,10165	7,316	2E-11	5,880526978	10,23767
X13	7,37759	1,50746	4,894	2,75E-06	4,396500189	10,35868
X20	1,09901	0,170458	6,447	1,83E-09	0,761915991	1,436096

Źródło: Opracowanie własne w Excelu.

Gdzie:

- Dolne/Górne 95% - przedział ufności,

- np. X5 - zapis zmiennej w tabeli danych dla celów przetwarzania komputerowego,

- np. E-19 (10^{-19}) - zapis komputerowy w postaci tzw. inżynierskiej.

Zwróćmy uwagę na uzyskanie tych samych wartości parametrów statystyki *t*-Studenta, prawdopodobieństwa popełnienia błędu w oszacowaniu danego parametru (wartość-*p*), jak w raporcie z programu *GRET*L uzyskanym w wyniku *eliminacji a posteriori*. Dodatkowym elementem w *Excelu* jest podanie przedziału ufności dla danego parametru.

Składniki resztowe (tzw. wyjście)

To kolejny raport z opcji *Regresja*; obejmuje części: reszty i standaryzacja reszt (dla obserwacji 1.1-16.9), percentyle. Dla poszczególnych obserwacji obliczane są reszty u_i . Następnie dokonana jest standaryzacja reszt według wzoru:

$$t_u = (u_i - \bar{u})/s$$

Gdzie: s - odchylenie standardowe reszt według raportu statystyki regresji, $\bar{u}$ - wartość średnia reszt.

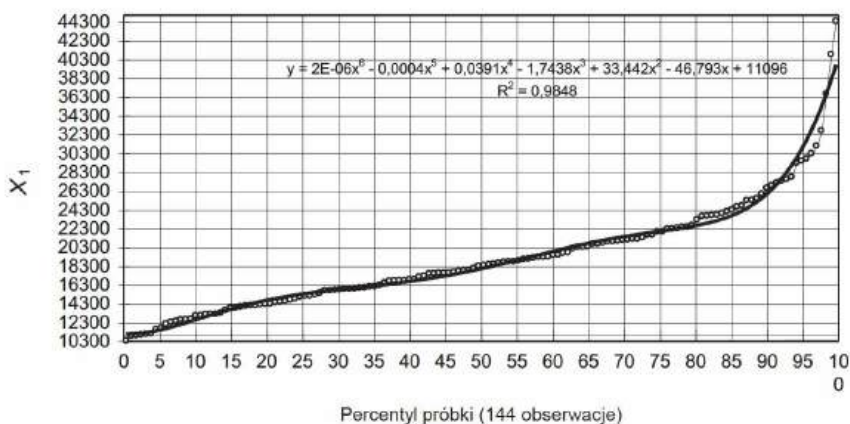
Oprócz tego jest możliwość uzyskania wyszczególnienia narastająco wartości zmiennej objaśnianej i obliczenia percentyli, przy czym percentylem nazywamy taką wartość cechy, która stanowi 1% zbiorowości próbnej. Fragment raportu ze standaryzacją reszt i percentylami obejmującego 10 pierwszych obserwacji oraz 5 końcowych podano w tabeli 3.5.

Tab. 3.5. Raport ze standaryzacji reszt oraz obliczenia percentyli (fragment)

Obserwacja	Przewidywane X_1	Składniki resztowe $u_i = x_i - \hat{x}_i$	Std. składniki resztowe (σ_u)	Percentyl	X_1
1	11061,47751	1620,522491	0,95828372	0,347	10383
2	13590,82679	906,1732099	0,53585867	1,042	10866
3	14824,37863	898,6213715	0,53139295	1,736	10928
4	15476,74703	2058,252968	1,21713232	2,431	10986
5	16849,02726	1642,972736	0,9715595	3,125	11056
6	18432,48335	1065,516651	0,63008521	3,819	11112
7	19722,06943	2551,930569	1,50906484	4,514	11580
8	20438,81594	3227,18406	1,90837088	5,208	11685
9	23983,30608	1537,693924	0,90930367	5,903	12146
10	23775,46812	-2855,468122	-1,68855947	6,597	12341
:	:	:	:	:	:
140	18269,01376	-3011,013763	-1,78054021	96,875	31115
141	17979,67365	-1475,673646	-0,87262845	97,569	32722
142	18451,54125	326,4587519	0,19304891	98,264	36636
143	19719,46356	-10,46355639	-0,00618754	98,958	40817
144	21450,42641	-445,4264132	-0,26339954	99,653	44381

Źródło: Opracowanie własne w *Excelu*.

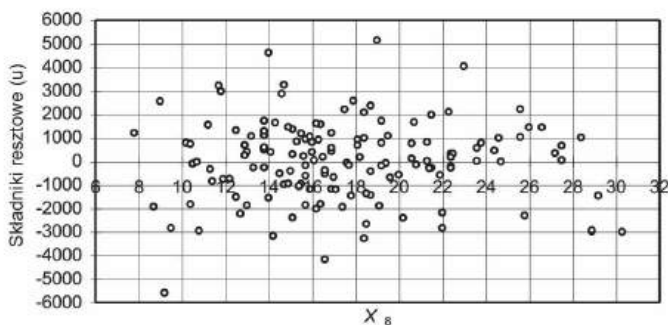
Oprócz ujęcia tabelarycznego percentyli dla zmiennej objaśnianej X_1 opcja *Regresja* programu *Excel* daje możliwość graficznego przedstawienia od 0-100% percentyli i odpowiadających im wartości X_1 . Przykładowo percentyl piąty to taka wartość poniżej której znajduje się 5%, a powyżej 95 % wszystkich obserwacji danej zbiorowości. Na rysunku 3.1 kółkami zaznaczono 144 obserwacje. Spróbowano również wyznaczyć funkcję wielomianową ułożenia się tych wartości w przedziale percentyli.



Źródło: Opracowanie własne w Excelu; gdzie: $y \rightarrow X_1$

Rys. 3.1. Struktura zbiorowości zmiennej X_1 - podział na pertencyle

W ramach opcji *Regresja* uzyskujemy również zestaw rozkładów punktowych reszt dla poszczególnych zmiennych. Przykład rozkładu składnika resztowego dla zmiennej X_8 podano na rysunku 3.2.

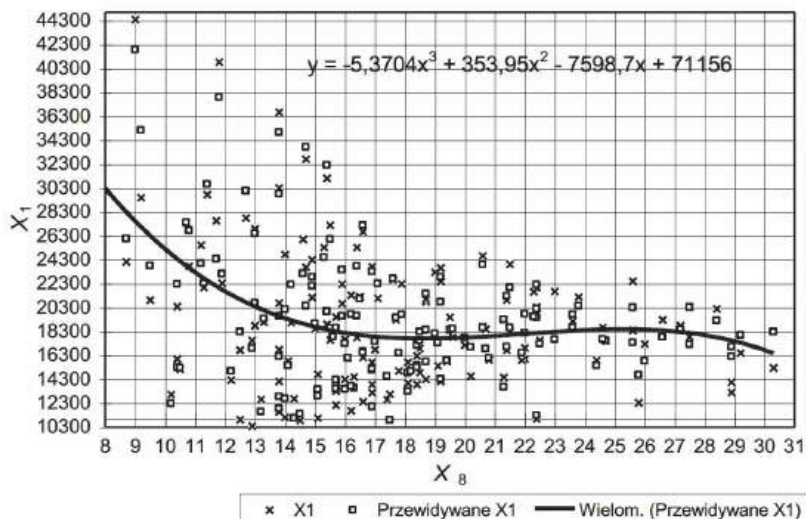


Źródło: Opracowanie własne w Excelu.

Rys. 3.2. Rozkład składnika resztowego zmiennej X_8

Na osi odciętej występuje wyszczególnienie wartości zmiennej X_8 , a na osi rzędnej są wartości reszt jako różnicy między wartością empiryczną a modelową $u_i = x_i - \bar{x}_i$

Rozkład linii dopasowanej (przykładowo dla zmiennej X_8)



Źródło: Opracowanie własne w Excelu; przy czym: $X_1 \rightarrow X_1$

Rys. 3.3. Rozkład linii dopasowanej zmiennej X_8

Na wykresie przedstawiono punkty stanowiące pary wartości empirycznych i modelowych zmiennej objaśnianej X_1 oraz przykładowej zmiennej objaśniającej X_8 . Dla wartości modelowych podjęto próbę wyznaczenia trendu wielomianowego trzeciego stopnia. Widzimy, że między tylko tymi zmiennymi występuje słabe dopasowanie modelu do danych empirycznych $R^2 \approx 11\%$, ale model ekonometryczny składa się z siedmiu zmiennych objaśniających, w wyniku oddziaływania których następuje poprawa współczynnika determinacji. Zwróćmy uwagę na raport *Statystyka regresji*, w którym $R^2 \approx 0,91$. Teraz zainteresujemy się raportem uzyskanym poprzez ścieżkę menu *Excela*:

Narzędzia/Analiza danych/Statystyka opisowa

Tab. 3.4. Statystyka opisowa zmiennych modelu PKB województw

Charakterystyka	X_1	X_5	X_6	X_8	X_{10}	X_{12}	X_{13}	X_{20}
Średnia	19143,59	7712,14	i	17,58	46,34	2079,53	1730,97	2716,26
Błąd standardowy	480,57	1014,99	59,47	0,41	3,42	29,89	21,76	174,66
Mediana	18360,00	3379,30	197,60	16,75	33,70	2010,39	1663,56	1904,99
Tryb (modalna)	#N/D!	4884,30	116,30	13,80	59,90	#N/D!	#N/D!	#N/D!
Odchylenie standardowe	5766,83	12179,88	713,65	4,86	41,09	358,65	261,10	2095,91
Wariancja próbki	33256287,29	148349520,93	509298,17	23,65	1688,47	128633,05	68171,00	4392854,08
Kurtoza	3,07	3,64	0,78	-0,16	25,49	0,69	5,30	2,29
Skośność	1,35	2,25	1,40	0,50	3,92	0,91	1,87	1,48

Charakterystyka	X_1	X_5	X_6	X_8	X_{10}	X_{12}	X_{13}	X_{20}
Zakres	33998,00	48639,10	2709,60	22,50	356,26	1740,73	1737,24	11440,79
Minimum	10383,00	399,20	36,00	7,80	11,00	1489,95	1305,01	558,40
Maksimum	44381,00	49038,30	2745,60	30,30	367,26	3230,68	3042,25	11999,19
Suma	2756677,00	1110547,50	83151,80	2531,20	6672,60	299452,54	249260,02	391141,33
Licznik	144,00	144,00	144,00	144,00	144,00	144,00	144,00	144,00
Największy(1)	44381,00	49038,30	2745,60	30,30	367,26	3230,68	3042,25	11999,19
Najmniejszy(1)	10383,00	399,20	36,00	7,80	11,00	1489,95	1305,01	558,40
Poziom ufności(95,0%)	949,94	2006,32	117,56	0,80	6,77	59,08	43,01	345,25

Źródło: Opracowanie własne w Excelu.

Uzyskany raport podaje podstawowe cechy charakterystyczne statystyki opisowej dla zmiennych omawianego modelu ekonometrycznego, które ze względu na występowanie również w innych pakietach programowych przypominamy sobie:

- średnia arytmetyczna;
- błąd standardowy oszacowania średniej na poziomie ufności 95%;
- mediana, czyli wartość poniżej której znajduje się 50% obserwacji;
- modalna (wartość najczęściej występująca); symbolem #N/D! zaznaczono nie występowanie modalnej dla danej zmiennej;
- odchylenie standardowe, jako pierwiastek z wariancji próby;
- wariancja próby, stanowiąca sumę kwadratów odchyleń wartości empirycznych od ich średniej, którą dzielimy przez $(n - 1)$;
- kurtosa K charakteryzująca względną szczytowość lub płaskość rozkładu empirycznego względem rozkładu normalnego; kurtosa dodatnia świadczy o szczytowości, a ujemna o płaskości rozkładu;
- skośność, a raczej współczynnik asymetrii rozkładu.

Dla naszego przykładu dla zmiennych modelu występuje asymetria prawostronna, czyli rozciąganie się „ogona” w kierunku do strony prawej, czyli wartości dodatnich osi odciętych. Ponadto w tabeli statystyki opisowej występują miary:

- zakres, stanowiący różnicę między wartością maksymalną a minimalną danej zmiennej;
- licznik (liczba obserwacji);
- największy/najmniejszy; po podaniu liczby wskazuje wartość obserwacji w kolejności największej, czy też najmniejszej w szeregu szczegółowym uporządkowanym narastająco;
- poziom ufności (95%), który jest kryterium na określenie przedziału ufności w zbiorowości generalnej w stosunku do obliczonej wartości średniej danej zmiennej modelu.

Macierz współczynników korelacji liniowej Pearsona uzyskujemy korzystając z menu:

Narzędzia/Analiza danych/Korelacja

Współczynnik korelacji obliczany jest według znanego nam już wzoru:

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

Tab. 3.5. Macierz współczynników korelacji zmiennych modelu

	X ₁	X ₅	X ₆	X ₈	X ₁₀	X ₁₂	X ₁₃	X ₂₀
X ₁	1							
X ₅	0,2435	1						
X ₆	0,5130	-0,0348	1					
X ₈	-0,2680	-0,1681	-0,2623	1				
X ₁₀	0,4593	0,4112	0,5455	-0,4572	1			
X ₁₂	0,8701	0,5133	0,3050	-0,3157	0,4431	1		
X ₁₃	0,8989	0,1515	0,4966	-0,4152	0,4912	0,8326	1	
X ₂₀	0,6866	0,6840	0,4311	-0,4838	0,7198	0,7223	0,6200	1

Źródło: Opracowanie własne w Excelu.

Jest to macierz symetryczna kwadratowa, stąd z prawej strony komputer nie drukuje wartości współczynnika korelacji liniowej Pearsona.

Zastosowanie menu *Narzędzia/Analiza danych/Kowariancja* daje zestawienie kowariancji między zmiennymi modelu. Kowariancja obliczana jest według ogólnego wzoru:

$$cov_{xy} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

Raport stanowi też macierz symetryczną kwadratową współczynników kowariancji między zmiennymi modelu. Na głównej przekątnej tej macierzy są wariancje próby dla określonych zmiennych.

Tab. 3.6. Macierz współczynników kowariancji zmiennych modelu

	X ₁	X ₅	X ₆	X ₈	X ₁₀	X ₁₂	X ₁₃	X ₂₀
X ₁	33256287,29							
X ₅	17100674,01	148349520,93						
X ₆	2111399,28	-302534,69	509298,17					
X ₈	-7516,40	-9960,34	-910,52	23,65				
X ₁₀	108840,46	205802,45	15996,05	-91,38	1688,47			
X ₁₂	1799680,92	2242311,61	78053,38	-550,76	6529,86	128633,05		
X ₁₃	1353436,55	481722,81	92535,09	-527,19	5270,41	77971,41	68171,00	
X ₂₀	8299189,62	17459956,22	644860,78	-4931,24	61990,11	542949,42	339261,87	4392854,08

Źródło: Opracowanie własne w Excelu.

Graficzną prezentację występowania określonych przedziałów (klas) danej zmiennej uzyskamy z menu: *Narzędzia/Analiza danych/Histogram*, przy czym w oknie dialogowym definiujemy:

- Wejście: zakres kolumny zmiennej
- Tytuły: zaznaczenie (v)
- Opcja wyjścia:
 - Nowy arkusz
 - Pareto (posortowany histogram)
 - Łączny udział procentowy
 - Wykres wyjściowy

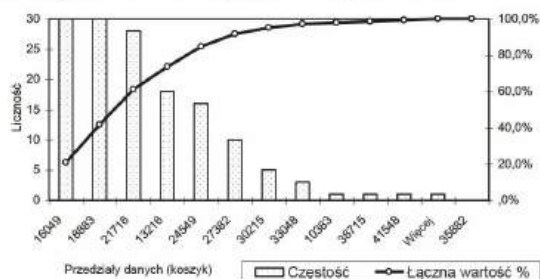
Procedurę wykonujemy odrębnie dla każdej zmiennej modelu. Przykład dla X₁ podano w tabeli 3.7.

Tab. 3.7. Przedziały zmiennej X_1

Przedział	Liczność	Łączna wartość %	Przedział	Liczność	Łączna wartość %
10383,0	1	0,69	16049,3	30	20,83
13216,2	18	13,19	18882,5	30	41,67
16049,3	30	34,03	21715,7	28	61,11
18882,5	30	54,86	13216,2	18	73,61
21715,7	28	74,31	24548,8	16	84,72
24548,8	16	85,42	27382,0	10	91,67
27382,0	10	92,36	30215,2	5	95,14
30215,2	5	95,83	33048,3	3	97,22
33048,3	3	97,92	10383,0	1	97,92
35881,5	0	97,92	38714,7	1	98,61
38714,7	1	98,61	41547,8	1	99,31
41547,8	1	99,31	Więcej	1	100,00
Więcej	1	100,00	35881,5	0	100,00
	144			144	

Źródło: Opracowanie własne w Excelu.

Oprócz przedziałów wartości danej zmiennej uzyskujemy również histogram z zaznaczonym na nim wykresem *Pareto*. Przykład dla zmiennej X_1 stanowi rysunek 3.4.



Źródło: Opracowanie własne w Excelu.

Rys. 3.4. Histogram i wykres Pareto zmiennej X_1

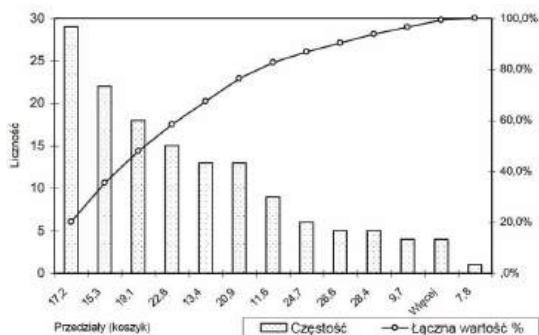
Wykonajmy jeszcze przykład dla zmiennej X_8 .

Tab. 3.8. Przedziały wartości zmiennej X_8

Przedział	Liczność	Łączna wartość %	Przedział	Liczność	Łączna wartość %
7,8	1	0,69	17,2	29	20,14
9,7	4	3,47	15,3	22	35,42
11,6	9	9,72	19,1	18	47,92
13,4	13	18,75	22,8	15	58,33
15,3	22	34,03	13,4	13	67,36
17,2	29	54,17	20,9	13	76,39
19,1	18	66,67	11,6	9	82,64
20,9	13	75,69	24,7	6	86,81
22,8	15	86,11	26,6	5	90,28

Przedział	Liczność	Łączna wartość %	Przedział	Liczność	Łączna wartość %
24,7	6	90,28	28,4	5	93,75
26,6	5	93,75	9,7	4	96,53
28,4	5	97,22	Więcej	4	99,31
Więcej	4	100,00	7,8	1	100,00
144			144		

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne w Excelu.

Rys. 3.5. Histogram i wykres Pareto zmiennej X_8

Ponadto stosując ścieżkę menu: *Narzędzia/Analiza danych/Ranga i percentyl* można przykładowo otrzymać dla zmiennej X_8 określenie dla każdej obserwacji jej rangi oraz wartości percentylu w strukturze zbiorowości próbnej. W tym względzie dla naszego przykładu podajemy następujące informacje:

Wejście:

Zakres wejściowy np. \$F\$1:\$F\$145

Grupowanie wg kolumn

Tytuły w pierwszym wierszu

Opcja wyjścia: Nowy arkusz

Raport jest długi i z tego względu zamieszczono jego początkowy i końcowy fragment.

Tab. 3.9. Nadanie rang i określenie percentyli

Obserwacja	X_8	Ranga	Percentyl
140	30,3	1	100,0%
141	29,2	2	99,3%
138	28,9	3	97,9%
139	28,9	3	97,9%
95	28,4	5	97,2%
77	27,5	6	95,8%
96	27,5	6	95,8%
:	:	:	:
63	10,4	137	4,1%
19	10,2	139	3,4%
10	9,5	140	2,7%

Obserwacja	X_8	Ranga	Percentyl
36	9,2	141	2,0%
18	9,0	142	1,3%
27	8,7	143	0,6%
90	7,8	144	0,0%

Źródło: Opracowanie własne w Excelu.

3.3. STATGRAPHICS

Pakiet *STATGRAPHICS* jest oprogramowaniem licencyjnym. Dla celów testowych i edukacyjnych można go pozyskać z Internetu na okres 30 dni. Firma umożliwia zainstalowanie wersji testowej ze swego serwera po podaniu danych adresowych użytkownika i adresu *e-mail*. Na ten adres przesyłany jest *kod aktywacji*. Celem rozpoczęcia procesu instalowania oprogramowania, po zainstalowaniu programu na dysku *C* własnego komputera możemy przystąpić do wykonania raportów statystyki opisowej zarówno ilościowej jak i graficznej. Przykładowa procedura uzyskania statystyki opisowej dla zmiennych przykładowego modelu ekonometrycznego *PKB* województw w *STATGRAPHICS* jest następująca:

1. Kopiowanie tylko szeregów zmiennych o nazwach realizacji komputerowej tj. X_1 , X_5 , X_6 , X_8 , X_{10} , X_{12} , X_{13} , X_{20} z arkusza kalkulacyjnego Excela jako pliku *GRETL.xls*. Zaznaczenie kolumn wszystkich zmiennych, a następnie wywołanie menu: *Edycja/Kopiuj*.

2. Wklejenie do arkusza programu *STATGRAPHICS*, po zaznaczeniu tylko pierwszej komórki i stosując menu: *Edit/Paste*. Zapisanie jako plik o nazwie *Dane.sgd*.

3. Zmiana nazw proponowanych przez program standardowych kolumn i dodanie opisu zmiennej poprzez menu: *Edit/Modify Column*.

4. Zaznaczenie kolumn ze zmiennymi X_1 , ..., X_{20}

Zażądanie statystyki opisowej dla zestawu zmiennych modelu ekonometrycznego wywołując menu:

Describe/NumericData/Multi-Variable Analysis ...

Następnie wypełniamy okno dialogowe *DATA* zaznaczając jako dane zmienne X_2 , ..., X_{20} i *OK*. Pojawia się kolejne okno *Tables and Graphs* (Tabele i grafiki). Program proponuje jako domyślne następujące tabele:

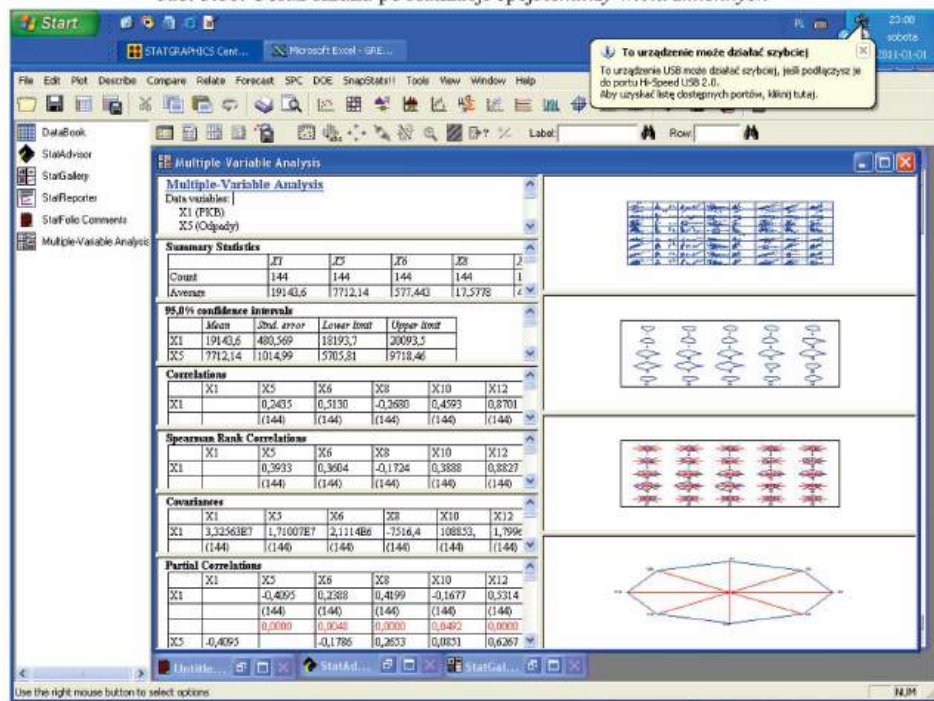
- analizy sumarycznej (*Analysis Summary*),
- statystyki sumarycznej (*Summary Statistics*),
- korelacji (*Correlation*).

Natomiast standardową grafiką jest macierz wykresów rozrzutu (*Scatterplot Matrix*) dla wszystkich zmiennych modelu ekonometrycznego. Jednak dla poznania szerszych możliwości *STATGRAPHICS* zaznaczamy wszystkie opcje w odniesieniu do tabel oraz grafiki. Tak więc zawartość okna z określonymi opcjami wyboru, w tym *korelacji rang Spearmana* jest następująca:

TABLES	GRAPHS
<i>Analysis Summary</i>	<i>Scatterplot Matrix</i>
<i>Summary Statistics</i>	<i>Summary Plots</i>
<i>Confidence Intervals</i>	<i>Key Glyph</i>
<i>Correlation</i>	
<i>Rang Correlation</i>	
<i>Covariances</i>	
<i>Partial Correlations</i>	

Po naciśnięciu klawisza *ENTER* uzyskano sumaryczne zestawienie tabel oraz grafiki, które następnie skopiowano do pliku *D:\Programy-stat.doc* edytora *Word*. Obraz ekranu ze zestawieniem sumarycznym skopiowanym do schowka systemowego klawiszem *PrSc* prezentuje tabela 3.10

Tab. 3.10. Obraz ekranu po realizacji opcji *Analizy wielu zmiennych*



Źródło: Opracowanie własne w *STATGRAPHICS*.

Analiza wielu zmiennych przedstawiona jednocześnie na ekranie obejmuje z jednej strony opis sumaryczny na sześciu tabelach, a z drugiej strony cztery zestawy grafik dotyczących zmiennych. Kliknięcie na tabelę wywołuje zaznaczenie jej całości, po czym możemy ją skopiować w powiększeniu do dokumentu *Word*. Podwójne kliknięcie na grafice stanowiącej zbiór wykresów wywołuje dodatkowe okno którego obraz możemy skopiować do naszego dokumentu *Word*. Ponowne podwójne kliknięcie powoduje powrót do grafiki zbiorczej - macierzy.

Analiza wielu zmiennych jednocześnie obsługiwana jest przez procedurę *StatAdvisor*. Oprócz wyników daje również *Komentarze* w języku angielskim, które tu ze względu na ich obszerność i szczegółowość opisu nie zamieszczono. Dla potrzeb procedury *StatAdvisor* wprowadzono opisy poszczególnych zmiennych modelu ekonometrycznego, a mianowicie:

- X1 - PKB
- X5 - Odpady
- X6 - Ścieki
- X8 - Stopa
- X10 - P-bud
- X12 - W-przem
- X13 - W-bud

X20 - I-przem.

W tabeli 3.11 opisy te zamieszczone są jako wstępne przed danymi ilościowymi. Jak już wspomniano na obrazie ekranu analizy wielu zmiennych występują tabele W tym *Summary Statistics*, którą w powiększeniu zamieszczono w tabeli 3.11.

Tab. 3.11. Statystyka sumaryczna (*Summary Statistics*)

	X1	X5	X6	X8	X10	X12	X13	X20
Count	144	144	144	144	144	144	144	144
Average	19143,6	7712,14	577,443	17,5778	46,3368	2079,53	1730,97	2716,26
Standard deviation	5766,83	12179,9	713,651	4,86359	41,0948	358,655	261,096	2095,91
Coeff. of variation	30,1241%	157,931%	123,588%	27,6689%	88,6871%	17,2469%	15,0838%	77,1616%
Minimum	10383,0	399,2	36,0	7,8	11,0	1489,95	1305,01	558,4
Maximum	44381,0	49038,3	2745,6	30,3	367,3	3230,68	3042,25	11999,2
Range	33998,0	48639,1	2709,6	22,5	356,3	1740,73	1737,24	11440,8
Std. skewness	6,58917	11,019	6,87825	2,44757	19,1943	4,44081	9,15037	7,24586
Std. kurtosis	7,51286	8,92362	1,91123	-0,388053	62,4413	1,67988	12,9851	5,59945

To make the variables more normal, you might try a transformation such as LOG(Y), SQRT(Y), or 1/Y.

Źródło: Opracowanie własne w *STATGRAPHICS*.

Tabela 3.11 pokazuje statystykę sumaryczną dla każdej zmiennej modelu ekonometrycznego. Czcionka pochyłą zaznaczono standaryzowany współczynnik skośności (*Std. Skewness*) oraz standaryzowany współczynnik kurtosis (*Std. Kurtosis*) obliczony dla przedziału poza (-2; 2). Pod tabelą znajduje się komentarz programu *STATGRAPHICS* proponujący zastosowanie transformacji modelu jako: LOG(Y), SQRT(Y) lub 1/Y, przy czym Y to u nas zmienna X1, w celu doprowadzenia do większej normalności zmiennych. Sytuacja taka wystąpiła być może ze specyficznego charakteru zmiennych jako dane panelowe. W kolejnej tabeli korelacje (*Correlations*) dla każdej relacji zmiennych występują trzy informacje:

- współczynnik korelacji liniowej Pearsona,
- licznosc próby,
- prawdopodobieństwo popełnienia błędu w oszacowaniu.

Tab. 3.12. Korelacje (*Corelations*)

	X1	X5	X6	X8	X10	X12	X13	X20
X1		0,2435	0,5130	-0,2680	0,4593	0,8701	0,8989	0,6866
		(144)	(144)	(144)	(144)	(144)	(144)	(144)
		0,0033	0,0000	0,0012	0,0000	0,0000	0,0000	0,0000
X5	0,2435		-0,0348	-0,1681	0,4112	0,5133	0,1515	0,6840
	(144)		(144)	(144)	(144)	(144)	(144)	(144)
	0,0033		0,6788	0,0440	0,0000	0,0000	0,0699	0,0000
X6	0,5130	-0,0348		-0,2623	0,5454	0,3050	0,4966	0,4311
	(144)	(144)		(144)	(144)	(144)	(144)	(144)
	0,0000	0,6788		0,0015	0,0000	0,0002	0,0000	0,0000
X8	-0,2680	-0,1681	-0,2623		-0,4572	-0,3157	-0,4152	-0,4838

	X1	X5	X6	X8	X10	X12	X13	X20
	(144)	(144)	(144)		(144)	(144)	(144)	(144)
	0,0012	0,0440	0,0015		0,0000	0,0001	0,0000	0,0000
X10	0,4593	0,4112	0,5454	-0,4572		0,4431	0,4912	0,7198
	(144)	(144)	(144)	(144)		(144)	(144)	(144)
	0,0000	0,0000	0,0000	0,0000		0,0000	0,0000	0,0000
X12	0,8701	0,5133	0,3050	-0,3157	0,4431		0,8326	0,7223
	(144)	(144)	(144)	(144)	(144)		(144)	(144)
	0,0000	0,0000	0,0002	0,0001	0,0000		0,0000	0,0000
X13	0,8989	0,1515	0,4966	-0,4152	0,4912	0,8326		0,6200
	(144)	(144)	(144)	(144)	(144)	(144)		(144)
	0,0000	0,0699	0,0000	0,0000	0,0000	0,0000		0,0000
X20	0,6866	0,6840	0,4311	-0,4838	0,7198	0,7223	0,6200	
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	
	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	

Źródło: Opracowanie własne w STATGRAPHICS.

Współczynnik korelacji rang Spearmana [-1; 1] dla każdych dwóch cech uporządkowanych malejąco obliczany jest według wzoru⁵²:

$$r_s = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)}$$

Gdzie: d_i - różnica między rangami odpowiadających sobie wartości cechy x_i i cech y_i ($i = 1, 2, \dots, n$). Korelacje rang Spearmana z podaniem trzech informacji analogicznie jak dla korelacji (zob. rysunek 3.12) przedstawiono w tabeli 3.13.

Tab. 3.13. Korelacje rang Spearmana (*Spearman Rang Correlation*)

	X1	X5	X6	X8	X10	X12	X13	X20
X1		0,3933	0,3604	-0,1724	0,3888	0,8827	0,8486	0,6360
		(144)	(144)	(144)	(144)	(144)	(144)	(144)
		0,0000	0,0000	0,0392	0,0000	0,0000	0,0000	0,0000
X5	0,3933		0,6356	-0,3284	0,7912	0,4551	0,2825	0,7448
	(144)		(144)	(144)	(144)	(144)	(144)	(144)
	0,0000		0,0000	0,0001	0,0000	0,0000	0,0007	0,0000
X6	0,3604	0,6356		-0,3141	0,7289	0,3777	0,3538	0,6099
	(144)	(144)		(144)	(144)	(144)	(144)	(144)
	0,0000	0,0000		0,0002	0,0000	0,0000	0,0000	0,0000
X8	-0,1724	-0,3284	-0,3141		-0,5118	-0,2438	-0,3384	-0,4496
	(144)	(144)	(144)		(144)	(144)	(144)	(144)
	0,0392	0,0001	0,0002		0,0000	0,0036	0,0001	0,0000
X10	0,3888	0,7912	0,7289	-0,5118		0,3787	0,3693	0,8865
	(144)	(144)	(144)	(144)		(144)	(144)	(144)
	0,0000	0,0000	0,0000	0,0000		0,0000	0,0000	0,0000

⁵² Praca zbiorowa, *Statystyka Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006, s. 338.

	X1	X5	X6	X8	X10	X12	X13	X20
X12	0,8827	0,4551	0,3777	-0,2438	0,3787		0,8534	0,5720
	(144)	(144)	(144)	(144)	(144)		(144)	(144)
	0,0000	0,0000	0,0000	0,0036	0,0000		0,0000	0,0000
X13	0,8486	0,2825	0,3538	-0,3384	0,3693	0,8534		0,5663
	(144)	(144)	(144)	(144)	(144)	(144)		(144)
	0,0000	0,0007	0,0000	0,0001	0,0000	0,0000		0,0000
X20	0,6360	0,7448	0,6099	-0,4496	0,8865	0,5720	0,5663	
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	
	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	0,0000	

Źródło: Opracowanie własne w STATGRAPHICS.

Tabela 3.14. pokazuje kowariancje między parami zmiennych. Gdy parę stanowi ta sama zmienna to obliczona wartość jest wariancją. Natomiast pierwiastek z wariancji jest odchyleniem standardowym (patrz tabela 3.11 „Statystyka sumaryczna...”).

Tab. 3.14. Kowariancje (Covariances)

	X1	X5	X6	X8	X10	X12	X13	X20
X1	3,32563E7	1,71007E7	2,1114E6	-7516,4	108853,	1,79968E6	1,35344E6	8,29917E6
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X5	1,71007E7	1,4835E8	-302535,	-9960,34	205838,	2,24231E6	481723,	1,746E7
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X6	2,1114E6	-302535,	509298,	-910,525	15996,2	78053,4	92535,1	644859,
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X8	-7516,4	-9960,34	-910,525	23,6545	-91,3718	-550,758	-527,195	-4931,23
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X10	108853,	205838,	15996,2	-91,3718	1688,78	6530,9	5270,86	61994,7
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X12	1,79968E6	2,24231E6	78053,4	-550,758	6530,9	128633,	77971,4	542949,
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X13	1,35344E6	481723,	92535,1	-527,195	5270,86	77971,4	68171,0	339261,
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)
X20	8,29917E6	1,746E7	644859,	-4931,23	61994,7	542949,	339261,	4,39285E6
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	(144)

Gdzie: (144) - liczność próby.

Źródło: Opracowanie własne w STATGRAPHICS.

Program STATGRAPHICS pokazuje także tzw. częściową korelację (patrz tabela 3.15).

Tab. 3.15. Częściowa korelacja (Partial Correlations)

	X1	X5	X6	X8	X10	X12	X13	X20
X1		-0,4095	0,2388	0,4199	-0,1677	0,5314	0,3870	0,4838
		(144)	(144)	(144)	(144)	(144)	(144)	(144)
		0,0000	0,0048	0,0000	0,0492	0,0000	0,0000	0,0000
X5	-0,4095		-0,1786	0,2653	0,0851	0,6267	-0,3537	0,6693
		(144)	(144)	(144)	(144)	(144)	(144)	(144)
	0,0000		0,0361	0,0017	0,3208	0,0000	0,0000	0,0000

	X1	X5	X6	X8	X10	X12	X13	X20
X6	0,2388	-0,1786		0,0290	0,3613	-0,1870	0,0230	0,1130
	(144)	(144)		(144)	(144)	(144)	(144)	(144)
	0,0048	0,0361		0,7361	0,0000	0,0281	0,7887	0,1869
X8	0,4199	0,2653	0,0290		-0,0594	-0,1159	-0,2923	-0,4228
	(144)	(144)	(144)		(144)	(144)	(144)	(144)
	0,0000	0,0017	0,7361		0,4891	0,1758	0,0005	0,0000
X10	-0,1677	0,0851	0,3613	-0,0594		-0,1033	0,2195	0,3645
	(144)	(144)	(144)	(144)		(144)	(144)	(144)
	0,0492	0,3208	0,0000	0,4891		0,2277	0,0097	0,0000
X12	0,5314	0,6267	-0,1870	-0,1159	-0,1033		0,4523	-0,1981
	(144)	(144)	(144)	(144)	(144)		(144)	(144)
	0,0000	0,0000	0,0281	0,1758	0,2277		0,0000	0,0198
X13	0,3870	-0,3537	0,0230	-0,2923	0,2195	0,4523		-0,0168
	(144)	(144)	(144)	(144)	(144)	(144)		(144)
	0,0000	0,0000	0,7887	0,0005	0,0097	0,0000		0,8451
X20	0,4838	0,6693	0,1130	-0,4228	0,3645	-0,1981	-0,0168	
	(144)	(144)	(144)	(144)	(144)	(144)	(144)	
	0,0000	0,0000	0,1869	0,0000	0,0000	0,0198	0,8451	

Źródło: Opracowanie własne w STATGRAPHICS.

Wartość prawdopodobieństwa (*P-values*) poniżej 0,05 świadczy o uznawanej za zerową korelację przy poziomie ufności $(1 - \alpha) = 0,95$, tj. 95%). Dla par zmiennych, gdzie *P-value* jest poniżej 0,05 zaznaczono czcionką pochylą. Z tabeli 3.15. wynika, że następujące pary zmiennych mają *P-values* poniżej 0,05:

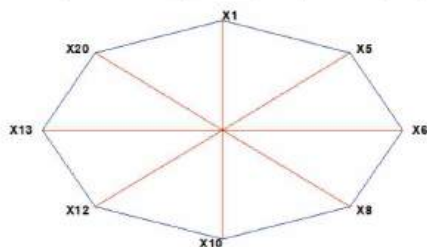
X1	X5	X5	X13
X1	X6	X5	X20
X1	X8	X6	X10
X1	X10	X6	X12
X1	X12	X8	X13
X1	X13	X8	X20
X1	X20	X10	X13
X5	X6	X10	X20
X5	X8	X12	X13
X5	X12	X12	X20

Zaznaczenie w *GRAPH* opcji *Scatterplot Matrix* i potwierdzenie daje raport graficzny macierzy zmiennych X1, X5, X6, X8, X10, X12, X13, X20 modelu ekonometrycznego. W tej macierzy dla każdej zmiennej występuje wykres pudełkowy oraz dla każdej pary zmiennych wykres rozrzutu w układzie X1-X_n. Pozostaje jednak potrzeba skopiowania i znacznego powiększenia obrazu w celem jego wydruku na dużym formacie.

Kolejna grafika nazwana *Star Plots* daje obraz 25. obszarów. Metoda *Star Plots* reprezentuje obserwacje wielowymiarowe o dowolnej liczbie zmiennych. Poszczególne obserwacje opisane są przez zmienną objaśnianą X₁ oraz zmienne objaśniające X₅, X₆, X₈, X₁₀, X₁₂, X₁₃, X₂₀. Obszar danej obserwacji przypominający gwiazdę powstaje przez odmierzenie długości zestandaryzowanych

zmiennych na promieniach wychodzących z jednego centralnego punktu w kierunku „rogów” danej gwiazdy. Reprezentuje to dobrze następną grafikę *Sunray Plots*, przy czym promienie dochodzą tylko do rogów gwiazdy i są przezroczyste. Kolejność ułożenia zmiennych w danym obszarze podaje grafika *Key Glyph*. Obszary (kolejne obserwacje wielowymiarowe) zestawione są w prostokącie.

Grafika *Key Glyph* wykonana jest dla wszystkich zmiennych modelu tj. $X_1, \dots, X_{20}$. Okręg podzielono na osiem części (tyle co zmiennych) tworząc siatkę wielu osi o tej samej podziałce wartości zmiennych po ich standaryzacji. Na osiach odmierzamy wartość danej zmiennej, a następnie łączymy końce poszczególnych zmiennych. Występuje tu analogia do wykresu liniowego biegunowego zwanego również radarowym lub radiogramem, który możemy wykonać w *Excelu*.



Źródło: Opracowanie własne w *STATGRAPHICS*.

Rys. 3.6. Wykres biegunowy zestandaryzowanych zmiennych (*Key Glyph*)

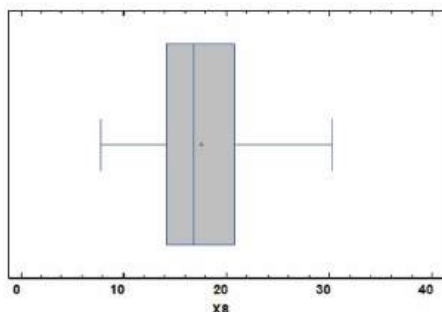
Program *STATGRAPHICS* umożliwia również wykonywanie statystyk ilościowych i graficznych dla pojedynczych zmiennych. W tym celu wykonano przykład standardowy dla jednej zmiennej tj. X_8 korzystając z menu: *Describe/Numeric Data/One-Variable Analysis ...* obejmującej opcje: *TABLES: Analysis Summary, Summary Statistics* oraz *GRAPH: Scatterplot, Box-and-Whisker Plot*. Powyższe raporty zapisano w przykładowym pliku: *D:\... \Programy-stat.doc*. Program na początku informuje nas o opcji i zastosowanej zmiennej: *One-Variable Analysis- X_8 ; Data variable X_8 (Stopa); 144 values ranging from 7,8 to 30,3* (przedział wartości od 7,8 do 30,3). Przykładową statystykę sumaryczną dla zmiennej X_8 podaje tabela 3.16.

Tab. 3.16. Statystyka sumaryczna (*Summary Statistics for X_8*)

<i>Count</i>	144
<i>Average</i>	17,5778
<i>Standard deviation</i>	4,86359
<i>Coeff. of variation</i>	27,6689%
<i>Minimum</i>	7,8
<i>Maximum</i>	30,3
<i>Range</i>	22,5
<i>Std. skewness</i>	2,44757
<i>Std. kurtosis</i>	-0,388053

Źródło: Opracowanie własne w *STATGRAPHICS*.

Ponadto program rysuje wykres pudełkowy między wartością $x_{min} = 7,8$, a wartością $x_{max} = 30,3$. Na podstawie zaznaczonej osi wartości możemy oszacować miary pozycyjne statystyki opisowej takie jak: kwartył pierwszy $Q_1 \approx 14$, mediana $Me \approx 16,5$, kwartył trzeci ≈ 21 .



Źródło: Opracowanie własne w *STATGRAPHICS*.

Rys. 3.7. Wykres pudełkowy zmiennej X8 (*Box-and-Whisker Plot*)

3.4. Środowisko R

Jak już nadmieniałem pierwszym etapem jest zainstalowanie *Base* wraz z niezbędnymi bibliotekami⁵³. Umożliwia to realizację podstawowych funkcji w zakresie analizowania danych oraz rysowania wykresów. Kolejne etapy to doinstalowanie pakietów z funkcjami rozszerzającymi wersję podstawową *Base*. Przykładowo dla wersji *Windows 95* pakiet *R* jest dostępny w postaci źródłowej oraz skompilowanej. Kopię pakietu instalacyjnego ściągamy poprzez Internet z serwera tzw. *mirror*, najlepiej zlokalizowanego w ośrodku w pobliżu komputera użytkownika (patrz <http://cran.r-project.org/mirrors.html>).

W kolejności uruchamiamy plik *R-2.7.1-win32.exe* i klikając *Next* doprowadzamy do utworzenia na własnym komputerze katalogu *C:\Program Files\R\R-2.7.1* struktury podkatalogów⁵⁴. Jednym z podkatalogów jest *bin* z plikami wykonywalnymi. Kolejny *doc* zawiera dokumentację pakietu *R*. W podkatalogu *library* instalowane są dalsze pakiety programowe funkcją *install.packages (...)*. Środowisko programowe *R*, zwane też *platformą R* można uruchomić w trzech trybach:

- tekstowym (plik *R.exe*), gdy realizujemy większe symulacje i niekonieczny jest interfejs graficzny;

- okienkowym (plik *Rgui.exe*), który zalecany jest jako wiodący;

- „nadzorowanym” (*windowśowym*) po dodatkowym załadowaniu pakietu *Rcmdr* wraz ze wszystkimi pakietami pomocniczymi niezbędnymi do jego działania stosując instrukcję:

install.packages („Rcmdr”,dependencies=TRUE).

Wróćmy jednak do przygotowania danych, które umożliwią nam później przetestowanie otrzymania raportów liczbowych i graficznych statystyki opisowej. W tym względzie skorzystamy ze zbioru na podstawie którego opracowano model ekonometryczny *PKB województw*⁵⁵. Procedura wstępna do importu z *Excelsa* zmiennych *X1*, *X5*, *X6*, *X8*, *X10*, *X12*, *X13* oraz *X20* do środowiska *R* jest następująca:

- skopiowanie z pliku (skoroszytu) *GRETl.xls* danych z arkusza *Model* w zakresie tylko wymienionych zmiennych,

- wklejenie do nowego skoroszytu i zapisanie w *D:\R\Dane.xls*,

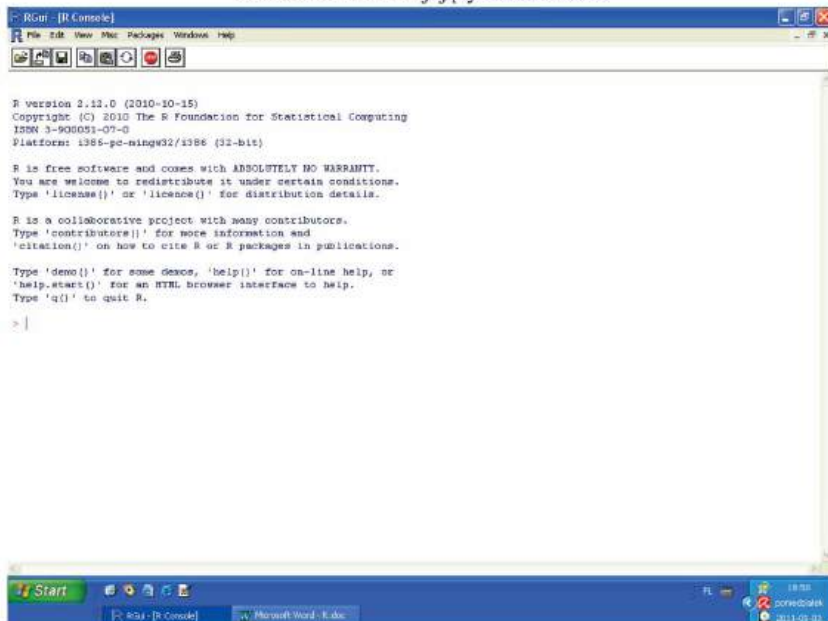
⁵³ Biecek P., *Przewodnik po pakiecie R*, Oficyna Wydawnicza GiS, Wrocław 2008.

⁵⁴ Podana procedura dotyczy czasu niniejszego opracowania.

⁵⁵ Wornalkiewicz Wł., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, rozdz. 8, Wydawnictwo Instytut Śląski, Opole 2013.

- pozostawienie tylko jednego arkusza w ramach nowego skoroszytu,
- nadanie nazwy arkuszowi *Dane1*.

Tab. 3.17. Ekran inicjujący środowisko *R*



Źródło: Opracowanie własne pakietem *R*.

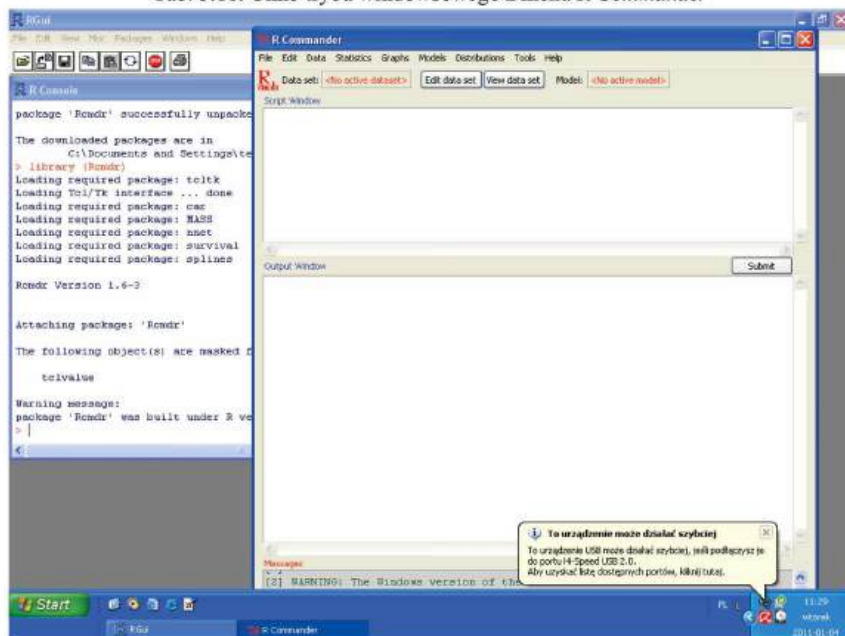
Po zainstalowaniu *R* ukazuje się jego ikona *R 2.12.0* na pulpicie (ekranie) komputera. Kliknięcie na tą ikonę wywołuje pojawienie się ekranu inicjującego *Rgui [R Console]*, którego obraz uzyskany poprzez klawisz *PrintScr* przedstawiono w tabeli 3.17. Występujący znak zachęty „>” oznacza, że środowisko programowe oczekuje na polecenie. Zwróćmy uwagę na menu główne tego trybu obejmujące rozwijalne opcje:

File, Edit, Views, Misc, Packages, Windows, Help.

Dla początkujących użytkowników *R* dogodna jest wersja *windowsowa*. Oprócz wywołania niezbędnego pakietu *Rcmdr* poprzez komendę *install* (o czym już wspomniano) istnieje alternatywna możliwość skorzystania z menu: *Packages/install package(s)*...Podajemy nazwę serwera (*mirror*) - *Poland Wrocław*, a następnie wybieramy pakiet *Rcmdr*. Po pewnym czasie pojawia się okno *R Commander* (patrz tabela 3.18), które umożliwia nam wybór odpowiedniej funkcji z menu⁵⁶.

⁵⁶ Walesiak M., Gatnar E., *Statystyczna analiza danych z wykorzystaniem pakietu R*, Wydawnictwo Naukowe PWN, Warszawa 2009, rozdz. 1.8.

Tab. 3.18. Okno trybu windowsowego z menu *R Commander*



Źródło: Opracowanie własne pakietem *R*.

Skorzystanie jednak z takiego udogodnienia wymaga załadowania *Rcmdr* do pamięci operacyjnej poprzez komendę: *library(Rcmdr)*. Menu główne pakietu *Rcmdr* oprócz opcji *File*, *Edit*, *Tools*, *Help* obejmuje ponadto interesujące nas z punktu widzenia statystyki opisowej następujące opcje:

- *Data* (założenie zbioru, załadowanie istniejących, import z ...);
- *Statistics* (podstawowa statystyka opisowa, testy sumaryczne, ...);
- *Graphs* (możliwości wykonania wykresów różnego typu dla pojedynczej zmiennej jak i grupowych);
- *Models* (analiza regresji);
- *Distributions* (tablice statystyczne).

Zwróćmy ponownie uwagę na tabelę 3.18, gdyż występują na ekranie *R Commander* trzy podokna opisowe:

- skryptowe (*Script Windows*) na którym pojawiają się polecenia odpowiadające naszym działaniom w menu głównym; stanowi to również samouczek komend, które później bez menu możemy wpisywać do tej części skryptowej celem ich bezpośredniej realizacji;
- realizacji (*Output Windows*) w której wyświetlane są wyniki/raporty z realizacji przez komputer danego polecenia;
- komunikatów (*Messages*), zawierające informacje o błędach, komentarze do zbiorów.

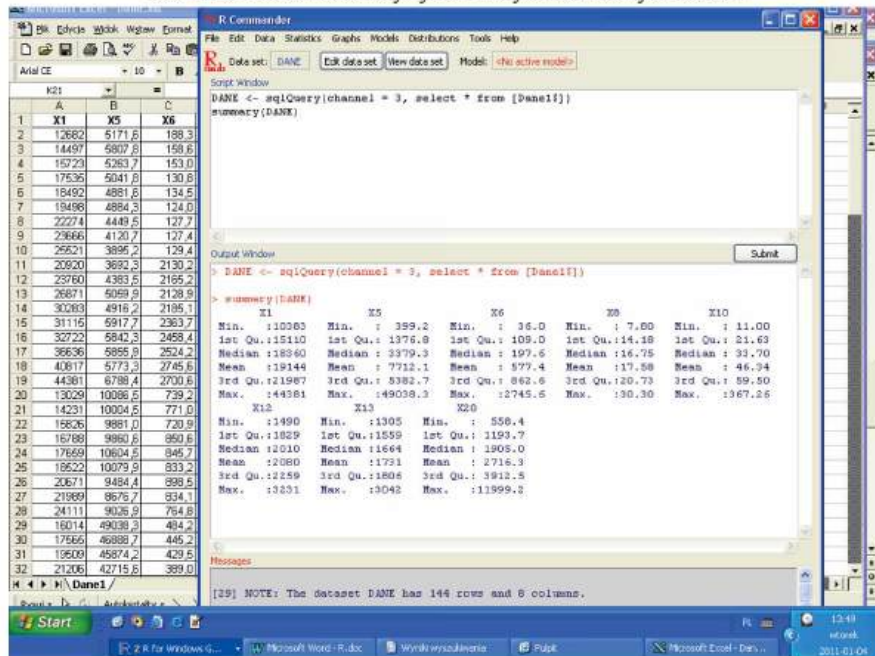
Mamy już aktywny tryb „nadzorowany” korzystania z funkcji środowiska *R*, możemy zatem dokonać importu pliku *Dane.xls*, traktowanego jako obiekt „tabela danych”. W tym względzie korzystamy z menu: *Data/Import data/from Excel, Access or dBase data set ...*

Podajemy nazwę obiektu *DataB* i OK. Następnie wyszukujemy w oknie „Otwieranie” plik *Excels*: *D:\R\Dane.xls*. Otwórz i w oknie skryptowym pojawia się polecenie w języku R:

```
>DANE <- sqlQuery(channel=3, select*from [Dane1$].
```

Ze skoroszytu wybrany został arkusz z danymi *Dane1* stanowiący wraz z nagłówkiem nazw zmiennych tabelę danych w pojęciu języka R. Mając aktywne okno danych przystąpmy teraz do naszego celu jakim jest raport sumaryczny podstawowej statystyki opisowej wszystkich zmiennych naszego modelu ekonometrycznego, korzystając z menu: *Statistics/Summaries/Active data set*. Efekt pracy programu występuje w oknie wyjściowym *Output Windows* (patrz tabela 3.19).

Tab. 3.19. Podstawowa statystyka sumaryczna zmiennych modelu



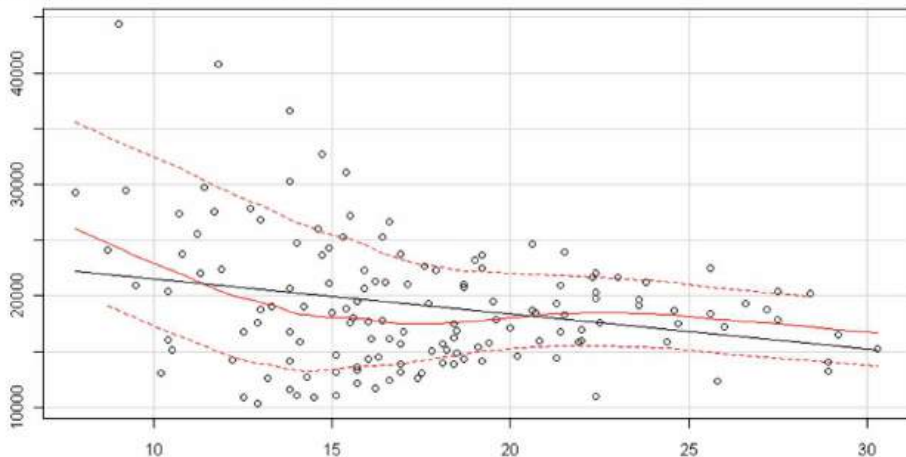
Źródło: Opracowanie własne pakietem R.

Dla każdej ze zmiennych *X1 ... X20* w tabeli 3.19 wyświetlone są następujące wartości:

- minimum (*Min.*),
- pierwszy kwartył (1 *st Qu.*),
- mediana (*Median*),
- średnia - arytmetyczna (*Mean*),
- trzeci kwartył (3 *rd Qu.*),
- maksimum (*Max.*).

Dane te pozwalają na narysowanie wykresu pudełkowego stanowiącego charakterystykę szeregu danej zmiennej. Zwróćmy uwagę na pokazanie na ekranie fragmentu danych *Excels* w zakresie zmiennych *X1, X5, X6*. Program R w oknie komunikatów podaje uwagę, informującą, że tabela danych *DANE* obejmuje 144 wiersze oraz 8 kolumn wartości zmiennych.

Informację o różnych możliwościach prezentacji graficznej szeregów szczegółowych zmiennych możemy znaleźć w rozdziale „Graficzne statystyki opisowe”⁵⁷. Ze względu na ramy niniejszej pracy ograniczono się do sporządzenia wykresu punktowego (rozrzutu) w układzie $X1 \sim X8$ (patrz wykres 3.8) korzystając z menu: *Graph/Scatterplot...*



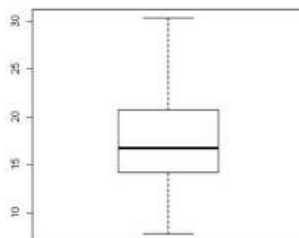
Źródło: Opracowanie własne pakietem R.

Rys. 3.8. Wykres rozrzutu zmiennych $X1 \sim X8$ (*Scatterplot*)

Ze względu na dane panelowe (przestrzenno-czasowe) dla 16. województw i 9. okresów (lata 1999-2007) występuje duże rozproszenie obserwacji. W tej sytuacji generowana jest podstawowa linia regresji oraz trzy regresje nieliniowe stanowiące podregresje skupisk obserwacji (górných, środkowych, dolnych). Dla tego rezultatu w oknie skryptowym pojawiło się polecenie języka R:

```
>scatterplot (X1~X8, reg.line=lm, smooth=TRUE, spread=TRUE, boxplots=FALSE,
+ span=0,5,data=DANE)
```

Na wykresie pominięto argumentem *boxplots=FALSE* możliwość wydruku wykresów pudełkowych dla zmiennych $X1$ na osi rzędnej oraz dla zmiennej $X8$ na osi odciętych. Dla lepszej czytelności wykres pudełkowy wykonano oddzielnie przykładowo dla zmiennej $X8$ (patrz rysunek 3.9) korzystając z menu: *Graph/boxplot*.



Źródło: Opracowanie własne pakietem R.

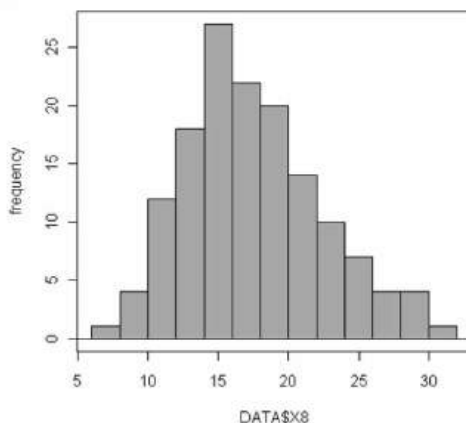
Rys. 3.9. Wykres pudełkowy zmiennej $X8$ (*boxplot*)

⁵⁷ Biecek M., *Przewodnik po pakiecie R*, op. cit.

Po realizacji tego wykresu w oknie skryptowym ukazało się polecenie: `boxplot (DATA$X8, ylab='X8')`

Zaprezentujemy jeszcze wykonanie histogramu dla zmiennej *X8* poprzez wywołanie menu: *Graph/Histogram ... Variable (X8)*.

Na rysunku 3.10 widzimy histogram, gdzie na osi odciętych program *R* dokonał podziału wartości zmiennej *X8* na 13 klas przedziałów wartości obserwacji, dla których obliczył częstości (*frequency*) na podstawie zbioru *DATA*.



Źródło: Opracowanie własne pakietem *R*.

Rys. 3.10. Histogram zmiennej *X8*

Program *R* w oknie skryptowym napisał polecenie:

`Hist (DATA$X8, scale="frequency", breaks="sturges", col="derkgrup")`.

* * *

Warto nadmienić, że podstawy języka *R*, o takiej samej nazwie jak pakiet znajdują się w publikacji „Statystyczna analiza danych z wykorzystaniem programu *R*”⁵⁸. Ponadto środowisko programowe *R* posiada wbudowany system pomocy, a jego elementami są⁵⁹:

- podręczniki użytkownika w podkatalogu: `..doc\manual`,
- pomoc dla poszczególnych funkcji poprzez polecenie: `help („<nazwa_funkcji>”)`,
- odpowiedź dotycząca argumentów wybranej funkcji za pomocą polecenia: `args`,
- wyświetlenie przykładów danej funkcji poleceniem: `example`.

⁵⁸ Ibidem, rozdz. 1.6, *Podstawy języka R*.

⁵⁹ Ibidem, rozdz. 1.5, *System pomocy*.

4. ZASTOSOWANIE KOMPUTERA W EKONOMETRII



4.1. Wstęp

Ekonometria jako odrębna dyscyplina naukowa rozwija się coraz prężniej. Powstają nowe metody i modele. Doskonalony jest aparat teoretyczny do estymacji i weryfikacji modeli. Na rynku wydawniczym pojawiają się publikacje sugerujące stosowanie nowoczesnej techniki obliczeniowej. Obszerny dorobek z zakresu oprogramowania znany już we wiodących gospodarczo krajach przenika do wydawanych u nas pozycji książkowych oraz coraz szerzej stosowany jest na uczelniach ekonomicznych. Nieliczne są jednak publikacje poświęcone w całości rozwiązywaniu problemów z zastosowaniem określonego pakietu komputerowego. Często występuje tylko nawiązanie do programów lub zastosowania podane są fragmentarycznie w formie dołączonego aneksu. Maddala⁶⁰ w swoim powszechnie stosowanym podręczniku ekonometrii stwierdza, że nie ma jednak pakietu, który obejmowałby rozwiązywanie wszystkich tematów objętych jego pracą. Wiele jest programów o różnorodnych symbolach, które w swej funkcjonalności nieznacznie różnią się między sobą oraz dotyczą metod i testów podstawowych.

Zamierzeniem jest, aby niniejszy artykuł był zainicjowaniem czegoś w rodzaju przewodnika po publikacjach i pakietach programowych zarówno tych ogólnie bezpłatnie dostępnych jak i zamkniętych wymagających indywidualnych licencji. Jest to bardzo obszerna tematyka znajdująca swoje odbicia wycinkowe w Internecie w zakresie przykładów zastosowań. Wystarczy sięgnąć do strony www.feweb.vu.nl/econometriclinks/software/software.html, aby się przekonać o ogromności dorobku w zakresie oprogramowania rozwijanego w miarę postępu w zakresie systemów operacyjnych, programów narzędziowych i sieciowych. Strona ta zawiera linki (połączenia) z dystrybutorami określonego oprogramowania.

W niniejszym materiale dokonano przeglądu dostępnych pozycji literaturowych pod kątem zastosowania ETO. Pozostawia się studiującym dalsze penetrowanie możliwości zainstalowania programów jak i rozpoznania ich funkcjonalności, tak by ekonometria nie kojarzyła się tylko z rozważaniami teoretycznymi i mnóstwem wzorów.

4.2. Programy *GRET*, *Eviews*, *WinQSB*, *Excel*

Omówienie obszarów zastosowania całej palety programów komputerowych przeznaczonych dla ekonometrii nie jest możliwe w ramach niniejszego materiału. Ponadto nie wszystkie stosowane są w Polsce. Z tego względu oparto się na przykładowych - powszechnie dostępnych.

GRET. Oprogramowanie *GRET* (*GNU Regression Econometrics Time-Series Library*) opracowane jest w wersji polskiej. Dostępne jest na stronie internetowej <http://www.gretl.sourceforge.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 autora⁶¹. Oprogramowanie wydane jest przez Fundację Wolnego Oprogramowania (*Free Software Founda-*

⁶⁰ Maddala G.S., *Ekonometria*. Wydawnictwo Naukowe PWN, Warszawa 2006.

⁶¹ Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu Gretl*, Wydawnictwo Naukowe PWN, Warszawa 2007.

tion) i jest ogólnie bezpłatnie dostępne. Może być przez użytkowników modyfikowany jego kod źródłowy, jednak ryzyko w zakresie skuteczności działania funkcji ponosi sam użytkownik. Obecnie lansowana jest wersja *GRET*L-a o numerze 1.6.5. Procedurę instalacji pliku *gretl.install.exe* oraz plików z danymi pochodzącymi z podręczników ekonometrii podano w rozdziale 1.2 wymienionej publikacji Kufla. Menu programu *GRET*L 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 oraz szacowanie statystyk,
- dodawanie zmiennych* - tworzenie nowych zmiennych na bazie istniejących,
- próba* - określenie zakresu obserwacji z pliku danych,
- zmienna* - udostępnianie statystyk opisowych i rozkładów oraz ich zgodności,
- model* - obszerny zestaw metod estymacji modeli ekonometrycznych i szeregów czasowych oraz modeli nieliniowych i równań współzależnych i panelowych.

Jak już wspomniano, oprogramowanie *GRET*L 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 *GRET*L-em pozyskany plik należy zapisać na dysku własnego komputera. W ramach tego programu występują przykłady modeli ekonometrycznych, a dla części z nich zestaw poleceń - komend. Wbudowano również tablice statystyczne rozkładów: normalnego, *t*-Studenta, chi-kwadrat, *F*-Fishera, Durбина-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 omawianej aplikacji 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 będące modyfikacją metody najmniejszych kwadratów. Bardzo pomocne są informacje z procesu estymacji oraz specjalne funkcje *GRET*L-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 *GRET*L 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 *GRET*L-a. Daje to możliwość zrozumienia raportów programu z realizacji budowy modeli: wielomianowych trendu, wahań sezonowych, autoregresyjnych.

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 *GRET*L proponuje w ramach modeli szeregow czasowych metody: Cochran'a-Orcutta, Hirdertha-Lu, Prais-Winstena.

Istnienie odstających (nietypowych, wpływowych) obserwacji jest przyczyną występowania heteroskedastyczności wariancji resztowej. Niespełnienie 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 (2MKN) 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 przestrzennie-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 programie *GRET*L (rozdział 12. publikacji⁶²).

Eviews. Statystyczny pakiet komputerowy *Eviews* posłużył Lipiec-Zajchowskiej do przeprowadzenia analiz, których wyniki są w formacie tego pakietu⁶³. W aneksie publikacji zawarto opisy komputerowych programów *Excel* oraz *Eviews*. Wyniki analiz dyskryminacyjnych otrzymano stosując pakiet *Statistica*. Pakiet *Eviews* zastosowano do budowy liniowych modeli regresji z jedną zmienną objaśniającą (niezależną). W ramach tego dokonano oszacowania współczynników korelacji, estymacji parametrów modelu oraz otrzymano podstawowe statystyki dotyczące dopasowania modelu do danych empirycznych, a mianowicie uzyskano:

- współczynniki przy zmiennych objaśniających;
- błędy standardowe oszacowania parametrów,
- wartości testu *t*-Studenta, poziomu istotności dla statystyk *t*;
- współczynnik determinacji, skorygowany współczynnik determinacji;
- odchylenie standardowe składnika resztowego;
- wartość testu Durbina-Watsona;
- średnia zmiennej zależnej, odchylenie standardowe zmiennej zależnej;
- wartość testu Fishera, poziom istotności dla statystyki *F*.

W wyżej wymienionej publikacji użyto *EViews* również do oszacowania parametrów liniowego dwuwymiarowego modelu ekonometrycznego uzyskując także statystykę dopasowania

⁶² Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu Gretl*, op. cit.

⁶³ Lipiec-Zajchowska M. (red.): *Wspomaganie procesów decyzyjnych. Tom II. Ekonometria*. Wydawnictwo C.H. Beck, Warszawa 2003.

modelu. Program zastosowano również do wielowymiarowych modeli regresji, gdy liczba zmiennych objaśniających jest większa od dwóch. W rozdziale 1.14.2 spotykamy przykład oszacowania tym programem nieliniowego wielowymiarowego modelu regresji z opóźnioną zmienną endogeniczną Y_{t-1} . Oprócz zastosowania pakietu do modeli przyczynowo-skutkowych we wspomnianej publikacji spotykamy również wykorzystanie *EViews* do modeli rozwoju zjawiska w czasie. Interesujący jest przykład doboru funkcji trendu spośród liniowej, potęgowej oraz wykładniczej.

Gdy występuje sezonowość to dla kwartałów możemy wprowadzić zmienne zero-jedynkowe i wówczas estymujemy parametry modelu obejmującego czas oraz zmienne. Zastosowanie to dotyczy zarówno modeli multiplikatywnych jak i addytywnych. Zjawiska w czasie opisywane są dla celów prognozowania za pomocą modeli:

- prostych średnich ruchomych, przykładowo dwu lub cztero-okresowych;
- ważonych średnich ruchomych;
- prostego wyrównywania wykładniczego z parametrem wyrównującym α ,
- adaptacyjnego wyrównywania wykładniczego z sygnałem adaptacyjnym $TST\alpha$,
- podwójnych średnich ruchomych;
- podwójnego wyrównywania wykładniczego (model Browna);
- Holta-Wintersa (bez sezonowości, z sezonowością multiplikatywną lub addytywną).

Omawiany pakiet umożliwia zbudowanie modeli wyrównywania wykładniczego (prostego, podwójnego) oraz Holta-Wintersa. Szybka realizacja komputerowa daje szansę wyboru adekwatnego modelu jak i najlepszych parametrów z punktu widzenia miary błędu wygasłych już prognoz jaką jest RMSE (pierwiastek średniego kwadratu błędu).

Modele autoregresyjne stosowane do prognoz krótkoterminowych wykorzystujemy, gdy utrudnione jest zebranie danych dotyczących zmiennych objaśniających. W klasycznych modelach autoregresyjnych np. gdy zmienne objaśniające stanowią zmienne opóźnione o 1, 2, 3 okresy estymujemy parametry w *EViews* korzystając z KMNK. Warto dodać, że oprócz estymacji parametrów modelu możemy tym pakietem przeprowadzić (podobnie jak w programie *GRET*) następujące badania:

- istotności parametrów (test t -Studenta),
- autokorelacji (test LM),
- normalności reszt (test JB),
- stałości wariancji (test White'a).

Na podstawie analizy wyników testów decydujemy o odrzuceniu lub nie modelu oszacowanego na podstawie KMNK biorąc pod uwagę kryteria założeń tej metody.

Gdy mamy do czynienia ze stacjonarnym szeregiem czasowym, który charakteryzuje się stałą średnią i wariancją, a kowariancję zależną od rzędu opóźnienia p zmiennej objaśnianej Y , a nie od czasu t , to budujemy modele Boxa i Jenkinsa. Do badania stacjonarności szeregu stosujemy funkcję autokorelacji lub rozszerzony test Dickey-Fullera zawarty w programie *EViews*. Jeśli szereg jest niestacjonarny to budowane są w oparciu o różnicę rzędu d zmiennej objaśnianej Y modele średnich ruchomych, a także modele mieszane. Po usunięciu przykładowo z modelu mieszanego nieistotnych zmiennych otrzymujemy poprawny statystycznie model nadający się do praktycznego stosowania.

Odrębną grupę stanowią zastosowania *EViews* do estymacji parametrów modeli ze zmiennymi dychotomicznymi, gdy Y ma wartość zero-jedynkową oraz występuje zestaw cech jakościowych X . Poszukujemy związku między X a prawdopodobieństwem realizacji określonego zdarzenia wyrażonego przez Y . W zależności od formuły wyrażającej tę zależność występują poniższe modele, których parametry możemy oszacować również programem *EViews* (który niestety nie jest ogólnie dostępny w sieci publicznej Internetu):

- liniowy model prawdopodobieństwa,
- logit (logarytm naturalny współczynnika szansy),
- probit, wykorzystujący dystrybucję rozkładu normalnego.

Procedura zastosowania pakietu *EViews* w wersji 3.11 z przykładami ekranów menu znajduje się w rozdziale 1. *Aneksu* publikacji⁶⁴. Możemy dowiedzieć się tam jak wprowadzać dane do modelu regresji, wykonać odpowiednie wykresy, przeprowadzić operacje na szeregach danych, dokonać estymacji parametrów, sporządzić wykres rozrzutu reszt oraz otrzymać miary statystyki opisowej. *EViews* jest też pomocny w modelach trendu.

Możemy prognozować bazując na modelu autoregresyjnym, czy też wyrównywania wykładniczego. Opcje programu - *histogram i statystyki opisowe, szereg rozdzielnicy* dają nam materiał do podstawowej analizy statystycznej. Warto jeszcze zaznaczyć, że pracując w systemie operacyjnym *Windows* mamy możliwość dokonania importu/eksportu plików między *EViews* a arkuszem kalkulacyjnym *Excel* lub edytorem tekstów *Word*. Zastosowanie omawianego pakietu spotykamy również w tomie I. *Statystyka*⁶⁵. Korzystanie z menu dla pozyskania statystyki opisowej oraz niżej wymienionych typów wykresów podano w rozdziale 1. *Aneksu*:

- liniowy, liniowy warstwowy;
- słupkowy, słupkowy warstwowy, liniowo-słupkowy;
- korelacyjny;
- kołowy.

W ramach statystyki opisowej otrzymujemy określenie: średniej wartości zmiennej, mediany, wartości maksymalnej oraz minimalnej, odchylenia standardowego, współczynnika skośności i kurtozy, testu Jargue-Bery.

WinQSB. Pakiet *Windows Quantitative System for Business* (WinQSB) służy do rozwiązywania zadań z programowania matematycznego. W czasie opracowania niniejszego materiału dostęp w Internecie do plików tego programu oraz jego opisu uzyskano wywołując hasło *WinQSB* w domenie [//www.onet.pl](http://www.onet.pl). Spośród różnych możliwości wejścia wybrano Zakład Metod Ilościowych w Zarządzaniu. Po otwarciu strony mamy do wyboru: *opis programu WinQSB, program WinQSB*. Opis instalujemy na własnym dysku jako plik *winqsb.doc* o pojemności 30 MB - 34 strony formatu A4, a w formie skompresowanej 1,8 MB. Program zapisany jest na 3. plikach: *winqsb(2).exe, winqsb.r00, winqsb.r01*. Pliki programu instalujemy również na dysku własnego komputera. Opis programu "Rozwiązywanie zadań za pomocą pakietu *WinQSB*" można było pobrać również z publicznej domeny o symbolu dostępu *miz.marc.pl*⁶⁶. Wyszczególnione są w nim funkcje i elementy poszczególnych pozycji menu, a mianowicie:

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 (rozwiązywanie i analiza): rozwiązywanie, rozwiązywanie krok po kroku, metoda graficzna, analiza parametryczna);

Results (wyniki, grafy);

Utilities (pomocnicze) (kalkulator, zegar, pomoc).

⁶⁴ Kufel T., *Ekonomometria. Rozwiązywanie problemów z wykorzystaniem programy Gretl*, op. cit.

⁶⁵ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom I, *Statystyka*, Wydawnictwo C.H. Beck, Warszawa 2003.

⁶⁶ Podana procedura dotyczy czasu tworzenia tej monografii.

Zamieszczone przykłady dydaktyczne dotyczą zasadniczo programowania liniowego i całkowitoliczbowego oraz zagadnienia transportowego z zastosowaniem algorytmu simpleks. Pakiet *WinQSB* dotyczy przede wszystkim zagadnień z badań operacyjnych. Natomiast z obszaru ekonometrii na uwagę zasługuje moduł "Prognozowanie i liniowa regresja" (FC - *Forecasting*) umożliwiający estymację modeli regresji i modeli trendu. Występuje analogia w zakresie rodzaju modeli do pakietu *Eviews*. Korzystając z modułu FC otrzymujemy raporty z oszacowaniem parametrów, wygastymi 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. Wspomniany moduł FC umożliwia nam określenie parametrów wyrównywania lub doboru najlepszej ich wartości z punktu minimalizacji błędu RMSE wygastych 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: α - *zmiennej*, β - *trendu*, γ - *sezonowości*, z krokiem co 0,01 z przedziału 0+1, co jest dogodne przy realizacji komputerowej.

Opis pakietu programowego wspomagającego podejmowanie decyzji *WinQSB* znajdujemy także w *Aneksie* tomu III. publikacji pod redakcją Lipiec-Zajchowskiej⁶⁷. Obejmuje on między innymi moduły:

Prognozowanie,
Programowanie dynamiczne,
Problem wyboru najkrótszej trasy oraz 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*. Moduł ten stanowi narzędzie do formułowania na podstawie szeregu czasowego zmiennej objaśnianej parametrów różnych modeli, a także wygastych i przewidywanych prognoz.

Excel. Pakiet ten jest powszechnym i uniwersalnym narzędziem wspomagającym analizy statystyczne, estymację oraz weryfikację modeli ekonometrii. Posługuje się arkuszem kalkulacyjnym tworzącym skoroszyt. Znając wzory na poszczególne statystyki, testy, czy równania modeli możemy przy pomocy wyrażeń napisać potrzebne formuły obliczeniowe. Ponadto *Excel* ma wbudowane funkcje statystyczne i matematyczne. Do grupy statystycznych należą następujące funkcje elementarne:

WSP.KORELACJI;
ŚREDNIA, NORMALIZUJ;
WARIANCJA, WARIANCJA.POPUL, KOWARIANCJA;
NACHYLENIE, ODCIĘTA;
ODCH.STANDARDOWE, ODCH.STANDARDOWE.POPUL.

⁶⁷ Lipiec-Zajchowska (red.), *Wspomaganie procesów decyzyjnych*, Tom III, *Badania operacyjne*, Wydawnictwo C.H. Beck, Warszawa 2003.

a także funkcje:

R.KWADRAT (współczynnik dopasowania),

REGLINP (regresji liniowej),

REGEXPP (regresji wykładniczej),

REGBLSTD (błędu standardowego reszt).

Istnieje również możliwość skorzystania z opcji *Excel-Analiza danych-Histogram* w celu sporządzenia wykresów oraz tabel częstości występowania wartości danych w określonych przedziałach. Funkcje matematyczne, a w szczególności z zakresu rachunku macierzowego, umożliwiają estymację parametrów równań liniowych, potęgowych, wykładniczych, wielowymiarowych. Należą do nich: *TRANSPONUJ* (transpozycja macierzy), *MACIERZ.ILOCZYN*, *MACIERZ.ODW* (macierz odwrotna).

Interesującym segmentem tego programu niezbędnym w ekonometrii i statystyce opisowej jest *Dodatki*. Między innymi dzięki niemu *Excel* znalazł zastosowanie w dydaktyce przedmiotu ekonometria, stąd fragmenty jego użycia spotykamy w różnych pozycjach literaturowych⁶⁸. Analizie ekonometryczno-statystycznej z zastosowaniem zbioru procedur *Dodatki* arkusza kalkulacyjnego *Microsoft Excel* poświęcona jest część trzecia publikacji Jajugi. Przedstawione zostały tam zastosowania opcji: korelacja, kowariancja, statystyka opisowa i regresja.

W *Excelu* ustawiamy się na aktywnej komórce i korzystamy z menu *Wstaw/Funkcje/Statystyczne* i wybieramy nazwę funkcji, przykładowo *kwartył*. Możemy też wybrać ścieżkę *Narzędzia/Dodatki/Analiza danych/Statystyka opisowa*. Jeśli mamy wprowadzone dane do określonej kolumny to w oknie dialogowym wpisujemy lub wybieramy między innymi:

- zakres wejściowy komórek zawierających dane;
- grupowanie wg (wierszy/kolumn);
- poziom ufności dla średniej;
- zakres wyjściowy (adres lewej górnej komórki dla tabeli wyników);
- statystyki podsumowujące dla: średniej, błędu standardowego, mediany, dominanty, odchylenia standardowego, wariancji, kurtozy, skośności, zakresu, minimum, maksimum.

W niniejszym materiale przyjęto, że modele użyte do prognozowania i symulacji są wytworem ekonometrii, stąd wzięto też pod uwagę literaturę z tego obszaru. W rozdziale piątym pracy pod redakcją Radzikowskiej⁷⁰ podano przykłady zastosowania arkusza kalkulacyjnego *Excel* w ustalaniu prognoz. Określono regułę prognozowania oraz ocenę dopuszczalności prognozy a funkcje *Excelsa* zastosowano do:

- współczynnika zmienności, współczynnika korelacji;
- oszacowania parametrów wielowymiarowego liniowego modelu regresji;
- statystyki regresji i analizy wariancji;
- określenia formuły modelu Holta i jego ekstrapolacji dla wyznaczenia prognoz;
- danych analizy harmonicznej, współczynnika determinacji; odchylenia standardowego.

Realizacja komputerowa umożliwia dobieranie parametrów wygładzania dla minimalizacji miary RMSE. W omawianej publikacji zademonstrowano wygodną w tym zakresie funkcję *SUMA.XMY.2*, która oblicza sumę kwadratów różnic pomiędzy wartościami w dwóch zakresach

⁶⁸ Lipiec-Zajchowska M. (red.): *Wspomaganie procesów decyzyjnych. Tom I. Statystyka*. Wydawnictwo C.H. Beck. Warszawa 2003.

⁶⁹ Jajuga K. (red.), *Ekonometria. Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu. 2002.

⁷⁰ Radzikowska B.(red.), *Metody prognozowania. Zbiór zadań*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu. 2004.

komórek. Skorzystano także z *Excels* dla obliczania błędu względnego prognozy, przewidywanej zmiennej objaśniającej (zależnej) stosując funkcję logarytmiczną trendu. Sformułowano wyrażenia *Excels* na określenie błędów *ex ante* wyznaczanych prognoz.

Z potrzebą zastosowania arkusza kalkulacyjnego *Excel* do budowy modeli prognostycznych, wyznaczania prognoz oraz oceny ich jakości spotykamy się też w pracy pod redakcją Cieślaka⁷¹. Omówiono w niej na przykładzie kolejne kroki postępowania (budowę i ekstrapolację, dobór parametrów, dopuszczalność prognozy) dla modelu wykładniczego. Zaprezentowano procedurę zbudowania prognoz punktowych i przedziałowych z zastosowaniem modelu analitycznego bazującego na szeregu czasowym i funkcji *REGLINP Excels*.

4.3. Informacje o innych programach

Tylko niektóre programy komputerowe zastosowane w ekonometrii, statystyce czy też prognozowaniu i symulacji doczekały się na rynku polskim odrębnych publikacji. Najczęściej książki z ekonometrii przytaczające i rozwijające sekwencje opisów teoretycznych z innych prac wzmiankują o możliwości zastosowania informatyki dla wspomagania budowy i weryfikacji modeli. Przeglądając dostępną literaturę znalaziono nawiązania do stosowania w różnych obszarach oprogramowania:

Eview 5 - program do analizy szeregów czasowych i modeli VAR;

R-CRAN - program statystyczno-ekonometryczny, w tym rozwiązywanie szeregów czasowych;

GAUSS - program macierzowy do obliczeń statystycznych;

MATLAB - laboratorium macierzowe,

Scilab - program typu *MATLAB* lub *GASS*,

Stat - wyznaczanie podstawowych statystyk oraz prognoz wg *Holta*;

DEMS - estymacja i prognozowanie (wykorzystujący prostotę i elastyczność *Excels*).

W przedmowie do wydania drugiego *Ekonometrii*⁷² czytamy: "Mamy obecnie do dyspozycji wiele programów (komputerowych) z których można wybierać do woli. Ponadto zmieniają się one tak szybko, że ich listy należałoby bardzo często uzupełniać i rozszerzać". Maddala w *Ekonometrii* (Dodatek D: Programy komputerowe) wymienia programy dostępne w końcu lat 1990. Od tego czasu opracowano dla nich nowe wersje uwzględniające postęp w zakresie systemów operacyjnych. Oprócz programu *SAS* (którym posłużył się Maddala do dyskusji na temat współczynnika determinacji) są i inne, które ten wybitny ekonometryk uznał za powszechnie stosowane, a należą do nich:

SHAZAM, *MICROFIT 4.0*, *MICRO-TSP*;

RATS, *CATS* w pakiecie *RATS* do analizy kointegracji,

LIMDEP (do analizy modeli z ograniczoną zmienną zależną, a także do prostych modeli logit, probit, tobit).

Według Maddali niepokojące jest jednak to, że kolejne programy dają różne odpowiedzi na to samo pytanie. Sprawdzenia dokonał na przykładzie estymacji parametrów modeli logitowych i probitowych używając programów *SAS* oraz *LIMDEP*. Po swojej analizie ten autor dla prostych modeli logitowych i probitowych proponuje program *SAS*, a dla skomplikowanych, w tym z ograniczoną zmienną zależną pakiet *LIMDEP*. W zakresie podstawowych modeli z niedużymi bazami zaleca *SHAZAM* oraz *MICROFIT*.

Z rozpoznania w Internecie wynika, że wyższe uczelnie ekonomiczne w coraz szerszym zakresie w nauczaniu statystyki, ekonometrii, jak również prognozowania i symulacji prowadzą

⁷¹ Cieślak M. (red.), *Prognozowanie gospodarcze. Metody i zastosowania*. Wydawnictwo Naukowe PWN, Warszawa 2005.

⁷² Maddala G.S., *Ekonometria*, op. cit.

ćwiczenia z tych przedmiotów w laboratoriach komputerowych. Do modelowania zjawisk gospodarczych stosowane są programy: *Statistica*, *EViews*, *EasyReg*, *GRET*L oraz inne pracujące w środowisku *Windows*, *Linux* lub w systemach sieciowych uczelni. Zdaniem jednak niektórych autorów do prowadzenia ćwiczeń laboratoryjnych wystarczający jest arsenał narzędziowy oferowany przez arkusz kalkulacyjny *Excel*⁷³.

Także firmy internetowe oferujące swoje usługi w zakresie rozwiązywania zadań z ekonometrii służą szeroka gamą oprogramowania, a mianowicie programami: *Statistica*, *Excel*, *SPSS*, *Eviews*, *Winstat*, *Microfit*, *Gretl*, *GiveWin*, *G*, *PDG*, *Statgraf*. Dla zrealizowania zamówienia, opracowania i sprawdzenia modelu, można podać własne dane jak też wskazać ich źródło, przykładowo Główny Urząd Statystyczny. Odnośnie doboru zmiennych objaśniających stosowane są przez te firmy takie metody jak: *Helwig*a (pojemności informacji), *grafowa*, *CIV*, *współczynnika korelacji*, *współczynnika zmienności*, *test t-Studenta*.

Estymacja parametrów modelu przeprowadzana jest najczęściej metodami: *klasyczną metodą najmniejszych kwadratów*, *największej wiarygodności*, *Gaussa-Seidela*, *Newtona-Raphsona*. W ofercie firm usługowych jest wykonywanie modeli: liniowych, nieliniowych, wielorównaniowych (rekurencyjnych, o równaniach współzależnych), VAR i innych o złożonej postaci analitycznej. Jak już wspomniano w usłudze jest weryfikacja modeli, a w ramach niej badanie autokorelacji, stałości wariancji (heteroskedastyczności), normalności reszt, współczynnika determinacji, odchylenia standardowego, przeprowadzenie testu *t-Studenta*, testu serii, określenie symetrii i losowość składnika losowego (reszt), a także ocena stabilności parametrów. Wykonane opracowania obejmują opisy, wzory oraz sformułowane wnioski.

Na stronie <http://feweb.vu.nl/econometriclinks/software.html> występuje bardzo obszerny indeks aktualizowanych na bieżąco informacji o programach z ekonometrii. Pozostawia się Czytelnikowi zapoznanie się z dokładnymi informacjami na podanej stronie *WWW* podającej autorów, wersje, linki do stron dystrybutorów. Ze względu na ramy niniejszego materiału przykładowo podano tylko mały fragment z tekstu w Internecie, obejmujący symbole programów i to tylko w zakresie liter A-G:

A+, ACML, ADMB, APL, AIMMS, ALOGIT, Alyuda, AMOS, AMPL, Apophenia, Arc, AREMOS, AutoBox, Autometrics, AutoSignal;
B34S, BACC, BATS, BETA, BIOGEME, BMDP, Brodgar, BUGS, BV4;
CalculationCenter 3, CART 5.0, Census X12, Caterpillar-SSA 3.30, CPLEX 10, Confort 2.0, CVar Expert 1.7;
DataDesk, Dataplore, Dataplot, DATAVIEW, DEA-Solver, DEMETRA, DYALOG, DYNARE NEW;
EasyFit, EasyReg, EcoWin, ECTS, EQS, Eviews UPDATED, Excel UPDATED, Excel 2007, EXPO;
FAME, ForecastPro, Fortran, FreeFore, FSQP;
GAMS, G@RCH UPDATED, GAUSS, GAUSSX, GiveWin, Gempack, GeoDa, Genstat, GLIM, GLIMMIX, GQOPT, graphpad, Gnuplot, GSL, GRET L (w ramach GRET L - wersja GRET L 1.7.4 (*Gnu Regression, Econometric and Time Series Library*)).

4.4. O oprogramowaniu modeli symulacyjnych

Symulacja jest elementem pośrednim między realizmem a idealizmem. Burzliwy rozwój informatyki, zastosowań ekonometrii, badań operacyjnych oraz osiągnięcia w matematyce i komuni-

⁷³ Dziechciarz J. red., *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2003.

kacji elektronicznej są bodźcem do rozwoju oprogramowania w zakresie symulowania wirtualnej rzeczywistości funkcjonalności obiektów gospodarczych, bankowych, logistycznych i innych.

Z punktu widzenia symulacji model jest zbiorem obiektów abstrakcyjnych, relacji pomiędzy tymi obiektami. Cechuje go homeomorfizm, czyli analogiczne zachowania jak w obiekcie rzeczywistym. Zatem badania symulacyjne wymagają odpowiednich metod budowy, weryfikacji, walidacji (ustalenia stopnia trafności) modelu. Jeśli mówimy obecnie o symulacji to jest to przede wszystkim symulacja komputerowa bazująca na odpowiednich modelach matematycznych. Przykładowo w technice nowo zaprojektowany samochód jeżdżący po różnych nawierzchniach to układy równań różniczkowych zachowania się zawieszenia. W kreowaniu modelu symulacyjnego wyróżniamy:

- grupowe budowanie modelu konceptualnego, przygotowanie danych źródłowych,
- wyodrębnienie elementarnych podsystemów, budowę prostych modeli (z reguły nieliniowych);
- formułowanie "modelu całościowego" dla rozwiązania analitycznego, w którym następnie dla rozwiązania matematycznego skomplikowanych równań z reguły nieliniowych wprowadza się z konieczności uproszczenia.

Stosowanie modelu symulacyjnego ma następujące zalety:

- umożliwia przyspieszenie, spowolnienie czasu dla lepszego poznania zjawiska;
- można stosować wielu kryteriów jednocześnie, nawet w warunkach ekstremalnych;
- jest możliwa przy obecnym "aparacie informatycznym" budowa skomplikowanych modeli bez dużej wiedzy matematycznej;
- koszt budowy modelu symulacyjnego są mniejsze od koszt budowy obiektu.

Są również wady, a do nich zaliczamy:

- wyniki badań są zasadne do określonych warunków;
- jest konieczna wielokrotność badań dla sformułowania wniosku, co wiąże się z długimi czasem badań;
- występują ograniczenia w aparacie matematycznym dla znalezienia rozwiązania optymalnego;
- może wystąpić łatwość nadużywania, czyli przeceniania "dobroci" modelu tylko po jego efektywnym interfejsie.

W przypadku symulacji systemów społeczno-gospodarczych stosowane są różne techniki matematyczne i informatyczne modelowania. Niektóre techniki są łączone tworząc hybrydy (krzyżówki), a jako przykłady można wymienić:

- zastosowanie symulacji i optymalizacji dla doboru struktury modelu;
- planowanie eksperymentu w układzie symulacja - wspomaganie decyzji (poszukiwanie optymalnych wartości zmiennych decyzyjnych w modelu przedsiębiorstwa).

W ujęciu branżowym można wymienić przykładowo:

- modelowanie symulacyjne w usprawnieniu obsługi klienta w banku (zdefiniowanie problemu, określenie etapów budowy modelu, propozycja zmiany systemu obsługi klienta);
- analiza zyskowności produktów ubezpieczeń majątkowych i osobowych;
- analiza symulacyjna efektywności inwestycji długookresowej.

Jednak największe trudności w algorytmizacji systemów gospodarczych występują w odwzorowanie "czynnika ludzkiego", czyli możliwości poznawczych i emocji samych decydentów⁷⁴. W tym względzie wykonuje się obserwacje ich zachowań przez "wciągnięcie" ich do wirtualnej rywalizacji w grze. Przykładem jest gra symulacyjna ze zwycięzcą, a w ramach niej:

- występuje prowadzący, który ilościowo i jakościowo analizuje oraz ocenia dany przebieg gry;

⁷⁴ *Symulacja systemów społeczno-gospodarczych*, Prace Naukowe Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej. Seria: Studia i Materiały. Wrocław 2005.

- biorą udział uczestnicy - specjaliści danej branży;
- ustalone są reguły funkcjonowania firmy (obiektu) zakodowane w grze;
- wyłania się konkluzja, uczestnik, a może nim być również student kierunku zarządzanie odczuwa pilną potrzebę pogłębienia swojej wiedzy przykładowo z zakresu planowania produkcji oraz poznania metod i narzędzi informatycznych, które ten proces ułatwiają.

Warto zauważyć, że bankructwo wirtualnej firmy spowodowane "niesprawnością" w grze nie pozbawia pracy i majątku, ale wyzwala w uczestnikach prawdziwe emocje: gniew, wstyd, radość, dumę, współczucie dla przegranych. Taka więc gra tworzy mini system społeczny bowiem długotrwale przebywanie w rzeczywistości alternatywnej (wirtualny rynek, wirtualny zarząd) ma wpływ na funkcjonowanie uczestników. Gry symulacyjne kreują cechy menedżerskie - dają poczucie władzy, uczą odpowiedzialności. W trakcie gry podejmuje się właściwe zdaniem użytkowników decyzje i ponosi konsekwencje tych decyzji. Ekonomiczne gry symulacyjne pozwalają eksperymentować na modelach wielu zmiennych decyzyjnych. Symulacja pomaga w rozpoznaniu i wykorzystaniu własnych zdolności i predyspozycji.

Doskonalenie swoich cech menedżerskich pozwala unikać eksperymentowania na żywych organizmach firm. W Internecie oferowane są różne kursy z tego zakresu. W niniejszym materiale skoncentrowano się na niektórych modelach symulacyjnych w zakresie zarządzania produkcją, które dość wcześnie, bo już w latach 70-tych pojawiły się w wiodących gospodarczo krajach i stosowane są również w dydaktyce.

Tab. 4.1. Modele symulacyjne w zakresie zarządzania produkcją

Autor(zy)	Nazwa modelu gry	Opis
Churchil (1970)	Joblot	Działanie wydziału produkcyjnego dla zadanego rynku, konkurencji, procesu technologicznego.
Mize i in. (1971)	PROSIM	Gra niekonkurencyjna - koncentracja na minimalizacji kosztu. Przedmiotem produkcji wyrobów opisanych wg BOM, proces zakupu surowców, planowania i sprzedaży wyrobów. Wyniki w formie raportów z całego procesu produkcyjnego oraz rozliczeniowego kosztów.
Goldratt i Cox (1984)	OPT	System produkcyjny. Jakość zarządzania produkcją oceniana na podstawie zysku.
Southern (1986)	CAMP	Praca fabryki - harmonogramowanie procesu produkcyjnego.
Biggs (1987)	DECIDE-P/OM	Niekonkurencyjna gra kierownicza. Wyniki mają postać rachunku wyników produkcji i rozliczeń finansowych
Smith (1990)	SIMAN-CI-NEMA	Animacja przemieszczania się części i formowania kolejek w strukturze technologicznej i przedmiotowej produkcji.
Henshaw i Jackson (1990)	The Executive Game	Gra dla menedżerów. Zmienne to wielkość produkcji, zakup materiałów, wydatki na badania.
Wiendahl i in. (1995)	TRAIN-F	Gra - dobór parametrów planowania i sterowania produkcją.
Garetti i Taisch (1995)	FMS Design Game	Gra do nauczania projektowania elastycznych systemów produkcyjnych ESP.
Skrzypek i Szubra (1996)	TEES-2	Symulacyjna gra decyzyjna zarządzania produkcją.
Basnet (1999)	MRP-SIM	Gra oparta na arkuszu kalkulacyjnym dla zrozumienia planowania i sterowania produkcją.

Źródło: Fragment wg tabeli 1 "Przegląd modeli symulacyjnych"⁷⁵.

Przedstawię teraz dwa z programów symulacyjnych, a mianowicie wymieniony w powyższym zestawieniu *TEES-2* oraz *Zarządzanie kapitałem*.

⁷⁵ Ibidem.

TESS-2. Jest to narzędzie do doskonalenia umiejętności w zakresie zarządzania firmą. Uczestnicy przygotowują biznes plan oraz realizację w warunkach zmieniającego się rynku i ostrej konkurencji. *TESS-2* posiada moduł w zakresie zarządzania finansami i kontroli wyników na podstawie wskaźników finansowych. Gra opiera się na kompleksowej symulacji decyzyjnej w obszarze zarządzania przedsiębiorstwem. Każda z grup prowadzi fikcyjną firmę - oferującą jeden produkt i działającą na jednym rynku. Symulacja obejmuje od 12 do 20 kwartalnych okresów rozliczeniowych.

Konkurencja jest bezpośrednia i ostra a wynik jednego zespołu wpływa bezpośrednio na wynik pozostałych. Wyniki gry ocenia się na podstawie parametrów finansowych, takich jak zysk, udział w rynku, wartość likwidacyjna firmy, wysokość wypłaconych dywidend. Po zakończeniu gry ogłaszany jest ranking poszczególnych zespołów.

Zarządzanie kapitałem. Program ten to model uniwersalnego banku komercyjnego. Uczestnicy gry działają w zespołach tworzących mini-zarządy banków komercyjnych. Zespoły konkurują ze sobą na rynkach finansowych przeprowadzając operacje finansowe, depozytowo-kredytowe, giełdowe, walutowe i inne. Zadaniem każdego zespołu jest najefektywniejsze prowadzenie instytucji finansowej w okresie 3-4 lat. Można powiedzieć, że gra jest:

- przewodnikiem po rynkach finansowych (rynek depozytowy, rynek kredytowy, rynek papierów wartościowych, rynek walutowy, rynek międzybankowy, rynek finansowy ochrony środowiska) i ich mechanizmach;
- kształceniem umiejętności oceny zdolności kredytowej firmy i osoby fizycznej oraz oceny ryzyka;
- instrumentem rozwijania umiejętności współdziałania w grupie i zespołowego podejmowania decyzji.

Podobnie możemy prowadzić z użyciem komputera symulację na modelu ekonometrycznym, przykładowo w celu udzielenia odpowiedzi na pytania:

Jakie byłyby wartości zmiennych endogenicznych, gdyby zmienne egzogeniczne przyjęły określone wartości?

Jakie powinny być wartości zmiennych endogenicznych, by uzyskać pożądane wartości zmiennych egzogenicznych?

W takiej symulacji powinniśmy mieć możliwość zmieniać zmienne, parametry, właściwości składnika losowego, przy występowaniu zakłóceń losowych o znanym rozkładzie w czasie eksperymentu. Wynik symulacji, czyli warianty rozwoju obiektu opisywanego przez model mogą być traktowane jako prognozy mające wpływ na zmianę cen oraz rozwój przedsiębiorstwa.

Na zakończenie niniejszego przeglądu autor ma nadzieję, że studium zachęcony gamą programów oraz możliwościami zastosowania ETO w ekonometrii będzie penetrował dalej istniejącą literaturę przedmiotu jak i rozpoznawał pakiety w celu nabrania perfekcji w budowaniu modeli nie tylko gospodarczych.

Tab. 2.2. Literatura pomocnicza w zakresie wspomagania komputerowego podejmowania decyzji

Lp.	Opis
1.	Bartosiewicz S. (red.), <i>Ekonometria wspomagana mikrokomputerem</i> , Wydawnictwo Akademii Ekonomicznej, Wrocław 1991.
2.	Bartosiewicz S., <i>Ekonometria. Technologia ekonometrycznego przetwarzania informacji</i> , PWE, Warszawa 1989.
3.	Cieślak M.(red.), <i>Prognozowanie gospodarcze. Metody i zastosowania</i> , Wydawnictwo Naukowe PWN, Warszawa 2005.

Lp.	Opis
4.	Dobosz M., <i>Wspomagana komputerowo statystyczna analiza wyników badań</i> , Warszawa 2001.
5.	Dziechciarz J. (red.), <i>Ekonometria. Metody, przykłady, zadania</i> , Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2003.
6.	Gajda J., <i>Ekonometria praktyczna</i> , Łódź 1996.
7.	Gajda J., <i>Ekonometria. Wykład i łatwe obliczenia w programie komputerowym</i> , Wydawnictwo C.H. Beck, Warszawa 2004.
8.	Jajuga K., <i>Ekonometria. Metody i analiza problemów ekonomicznych</i> , Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2002.
9.	Koślacz P. (red.), <i>Nowoczesne pakiety komputerowe w podejmowaniu decyzji</i> , UW, Warszawa 1998.
10.	Kowerski M., <i>Ekonometria. Przewodnik po wykładach</i> , Wyższa Szkoła Zarządzania i Administracji w Zamościu, 1998.
11.	Krocosz-Krynke Zofia, <i>Symulacja w nauczaniu zarządzania produkcją</i> , referat w ramach publikacji na konferencję "Symulacja systemów gospodarczych", Prace Naukowe Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej, 2005
12.	Kufel T., <i>Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET</i> , Wydawnictwo Naukowe PWN, Warszawa 2007.
13.	Lipiec-Zajchowska M. (red.), <i>Wspomaganie procesów decyzyjnych</i> , Tom III. <i>Badania operacyjne</i> , Wydawnictwo C.B. Beck Warszawa 2003.
14.	Lipiec-Zajchowska M. (red.), <i>Wspomaganie procesów decyzyjnych.</i> , Tom II. <i>Ekonometria</i> , Wydawnictwo C.H. Beck, Warszawa 2003.
15.	Lipiec-Zajchowska M., <i>Wspomaganie procesów decyzyjnych</i> , Tom I. <i>Statystyka</i> , Wydawnictwo C.H. Beck, Warszawa 2003.
16.	Lipiec-Zajchowska, M. Sikora E., Zajchowski J., <i>Modele podejmowania decyzji w pakietach komputerowych</i> , UW, Warszawa 1995.
17.	Maddala G.S., <i>Ekonometria</i> , Wydawnictwo Naukowe PWN, Warszawa 2006.
18.	Nowak E., <i>Zarys metod ekonometrycznych. Zbiór zadań</i> , Wydawnictwo Naukowe PWN, Warszawa 1990.
19.	Piłatowska M.: <i>Repetitorium ze statystyki</i> , Wydawnictwo Naukowe PWN, Warszawa 2014.
20.	Podgórski J., <i>Statystyka z komputerem</i> , Warszawa 1995.
21.	Praca zbiorowa, <i>Badania operacyjne w przykładach i zadaniach</i> , Wydawnictwo Naukowe PWN, Warszawa 2004.
22.	Praca zbiorowa, <i>Przykłady i zadania z badań operacyjnych i ekonometrii</i> , Wydawnictwo Akademii Ekonomicznej w Poznaniu, 2003.
23.	Praca zbiorowa, <i>Wybrane zagadnienia z prognozowania</i> , ECONOMICUS, Szczecin 2006.
24.	Radzikowska B., <i>Metody prognozowania. Zbiór zadań</i> , Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2004.
25.	Skrzypek J. Szubra M., <i>Symulacyjna gra decyzyjna TESS-6 jako narzędzia e-learningu</i> , Symulacja systemów gospodarczych. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
26.	Skrzypek J. Szubra M., <i>TESS-2. Symulacyjna gra decyzyjna</i> , Poltex, Warszawa 1996.
27.	Strzała K., Przechlewski T., <i>Ekonometria Inaczej</i> , UG 1994.
28.	Szczurkowski L., <i>Komputerowa gra ekonomiczna. Symulacja systemów gospodarczych</i> , Oficyna wydawnicza Politechniki Wrocławskiej, Wrocław.
29.	Trzaskalik T., <i>Modelowanie optymalizacyjne. Seria: Metody badań operacyjnych</i> , Absolwent, Łódź 1997.
30.	Trzaskalik T., <i>Badania operacyjne z komputerem, Seria: Metody badań operacyjnych</i> , Absolwent, Łódź 1997.

Źródło: Opracowanie własne na podstawie informacji w Internecie.

5. WYBÓR KOMBINACJI ZMIENNYCH



5.1. Wstęp

W początkowym etapie budowy liniowego modelu ekonometrycznego nasuwa się nam wiele przyczyn od których zależy dane badane zjawisko. Rodzi się pytanie, jakimi zmiennymi ilościowymi je opisać oraz które z nich możemy uznać jako potencjalne zmienne objaśniające. Kryterium wstępnej klasyfikacji jest współczynnik zmienności lub rozstępu. Po tej selekcji otrzymujemy zestaw zmiennych, które mogą tworzyć wiele wzajemnych kombinacji. W kolejnym kroku poszukujemy metody, która wskazała by nam kombinację zmiennych istotnych statystycznie dającą dobre dopasowanie modelu do danych rzeczywistych. Aby ułatwić to zadanie w niniejszym artykule zaprezentowano coś w rodzaju mikromonografii metod doboru zmiennych objaśniających. Przegląd obejmuje opublikowane w dostępnych pozycjach książkowych. Oprócz najstarszych metod regresji krokowej dominują ostatnio propozycje bazujące na analizie korelacji lub badaniu pojemności nośników informacji jakimi są poszczególne kombinacje z zestawu potencjalnych zmiennych objaśniających.

5.2. Model ekonometryczny

Liniowy model ekonometryczny określamy w postaci:

$$Y = f(X_1, X_2, \dots, X_n) + u$$

$$Y = a_1 X_1 + a_2 X_2 + \dots + a_n X_n + u$$

Gdzie:

Y - zmienna objaśniana (endogeniczna),

$X_1, X_2, \dots, X_n$ - zmienne objaśniające (egzogeniczne),

$a_1, a_2, \dots, a_n$ - oszacowane na podstawie szeregów obserwacji parametry strukturalne modelu,

u - składnik losowy, uwzględniający różnice jakie występują między wartościami z obserwacji a teoretycznymi wyznaczonymi przez model.

Procedura formułowania modelu ekonometrycznego sprowadza się zasadniczo do:

- określenia zmiennych objaśnianych i objaśniających;
- wyboru postaci analitycznej modelu i wskazania źródeł danych statystycznych;
- estymacji, czyli oszacowania parametrów modelu na podstawie danych statystycznych korzystając z określonej metody;
- weryfikacji modelu.

Zmienne objaśniające (egzogeniczne) muszą mieć:

- wysoką zmienność, bo w przeciwnym wypadku są to zmienne *quasi-stale*;
- być silnie skorelowane ze zmienną objaśnianą;
- nie być skorelowane ze sobą.

5.3. Metoda wskaźnika wyboru wektora zmiennych objaśniających

Metody doboru zmiennych do modelu ekonometrycznego przedstawiają przede wszystkim relacje między zmiennymi w przeszłości. Odnośnie modelu dla celów prognostycznych na zmienną objaśnianą mają wpływ w przyszłości także inne zmienne objaśniające, przykładowo mające inną siłę

lub kierunku oddziaływania. M. Cieślak⁷⁶ proponuje dla oceny znaczenia zbioru zmiennych stosowanie: metody wskaźnika wyboru wektora zmiennych objaśniających, wyboru zmiennych na podstawie błędu prognoz *ex ante*.

$$\mathbf{y} = \begin{bmatrix} y_1 \\ y \\ \dots \\ y_n \end{bmatrix} \quad \mathbf{X} = \begin{bmatrix} x_{11} & x_{21} & \dots & x_{m1} \\ x_{12} & x_{22} & \dots & x_{m2} \\ \dots & \dots & \dots & \dots \\ x_{1n} & x_{2n} & \dots & x_{mn} \end{bmatrix}$$

Odnosnie pierwszego sposobu, niech wektor $\mathbf{y}$ przedstawia zbiór zaobserwowanych w próbie wartości zmiennej objaśnianej (prognozowanej), a macierz $\mathbf{X}$ zbiór zmiennych objaśniających $X_1, X_2, \dots, X_m$.

W tej metodzie s pierwszych obserwacji jest mniejsze od n . Obliczamy współczynniki korelacji $\mathbf{R}^1_0$ między Y , a każdymi ze zmiennych $X_1, X_2, \dots, X_m$ oraz między potencjalnymi zmiennymi objaśniającymi $\mathbf{R}^1$. Bazując na współczynnikach korelacji obliczamy indywidualne pojemności nośników informacji h^1_{Li} dla poszczególnych zmiennych X biorąc pod uwagę L kombinacji, których jest $2^m - 1$, przy czym $i, j = 1, \dots, m$. Przykładowo dla trzech zmiennych $2^3 - 1 = 7$. Metoda ta wykazuje analogię do metody *a priori* Hellwiga doboru zmiennych objaśniających do modelu ekonometrycznego.

$$\mathbf{R}^1_0 = \begin{bmatrix} r^1_{11} \\ r^1_{12} \\ \dots \\ r^1_{1n} \end{bmatrix} \quad \mathbf{R}^1 = \begin{bmatrix} 1 & r^1_{12} & \dots & r^1_{1m} \\ r^1_{21} & 1 & \dots & r^1_{2m} \\ \dots & \dots & \dots & \dots \\ r^1_{m1} & r^1_{m2} & \dots & 1 \end{bmatrix} \quad h^1_{Li} = \frac{(r^1_i)^2}{1 + \sum_{j \neq i} (r^1_{ij})^2}$$

Następnie obliczamy pojemności integralne każdej z możliwych kombinacji potencjalnych zmiennych objaśniających:

$$H^1_L = \sum_{i=1}^m h^1_{Li}$$

Podejściem odmiennym w stosunku do znanej metody Hellwiga jest wprowadzenie „ruchomej próby”, czyli w kolejnym etapie dodanie następnej obserwacji $s + 1$, a zrezygnowanie z pierwszej. Dla „nowej” próby obliczamy $\mathbf{R}^2_0$ oraz $\mathbf{R}^2$ i w kolejności H^2_L dla wszystkich kombinacji.

Postępując tak dalej otrzymujemy $n - s + 1$ pojemności integralnych, które możemy przedstawić w postaci macierzy:

$$\mathbf{H} = \begin{bmatrix} H^1_L & H^1_L & \dots & H^1_{2^n-1} \\ H^2_L & H^2_L & \dots & H^2_{2^n-1} \\ \dots & \dots & \dots & \dots \\ H^{n-s+1}_L & H^{n-s+1}_L & \dots & H^{n-s+1}_{2^n-1} \end{bmatrix}$$

W powyższej macierzy poszczególne kolumny stanowią szeregi czasowe kombinacji zmiennych objaśniających. Wprowadzamy kryterium wyboru $G_L = H^0_L W_L$, przy czym:

$$W_L = 1 + (n - s + 1) \beta_L \quad H_L = \alpha_L + \beta_L t ; \quad t = 1, \dots, n - s + 1.$$

Gdzie β_L - współczynnik kierunkowy trendu pojemności integralnych poszczególnych kombinacji zmiennych objaśniających X , H^0_L - pojemność integralna L -tej kombinacji określona na podstawie n obserwacji. Przyjmuje się jako optymalną kombinację spełniającą warunek: $G = \max G_L$.

⁷⁶ Cieślak M. (red. nauk.), *Prognozowanie gospodarcze Metody i zastosowania*, PWN Warszawa 2005, rozdz. 5.2.

5.4. Bazowanie na mierze błędów *ex ante*

Kolejnym sposobem proponowanym w publikacji „” jest wybór zmiennych objaśniających bazujący na błędzie *ex ante*, który dla modelu trendu liniowego obliczany jest wg poniższego wzoru. Dla modeli nieliniowych dokonujemy wpraw transpozycji na model liniowy.

$$\gamma_T = \left[\frac{(T - \bar{t})^2}{\sum_{t=1}^n (t - \bar{t})^2} + \frac{1}{n} + 1 \right]^{0,5} * s$$

We wzorze tym: T - okres prognozowany, $\bar{t}$ - średnia wartość zmiennej czasowej o n obserwacjach, s - odchylenie standardowe reszt obliczane wg poniższego wzoru gdzie $y^{\wedge}$ - wartość teoretyczna (prognozowana) na podstawie określonego modelu kombinacji zmiennych objaśniających.

$$s = \left[\frac{1}{n - m - 1} \sum_{t=1}^n (y_t - y^{\wedge})^2 \right]^{0,5}$$

Przyjmujemy, że prognoza według danej kombinacji zmiennych X jest dopuszczalna dla $t > n$ jeśli alternatywnie spełnione są warunki:

$$\gamma_T \leq \gamma^* r$$

$$\eta_T \leq \eta^* r$$

Wartości z gwiazdką zostają określone przez odbiorcę prognozy. O wyborze kombinacji może zdecydować minimalna wartość średniego bezwzględnego błędu prognozy *ex ante*, czyli γ_T lub średni procentowy względny błąd prognozy *ex ante* określony jako:

$$\eta_T = \frac{\gamma_T}{y_T} * 100$$

przy czym y_T - wartość zmiennej objaśnianej ustalona po zakończeniu okresu prognozowania T .

5.5. Analiza korelacji

Z tematyką doboru zmiennych objaśniających do modelu ekonometrycznego spośród potencjalnych zmiennych $X_1, X_2, X_j, \dots, X_m$ spotykamy się również w rozdziale 4.2. publikacji Jajugi⁷⁷. Pierwszym krokiem jest zbadanie wystarczającego zróżnicowania danej zmiennej X_j . Miarą zmienności jest współczynnik zmienności $V_j = S_x/X$ czyli stosunek odchylenia standardowego danej zmiennej do średniego poziomu wartości tej zmiennej. Wymagany jest jednak warunek, aby $V_j < V^*$ przy czym krytyczna wartość V^* zwykle przyjmowana jest jako 0,1. We wspomnianej wcześniej pracy pod redakcją K. Jajugi proponuje się dwa sposoby wyboru zmiennych objaśniających do modelu ekonometrycznego: analiza korelacji, metoda pojemności nośników informacji. Pierwsza stanowi analogię do metody wcześniej opisanej, proponowanej dla celów prognostycznych przez Marię Cieślak.

Występuje tu macierz korelacji R między zmiennymi objaśniającymi X_j oraz X_1 (określona na bazie wszystkich n obserwacji próby) oraz wektor korelacji R_0 .

$$R_0 = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix}$$

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix}$$

⁷⁷ Jajuga K. (red.), *Ekonometria Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu. 2002

W kolejnym kroku budowana jest macierz rozszerzona $\mathbf{R}^*$ oraz określony jest współczynnik korelacji wielokrotnej R bazujący na wyznacznikach $\det \mathbf{R}^*$ oraz $\det \mathbf{R}$.

$$R = \sqrt{1 - \frac{\det \mathbf{R}^*}{\det \mathbf{R}}} \quad \mathbf{R}^* = \begin{bmatrix} 1 & \mathbf{R}_0^T \\ \mathbf{R}_0 & \mathbf{R} \end{bmatrix}$$

Współczynnik R kształtuje się w przedziale $[0;1]$. Gdy $R = 0$ to nie ma zależności liniowej, a gdy $R = 1$ to istnieje funkcyjna zależność liniowa między zmienną objaśnianą a określoną kombinacją zmiennych objaśniających. Wartość R stanowi podstawę doboru zmiennych w liniowym modelu ekonometrycznym. Spośród kombinacji zmiennych objaśniających do modelu przyjmuje się tą dla której współczynnik korelacji wielokrotnej ma wartość największą.

5.6. Metoda pojemności nośników informacji

Kolejny sposób zaprezentowany w wymienionej publikacji pod redakcją K. Jajugi to metoda pojemności nośników informacji dla wyboru potencjalnych zmiennych, o której wspomniano wcześniej przy omawianiu propozycji zawartych w publikacji pod redakcją naukową M. Cieślak. Zmienne powinny być możliwie silnie skorelowane ze zmienną objaśnianą, natomiast słabo skorelowane ze sobą. Przyjmuje się, że nośnikiem informacji są wszystkie z m potencjalnych zmiennych objaśniających, dla których istnieje $k = 2^m - 1$ kombinacji tych zmiennych o liczbie m_k . Dla całej próby n obserwacji obliczamy indywidualne pojemności nośników informacji.

W poniższym wzorze: j - numer zmiennej w rozpatrywanej kombinacji ($j = 1, 2, \dots, m_k$); r_j - wartość z wektora korelacji $\mathbf{R}_0$; r_{ij} - wartość z macierzy korelacji $\mathbf{R}$. W kolejności oblicza się również pojemności integralne kombinacji nośników informacji dla $k = 1, 2, \dots, 2^m - 1$.

H_k stanowi kryterium wyboru odpowiedniej kombinacji zmiennych objaśniających, przy czym wybieramy zestaw zmiennych dla których H_k jest maksymalne.

$$h_{Li}^1 = \frac{r_j^2}{\sum_{i=1}^{m_k} I r_{ij} I} \quad H_k = \sum_{j=1}^{m_k} h_{kj}$$

5.7. Metoda grafowa i inne

Szersze zajęcie się tematem wyboru zmiennych do modelu ekonometrycznego spotykamy w rozdziale 2. pracy pod redakcją J. Dziechciarza⁷⁸ Oprócz klasycznego współczynnika zmienności V_j o którym była mowa wcześniej autor proponuje dla uznania, że dana zmienna jest potencjalną zmienną objaśniającą bazowanie na współczynniku opartym na rozstępie W_j . Uważa jednak to jak działanie wstępne wymagające dalszej analizy korelacji zmiennych. Posługuje się w tym względzie w/w współczynnikiem korelacji liniowej r_{ji} Pearsona dla utworzenia macierzy korelacji $\mathbf{R}$ składającej

$$W_j = 2 * \frac{\max_i x_{ij} - \min_i x_{ij}}{\max_i x_{ij} + \min_i x_{ij}} \quad r_{ji} = \frac{Cov_{ji}}{s_j s_i} = \frac{\sum_{i=1}^n (x_{ij} - \bar{x}_j)(x_{it} - \bar{x}_t)}{\sqrt{\sum_{i=1}^n (x_{ij} - \bar{x}_j)^2 \sum_{i=1}^n (x_{it} - \bar{x}_t)^2}}$$

się z współczynników r_{ji} .

⁷⁸ Dziechciarz J. (red.), *Ekonometria Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2003.

$$R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1l} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2l} & \dots & r_{2m} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{jl} & r_{j2} & \dots & r_{jl} & \dots & r_{jm} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r_{ml} & r_{m2} & \dots & r_{ml} & \dots & 1 \end{bmatrix}$$

Potrzeba nam jeszcze określić skorelowanie par dwóch zmiennych - określonej objaśniającej X_i oraz objaśnianej Y . Możemy arbitralnie obrać krytyczną wartość współczynnika korelacji lub uzależnić ją od liczności próby obserwacji. W tym drugim przypadku posługujemy się hipotezą $H_0: r_{xy} = 0$, czyli zakładamy, że zmienne X_i oraz Y nie są istotnie skorelowane i obliczamy statystykę:

$$t = \frac{I r_{xy} I}{\sqrt{1 - r_{xy}^2}} * \sqrt{n - 2}$$

Obliczoną wartość statystyki t porównujemy z t_{α} odczytanym z rozkładu t -Studenta. Przykładowo dla poziomu istotności $\alpha = 0,05$ i ilości stopni swobody $iss = n - 2 = 28 - 2 = 26$ $t_{0,05} = 2,0555$. Gdy $t > t_{\alpha}$ to hipotezę zerową odrzucamy. Zdaniem wymienionego wyżej autora, gdy próba $n > 120$ to należy zastosować rozkład normalny. Statystykę t stosujemy także dla par zmiennych objaśniających X_j oraz X_i . W ten sposób możemy wyeliminować nieistotnie statystycznie skorelowane najpierw zmienne X_j ze zmienną objaśnianą Y , a następnie zbędne pary zmiennych objaśniających. Pozwoli to nam na wstępne sformułowanie liniowego modelu ekonometrycznego.

Alternatywnym podejściem jest oparcie się na współczynniku korelacji wielorakiej, który poznaliśmy przy propozycjach K. Jajugi. W tym względzie określamy dla zmiennych macierz korelacji R , wektor korelacji R_0 . Następnie budujemy macierz rozszerzoną R^* oraz obliczamy współczynnik korelacji wielorakiej bazując na wyznacznikach $\det R^*$ oraz $\det R$.

$$R_0 = \begin{bmatrix} r_1 \\ r_1 \\ \dots \\ r_1 \end{bmatrix} \quad R = \sqrt{1 - \frac{\det R^*}{\det R}}$$

Wartość tego współczynnika stanowi podstawę do doboru jednakowo licznych kombinacji zmiennych objaśniających. Gdy mamy dwa zestawy potencjalnych zmiennych objaśniających to przyjmujemy ten dla którego współczynnik korelacji wielorakiej jest większy. Przykładowo dla 5 zmiennych X_j otrzymujemy 10 kombinacji dwuelementowych oraz 10 kombinacji trzelementowych. Dla poszczególnych kombinacji dwuelementowych obliczamy macierz R , wektor R_0 , macierz rozszerzoną R^* , wyznaczniki macierzy R^* oraz R . Następnie określamy współczynniki korelacji wielorakiej R między Y a poszczególnymi kombinacjami dwuelementowymi np. dla $\{X_1, X_2\}$.

Przyglądamy się wynikom i wybieramy kombinację dwuelementową zmiennych do modelu ekonometrycznego, dla której współczynnik R jest największy. Pozostaje nam jednak postąpić podobnie z kombinacjami trzelementowymi i dopiero bazując na wynikach z tych prac zdecydować o składzie zmiennych w modelu.

Nieco inną propozycję J. Dziechciarza doboru zmiennych objaśniających do modelu ekonometrycznego znajdujemy w rozdziale 2.3. Polega ono na określeniu krytycznej wartości współczynnika korelacji wg wzoru:

$$r^* = \sqrt{\frac{t_{\alpha}^2}{t_{\alpha}^2 + n - 2}}$$

przy czym t_α to statystyka wyznaczona na podstawie rozkładu t -Studenta. Przykładowo dla $\alpha = 0,05$ i $iss = n - 2 = 28 - 2 = 26$ $t_\alpha = 2,0555$. Postępowanie dla wyboru zmiennych objaśniających jest następujące:

- sprawdzamy $|r_{jI}| \leq r^*$ i eliminujemy zmienne X_j nieistotnie skorelowane z Y .
- z pozostałych zmiennych wybieramy X_h według kryterium: badamy relacje powyższej zmiennej z pozostałymi zmiennymi i eliminujemy te dla których $|r_{jI}| > r^*$, gdyż są one zbyt silnie skorelowane ze zmienną X_h .

Powyższe postępowanie prowadzimy do momentu wyczerpania się potencjalnych zmiennych objaśniających. Model ekonometryczny tworzy więc zmienna X_h oraz nie wyeliminowane zmienne.

W pracy J. Dziechciarza spotykamy również omówienie doboru zmiennych objaśniających do modelu metodą wskaźników pojemności informacji. Oprócz nieco innych oznaczeń dla obliczenia takich wielkości jak indywidualna oraz integralna pojemność nośników informacji zaprezentowana procedura jest identyczna z już wcześniej omówioną.

Na uwagę zasługuje jednak metoda grafowa doboru zmiennych objaśniających przedstawiona w rozdziale 2.5. publikacji Dziechciarza *Ekonometria Metody, przykłady, zadania*. Rozpoczynamy od określenia wektora R_0 oraz macierzy R . Następnie wyznaczamy r^* korzystając z powyższego wzoru lub reguły minimaksowej dla $j \neq i$:

$$r^* = \min_i \max_j |r_{ij}|$$

W metodzie grafowej, gdy $|r_{ij}| \leq r^*$ to w miejscu elementów macierzy R spełniających ten warunek wpisujemy zera, tworząc w ten sposób macierz R^* . W kolejnym etapie rysujemy graf, którego wierzchołkami są zmienne, a wiązkami niezerowe współczynniki korelacji. Do modelu wchodzi zmienne odosobnione (nie związane z innymi elementami grafu) oraz te zmienne, które mają najwięcej wzajemnych wiązań. Jeśli ilość wiązań jest taka sama to wybieramy zmienną o największym r_{ji} .

Powyższą metodę zaprezentowano na przykładzie⁷⁹ pięciu zmiennych objaśniających X_1, X_2, X_3, X_4, X_5 zbioru $n = 30$ obserwacji skorelowanych wzajemnie oraz skorelowanych ze zmienną objaśnianą Y . Niech macierz symetryczna R i wektor R_0 będą następujące:

$$R = \begin{bmatrix} 1 & 0,8 & 0,3 & 0,6 & 0,1 \\ 0,8 & 1 & 0,4 & 0,9 & 0,6 \\ 0,3 & 0,4 & 1 & 0,2 & 0,5 \\ 0,6 & 0,9 & 0,2 & 1 & 0,7 \\ 0,1 & 0,6 & 0,5 & 0,7 & 1 \end{bmatrix} \quad R_0 = \begin{bmatrix} 0,6 \\ 0,8 \\ 0,5 \\ 0,9 \\ 0,8 \end{bmatrix}$$

Dla $\alpha = 0,05$, $iss = n - 2 = 30 - 2 = 28$ $t_\alpha = 2,0484$

$$r^* = \sqrt{\frac{2,0484^2}{2,0484^2 + 30 - 2}} = 0,36$$

lub dla $j \neq i$

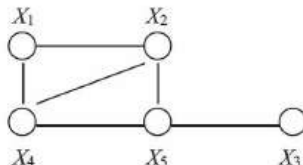
$$r^* = \min_i \max_j |r_{ij}| = 0,5$$

⁷⁹ Przykład ten jest analogią do przykładu 2.5 zamieszczonego w publikacji: Dziechciarz J. (red.), *Ekonometria Metody, przykłady, zadania*, op. cit.

Dla $r^* = 0,36$ i $|r_{ij}| \leq r^*$

$$R' = \begin{bmatrix} 1 & 0,8 & 0 & 0,6 & 0 \\ 0,8 & 1 & 0,4 & 0,9 & 0,6 \\ 0 & 0,4 & 1 & 0 & 0,5 \\ 0,6 & 0,9 & 0 & 1 & 0,7 \\ 0 & 0,6 & 0,5 & 0,7 & 1 \end{bmatrix}$$

Macierz R' o współczynnikach korelacji zerowych dla $|r_{ij}| \leq 0,36$ umożliwia zbudowanie grafu powiązań między zmiennymi.

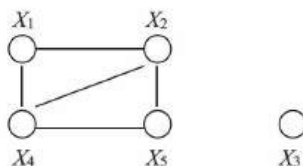


Zmienne X_2 , X_4 i X_5 mają po 3 wiązadła, przy czym najsilniej skorelowana według R_0 ze zmienną objaśnianą jest X_4 ($r_4 = 0,9$), zatem model jest następujący:

$$Y^{\wedge} = a_0 + a_1 X_2 + a_2 X_4$$

Zbudujmy graf w oparciu o kryterium $r^* = 0,5$ uzyskane z reguły minimaksowej. Kryterium $|r_{ij}| \leq 0,5$ przyjęto jako zero.

$$R' = \begin{bmatrix} 1 & 0,8 & 0 & 0,6 & 0 \\ 0,8 & 1 & 0 & 0,9 & 0,6 \\ 0 & 0 & 1 & 0 & 0 \\ 0,6 & 0,9 & 0 & 1 & 0,7 \\ 0 & 0,6 & 0 & 0,7 & 1 \end{bmatrix}$$



Zmienna X_3 jest odosobniona (izolowana), a więc wchodzi do modelu opisującego badane zjawisko, natomiast zmienne X_2 , X_4 mają po 3 wiązadła. Według R_0 najsilniej skorelowana jest zmienna X_4 , tak więc liniowy model ekonometryczny jest postaci:

$$Y^{\wedge} = a_0 + a_1 X_3 + a_2 X_4$$

5.8. Testy porównywania modeli

W literaturze „Ekonometrii” spotykamy testy umożliwiające porównywanie modeli służących do prognozowania. Jak się wydaje mogą one posłużyć również do wyboru wariantów wg tego samego rodzaju liniowego modelu ekonometrycznego, utworzonych z różnych kombinacji zmiennych objaśniających. Gdy zastosujemy transformację liniową to po estymacji parametrów modelu możemy zastosować testy bazujące na sumie kwadratów reszt również do modeli nieliniowych. Zamieszczone poniżej testy Akaike’a (AK) oraz Schwarza (SCH) dla n obserwacji stosowane są gdy występują

podobne wartości miary oceny dopasowania modeli do danych empirycznych oraz jeśli spełnione są testy istotności i stabilności parametrów⁸⁰:

$$AK = n \ln (\Sigma z_n^2) + 2k$$

$$SCH = n \ln (\Sigma z_n^2) + k \ln(n),$$

przy czym:

k - liczba parametrów modelu o n obserwacjach,

Σz_n^2 - suma kwadratów reszt jako różnicy wartości z obserwacji Y_n i wartości teoretycznej $Y_n^\wedge$ obliczonej na podstawie modelu (przykładowo danej kombinacji zmiennych objaśniających)

$$z_n = Y_n - Y_n^\wedge$$

Z grupy badanych różnych modeli powinniśmy wybrać model o najniższej wartości kryterium AK lub SCH obliczonego dla wszystkich modeli⁸¹.

5.9. Metoda regresji krokowej oraz minimalizacji resztowej sumy kwadratów

Tematyka doboru zmiennych objaśniających znajduje się również w rozdziale 12.7 dzieła G.S. Maddala „Ekonometria”⁸². Spośród k potencjalnych zmiennych objaśniających wybieramy podzbiór według określonego kryterium. Powyższy autor nadmienia, że w latach 60-tych powstawały metody regresji krokowej, które polegały na eliminowaniu zmiennych o statystykach t niższych od z góry ustalonych, ze wstępnego równania regresji o wszystkich k zmiennych objaśniających. Podejściem alternatywnym było startowanie z pojedynczą zmienną o najwyższym r_j względem zmiennej objaśnianej Y . W kolejnych krokach wybierano do modelu zmienną o najwyższym współczynniku korelacji cząstkowej z Y . Metody te nazwane przez powyższego autora „mechanicznymi” nie są obecnie stosowane, lecz wykorzystuje się je do modeli regresji jako jedno z kryteriów statystycznych, które minimalizuje wyrażenia podane w tabeli 5.1.

Tab. 5.1. Wybrane kryteria wyboru między modelami regresji

Kryterium	Wyrażenie
R^{-2} Thiela	$RSS_j / (n - k_j)$
S_p Hockinga	$RSS_j / [(n - k_j)(n - k_j - 1)]$
C_p Mallowsa	$RSS_j + 2 k_j \hat{\sigma}_m^2$
PC Amemiya	$RSS_j (n + k_j) / (n - k_j)$
AIC Akaike'a	$RSS_j \exp[2(k_j + 1)/n]$

Źródło: Według tabeli 12.3 w publikacji: Maddala G.S., *Ekonometria*, op. cit.

Powyższe wyrażenia bazują na resztowej sumie kwadratów RSS modelu j (utworzonego przykładowo na bazie określonej kombinacji zmiennych objaśniających). Określenie $\hat{\sigma}_m^2$ - oznacza wariancję resztową $RSS_j / (n - k)$ w modelu o wszystkich k zmiennych objaśniających. Na podstawie wartości z minimalizowanych wyrażeń dla określonego kryterium podanego w powyższej tabeli wybieramy model (kombinację zmiennych objaśniających) najlepiej dopasowaną do wartości empirycznej Y z n obserwacji, które stanowiły podstawę estymacji parametrów niezbędnej do obliczenia modelowej $Y^\wedge$.

⁸⁰ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom II, *Ekonometria*. C.H. Beck Warszawa 2003.

⁸¹ Ibidem, rozdz. 1.10. *Testy umożliwiające porównywanie modeli*.

⁸² Maddala G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.

5.10. Przykłady zastosowania metody Hellwiga

W metodzie pojemności nośników informacji tzw. metodzie Hellwiga, postępowanie jest analogiczne do wcześniej omówionej, lecz dla realizacji przykładu zostało ponownie przytoczone:

a) Określenie macierzy współczynników korelacji pomiędzy zmiennymi objaśniającymi oraz wektora współczynników korelacji zmiennych objaśniających ze zmienną objaśnianą.

b) Ustalenie ilości kombinacji zmiennych objaśniających wg wzoru $L = 2^n - 1$ gdzie: n - liczba zmiennych objaśniających.

c) obliczenie dla każdej kombinacji I i określonej zmiennej j indywidualnego wskaźnika pojemności informacji z zależności:

$$h_{ij} = \frac{r_{ij}^2}{1 + \sum_{I=1}^{m_i} I r_{ij} I}$$

Gdzie: $I = 1, \dots, L$ - numery kombinacji, $j = 1, \dots, m_j$, m_i - liczba zmiennych w kombinacji " I ", r_{ij} - współczynnik korelacji między zmiennymi o indeksach " i " oraz " j ".

d) obliczenie integralnego wskaźnika pojemności poszczególnych kombinacji:

$$H_i = \sum_{j=1}^{m_j} h_{ij}$$

e) Wybranie kombinacji zmiennych, dla których H_i jest największe: $H^* = \max \{H_i\}$

Dla zmiennych objaśniających X_1, X_2, X_3, X_4 dany jest poniższy wektor i macierz korelacji o wartościach umownych. Ze względu na ramy tego materiału i pracochłonność pełnej procedury obliczeń, w tym przykładzie określono tylko liczbę kombinacji oraz przykładowo integralny wskaźnik pojemności informacji dla kombinacji piątej.

$$[r_j] = 1 \ 0,45 \ -0,70 \ 0,16 \ 0,6 \ 1$$

$$[r_{ij}] = \begin{vmatrix} 1 & -0,6 & 0,15 & 0,44 \\ -0,6 & 1 & -0,11 & -0,5 \\ 0,15 & -0,11 & 1 & -0,05 \\ 0,44 & -0,5 & -0,05 & 1 \end{vmatrix}$$

Liczba kombinacji czterech zmiennych objaśniających $L = 2^n - 1 = 2^4 - 1 = 15$.

Tab. 5.2. Kombinacje zmiennych objaśniających

1: $\{X_1\}$	9: $\{X_2, X_4\}$
2: $\{X_2\}$	10: $\{X_3, X_4\}$
3: $\{X_3\}$	11: $\{X_1, X_2, X_3\}$
4: $\{X_4\}$	12: $\{X_1, X_2, X_4\}$
5: $\{X_1, X_2\}$	13: $\{X_1, X_3, X_4\}$
6: $\{X_1, X_3\}$	14: $\{X_2, X_3, X_4\}$
7: $\{X_1, X_4\}$	15: $\{X_1, X_2, X_3, X_4\}$
8: $\{X_2, X_3\}$	

Źródło: Opracowanie własne.

Indywidualne wskaźniki pojemności informacji dla kombinacji piątej h_{51} i h_{52} :

$$h_{51} = \frac{(0,45)^2}{1 + 0,6} = 0,127$$

$$h_{52} = \frac{(-0,7)^2}{1 + 0,6} = 0,306$$

Integralny wskaźnik dla kombinacji piątej: $H_5 = 0,127 + 0,306 = 0,433$

Gdybyśmy rozpatrzyli wszystkie kombinacje, to wtedy możemy określić maksymalną wartość integralnej pojemności wskaźnika informacji H^* , dla której zestaw zmiennych stanowi najlepsze rozwiązanie dla modelu ekonometrycznego. Mając określone zmienne objaśniające zapisujemy postać analityczną modelu regresji liniowej:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + \dots + a_n X_n + u$$

i przystępujemy do oszacowania poszczególnych parametrów a_i tego modelu o postaci ogólnej:

$$\hat{Y} = aX + a_0$$

Znając powyższą procedurę metody pojemności informacji możemy spróbować samodzielnie wybrać najlepszy zestaw spośród trzech zmiennych objaśniających X_1, X_2, X_3 mając dany kolejny wektor i macierz korelacji:

$$[r_{ij}] = \begin{bmatrix} 1 & 0,5 & -0,6 & 0,8 \\ 0,5 & 1 & -0,2 & 0,1 \\ -0,6 & -0,2 & 1 & -0,1 \\ 0,8 & 0,1 & -0,1 & 1 \end{bmatrix}$$

$$[r_{ij}] = \begin{bmatrix} 1 & -0,6 & 0,2 & 0,5 \\ -0,6 & 1 & -0,2 & -0,5 \\ 0,2 & -0,2 & 1 & -0,1 \\ 0,5 & -0,5 & -0,1 & 1 \end{bmatrix}$$

W kolejnym zadaniu mamy cztery zmienne objaśniające charakteryzujące wstępnie poziom życia mieszkańców w okresie piętnastu lat:

X_1 - liczba przestępstw w tys., X_2 - dochody ludności w bln zł,

X_3 - spożycie alkoholu w litrach spirytusu/osobę, X_4 - liczba rozwodów na 100 osób.

Tab. 5.3. Wartości zmiennych

n	X_1	X_2	X_3	X_4
1	9	4,4	424	1,1
2	9,2	4,5	414	1,1
3	9,5	4,9	432	1,1
4	9,7	5,1	414	1,2
5	10,1	5	410	1,3
6	10,8	4,9	394	1,3
7	10,5	4,8	375	1,2
8	10,7	4,6	366	1,2
9	11,1	4,2	348	1,1
10	10,3	4,2	333	1,1
11	9,8	4,3	338	1,1
12	10,1	4,5	380	1,1
13	11,6	3,7	436	1,3
14	11,9	3,7	466	1,4
15	12,3	3,8	539	1,4
16	12,4	3,9	569	1,5

Źródło: Opracowanie własne bazujące na fragmencie tabeli 6.6. część 2 publikacji: Jajuga K. (red.), *Ekonometria Metody i analiza problemów ekonomicznych*, op. cit.

5.11. Przykład określenia zmiennych objaśniających

W konstruowaniu jednorównaniowych modeli zależności istotną rolę odgrywa właściwy dobór potencjalnych zmiennych objaśniających spośród zasugerowanych. W tym celu jak już wspomniano, posługujemy się współczynnikiem zmienności a następnie możemy zastosować przykładowo analizę korelacyjną. Z tego względu zajmiemy się, w kolejnym przykładzie, zmierzaniem poziomu zróżnicowania obserwacji oraz określeniem współczynnika korelacji wielorakiej. Współczynnik ten jest miarą zależności między zmienną objaśnianą, a liniową kombinacją zmiennych objaśniających. We wstępnym formułowaniu zależności zmiennej Y od zmiennych X_m występuje przeważnie nadmierna ilość zmiennych objaśniających słabo związanych ze zmienną objaśnianą. Zmienne dla których współczynnik zmienności $V_j < V^*$ uznaje się za mało zróżnicowane eliminuje się ze zbioru potencjalnych zmiennych objaśniających. Dla przypomnienia współczynnik zmienności jest ilorazem odchylenia standardowego do średniej wartości obserwacji danej zmiennej X_j . Według publikacji *Ekonometria Metody i analiza problemów ekonomicznych* V_x oraz s_x oblicza się wg wzorów⁸³:

$$V_x = \frac{s_x}{\bar{x}} \quad s_x^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n} \quad \bar{X} = \frac{\sum_{i=1}^n x_i}{n}$$

Obliczenie współczynników zmienności z zastosowaniem arkusza kalkulacyjnego *Excel* przedstawiono na przykładzie opisu wartości produkcji stolarki okiennej w mln zł w firmie X przyjmując cztery zmienne objaśniające: X_1 - zatrudnienie w osobach, X_2 - wartość maszyn i urządzeń w tys. zł, X_3 - czas przestoju maszyn w dnach, X_4 - nakłady inwestycyjne. Dobierzmy zmienne objaśniające przyjmując wartość krytyczną $V^* = 12\%$.

Tab. 5.4. Obliczenia pomocnicze do określenia średniej zmiennych

i	Y	X_1	X_2	X_3	X_4
1	12	8	7	12	10
2	10	6	8	14	12
3	18	11	10	16	11
4	16	10	12	18	14
5	12	8	8	18	10
6	14	10	8	18	12
7	18	11	12	20	13
8	20	12	16	24	12
9	20	12	16	26	12
10	22	14	18	26	10
Suma	162	102	115	192	116
Średnia	16,2	10,2	11,5	19,2	11,6

Źródło: Opracowanie własne - dane umowne.

⁸³ Jajuga K. (red.), *Ekonometria Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, rozdz. 1.3.

Tab. 5.5. Obliczenia pomocnicze sumy kwadratu różnic i odchylenia standardowego

B	C	D	E	F	G	H	I	J
i	$(x_1 - \bar{x}_1)$	$(x_1 - \bar{x}_1)^2$	$(x_2 - \bar{x}_2)$	$(x_2 - \bar{x}_2)^2$	$(x_3 - \bar{x}_3)$	$(x_3 - \bar{x}_3)^2$	$(x_4 - \bar{x}_4)$	$(x_4 - \bar{x}_4)^2$
1	-2,2	4,84	-4,5	20,25	-7,2	51,84	-1,6	2,56
2	-4,2	17,64	-3,5	12,25	-5,2	27,04	0,4	0,16
3	0,8	0,64	-1,5	2,25	-3,2	10,24	-0,6	0,36
4	-0,2	0,04	0,5	0,25	-1,2	1,44	2,4	5,76
5	-2,2	4,84	-3,5	12,25	-1,2	1,44	-1,6	2,56
6	-0,2	0,04	-3,5	12,25	-1,2	1,44	0,4	0,16
7	0,8	0,64	0,5	0,25	0,8	0,64	1,4	1,96
8	1,8	3,24	4,5	20,25	4,8	23,04	0,4	0,16
9	1,8	3,24	4,5	20,25	6,8	46,24	0,4	0,16
10	3,8	14,44	6,5	42,25	6,8	46,24	-1,6	2,56
Suma		49,6		142,5		209,6		16,4

s ₁	s ₂	s ₃	s ₄
2,23	3,77	4,58	1,28
V1	V2	V3	V4
21,83	32,83	23,84	11,04
T	T	T	N

Przykłady wyrażeń:
=PIERWIASTEK(D28/10)
=(C31/D14)*100
=JEŻELI(C33>12;"T";"N")

Źródło: Obliczenia własne z zastosowaniem wyrażeń *Excela*.

Na końcu tabeli zastosowano funkcję badania logicznego, czy obliczony współczynnik zmienności V_j jest większy od przyjętego 12%. Zatem potencjalną zmienną X_4 z punktu widzenia współczynnika zmienności możemy pominąć w konstruowaniu modelu ekonometrycznego. Zastosowanie współczynnika zmienności V_j jest tylko wstępną selekcją zmiennych objaśniających. Dalszy wybór zmiennych spośród X_1, X_2, X_3 możemy przeprowadzić stosując *analizę korelacyjną*, która bazuje na współczynnikach korelacji między zmiennymi. W przykładzie zastosowano wzory z publikacji *Ekonometria Metody i analiza problemów ekonomicznych*⁸⁴. Wprowadźmy oznaczenia oraz przypomnijmy sobie wzory omówione wcześniej:

r_j - współczynnik korelacji między zmienną objaśnianą, a zmienną objaśniającą X_j ,

r_{ji} - współczynnik korelacji między zmiennymi objaśniającymi X_j oraz X_i ,

R_0 - wektor współczynników korelacji między zmienną objaśnianą Y a zmiennymi objaśniającymi X_j ,

R - macierz korelacji między zmiennymi objaśniającymi,

R^* - macierz rozszerzona,

R_k - współczynnik korelacji wielorakiej.

$$R_0 = \begin{bmatrix} r_1 \\ r_2 \\ \dots \\ r_n \end{bmatrix} \quad R = \begin{bmatrix} 1 & r_{12} & \dots & r_{1m} \\ r_{21} & 1 & \dots & r_{2m} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & 1 \end{bmatrix} \quad R^* = \begin{bmatrix} 1 & R_0^T \\ R_0 & R \end{bmatrix} \quad R_k = \sqrt{1 - \frac{\det R^*}{\det R}}$$

Przy dwóch zestawach potencjalnych zmiennych X_j wybiera się ten zestaw, dla którego współczynnik korelacji wielorakiej ma większą wartość.

⁸⁴ Jajuga K. (red.), *Ekonometria Metody i analiza problemów ekonomicznych*, op. cit., rozdz. 4.2.

Bazując na tabeli 5.4 w zakresie zmiennych X_1, X_2, X_3 obliczamy współczynniki korelacji między daną zmienną objaśniającą a y , dla otrzymania wektora R_0 , macierzy R , macierzy rozszerzonej R^* i obliczenia współczynnika korelacji wielorakiej. Współczynniki korelacji między zmienną objaśnianą a kolejnymi zmiennymi objaśniającymi obliczamy wg wzoru:

$$r_{xy} = \frac{\sum_{i=1}^n (x_{iy} - \bar{x}_i)(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_{iy} - \bar{x}_i)^2 * \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Tab. 5.6. Obliczenia pomocnicze do współczynników korelacji liniowej między X_j a y

4	16	10	12	18	-0,2	0,04	-0,2	0,04	0,04
5	12	8	8	18	-4,2	17,64	-2,2	4,84	9,24
6	14	10	8	18	-2,2	4,84	-0,2	0,04	0,44
7	18	11	12	20	1,8	3,24	0,8	0,64	1,44
8	20	12	16	24	3,8	14,44	1,8	3,24	6,84
9	20	12	16	26	3,8	14,44	1,8	3,24	6,84
10	22	14	18	26	5,8	33,64	3,8	14,44	22,04
Suma	162	102	115	192		147,6		49,6	83,6
Średnia	16,2	10,2	11,5	19,2					
	$\bar{Y}$	$\bar{X}_1$	$\bar{X}_2$	$\bar{X}_3$					

R_{xy}	R_{x_1y}	R_{x_2y}
0,98	0,92	0,85

Tab. 5.6. Obliczenia pomocnicze do współczynników korelacji liniowej między X_j a y (c.d.)

$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$	$(x_2 - \bar{x}_2) * (y - \bar{y})$	$x_3 - \bar{x}_3$	$(x_3 - \bar{x}_3)^2$	$(x_3 - \bar{x}_3) * (y - \bar{y})$
-4,5	20,25	18,9	-7,2	51,84	30,24
-3,5	12,25	21,7	-5,2	27,04	32,24
-1,5	2,25	-2,7	-3,2	10,24	-5,76
0,5	0,25	-0,1	-1,2	1,44	0,24
-3,5	12,25	14,7	-1,2	1,44	5,04
-3,5	12,25	7,7	-1,2	1,44	2,64
0,5	0,25	0,9	0,8	0,64	1,44
4,5	20,25	17,1	4,8	23,04	18,24
4,5	20,25	17,1	6,8	46,24	25,84
6,5	42,25	37,7	6,8	46,24	39,44
	142,5	133		209,6	149,6

Źródło: Powyższa tabela oraz obliczenie współczynników korelacji między zmiennymi objaśniającymi a y - opracowanie własne.

W obliczeniu R_{xy} zastosowano wyrażenie *Excela* przykładowo dla zmiennej X_1 : =O48/PIER-WIASTEK((N48*L48)), przy czym adres O48 oznacza komórkę arkusza z wartością obliczoną dla:

$$\sum_{i=1}^n (x_{iy} - \bar{x}_i)(y_i - \bar{y})$$

a adresy N48 oraz L48 wartości komórek z obliczeniami:

$$\sum_{i=1}^n (y_i - \bar{y})^2 \quad \sum_{i=1}^n (x_{iy} - \bar{x}_i)^2$$

Następnie obliczamy współczynniki korelacji między zmiennymi objaśniającymi j -tą a l -tą.

$$r_{x_j x_l} = \frac{\sum_{i=1}^n (x_{ji} - \bar{x}_j)(x_{li} - \bar{x}_l)}{\sqrt{\sum_{i=1}^n (x_{ji} - \bar{x}_j)^2 * \sum_{i=1}^n (x_{li} - \bar{x}_l)^2}}$$

Tab. 5.7. Obliczenia pomocnicze dla określenia współczynnika korelacji między zmiennymi objaśniającymi

i	$x_1 - \bar{x}_1$	$(x_1 - \bar{x}_1)^2$	$x_2 - \bar{x}_2$	$(x_2 - \bar{x}_2)^2$	$x_3 - \bar{x}_3$	$(x_3 - \bar{x}_3)^2$	$(x_1 - \bar{x}_1) * (x_2 - \bar{x}_2)$	$(x_1 - \bar{x}_1) * (x_3 - \bar{x}_3)$	$(x_2 - \bar{x}_2) * (x_3 - \bar{x}_3)$
1	-2,2	4,84	-4,5	20,25	-7,2	51,84	9,9	15,84	32,4
2	-4,2	17,64	-3,5	12,25	-5,2	27,04	14,7	21,84	18,2
3	0,8	0,64	-1,5	2,25	-3,2	10,24	-1,2	-2,56	4,8
4	-0,2	0,04	0,5	0,25	-1,2	1,44	-0,1	0,24	-0,6
5	-2,2	4,84	-3,5	12,25	-1,2	1,44	7,7	2,64	4,2
6	-0,2	0,04	-3,5	12,25	-1,2	1,44	0,7	0,24	4,2
7	0,8	0,64	0,5	0,25	0,8	0,64	0,4	0,64	0,4
8	1,8	3,24	4,5	20,25	4,8	23,04	8,1	8,64	21,6
9	1,8	3,24	4,5	20,25	6,8	46,24	8,1	12,24	30,6
10	3,8	14,44	6,5	42,25	6,8	46,24	24,7	25,84	44,2
Suma		49,6		142,5		209,6	73	85,6	160
Średnia									

$R_{x_1 x_2}$	$R_{x_1 x_3}$	$R_{x_2 x_3}$
0,87	0,84	0,93

Źródło: Opracowanie własne w Excelu.

Na podstawie powyższych obliczeń wektor, macierz korelacji oraz macierz rozszerzona są następujące:

$$R_0 = \begin{bmatrix} 0,98 \\ 0,92 \\ 0,85 \end{bmatrix} \quad R = \begin{bmatrix} 1 & 0,87 & 0,84 \\ 0,87 & 1 & 0,93 \\ 0,84 & 0,93 & 1 \end{bmatrix} \quad R^* = \begin{bmatrix} 1 & 0,98 & 0,92 & 0,85 \\ 0,98 & 1 & 0,87 & 0,84 \\ 0,92 & 0,87 & 1 & 0,93 \\ 0,85 & 0,84 & 0,93 & 1 \end{bmatrix}$$

Wyznaczniki macierzy R^* oraz macierzy R z zastosowanie funkcji *Excela* np. dla określonego R^* w arkuszu kalkulacyjnym *Excel*.

Tab. 5.8. Obliczenia pomocnicze i określenie wyznaczników

	C	D	E	F
55	1	0,98	0,92	0,85
56	0,98	1	0,87	0,84
57	0,92	0,87	1	0,93
58	0,85	0,84	0,93	1

0,00047 det R^* =WYZNACZNIK.MACIERZY(C55:F58)

0,03189 det R =WYZNACZNIK.MACIERZY(D56:F58)

Źródło: Opracowanie własne w Excelu.

Dla wyznaczenia det R można też zastosować regułę Sarrusa dla macierzy 3-go stopnia.

1	0,87	0,84	1	0,87
0,87	1	0,93	0,87	1
0,84	0,93	1	0,84	0,93

det R = 0,03189

$$= (1 \cdot 1 \cdot 1 + 0,87 \cdot 0,93 \cdot 0,84 + 0,84 \cdot 0,87 \cdot 0,93) - (0,84 \cdot 1 \cdot 0,84 + 1 \cdot 0,93 \cdot 0,93 + 0,87 \cdot 0,87 \cdot 1)$$

Natomiast obliczenie $\det \mathbf{R}^*$ dla macierzy 4-go stopnia wymaga wcześniejszego doprowadzenia do macierzy stopnia trzeciego poprzez wykorzystanie rozwinięcia Laplace'a względem wskazanego wiersza lub kolumny⁸⁵ podanego na stronie 65 lub algorytmu Chio' wymienionego na stronie 72 podanej w przypisie publikacji.

Określenie współczynnika korelacji wielorakiej

$$R_k = \sqrt{1 - \frac{\det \mathbf{R}^*}{\det \mathbf{R}}} = \sqrt{1 - \frac{0,00047}{0,03189}} = 0,9926$$

Współczynnik ten wskazuje na duże natężenie zależności liniowej między zmienną objaśnianą Y , a trzema zmiennymi objaśniającymi X_1, X_2, X_3 . Przyjęcie trzech zmiennych objaśniających X_1, X_2, X_3 jednocześnie w relacji ze zmienną objaśnianą Y nie daje odpowiedzi na pytanie, którą z nich wyeliminować. Dlatego współczynnik korelacji wielorakiej oblicza się kolejno według procedury podanej w powyższym przykładzie dla zestawów zmiennych objaśniających: $X_1 X_2$; $X_1 X_3$; $X_2 X_3$. Jak już wspomniano do modelu przyjmuje się ten zestaw, dla którego współczynnik korelacji wielorakiej ma wartość największą.

Zostawmy na chwilę pracochłonną analizę korelacyjną, a spróbujmy teraz zastosować metodę pojemności nośników informacji (o której wspomniano wcześniej). Bazuje ona na wektorze i macierzy umownych trzech zmiennych objaśniających X_1, X_2, X_3 i procedura jest następująca:

Określenie możliwych kombinacji zmiennych objaśniających: $L = 2^n - 1 = 2^3 - 1 = 7$.

Tab. 5.9. Kombinacje trzech zmiennych objaśniających

Kombinacja (k)	Liczba zmiennych (m_k)
1: $\{X_1\}$	1
2: $\{X_2\}$	1
3: $\{X_3\}$	1
4: $\{X_1, X_2\}$	2
5: $\{X_1, X_3\}$	2
6: $\{X_2, X_3\}$	2
7: $\{X_1, X_2, X_3\}$	3

Źródło: Opracowanie własne.

Obliczenie indywidualnych pojemności nośników informacji h_{ij} dla każdej zmiennej w kombinacji k , a następnie integralnej pojemności nośników informacji H_k korzystając ze wzorów:

$$h_{ij} = \frac{r_j^2}{1 + \sum_{j=1}^{m_k} I r_{ij} I} \quad H_k = \sum_{j=1}^{m_k} h_{ij}$$

Przy czym: r_j - wartość z wektora $\mathbf{R}_0$, r_{ij} - wartość z macierzy $\mathbf{R}$, m_k - liczba zmiennych w k -tej kombinacji.

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,92 \\ 0,85 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 0,87 & 0,84 \\ 0,87 & 1 & 0,93 \\ 0,84 & 0,93 & 1 \end{bmatrix}$$

⁸⁵ Jurliewicz T., Skoczylas J., *Algebra liniowa 1. Przykłady i zadania*. Wydawnictwo GiS. Wrocław 2003.

Tab. 6.10. Obliczenia pomocnicze do wyboru kombinacji zmiennych

k	Zmienne	r_j	r_j^2	$ r_{ij} $	$\sum r_{ij} $	h_{ij}	H_k
1	X_1	0,98	0,960	1	1	0,960	0,960
2	X_2	0,92	0,846	1	1	0,846	0,846
3	X_3	0,85	0,723	1	1	0,723	0,723
4	$X_1 X_2$	0,98	0,960	$11 + 0,87$	1,87	0,513	
4	$X_1 X_2$	0,92	0,846	$11 + 0,871$	1,87	0,452	0,965
5	$X_1 X_3$	0,98	0,960	$11 + 0,841$	1,84	0,522	
5	$X_1 X_3$	0,85	0,723	$11 + 0,841$	1,84	0,393	0,915
6	$X_2 X_3$	0,92	0,846	$11 + 0,931$	1,93	0,438	
6	$X_2 X_3$	0,85	0,723	$11 + 0,931$	1,93	0,375	0,813
7	$X_1 X_2 X_3$	0,98	0,960	$11 + 0,87 + 0,841$	2,71	0,354	
7	$X_1 X_2 X_3$	0,92	0,846	$11 + 0,87 + 0,841$	2,71	0,312	
7	$X_1 X_2 X_3$	0,85	0,723	$11 + 0,87 + 0,841$	2,71	0,267	0,933

Źródło: Opracowanie własne.

Wybór zestawu zmiennych dla których H_k jest największe.

$$H_{\max} = \{0,960; 0,846; 0,723; 0,965; 0,915; 0,813; 0,933\} = 0,965$$

Model zatem jest następujący: $Y^{\wedge} = a_1 * X_1 + a_2 * X_2$

Powróćmy do wcześniejszego przykładu i sprawdźmy analizą korelacyjną zestawu zmiennych $X_1 X_2, X_1 X_3, X_2 X_3$. Interesuje nas z pewnością, czy kombinacja zmiennych $X_1 X_2$ w tej metodzie również będzie najlepsza. Zatem rozpatrzmy zestaw $X_1 X_2$.

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,92 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 0,87 \\ 0,87 & 1 \end{bmatrix} \quad \mathbf{R}^* = \begin{bmatrix} 1 & 0,98 & 0,92 \\ 0,98 & 1 & 0,87 \\ 0,92 & 0,87 & 1 \end{bmatrix}$$

Zastosujmy funkcję *Excela* (=WYZNACZNIK.MACIERZY(...)) W nawiasie obszar adresowy macierzy $\mathbf{R}^*$ lub $\mathbf{R}$: $\det \mathbf{R}^* = 0,0051$ $\det \mathbf{R} = 0,2431$ i obliczamy współczynnik korelacji wielorakiej.

$$R_k = \sqrt{1 - \frac{\det \mathbf{R}^*}{\det \mathbf{R}}} = \sqrt{1 - \frac{0,0051}{0,2431}} = 0,9895$$

A teraz sprawdźmy dla zestawu $X_1 X_3$:

$$\mathbf{R}_0 = \begin{bmatrix} 0,98 \\ 0,85 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 0,84 \\ 0,84 & 1 \end{bmatrix}$$

$$\det \mathbf{R}^* = 0,0109 \quad \det \mathbf{R} = 0,2944$$

$$R_k = \sqrt{1 - \frac{0,0109}{0,2944}} = 0,9813$$

I jeszcze trzeci zestaw $X_2 X_3$:

$$\mathbf{R}_0 = \begin{bmatrix} 0,92 \\ 0,85 \end{bmatrix} \quad \mathbf{R} = \begin{bmatrix} 1 & 0,93 \\ 0,93 & 1 \end{bmatrix}$$

$$\det \mathbf{R}^* = 0,0207 \quad \det \mathbf{R} = 0,1351$$

$$R_k = \sqrt{1 - \frac{0,0207}{0,1351}} = 0,9202$$

Z porównania par zmiennych najwyższy współczynnik korelacji wielorakiej ma zestaw $X_1 X_2$ dla którego $R_k = 0,9895$. Jest to zgodne z wynikiem metody pojemności nośników informacji, lecz nieco mniejsze od zestawu trzech zmiennych $X_1 X_2 X_3$ o $R_k = 0,9926$. Reasumując przyjmijmy model liniowy o trzech zmiennych objaśniających:

$$\hat{Y} = a_1 X_1 + a_2 X_2 + a_3 X_3$$

5.12. Procedura wyboru zmiennych dla modelu regresji z 4-ma zmiennymi objaśniającymi

W poniższej tabeli: Y - produkcja przedsiębiorstwa w tys. szt, X_1 (tys. osób) - zatrudnienie, X_2 (kW) - moc zainstalowanych maszyn, X_3 (tys. ton) - dostawa surowców, X_4 (dni) - przestoje z powodu awarii.

Tab. 5.11. Dane umowne zmiennych

t	Y	X_1	X_2	X_3	X_4
1	18,5	4,4	10,5	2,80	8
2	19,2	4,3	10,0	2,83	10
3	20,0	4,5	11,7	3,01	15
4	20,8	4,5	12,1	3,17	12
5	21,7	4,4	13,0	3,42	7
6	20,8	4,6	13,9	3,56	10
7	23,7	4,6	15,1	3,73	11
8	25,0	4,6	13,9	3,81	13
9	26,7	4,5	18,1	4,07	8
10	28,3	4,5	20,0	4,32	17

Źródło: Opracowanie własne.

Dla modelu: $\ln Y = f(\ln X_1, \dots, \ln X_n, c)$ czyli $\ln Y = c + a_1 \ln X_1 + \dots + a_n \ln X_n$ określamy podstawowe parametry zmiennej objaśnianej Y , zmiennych objaśniających X_i w postaci zlogarytmowanej obliczając wcześniej dane podane w poniższej tabeli.

Tab. 5.12. Ramka do obliczenia cech statystyki opisowej zmiennych zlogarytmowanych

$\ln \dots$	średnia	odchylenie standardowe	maksimum	minimum
$\ln Y$				
$\ln X_1$				
$\ln X_2$				
$\ln X_3$				
$\ln X_4$				

Źródło: Opracowanie własne.

Wyznaczamy współczynniki kowariancji i korelacji pomiędzy logarytmem zmiennej objaśnianej a logarytmami zmiennych objaśniających

Tab. 5.13. Ramka do obliczenia kowariancji i korelacji

Lp.	Zestaw	Kowariancja	Korelacja
1	$\ln Y, \ln Y$		
2	$\ln Y, \ln X_1$		
3	$\ln Y, \ln X_2$		
4	$\ln Y, \ln X_3$		
5	$\ln Y, \ln X_4$		
6	$\ln X_1, \ln X_1$		
7	$\ln X_1, \ln X_2$		
8	$\ln X_1, \ln X_3$		
9	$\ln X_1, \ln X_4$		
10	$\ln X_2, \ln X_2$		
11	$\ln X_2, \ln X_3$		
12	$\ln X_2, \ln X_4$		
13	$\ln X_3, \ln X_3$		
14	$\ln X_3, \ln X_4$		
15	$\ln X_4, \ln X_4$		

Źródło: Opracowanie własne.

Do obliczenia korelacji między zmienną objaśnianą Y a zmiennymi objaśniającymi X_j lub między zmiennymi objaśniającymi dla $n = 10$ obserwacji zastosujemy wzory (przy czym wartości zmiennych zgodnie z wymogiem modelu podamy w postaci logarytmów):

$$\text{cov}_{yx} = \frac{1}{n} \sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x}) \quad r_{yx} = \frac{\text{cov}_{yx}}{\sqrt{\text{var}_y \cdot \text{var}_x}} = \frac{\text{cov}_{yx}}{S_y \cdot S_x}$$

Gdy zastosujemy analizę korelacyjną dla doboru zmiennych objaśniających to postępujemy następująco:

- a) Wyznaczenie wartości krytycznej współczynnika korelacji:

$$r^* = \sqrt{t^2 / (t^2 + n - 2)}$$

Dla $n = 10$, $iss = n - 2 = 8$, $P = \alpha = 0,01$ wg tablicy wartości krytycznych rozkładu t -Studenta $t^* = 3,3554$.

- b) Wyeliminowanie zmiennej objaśniającej dla której korelacja między $\ln Y$ a $\ln X_i$ jest statystycznie nieistotna według zasady: $|r_{ji}| \leq r^*$.

- c) Wybór z pośród pozostałych korelacji $|r_{ji}|$ zmiennej o korelacji maksymalnej między ($\ln Y$, $\ln X_i$) $|r_{hi}| = \max |r_{ji}|$,

- d) Eliminacja zmiennych o wzajemnej korelacji większej od krytycznej (dla zmiennej o korelacji max.) $|r_{hi}| > r^*$,

- e) Wybór następnej zmiennej objaśniającej przez powrót do punktu c),

- f) Gdy takiej nie ma to sformułowanie modelu.

Oszacowanie parametrów klasyczną metodą najmniejszych kwadratów dla wielu zmiennych, przy czym wartości zmiennych występują w postaci logarytmów.

Określenie funkcji produkcji $\ln Y^{\wedge} = a_0 + a_1 \ln X_1 + \dots + a_4 \ln X_4$ po zamianie jej w postać funkcji wykładniczej:

$$Y^{\wedge} = e^{(a_0 + a_1 \ln X_1 + \dots + a_4 \ln X_4)}$$

Określenie przez eksperta zmiany wielkości zmiennych objaśniających lub przez wyznaczenie ich trendu,

Rozwiązanie tego przykładu przeprowadzono z zastosowaniem wyrażeń *Excela*. Jednak ze względu na obszerność tabel obliczeń pomocniczych zaprezentowano jedynie powyższe etapy a-f. W rezultacie otrzymano następującą tabelę kowariancji i korelacji.

Tab. 5.14. Kowariancja i korelacja między zmiennymi zlogarytmowanymi

		Kowariancja	Korelacja
1	ln Y, ln Y	0,0185533	1
2	ln Y, ln X1	0,0018058	0,6273325
3	ln Y, ln X2	0,0274461	0,9586453
4	ln Y, ln X3	0,0191587	0,9863685
5	ln Y, ln X4	0,0311507	0,8414850
6	ln X1, ln X1	0,0004466	1
7	ln X1, ln X2	0,0025172	0,5666754
8	ln X1, ln X3	0,0019432	0,6448232
9	ln X1, ln X4	0,0027249	0,4744211
10	ln X2, ln X2	0,0441800	1
11	ln X2, ln X3	0,0290493	0,9691856
12	ln X2, ln X4	0,0448532	0,7851803
13	ln X3, ln X3	0,0203344	1
14	ln X3, ln X4	0,0308671	0,7851803
15	ln X4, ln X4	0,0738622	1

Źródło: Opracowanie własne w *Excelu*.

Ad. a) Wartość krytyczna współczynnika korelacji:

$$r^* = \sqrt{3,3554^2 / (3,3554^2 + 8 - 2)} = 0,76$$

Ad. b) Wylimitowanie zmiennej objaśniającej dla której korelacja między $\ln Y$ a $\ln X_i$ jest statystycznie nieistotna wg zasady:

$$|r_{YI}| \leq r^* \quad |r_{Y1}| = 0,627, \text{ czyli jest } < r^* \text{ wylimitowaniu podlega zmienna } X_1$$

Ad. c) Wybór z pozostałych zmiennych o korelacji maksymalnej między $(\ln Y, \ln X_i)$

$$|r_{YI}| = \max |r_{YI}| = 0,986 \quad \text{Jest to zmienna } X_3.$$

Ad. d) Eliminacja zmiennych o wzajemnej korelacji większej od krytycznej

$$|r_{ij}| > r^* \quad \text{Są nimi } X_4 \text{ o korelacji } 0,785 \text{ oraz } X_2 \text{ o korelacji } 0,969.$$

Ad. e) Pozostała tylko zmienna X_3 i ona wchodzi do modelu regresji liniowej.

$$\text{Ad. f) } \ln Y = f(\ln X_3) \quad Y' = a_1 * X_3 + a_0$$

* * *

Powstają coraz nowsze lepiej oparte statystycznie metody wyboru kombinacji zmiennych objaśniających. Zyskują one zaufanie ekonometryków. Starsze metody nazwane przez G.S. Maddalę jako "mechaniczne" są stopniowo zaniechane. Pozostaje więc nadzieja, że zebrany i zaprezentowany tu materiał będzie pomocny w przyjęciu właściwej metody i przyczyni się do doskonalenia umiejętności budowy liniowych modeli ekonometrycznych - po wcześniejszej analizie wyboru kombinacji zmiennych objaśniających.

6. DOBÓR MODELI TRENDU DLA ZMIENNYCH OBJAŚNIAJĄCYCH MODELU REGRESJI ZMIENNEJ OBJAŚNIANEJ



6.1. Wprowadzenie

Dobór modelu trendu dla zmiennych, które tu ze względu na obszerność nie zostały tu opublikowane, przeprowadzamy korzystając z testu F istotności statystycznej dwóch wariancji⁸⁶. Dla tej realizacji musimy wpierw wygenerować nowe zmienne objaśniające trendu, dla okresów czasowych, a mianowicie kwadrat T^2 , sześcian T^3 i czwartą potęgą zmiennej czasowej T^4 . Posługujemy się poniższym menu programu *GRET*:

Dodawanie zmiennych/generowanie nowej zmiennej.

Wprowadzamy nazwę i formułę dla nowej zmiennej $genrT2=T^2$ a następnie kolejnych formuł: $genrT3=T^3$ $genrT4=T^4$; przy czym $genrTi$ – to nazwa zmiennej. Uzyskujemy w bazie danych o nazwie *Korelacja1.gdt*, utworzonej przez autora niniejszej publikacji, trzy nowe zmienne, które posłużą nam do formułowania modeli trendów dla poniższych zmiennych modelu $\hat{Y}$ w celu obliczenia prognoz na okres 17.

$X21, X36, X47, X57, X60, X61, X62, X63, X70, X72, X73, X74$.

Wymienione zmienne zostały opisane w rozdziale „Macierz współczynników korelacji” publikacji „*Modele ekonometryczne PKB obiektów struktury terytorialnej*”⁸⁷. Zmiennym objaśniającym, których wartości danych pobrano podczas pisania tego artykułu, z *Roczników statystycznych GUS-u* z okresu 1991-2007. Nadano im nazwy dogodne w przetwarzaniu komputerowym, a ich znaczenie jest następujące:

$X21$ - zwolnienia z pracy (w tys.), w drodze wypowiedzenia przez zakład pracy;

$X36$ - wynagrodzenia (w mln), osobowe;

$X47$ - przeciętny miesięczny dochód rozporządzalny na 1 osobę – z braku danych przyjęto rok 1992 na poziomie roku 1993 (w zł);

$X57$ - studenci- wyższe (w dziesiątkach);

$X60$ - absolwenci – wyższe (w dziesiątkach);

$X61$ - absolwenci szkół dla dorosłych (w tys.);

$X62$ - uczący się języka angielskiego w szkołach podstawowych i średnich dla dzieci i młodzieży oraz policealnych, bez szkół wyższych – nauczanie obowiązkowe (w tys.);

$X63$ - uczący się języka angielskiego w szkołach podstawowych i średnich dla dzieci i młodzieży oraz policealnych, bez szkół wyższych – nauczanie dodatkowe (w tys.);

$X70$ - absolwenci szkół wyższych – inżynierijno-techniczne, stacjonarne;

$X72$ - wskaźnik – produkcja sprzedana przemysłu;

$X73$ - wskaźnik – produkcja budowlano-montażowa;

$X74$ - wskaźnik – usługi transportu – gospodarki magazynowej i łączności.

⁸⁶ Bliższe informacje na ten temat znajdziemy w monografii: Wornalkiewicz W., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski, Opole 2013, rozdział 5.

⁸⁷ Ibidem, rozdz. 3.

6.2. Dobór stopnia wielomianu dla modeli trendów zmiennych objaśniających wygenerowanego modelu zmiennej Y

Zmienna $X21$ już wcześniej występowała, lecz dla porządku rzeczy związanego z doбором stopnia wielomianu trendu wywołujemy menu:

Model/klasyczna metoda najmniejszych kwadratów

a w oknie specyfikacji modelu określamy: zmienna zależna - $X21$, zmienna niezależna - T .

Uzyskujemy model liniowy trendu dla 16. obserwacji w którym zwracamy szczególną uwagę na błąd standardowy reszt wynoszący 74,8854. Wprowadzamy dla kolejnego modelu wygenerowany kwadrat zmiennej czasowej $generT2$ i wówczas stosując *KMNK* otrzymujemy model kwadratowy, dla którego raport *GRET*L-a podano w tabeli 6.1.

Tab. 6.1. Estymacja parametrów trendu wielomianu 2-go stopnia

Zmienna	Współczynnik	Błąd standardowy	t -Studenta	Wartość p
const	502,601	65,7080	7,649	3,64E-06 ***
T	-8,90926	17,7900	-0,5008	0,6249
genrT2	-0,539006	1,01732	-0,5298	0,6052

Średnia arytmetyczna zmiennej zależnej = 376,475

Odchylenie standardowe zmiennej zależnej = 112,415

Suma kwadratów reszt = 76850

Błąd standardowy reszt = 76,8865

Współczynnik determinacji (R^2) = 0,59458

Skorygowany współczynnik = 0,53221

Statystyka F (2, 13) = 9,53278 (wartość p = 0,00283)

Stopnie swobody 13

Źródło: Opracowanie własne programem *GRET*L.

Teraz wybór stopnia wielomianu trendu wymaga porównania wariancji resztowej modelu trendu liniowego z wariancją resztową modelu kwadratowego za pomocą testu F przy hipotezach⁸⁸:

$$H_0: \sigma_I^2 = \sigma_{II}^2$$

$$H_1: \sigma_I^2 > \sigma_{II}^2$$

Obliczamy F_{obl} według wzoru:

$$F_{obl} = \frac{S^2_{I,e}}{S^2_{II,e}}$$

Dla trendu liniowego błąd standardowy reszt $S_e = 74,8854$, natomiast dla trendu kwadratowego $S_e = 76,8865$.

$$F_{obl} = \frac{(74,8854)^2}{(76,8865)^2} = \frac{5607,82}{5911,53} = 0,949$$

Przyjmijmy do badania poziom istotności $\alpha = 0,05$. Wartość krytyczną F_α dla $\alpha = 0,05$, przy stopniu swobody licznika $s_1 = n - k = 16 - 2 = 14$ oraz stopniu swobody mianownika $s_2 = n - 3 = 16 - 3 = 13$ odczytujemy korzystając z menu: *Narzędzia/tablice statystyczne*.

⁸⁸ Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*L, Wydawnictwo Naukowe PWN, Warszawa 2004, rozdz. 6. Podstawowe modele procesów ekonomicznych.

$F(14, 13)$

prawostronne prawdopodobieństwo = 0,05

prawdopodobieństwo dopełnienia = 0,95

krytyczna wartość $F_\alpha = 2,55362$

Sprawdzamy warunek:

$$F_{obl} < F_\alpha$$

Ponieważ nie występuje istotny spadek wariancji resztowej to właściwym modelem trendu pozostaje trend liniowy postaci:

$$\hat{X}_{21} = a_0 + a_1 * t = 530,09 - 18,0724 * t$$

Analiza powyższa potwierdza nam zatem wcześniej określony model oraz prognozę dla zmiennej objaśniającej X_{21} na rok 2008 (okres 17 wynoszącą 222,86). W sposób analogiczny postąpiono z dalszymi zmiennymi.

Zmienna X36

a) Trend liniowy

Tab. 6.2. Oszacowanie trendu liniowego zmiennej X_{36}

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	16952,3	9223,80	1,838	0,0874 *
$T = t$	17428,0	953,903	18,27	3,65E-011 ***
Średnia arytmetyczna zmiennej zależnej = 165090				
Odchylenie standardowe zmiennej zależnej = 84695,9				
Suma kwadratów reszt = 4,33127e+009				
Błąd standardowy reszt = 17589,1				
Współczynnik determinacji (R^2) = 0,95975				
Skorygowany współczynnik determinacji = 0,95687				
Stopnie swobody = 14				

b) Trend kwadratowy

Tab. 6.3. Oszacowanie trendu kwadratowego zmiennej X_{36}

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	-7827,50	12945,2	-0,6047	0,5558
T	25687,9	3504,81	7,329	5,75E-06 ***
genrT2	-485,879	200,422	-2,424	0,0307 **
Średnia arytmetyczna zmiennej zależnej = 165090				
Odchylenie standardowe zmiennej zależnej = 84695,9				
Suma kwadratów reszt = 2,98279e+009				
Błąd standardowy reszt = 15147,5				
Współczynnik determinacji (R^2) = 0,97228				
Skorygowany współczynnik determinacji = 0,96801				
Statystyka $F(2, 13) = 227,981$ (wartość $p < 0,00001$)				
Ilość stopni swobody 13				

$$F_{obl} = \frac{S_{I,e}^2}{S_{II,e}^2} = \frac{(17589,1)^2}{(15147,5)^2} = \frac{309376438,8}{229446756,3} = 1,35$$

$$F_{obl} = 1,35 < F_a = 2,55362$$

Oznacza to, że możemy stosować model liniowy.

Zmienna X47

a) Trend liniowy

Tab. 6.4. Oszacowanie trendu liniowego zmiennej X47

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	124,072	18,4345	6,730	9,62E-06 ***
T	49,2405	1,90645	25,83	3,28E-013 ***
Średnia arytmetyczna zmiennej zależnej = 542,616				
Odchylenie standardowe zmiennej zależnej = 236,879				
Suma kwadratów reszt = 17300,4				
Błąd standardowy reszt = 35,1531				
Współczynnik determinacji (R^2) = 0,97945				
Skorygowany współczynnik determinacji = 0,97798				
Stopnie swobody = 14				

b) Trend kwadratowy

Tab. 6.5. Trend kwadratowy zmiennej X47

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	74,0116	25,7443	2,875	0,0130 **
T	65,9273	6,97009	9,459	3,42E-07 ***
genrT2	-0,981572	0,398584	-2,463	0,0285 **
Średnia arytmetyczna zmiennej zależnej = 542,616				
Odchylenie standardowe zmiennej zależnej = 236,879				
Suma kwadratów reszt = 11797				
Błąd standardowy reszt = 30,1241				
Współczynnik determinacji (R^2) = 0,98598				
Skorygowany współczynnik determinacji = 0,98383				
Statystyka F (2, 13) = 457,252 (wartość p < 0,00001)				
Ilość stopni swobody 13				

$$F_{obl} = \frac{(35,1531)^2}{(30,1241)^2} = \frac{1235,74}{907,74} = 1,36$$

Wartość krytyczna = 2,55362 - wybieramy trend liniowy.

Zmienna X57. W zakresie tej zmiennej biorąc pod uwagę błąd standardowy reszt S_e obliczony przez program *GRET*L rozpatrzono najpierw trendy: liniowy; ($S_e = 127,462$), kwadratowy ($S_e = 76,1248$).

$$F_{obl} = \frac{S_{I,e}^2}{S_{II,e}^2} = \frac{(127,462)^2}{(76,1248)^2} = \frac{16246,56}{5794,985} = 2,8$$

$$F_{obl} = 2,8 > F_a = 2,55362$$

Występuje zatem istotny spadek wariancji resztowej i właściwym modelem jest wielomian drugiego stopnia, czyli trend kwadratowy. Musimy porównać go z kolejnym modelem wyższego stopnia.

Tab. 6.6. Trend wielomianowy 3-go stopnia zmiennej $X57$

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	426,324	31,8676	13,38	1,43E-08 ***
T	43,2767	15,7429	2,749	0,0176 **
genr72	16,9197	2,11837	7,987	3,82E-06 ***
genr73	-0,865891	0,0820773	-10,55	2,00E-07 ***
Średnia arytmetyczna zmiennej zależnej = 1375,19				
Odchylenie standardowe zmiennej zależnej = 539,294				
Suma kwadratów reszt = 7332,07				
Błąd standardowy reszt = 24,7185				
Współczynnik determinacji = 0,99832				
Skorygowany współczynnik determinacji = 0,99790				
Statystyka F (3, 12) = 2375,99 (wartość p < 0,00001)				
Ilość stopni swobody 12				

$$F_{obl} = \frac{S_{I,e}^2}{S_{II,e}^2} = \frac{(76,1248)^2}{(24,7185)^2} = \frac{5794,985}{611} = 9,48$$

Odczytujemy w GRETL wartość krytyczną F_a

$F(13, 12)$
prawostronne prawdopodobieństwo = 0,05
prawdopodobieństwo dopełnienia = 0,95
krytyczna wartość = 2,66018

Porównanie trendu kwadratowego i trendu wielomianowego trzeciego stopnia wskazuje na trend 3-go stopnia.

$$F_{obl} = 9,48 > F_a = 2,66018$$

W kolejności określamy trend wielomianowy 4-go stopnia dla którego błąd standardowy reszt wynosi 15,8574.

$$F_{obl} = \frac{S_{I,e}^2}{S_{II,e}^2} = \frac{(24,7185)^2}{(15,8574)^2} = \frac{611}{251,457} = 2,43$$

$F(12, 11)$
prawostronne prawdopodobieństwo = 0,05
prawdopodobieństwo dopełnienia = 0,95
krytyczna wartość = 2,78757

$$F_{obl} = 2,43 < F_a = 2,78757$$

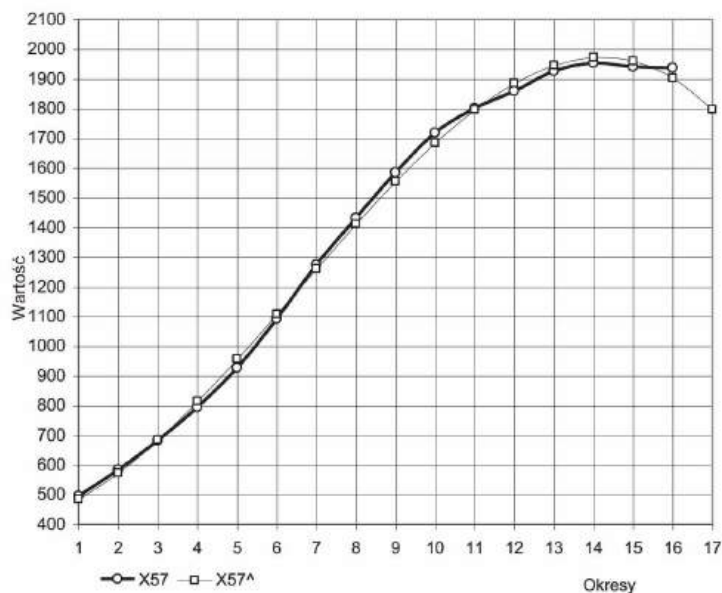
Porównanie powyższych trendów wielomianowych wskazuje, że właściwszym modelem jest model wielomianowy 3-go stopnia.

Dla przykładu porównajmy wartości empiryczne i modelowe (wyrównane) zmiennej X_{57} dla obserwacji 1-16 obliczając prognozę na okres 17.

Tab. 6.7. Dane rzeczywiste, modelowe i prognoza na okres 17

T	X_{57}	$\hat{X}_{57}$
1	495,7	485,65
2	584,0	573,63
3	682,2	685,05
4	794,6	814,73
5	927,5	957,46
6	1091,8	1108,06
7	1274,0	1261,33
8	1431,9	1412,06
9	1584,8	1555,08
10	1718,7	1685,17
11	1800,5	1797,15
12	1858,7	1885,82
13	1926,1	1945,99
14	1953,8	1972,45
15	1941,4	1960,02
16	1937,4	1903,50
17		1797,70

Źródło: Opracowanie własne w *GRET*L.



Źródło: Opracowanie własne w *Excel*-u, przy czym $X_{57}^{\wedge}$ to wartości modelowe.

Rys. 6.1. Wartości empiryczne i modelowe zmiennej X_{57}

Dalej kontynuujemy nasze badania dla kolejnych zmiennych objaśniających wchodzących do modelu zmiennej zależnej Y .

Zmienna X60. Rozpatrzono trendy: liniowy ($S_e = 34,3979$), kwadratowy ($S_e = 34,9084$).

Tab.6.8. Trend liniowy zmiennej X60

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	3,49250	18,0384	0,1936	0,8493
T	27,9656	1,86549	14,99	5,13E-010 ***
Średnia arytmetyczna zmiennej zależnej = 241,2				
Odchylenie standardowe zmiennej zależnej = 137,227				
Suma kwadratów reszt = 16565				
Błąd standardowy reszt = 34,3979				
Współczynnik determinacji = 0,94136				
Skorygowany współczynnik determinacji = 0,93717				
Stopnie swobody = 14				

Według programu *GRET* statystyka F dla trendu kwadratowego jest następująca:

$F(2, 13) = 109,4$ (wartość $p < 0,00001$).

$$F_{obl} = \frac{(34,3979)^2}{(34,9084)^2} = \frac{1183,22}{1218,6} = 0,97$$

Właściwym modelem trendu jest więc trend liniowy.

Zmienna X61. Najpierw porównano trendy: liniowy ($S_e = 16,7686$), kwadratowy ($S_e = 9,23893$).

Tab. 6.9. Trend kwadratowy zmiennej X61

Zmienne	Współczynnik	Błąd standardowy	t-Student	Wartość p
const	31,9339	7,89568	4,044	0,0014 ***
T	13,6919	2,13770	6,405	2,33E-05 ***
genrT2	-0,703501	0,122244	-5,755	6,66E-05 ***
Średnia arytmetyczna zmiennej zależnej = 82,5375				
Odchylenie standardowe zmiennej zależnej = 18,1787				
Suma kwadratów reszt = 1109,65				
Błąd standardowy reszt = 9,23893				
Współczynnik determinacji = 0,77614				
Skorygowany współczynnik determinacji = 0,74170				
Statystyka $F(2, 13) = 22,5363$ (wartość $p = 5,95e-005$)				
Ilość stopni swobody 13				

$$F_{obl} = \frac{(16,7686)^2}{(9,23893)^2} = \frac{289,19}{85,36} = 3,39$$

Występuje istotny spadek wariancji resztowej, zatem dominuje trend kwadratowy. W kolejności porównujemy trend kwadratowy z trendem wielomianowym 3-go stopnia dla którego według programu

$$F_{obl} = \frac{(9,23893)^2}{(6,0056)^2} = \frac{85,36}{36,07} = 2,37$$

GRETl $S_e = 6,0056$.

Wartość krytyczna = 2,66018, tak więc pozostajemy przy modelu trendu kwadratowego.

Zmienna X62. Porównujemy trend liniowy z kwadratowym dla którego $Se = 317,911$.

Tab. 6.10. Trend liniowy zmiennej X62

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	1345,35	206,509	6,515	1,37E-05 ***
T	200,838	21,3567	9,404	1,99E-07 ***
Średnia arytmetyczna zmiennej zależnej = 3052,47				
Odchylenie standardowe zmiennej zależnej = 1029,09				
Suma kwadratów reszt = 2,17107e+006				
Błąd standardowy reszt = 393,798				
Współczynnik determinacji = 0,86333				
Skorygowany współczynnik determinacji = 0,85357				
Ilość stopni swobody = 14				

$$F_{obl} = \frac{(393,798)^2}{(317,911)^2} = \frac{155076,86}{101067,4} = 1,53$$

Wartość krytyczna wynosi = 2,55362 – wybieramy zatem model trendu liniowego.

Zmienna X63. Postępujemy analogicznie jak dla zmiennej X62.

Tab. 6.11. Trend liniowy zmiennej X63

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	150,022	82,4971	1,819	0,0904 *
T	39,3701	8,53165	4,615	0,0004 ***
Średnia arytmetyczna zmiennej zależnej = 484,669				
Odchylenie standardowe zmiennej zależnej = 241,313				
Suma kwadratów reszt = 346476				
Błąd standardowy reszt = 157,316				
Współczynnik determinacji = 0,60334				
Skorygowany współczynnik determinacji = 0,57501				
Ilość stopni swobody = 14				

Dla trendu kwadratowego błąd standardowy reszt wynosi 132,628.

$$F_{obl} = \frac{(157,316)^2}{(132,628)^2} = \frac{24748,32}{17590,19} = 1,41$$

Wybieramy model trendu liniowego.

Zmienna X70. Porównujemy trend liniowy o błędzie standardowym równym 1803,78 z kwadratowym dla którego $Se = 1498,27$

$$F_{obl} = \frac{(1803,78)^2}{(1498,27)^2} = \frac{3253622,29}{2244812,99} = 1,45$$

Tak więc właściwym modelem trendu jest trend liniowego o parametrach i statystyce opisowej podanej w tabeli 6.12.

Tab. 6.12. Trend liniowy zmiennej X70

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	9907,88	945,909	10,47	5,25E-08 ***
T	483,919	97,8235	4,947	0,0002 ***
Średnia arytmetyczna zmiennej zależnej = 14021,2				
Odchylenie standardowe zmiennej zależnej = 2888,72				
Suma kwadratów reszt = 4,55506e+007				
Błąd standardowy reszt = 1803,78				
Współczynnik determinacji = 0,63609				
Skorygowany współczynnik determinacji = 0,61010				
Ilość stopni swobody = 14				

Zmienna X72. Na podstawie błędów standardowych trendu liniowego i kwadratowego obliczamy statystykę F_{obl}

$$F_{obl} = \frac{(5,70326)^2}{(2,79113)^2} = \frac{32,53}{7,79} = 4,18$$

Tab. 6.13. Trend kwadratowy zmiennej X72

Zmienne	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	141,993	2,38532	59,53	3,14E-017 ***
T	-6,35568	0,645810	-9,841	2,17E-07 ***
genr72	0,248985	0,0369305	6,742	1,38E-05 ***
Średnia arytmetyczna zmiennej zależnej = 111,25				
Odchylenie standardowe zmiennej zależnej = 11,5115				
Suma kwadratów reszt = 101,275				
Błąd standardowy reszt = 2,79113				
Współczynnik determinacji = 0,94905				
Skorygowany współczynnik determinacji = 0,94121				
Statystyka $F(2, 13) = 121,075$ (wartość $p < 0,00001$)				
Ilość stopni swobody = 13				

Wartość krytyczna wynosi $= 2,55362$. Występuje istotny spadek wariancji resztowej, a właściwym modelem trendu będzie trend kwadratowy. Porównajmy go jeszcze z trendem wielomianu 3-go stopnia o błędzie standardowym równym 2,84114.

$$F_{obl} = \frac{(2,79113)^2}{(2,84114)^2} = \frac{7,79}{8,09} = 0,96$$

Pozostajemy więc przy modelu trendu kwadratowego.

Tab. 6.14. Trend liniowy zmiennej X73

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	123,093	2,23931	54,97	9,32E-018 ***
T	-1,49324	0,231584	-6,448	1,53E-05 ***
Średnia arytmetyczna zmiennej zależnej = 110,4				
Odchylenie standardowe zmiennej zależnej = 8,21949				
Suma kwadratów reszt = 255,284				
Błąd standardowy reszt = 4,2702				
Współczynnik determinacji = 0,74809				
Skorygowany współczynnik = 0,73010				
Ilość stopni swobody = 14				

Porównujemy z trendem kwadratowym dla którego $Se = 3,89557$

$$F_{obl} = \frac{(2,84114)^2}{(3,89557)^2} = \frac{8,09}{15,175} = 0,53$$

Wybieramy model trendu liniowego.

Zmienna X74. Dla trendu liniowego błąd standardowy reszt wynosi 5,3436 a dla kwadratowego podano w tabeli 6.15.

Tab. 6.15. Trend kwadratowy zmiennej X74

Zmienne	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	144,829	2,29038	63,23	1,43E-017 ***
T	-6,35617	0,620103	-10,25	1,35E-07 ***
genrT2	0,231600	0,0354605	6,531	1,91E-05 ***
Średnia arytmetyczna zmiennej zależnej = 112,456				
Odchylenie standardowe zmiennej zależnej = 12,6207				
Suma kwadratów reszt = 93,3732				
Błąd standardowy reszt = 2,68003				
Współczynnik determinacji = 0,96092				
Skorygowany współczynnik determinacji = 0,95491				
Statystyka F (2, 13) = 159,822 (wartość p < 0,00001)				
Ilość stopni swobody = 13				

$$F_{obl} = \frac{(5,3436)^2}{(2,68003)^2} = \frac{28,554}{7,183} = 3,98$$

$$F_{obl} > F_{\alpha} = 2,55362$$

Zachodzi potrzeba porównania jeszcze z trendem wielomianu 3-go stopnia, dla którego $S_e = 2,78159$.

$$F_{obl} = \frac{(2,68003)^2}{(2,78159)^2} = \frac{7,183}{7,737} = 0,94$$

Pozostajemy więc przy modelu trendu kwadratowego.

6.3. Generowanie zmiennych o równaniach trendów

W wyniku doboru stopnia wielomianu dla zmiennych modelu regresji produktu krajowego brutto przeprowadzonego we wcześniejszym podrozdziale uzyskaliśmy modele trendów na podstawie których możemy określić wartości prognoz F_{17} , F_{18} odpowiadające latom 2008, 2009. Dla porównania podano również wartości zmiennej Y dla $T = 16$.

Tab. 6.16. Zestawienie modeli trendów oraz obliczone wartości prognoz na okres 17 i 18

Zmienna	Y_{16}	Model regresji/trendu	F_{17}	F_{18}
Y	30834	$a_0 + a_1X_{21} + a_2X_{36} + a_3X_{47} + a_4X_{57} + a_5X_{60} + a_6X_{61} + a_7X_{62} + a_8X_{63} + a_9X_{70} + a_{10}X_{72} + a_{11}X_{73} + a_{12}X_{74}$		
X_{21}	215,4	$530,09 - 18,07 * t$	222,9	204,8
X_{36}	295389	$16952,3 + 17428 * t$	313233	330656
X_{47}	928,87	$124,07 + 49,24 * t$	961,15	1010,39
X_{57}	1937,4	$426,32 + 43,28 * t + 16,92 * t^2 - 0,866 * t^3$	1797,7	1636,9
X_{60}	410,1	$3,49 + 27,97 * t$	479,0	506,9
X_{61}	63,4	$31,93 + 13,69 * t - 0,7 * t^2$	62,4	51,6
X_{62}	4143,5	$1345,35 + 200,84 * t$	4759,6	4960,5
X_{63}	669,1	$150 + 39,37 * t$	819,3	858,7
X_{70}	16628	$9907,88 + 483,92 * t$	18135	18618
X_{72}	102,3	$141,99 - 6,36 * t + 0,25 * t^2$	106,1	108,5
X_{73}	107,8	$123,09 - 1,49 * t$	97,8	96,3
X_{74}	102,5	$144,83 - 6,36 * t + 0,23 * t^2$	103,2	104,9

Źródło: Opracowanie własne w Excelu.

Zamieszczone w tabeli 6.16 modele trendów dla zmiennych X_{21} , X_{36} , ..., X_{74} określono po zastosowaniu w *GRET*L metody *KMNK* oraz stosując test istotności dwóch wariancji resztowych. Modele trendów zmiennych mogą być zapisane w *GRET*L jako nowe zmienne endogeniczne, bazujące na okresach czasowych t (1-16).

Dla określenia wartości zmiennych w okresach przyszłych należy traktować je w *GRET*L jako zmienne zależne. Należy przyjąć zakres próby 16 obserwacji, a następnie po wykonaniu modelu określić prognozę przykładowo na lata 2008, 2009. Niżej przytoczono ponownie, z powyższej tabeli, zarówno model regresji jak i modele trendów zmiennych biorących udział w modelu regresji:

$$\hat{Y} = a_0 + a_1X_{21} + a_2X_{36} + a_3X_{47} + a_4X_{57} + a_5X_{60} + a_6X_{61} + a_7X_{62} + a_8X_{63} + a_9X_{70} + a_{10}X_{72} + a_{11}X_{73} + a_{12}X_{74}$$

$$\hat{X}_{21} = 530,09 - 18,07 * t; \hat{X}_{36} = 16952,3 + 17428 * t; \hat{X}_{47} = 124,07 + 49,24 * t;$$

$$\hat{X}_{57} = 426,32 + 43,28 * t + 16,92 * t^2 - 0,866 * t^3; \hat{X}_{60} = 3,49 + 27,97 * t; \hat{X}_{61} = 31,93 + 13,69 * t - 0,7 * t^2$$

$$\hat{X}_{62} = 1345,35 + 200,84 * t; \hat{X}_{63} = 150 + 39,37 * t; \hat{X}_{70} = 9907,88 + 483,92 * t;$$

$$\hat{X}_{72} = 141,99 - 6,36 * t + 0,25 * t^2; \hat{X}_{73} = 123,09 - 1,49 * t; \hat{X}_{74} = 144,83 - 6,36 * t + 0,23 * t^2$$

przy czym w bazie danych *Korelacja1.gdt* nazwy zmiennych występują jako:

$$X_{21}T, \dots, X_{74}T, \text{ gener}T2, \text{ gener}T3$$

Jak już nadmieniono, generowanie nowej zmiennej realizujemy korzystając z następującego menu *GRET*L-a:

Dodawanie zmiennych/generowanie nowej zmiennej

Wprowadzamy nazwę i formułę dla szeregu teoretycznego przykładowo dla zmiennej X_{21} otrzymanego na podstawie powyższego modelu trendu liniowego:

$$\text{genr}X_{21}T = (530.09) - (18.07) * (T)$$

Zmienna - okresy czasowe ma symbol T , jej kwadraty $\text{genr}T^2$, sześciany $\text{genr}T^3$.

Formuły dla wprowadzenia dalszych zmiennych a następnie wygenerowania wartości są następujące:

$$\text{genr}X_{36}T = (16952.3) + (17428) * (T)$$

$$\text{genr}X_{47}T = (124.07) + (49.24) * (T)$$

$$\text{genr}X_{57}T = (426.32) + (43.28) * (T) + (16.92) * (\text{genr}T^2) - (0.866) * (\text{genr}T^3)$$

$$\text{genr}X_{60}T = (3.49) + (27.97) * (T)$$

$$\text{genr}X_{61}T = (31.93) + (13.69) * (T) - (0.7) * (\text{genr}T^2)$$

$$\text{genr}X_{62}T = (1345.35) + (200.84) * (T)$$

$$\text{genr}X_{63}T = (150) + (39.37) * (T)$$

$$\text{genr}X_{70}T = (9907.88) + (483.92) * (T)$$

$$\text{genr}X_{72}T = (141.99) - (6.36) * (T) + (0.25) * (\text{genr}T^2)$$

$$\text{genr}X_{73}T = (123.09) - (1.49) * (T)$$

$$\text{genr}X_{74}T = (144.83) - (6.36) * (T) + (0.23) * (\text{genr}T^2)$$

Tab. 6.17. Raport wygenerowanych wartości dla 16 obserwacji

Obs.	genrX ₂₁ T	genrX ₃₆ T	genrX ₄₇ T	genrX ₅₇ T	genrX ₆₀ T	genrX ₆₁ T
1	512,02	34380,3	173,31	486,52	31,46	44,92
2	493,95	51808,3	222,55	580,56	59,43	56,51
3	475,88	69236,3	271,79	708,44	87,40	66,70
4	457,81	86664,3	321,03	870,16	115,37	75,49
5	439,74	104092,3	370,27	1065,72	143,34	82,88
6	421,67	121520,3	419,51	1295,12	171,31	88,87
7	403,60	138948,3	468,75	1558,36	199,28	93,46
8	385,53	156376,3	517,99	1855,44	227,25	96,65
9	367,46	173804,3	567,23	2186,36	255,22	98,44
10	349,39	191232,3	616,47	2551,12	283,19	98,83
11	331,32	208660,3	665,71	2949,72	311,16	97,82
12	313,25	226088,3	714,95	3382,16	339,13	95,41
13	295,18	243516,3	764,19	3848,44	367,10	91,60
14	277,11	260944,3	813,43	4348,56	395,07	86,39
15	259,04	278372,3	862,67	4882,52	423,04	79,78
16	240,97	295800,3	911,91	5450,32	451,01	71,77

Tab. 6.18. Raport wartości wygenerowanych przez GRETl dla 16 obserwacji c.d.

Obs.	genrX ₆₂ T	genrX ₆₃ T	genrX ₇₀ T	genrX ₇₂ T	genrX ₇₃ T	genrX ₇₄ T
1	1546,19	189,37	10391,80	135,88	121,60	138,70
2	1747,03	228,74	10875,72	130,27	120,11	133,03
3	1947,87	268,11	11359,64	125,16	118,62	127,82
4	2148,71	307,48	11843,56	120,55	117,13	123,07
5	2349,55	346,85	12327,48	116,44	115,64	118,78
6	2550,39	386,22	12811,40	112,83	114,15	114,95

Obs.	genrX62T	genrX63T	genrX70T	genrX72T	genrX73T	genrX74T
7	2751,23	425,59	13295,32	109,72	112,66	111,58
8	2952,07	464,96	13779,24	107,11	111,17	108,67
9	3152,91	504,33	14263,16	105,00	109,68	106,22
10	3353,75	543,70	14747,08	103,39	108,19	104,23
11	3554,59	583,07	15231,00	102,28	106,70	102,70
12	3755,43	622,44	15714,92	101,67	105,21	101,63
13	3956,27	661,81	16198,84	101,56	103,72	101,02
14	4157,11	701,18	16682,76	101,95	102,23	100,87
15	4357,95	740,55	17166,68	102,84	100,74	101,18
16	4558,79	779,92	17650,60	104,23	99,25	101,95

Źródło: Opracowanie własne w *GRET*L.

6.4. Określenie błędów prognozy

Wprowadzamy do bazy danych *Korelacja1.gdt* dodatkowe obserwacje 17 oraz 18 dla okresów przyszłych korzystając z menu: *Dane/dodaj obserwacje*.

Określamy liczbę dodawanych obserwacji = 2, a następnie wprowadzamy dane z poniższej tabeli dla odpowiednich zmiennych.

Tab. 6.19. Prognozy zmiennych objaśniających

Zmienna	Wartość dla $T = 16$	F_{17}	F_{18}
Y	30834		
X21	215,4	222,9	204,8
X36	295389	313233	330656
X47	928,87	961,15	1010,39
X57	1937,4	1797,7	1636,9
X60	410,1	479,0	506,9
X61	63,4	62,4	51,6
X62	4143,5	4759,6	4960,5
X63	669,1	819,3	858,7
X70	16628	18135	18616
X72	102,3	106,1	108,5
X73	107,8	97,8	96,3
X74	102,5	103,2	104,9

Źródło: Opracowanie własne w *Excelu* według tabeli 6.16.

Definiujemy w *GRET*L model do procedury *eliminacji a posteriori* bazującej na szeregach obserwacji zmiennych wyszczególnionych w tabeli 6.17 i 6.18. Przyjmujemy zakres próby 16. obserwacji, a zakres prognozy przyszłe dwie obserwacje. Po wywołaniu menu:

Model/klasyczna metoda najmniejszych kwadratów

uzyskujemy *Model 1* dla zmiennej zależnej *Y* dla którego raport podano w tabeli 6.20.

Tab. 6.20. Zainicjowanie modelu zmiennej Y przed określeniem prognozy

Zmienne	Współczynnik	Błąd standardowy	t -Studenta	Wartość p
const	7013,05	260,370	26,93	0,0001 ***
X21	-25,8925	0,0670563	-386,1	3,83E-08 ***
X36	0,131012	0,000476206	275,1	1,06E-07 ***
X47	-24,6239	0,252052	-97,69	2,36E-06 ***
X57	7,00137	0,141715	49,40	1,83E-05 ***
X60	7,95647	0,119362	66,66	7,44E-06 ***
X61	-48,2390	0,675869	-71,37	6,06E-06 ***
X62	-1,24159	0,0317142	-39,15	3,67E-05 ***
X63	-3,05748	0,0535376	-57,11	1,18E-05 ***
X70	-0,0412958	0,00292043	-14,14	0,0008 ***
X72	36,6535	0,914877	40,06	3,42E-05 ***
X73	-67,6121	1,07809	-62,71	8,93E-06 ***
X74	109,446	1,06575	102,7	2,04E-06 ***

Źródło: Opracowanie własne w *GRET*L.

Ponowna estymacja parametrów modelu wynikała z warunku *GRET*L-a korzystania z menu prognozy na bazie aktualnego modelu na ekranie. Następnie wywołujemy menu:

Analiza/prognoza... (określamy obserwację startową i końcową)

i otrzymujemy raport:

Obs	Y	Prognoza	Błąd <i>ex ante</i>	95% przedział ufności
1	2996,00	2994,18		
2	4051,00	4047,89		
3	5459,00	5467,97		
4	7412,00	7407,52		
5	9395,00	9392,22		
6	14211,00	14212,62		
7	12144,00	12146,74		
8	15925,00	15924,95		
9	17723,00	17726,25		
10	19392,00	19389,83		
11	20169,00	20167,80		
12	21336,00	21332,78		
13	24153,00	24157,88		
14	25698,00	25695,86		
15	27803,00	27802,30		
16	30834,00	30834,23		
17		31405,07	53,311	31235,41 - 31574,73
18		32548,52	83,853	32281,66 - 32815,38

Powyższy raport zawiera numer okresu, realizację, prognozę, średni błąd prognozy *ex ante* (V_T), oraz dolną i górną granicę przedziału ufności prognozy. Dla okresów 17 i 18 możemy obliczyć względne błędy prognoz według wzoru:

$$V_T^* = \frac{V_T}{Y_{T_p}} * 100$$

$$V_{17}^* = \frac{V_{17}}{Y_{17_p}} * 100 = \frac{53,311}{31405,07} * 100 = 0,17\%$$

$$V_{18}^* = \frac{83,853}{32548,52} * 100 = 0,26\%$$

Zwróćmy uwagę, że w przedstawionym zadaniu dotyczącym produktu krajowego brutto *PKB* względne błędy *ex ante* nie przekraczają poziomu istotności 5%, czyli prognoza może być praktycznie wykorzystywana.

7. SYMULACJA DOBORU ZMIENNYCH OBJAŚNIAJĄCYCH DLA MODELU EKONOMETRYCZNEGO



7.1. Słowo wstępne

Dobór zmiennych objaśniających do modelu ekonometrycznego przykładowo Produktu Krajowego Brutto (PKB) określonej jako Y może odbywać się różnymi metodami. Nie uzyskuje się jednak jednolitego rezultatu – struktury modelu. Istotną rolę odgrywa analiza współczynników korelacji między zmienną zależną (objaśnianą) a zmiennymi niezależnymi (objaśniającymi), a także współczynników korelacji wzajemnych między poszczególnymi zmiennymi objaśniającymi. We wcześniejszym badaniu zależności 80. zmiennych przedstawiono wybrane zmienne z roczników statystycznych oraz obliczono korzystając z programu *GRET*L współczynniki korelacji tworzące wektor oraz macierz korelacji⁸⁹. W wyniku analizy współczynników korelacji dla poziomu istotności $\alpha = 0,1$ sformułowano wstępnie poniższy model ekonometryczny:

$$\hat{Y} = a_0 + a_1 X_2 + a_2 X_{47}$$

Niniejsze rozważania mają na celu wykorzystanie w pełni programu *GRET*L do redukcji zmiennych, spośród potencjalnych 79. zmiennych objaśniających stosując metodę eliminacji *a posteriori*. Polega ona na usuwaniu kolejnych zmiennych z modelu biorąc pod uwagę podpowiedzi tego programu według kryterium największego błędu udziału danego parametru, a przez niego danej zmiennej objaśniającej sygnalizowane w raportach ilością gwiazdek.

Takie podejście nie może być jednak w pełni zastosowane w naszym badaniu, gdyż liczba zmiennych (79) znacznie przekracza liczbę obserwacji (16). Z tego względu podjęto próbę określenia modelu dla zmiennej Y , a następnie dla zmiennych w nim uczestniczących jako zmienne endogenne sposobem kombinowanym pomysłu autora⁹⁰. Polega on na zabazowaniu na modelu wstępnym a następnie dodawaniu „porcji” zmiennych i postępowanie dalej wg zaleceń *GRET*L-a. Procedura postępowania w odniesieniu do zmiennej Y jest następująca:

- Wprowadzenie najpierw zmiennych X_2 oraz X_{47} oraz dodatkowo $X_1, X_3 \div X_{11}$. W odpowiedzi po kilku eliminacjach *a posteriori* uzyskano:

$$\hat{Y} = a_0 + a_1 X_{10} + a_2 X_{47}$$

- Dodanie zmiennych $X_{12} \div X_{21}$ a po eliminacjach:

$$\hat{Y} = a_0 + a_1 X_{13} + a_2 X_{14} + a_3 X_{15} + a_4 X_{16} + a_5 X_{18} + a_6 X_{21} + a_7 X_{47}$$

- Uzupełnienie o zmienne $X_{22} \div X_{26}$ i w wyniku eliminacji otrzymanie modelu:

$$\hat{Y} = a_0 + a_1 X_{13} + a_2 X_{15} + a_3 X_{16} + a_4 X_{18} + a_5 X_{21} + a_6 X_{24} + a_7 X_{26}$$

- Wprowadzenie zmiennych $X_{27} \div X_{31}$. Model nie ulega zmianie. Dodanie $X_{32} \div X_{36}$ i uzyskanie w kilku eliminacjach modelu:

$$\hat{Y} = a_0 + a_1 X_{13} + a_2 X_{15} + a_3 X_{16} + a_4 X_{18} + a_5 X_{21} + a_6 X_{26} + a_7 X_{36}$$

- Uzupełnienie o $X_{37} \div X_{41}$, a po eliminacji *a posteriori* otrzymanie modelu:

$$\hat{Y} = a_0 + a_1 X_{13} + a_2 X_{15} + a_3 X_{21} + a_4 X_{26} + a_5 X_{36} + a_6 X_{37} + a_7 X_{39}$$

⁸⁹ Wornalkiewicz W., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, op. cit., rozdział 4.5.

⁹⁰ Autorem tym jest Władysław Wornalkiewicz.

- Wprowadzenie zmiennych $X42 \div X46$, po czym mimo eliminacji model nie ulega zmianie. Dodanie zmiennych $X47 \div X51$ i po eliminacji uzyskanie modelu:

$$\hat{Y} = a_0 + a_1 X_{21} + a_2 X_{26} + a_3 X_{36} + a_4 X_{37} + a_5 X_{47} + a_6 X_{49}$$

- Dodanie zmiennych $X52 \div X56$, a mimo eliminacji zmiennych model nie ulega zmianie. Dodanie zmiennych $X57 \div X61$ i po eliminacji uzyskanie modelu:

$$\hat{Y} = a_0 + a_1 X_{21} + a_2 X_{36} + a_3 X_{37} + a_4 X_{47} + a_5 X_{49} + a_6 X_{57} + a_7 X_{58} + a_8 X_{60} + a_9 X_{61}$$

- Wprowadzenie dodatkowo zmiennych $X62 \div X66$ i w drodze eliminacji zmiennych model jest następujący:

$$\hat{Y} = a_0 + a_1 X_{21} + a_2 X_{36} + a_3 X_{47} + a_4 X_{49} + a_5 X_{57} + a_6 X_{60} + a_7 X_{61} + a_8 X_{62} + a_9 X_{63}$$

- Dodanie dodatkowo zmiennych $X67 \div X71$ i po eliminacji uzyskanie modelu:

$$\hat{Y} = a_0 + a_1 X_{21} + a_2 X_{36} + a_3 X_{47} + a_4 X_{49} + a_5 X_{57} + a_6 X_{60} + a_7 X_{61} + a_8 X_{62} + a_9 X_{63} + a_{10} X_{70}$$

- Wprowadzenie zmiennych $X72 \div X79$ i uzyskanie modelu końcowego w wyniku 87. kroków eliminacji *a posteriori* przy wykorzystaniu zmiennej zależnej Y .

Otrzymany z *GRETL* raport końcowy zamieszczono w tabeli 7.1.

Tab. 7.1. Estymacja parametrów modelu zmiennej Y metodą KMNK w wyniku dodawania zmiennych objaśniających i eliminacji *a posteriori*

Zmienna	Współczynnik	Błąd standardowy	<i>t</i> -Studenta	Wartość <i>p</i>	
const	7013,05	260,370	26,93	0,0001	***
X_{21}	-25,8925	0,0670563	-386,1	3,83E-08	***
X_{36}	0,131012	0,000476206	275,1	1,06E-07	***
X_{47}	-24,6239	0,252052	-97,69	2,36E-06	***
X_{57}	7,00137	0,141715	49,40	1,83E-05	***
X_{60}	7,95647	0,119362	66,66	7,44E-06	***
X_{61}	-48,2390	0,675869	-71,37	6,06E-06	***
X_{62}	-1,24159	0,0317142	-39,15	3,67E-05	***
X_{63}	-3,05748	0,0535376	-57,11	1,18E-05	***
X_{70}	-0,0412958	0,00292043	-14,14	0,0008	***
X_{72}	36,6535	0,914877	40,06	3,42E-05	***
X_{73}	-67,6121	1,07809	-62,71	8,93E-06	***
X_{74}	109,446	1,06575	102,7	2,04E-06	***

Średnia arytmetyczna zmiennej zależnej = 16168,8

Odchylenie standardowe zmiennej zależnej = 8710,92

Suma kwadratów reszt = 187,405

Błąd standardowy reszt = 7,90369

Współczynnik determinacji $R^2 = 1,00000$

Skorygowany współczynnik $R^2 = 1,00000$

Statystyka $F(12, 3) = 1,51837e+006$ (wartość $p < 0,00001$).

Źródło: Opracowanie własne w programie *GRETL*.

Test na normalność rozkładu reszt:

Hipoteza zerowa H_0 : składnik losowy ma rozkład normalny

Statystyka testu: $\chi_{(2)} = 4,46213$ z wartością $p = 0,107414$

Wartość krytyczna $\chi^2_{(\alpha, r)}$ wg rozkładu χ^2 przy poziomie istotności $\alpha = 0,1$ o dwóch stopniach swobody wynosi 4,605. Zatem spełnione jest hipoteza H_0 , gdyż wartość statystyki jest mniejsza niż wartość krytyczna.

7.2. Modele zmiennych endogenicznych modelu Y (PKB)

Dobór zmiennych dla zmiennych endogenicznych równania dla Y

W macierzy współczynników korelacji oprócz zmiennej Y występuje 79 zmiennych objaśniających. Dla sformułowania wielorównaniowego modelu ekonometrycznego PKB przyjęto, że 12 zmiennych w równaniu dla Y stanowią zmienne endogeniczne, a więc tworzą kolejne równania modelu ekonometrycznego. Jak już wspomniano w równaniu dla Y występują zmienne wymienione w rozdziale „Dobór trendu dla zmiennych objaśniających modelu regresji zmiennej objaśnianej” o symbolach: $X_{21}, X_{36}, X_{47}, X_{57}, X_{60}, X_{61}, X_{62}, X_{63}, X_{70}, X_{72}, X_{73}, X_{74}$.

Postępowanie dla wyłonienia zmiennych objaśniających dla 12. powyższych zmiennych, traktowanych jako endogeniczne, modelu ekonometrycznego Y było różne i tak przykładowo dla X_{21} polegało na przeprowadzeniu analizy współczynników korelacji o procedurze:

a) Wyznaczenie wartości krytycznej współczynnika korelacji. Współczynniki korelacji liniowej uzyskujemy dla obserwacji z próby 1+16. *GRETL* podaje nam wartość krytyczną $r^* = 0,4259$ przy dwustronnym 10% obszarze krytycznym dla $n = 16$. Wartość tą możemy obliczyć także ze wzoru:

$$r^* = \sqrt{\frac{t_{\alpha}^2}{t_{\alpha}^2 + n - 2}} = \sqrt{\frac{(1,7613)^2}{(1,7613)^2 + 16 - 2}} = 0,4259$$

Przy czym:

$t_{\alpha} = 1,7613$ - statystyka wyznaczona na podstawie rozkładu t -Studenta dla poziomu istotności $\alpha = 0,1$ oraz liczby stopni swobody $iss = n - 2 = 16 - 2 = 14$

b) Sprawdzamy czy $|r_j| \leq r^*$ i eliminujemy zmienne X_j nieistotnie skorelowane z Y :

$$|r_j| \leq r^* = 0,4259$$

c) Z pozostałych potencjalnych zmiennych wybieramy X_h wg kryterium:

$$|r_h| = \max_j |r_j|$$

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

Analiza współczynników korelacji oraz eliminacja a posteriori

Zmienna X_{21}

ad b) Wyeliminowano następujące zmienne: $X_2, X_3, X_4, X_{14}, X_{15}, X_{16}, X_{18}, X_{22}, X_{27}, X_{28}, X_{30}, X_{31}, X_{32}, X_{52}, X_{58}, X_{61}, X_{67}, X_{69}, X_{71}, X_{78}$,

ad c) Wyłoniono zmienną najsilniej skorelowaną z X_{21} , jest nią X_{44} o współczynniku korelacji

$$|r_h| = |-0,9027| = 0,9027$$

ad. d) Pozostały nam zmienne X_{46}, X_{61}, X_{66} o współczynnikach korelacji odpowiednio:

$$-0,4428; -0,1759; 0,0506$$

Powracamy do etapu c procedury i z pozostałych 3. zmiennych wybieramy najsilniej skorelowaną ze zmienną objaśnianą X_{21} , a jest nią X_{46} . Pozostały nam zmienne X_{61} , X_{66} o współczynnikach korelacji względem X_{46} odpowiednio: 0,8536; 0,5491 do których stosujemy kryterium eliminacji:

$$|r_{hi}| > r^* = 0,4259$$

Obie zmienne podlegają eliminacji i model wstępny po analizie współczynników korelacji jest następujący:

$$X_{21} = f(c, X_{44}, X_{46})$$

Alternatywnie sprawdzamy za pomocą programu *GRETL*, w wielu krokach, stosując eliminację *a posteriori*. Etap ten jest dodatkowym upewnieniem się co do trafności obranego modelu, przy czym jako kryterium wyboru modelu spośród z obu sposobów doboru zmiennych przyjmijmy kryterium informacyjne Akeike'a. Uzyskany raport z *GRETL*-a przy zastosowaniu KMNK i 16. obserwacji dla zmiennej X_{21} podano w tabeli 7.2.

Tab. 7.2. Parametry oraz statystyka opisowa zmiennej X_{21}

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p	
const	589,326	34,2541	17,20	2,52E-010	***
X_{44}	-1,36093	0,198542	-6,855	1,16E-05	***
X_{46}	-0,380021	0,420717	-0,9033	0,3828	

Średnia arytmetyczna zmiennej zależnej = 376,475

Odchylenie standardowe zmiennej zależnej = 112,415

Suma kwadratów reszt = 33026,7

Błąd standardowy reszt = 50,4035

Współczynnik determinacji $R^2 = 0,82577$

Skorygowany współczynnik $R^2 = 0,79896$

Statystyka $F(2, 13) = 30,8068$ (wartość $p = 1,17\text{e-}005$)

Logarytm wiarygodności = -83,7629

Kryterium informacyjne Akaike'a (AIC) = 173,526.

Źródło: Opracowanie własne w programie *GRETL*.

GRETL nie proponuje usunięcia któreś ze zmiennych. Statystyka F oceny całości modelu wykazuje niskie prawdopodobieństwo popełnienia błędu, chociaż statystyka t -Studenta dla zmiennej X_{46} jest powyżej poziomu istotności $\alpha = 0,1$. Otrzymaliśmy taki sam model dla zmiennej X_{21} jak po analizie współczynników korelacji. Zatem nie ma potrzeby stosowania kryterium Akaike'a, które w literaturze określane jest wg wzoru⁹¹:

$$AK = N \ln(\Sigma \hat{\epsilon}_i^2) + 2k = 16 * \ln 33026,7 + 2 * 3 = 172,4811$$

Natomiast w *GRETL* obliczany jest wg formuły:

$$AIC = -2l(\hat{\theta}) + 2k = -2 * (-83,7629) + 2k = 167,5258 + 6 = 173,526$$

gdzie logarytm wiarygodności oznaczany jest symbolem:

$$l(\hat{\theta})$$

Tak więc model jest następujący:

$$\hat{X}_{21} = 589,326 - 1,36093 * X_{44} - 0,380021 * X_{46}$$

⁹¹ Lipiec-Zajchowska (red) Wspomaganie procesów decyzyjnych Tom II. Ekonometria Wydawnictwo C.H. Beck, Warszawa 2003.

*Zmienna X36*⁹²

ad b) Wyeliminowano następujące zmienne: $X3, X4, X14, X18, X19, X20, X27, X28, X29, X30, X31, X66, X67, X78$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z $X36$, jest nią $X47$ o współczynniku korelacji

$$r_h = 0,9918$$

Dalsze eliminowanie zmiennych metodą analizy współczynników korelacji jest pracochłonne. Z tego względu posłużono się dalej, zamiast realizować etap *d*, programem *GRETL* przyjmując wyjściowo model ze zmienną $X47$ i nie uwzględniając w kolejnych dodawanych porcjami zmiennych spośród 79 już wyeliminowanych w etapie *b* (zob. tabela 7.3).

Tab. 7.3. Parametry oraz statystyka opisowa zmiennej $X36$

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p	
const	66764,8	7656,74	8,720	0,0032	***
X47	401,587	2,37322	169,2	4,55E-07	***
X6	3581,50	28,0374	127,7	1,06E-06	***
X15	7,50611	0,542997	13,82	0,0008	***
X16	-41,5674	1,57651	-26,37	0,0001	***
X17	-124,367	4,48927	-27,70	0,0001	***
X21	98,2313	0,940700	104,4	1,94E-06	***
X23	-418,541	2,86156	-146,3	7,05E-07	***
X45	-76,5408	2,10946	-36,28	4,60E-05	***
X49	-46,8263	2,38557	-19,63	0,0003	***
X57	-18,0817	2,17551	-8,311	0,0036	***
X63	-2,77458	0,377048	-7,359	0,0052	***
X65	-0,0206072	0,00280578	-7,345	0,0052	***

Średnia arytmetyczna zmiennej zależnej = 165090

Odchylenie standardowe zmiennej zależnej = 84695,9

Suma kwadratów reszt = 39536,5

Błąd standardowy reszt = 114,799

Współczynnik determinacji $R^2 = 1,00000$

Skorygowany współczynnik $R^2 = 1,00000$

Statystyka $F(12, 3) = 680390$ (wartość $p < 0,00001$)

Logarytm wiarygodności = -85,2022

Kryterium informacyjne Akaike'a (AIC) = 196,404

Źródło: Opracowanie własne w programie *GRETL*.

W modelu zmiennej $X36$ występują także zmienne endogeniczne: $X21, X47, X57, X63$, które w raporcie zaznaczono czcionką pogrubioną.

$$\hat{X}_{36} = f(c, \mathbf{X47}, X6, X15, X16, X17, \mathbf{X21}, X23, X45, X49, \mathbf{X57}, \mathbf{X63}, X65)$$

$$\hat{X}_{36} = 66764,8 + 401,587 X47 + 3581,5 X6 + 7,50611 X15 - 41,5674 X16 - 124,367 X17 + 98,2313 X21 - 418,541 X23 - 76,5408 X45 - 46,8263 X49 - 18,0817 X57 - 2,77458 X63 - 0,0206072 X65$$

⁹² W opracowaniu poszczególnych modeli zmiennych endogenicznych modelu zależnego Y bazowano na: Wornalkiewicz W., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski, Opole 2013, rozdz. 6.

Zmienna X47

ad b) Wyeliminowano następujące zmienne: X2, X3, X4, X14, X18, X27, X28, X29, X30, X31, X66, X67, X78.

ad c) Wyłoniono zmienną najsilniej skorelowaną z X47, jest nią X36 o współczynniku korelacji $r_h = 0,9918$

Eliminację *a posteriori* w GRETL wykonano analogicznie jak dla zmiennej X36 uzyskując raport końcowy modelu zaprezentowany w tabeli 7.4.

Tab. 7.4. Parametry oraz statystyka opisowa zmiennej X47

Zmienna	Współczynnik	Błąd standardowy	t-Studenta	Wartość p
const	-354,736	1,29933	-273,0	1,34E-05 ***
X36	0,00215128	2,13087E-06	1010	9,81E-07 ***
X6	-8,43250	0,0179360	-470,1	4,52E-06 ***
X7	0,486285	0,00228455	212,9	2,21E-05 ***
X13	0,439346	0,000590508	744,0	1,81E-06 ***
X15	0,0479328	9,86960E-05	485,7	4,24E-06 ***
X19	-0,511685	0,000927095	-551,9	3,28E-06 ***
X21	-0,281815	0,000460028	-612,6	2,66E-06 ***
X23	1,59621	0,00208131	766,9	1,70E-06 ***
X24	-3,70057E-05	3,89575E-08	-949,9	1,11E-06 ***
X25	0,0331699	7,59044E-05	437,0	5,24E-06 ***
X51	-0,00664673	7,18758E-05	-92,48	0,0001 ***
X52	-0,00813845	0,000135413	-60,10	0,0003 ***
X54	-1,72707	0,0122544	-140,9	5,03E-05 ***

Średnia arytmetyczna zmiennej zależnej = 542,616

Odchylenie standardowe zmiennej zależnej = 236,879

Suma kwadratów reszt = 0,00103721

Błąd standardowy reszt = 0,0227729

Współczynnik determinacji $R^2 = 1,00000$

Skorygowany współczynnik $R^2 = 1,00000$

Statystyka $F(13, 2) = 1,24843e+008$ (wartość $p < 0,00001$)

Logarytm wiarygodności = 54,4475

Kryterium informacyjne Akaike'a (AIC) = -80,895

Źródło: Opracowanie własne w programie GRETL.

Model o parametrach jak w tabeli 7.4 jest następujący:

$$X^{47} = f(c, \mathbf{X36}, X6, X7, X13, X19, \mathbf{X21}, X23, X24, X25, X51, X52, X54)$$

W tym modelu występują dwie zmienne endogeniczne: X21 i X36.

Zmienna X57

ad b) Wyeliminowano następujące zmienne: X3, X4, X14, X18, X19, X20, X27, X28, X29, X30, X31, X66, X67, X78.

ad c) Wyłoniono zmienną najsilniej skorelowaną z X57, jest nią X1 o współczynniku korelacji $r_h = 0,9928$, i następnie poprzez eliminację *a posteriori* otrzymano model końcowy (zob. tab. 7.5).

Tab. 7.5. Parametry oraz statystyka opisowa zmiennej X_{57}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	-606,140	134,507	-4,506	0,0108	**
X_8	-1,21046	0,0136098	-88,94	9,58E-08	***
X_{12}	-0,930553	0,0142033	-65,52	3,25E-07	***
X_{23}	-1,25205	0,0193109	-64,84	3,39E-07	***
X_{24}	5,89525E-05	9,58333E-07	61,52	4,18E-07	***
X_{26}	3,94232	0,0550540	71,61	2,28E-07	***
X_{34}	-8,85757	0,148814	-59,52	4,77E-07	***
X_{35}	0,0270387	0,000557579	48,49	1,08E-06	***
X_{37}	0,0213594	0,000567893	37,61	2,98E-06	***
X_{39}	0,648024	0,0109195	59,35	4,83E-07	***
X_{46}	-0,323673	0,0230106	-14,07	0,0001	***
X_{49}	-0,115892	0,0133104	-8,707	0,0010	***

Średnia arytmetyczna zmiennej zależnej = 1375,19

Odchylenie standardowe zmiennej zależnej = 539,294

Suma kwadratów reszt = 1,989

Błąd standardowy reszt = 0,705159

Wsp. determinacji $R^2 = 1,00000$

Skorygowany wsp. $R^2 = 1,00000$

Statystyka $F(11, 4) = 797582$ (wartość $p < 0,00001$)

Logarytm wiarygodności = -6,02334

Kryterium informacyjne Akaike'a (AIC) = 36,0467.

Źródło: Opracowanie własne w programie GRETL.

Model o parametrach wymienionych w tabeli 7.5 nie zawiera zmiennej endogenicznej i jego postać jest następująca:

$$X_{57} = f(c, X_8, X_{12}, X_{23}, X_{24}, X_{26}, X_{34}, X_{35}, X_{37}, X_{39}, X_{46}, X_{49})$$

Zmienna X_{60}

ad b) Wyeliminowano następujące zmienne: $X_3, X_4, X_{14}, X_{18}, X_{19}, X_{20}, X_{27}, X_{28}, X_{29}, X_{30}, X_{31}, X_{66}, X_{67}, X_{78}$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z X_{60} , jest nią X_{25} o współczynniku korelacji $r_h = 0,9889$, i następnie poprzez eliminację *a posteriori* otrzymano model końcowy (zob. tab. 7.6).

Tab. 7.6. Parametry oraz statystyka opisowa zmiennej X_{60}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	1799,38	9,51508	189,1	2,80E-05	***
X_2	0,129847	0,000991327	131,0	5,83E-05	***
X_{12}	-1,27182	0,00569868	-223,2	2,01E-05	***
X_8	-1,22551	0,00463034	-264,7	1,43E-05	***
X_{41}	0,0133802	0,000103353	129,5	5,97E-05	***
X_{17}	-1,60751	0,0100240	-160,4	3,89E-05	***
X_{46}	-0,556619	0,00800182	-69,56	0,0002	***
X_{48}	-0,502649	0,00378684	-132,7	5,68E-05	***

X50	-0,0254642	0,000256884	-99,13	0,0001	***
X51	-0,142187	0,000872142	-163,0	3,76E-05	***
X59	0,907679	0,0217326	41,77	0,0006	***
X61	-0,319722	0,0207444	-15,41	0,0042	***
X62	-0,00881339	0,000628578	-14,02	0,0050	***
X64	-0,000538683	1,89860E-05	-28,37	0,0012	***

Średnia arytmetyczna zmiennej zależnej = 241,2

Odchylenie standardowe zmiennej zależnej = 137,227

Suma kwadratów reszt = 0,100726

Błąd standardowy reszt = 0,224417

Wsp. determinacji $R^2 = 1,00000$

Skorygowany wsp. $R^2 = 1,00000$

Statystyka $F(13, 2) = 431437$ (wartość $p < 0,00001$)

Logarytm wiarygodności = 17,8405

Kryterium informacyjne Akaike'a (AIC) = -7,68099

Źródło: Opracowanie własne w programie *GRET*.

Model, w którym występuje zmienna endogeniczna X62 jest następujący:

$$X_{60}^* = f(c, X2, X8, X12, X17, X41, X46, X48, X50, X51, X59, X61, \mathbf{X62}, X64)$$

Zmienna X61

ad b) Wyeliminowano następujące zmienne: X4, X5, X8, X14, X16, X20, X21, X26, X27, X28, X29, X33, X35, X37, X40, X44, X50, X52, X63, X69, X77, X79.

ad c) Wyłoniono zmienną najsilniej skorelowaną z X61, jest nią X58 o współczynniku korelacji $r_h = 0,9256$, i następnie poprzez eliminację *a posteriori* otrzymano model końcowy (zob. tab. 7.7).

Tab. 7.7. Parametry oraz statystyka opisowa zmiennej X61

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	-1517,33	9,85766	-153,9	4,22E-05	***
X58	0,542607	0,00200718	270,3	1,37E-05	***
X2	0,0640580	0,000448190	142,9	4,89E-05	***
X31	0,160760	0,00106406	151,1	4,38E-05	***
X59	0,550401	0,00850740	64,70	0,0002	***
X64	0,000254208	6,84162E-06	37,16	0,0007	***
X66	0,00118056	1,37455E-05	85,89	0,0001	***
X67	-0,00188392	1,74550E-05	-107,9	8,58E-05	***
X70	0,00269796	5,26664E-05	51,23	0,0004	***
X72	0,571061	0,00858318	66,53	0,0002	***
X73	0,331739	0,0200231	16,57	0,0036	***
X74	0,956332	0,00678777	140,9	5,04E-05	***
X78	1,29934	0,153457	8,467	0,0137	**
X76	-0,0604667	0,00751658	-8,044	0,0151	**

Średnia arytmetyczna zmiennej zależnej = 82,5375

Odchylenie standardowe zmiennej zależnej = 18,1787

Suma kwadratów reszt = 0,00287146

Błąd standardowy reszt = 0,037891

Wsp. determinacji $R^2 = 1,00000$
 Skorygowany wsp. $R^2 = 1,00000$
 Statystyka $F(13, 2) = 265582$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = 46,3012
 Kryterium informacyjne Akaike'a (AIC) = -64,6023

Źródło: Opracowanie własne w programie *GRETL*.

Model jest następujący:

$$X_{61}^{\wedge} = f(c, X2, X31, X58, X59, X64, X66, X67, \mathbf{X70, X72, X73, X74, X76, X78})$$

W modelu występują cztery zmienne endogeniczne: $X70, X72, X73, X74$.

Zamieszczenie wszystkich tabel z modelami wynikowymi zmiennych uzyskanych z *GRETL*-a jest zbyt obszerne. Z tego powodu dla kolejnych zmiennych ograniczono się do podania procedury i równania danego modelu z wyszczególnieniem tylko zmiennych. Natomiast równania z parametrami poszczególnych zmiennych zestawiono w tabeli 7.8.

Zmienna $X62$

ad b) Wyeliminowano następujące zmienne: $X4, X14, X18, X19, X20, X27, X28, X29, X30, X67, X78$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z $X62$, jest nią $X57$ o współczynniku korelacji $r_h = 0,9796$. Następnie dokonano estymacji KMNK z wykorzystaniem 16. obserwacji dla zmiennej zależnej $X62$ z zastosowaniem *GRETL*-a. Model jest następujący:

$$X_{62}^{\wedge} = f(c, X32, X40, X42, X46, X48, X49, X51, X54, \mathbf{X60, X63, X64, X65, X70})$$

Jak widzimy, w wyniku aż 255 iteracji *GRETL* wyeliminował zmienną endogeniczną $X57$, lecz w modelu pozostają jeszcze zmienne endogeniczne: $X60, X63$ oraz $X70$.

Zmienna $X63$

ad b) Wyeliminowano następujące zmienne: $X2, X3, X4, X14, X18, X22, X24, X27, X28, X29, X30, X31, X32, X45, X46, X58, X61, X66, X67, X71$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z $X63$, jest nią $X79$ o współczynniku korelacji $r_h = 0,9177$, którą jednak według podpowiedzi *GRETL*-a w wyniku eliminacji *a posteriori* usunięto. Model jest następujący:

$$X_{63}^{\wedge} = f(c, X8, X11, X12, X15, X17, X47, X48, X50, X51, X52, X53, \mathbf{X62, X74})$$

W modelu występują dwie zmienne endogeniczne: $X62$ i $X74$.

Zmienna $X70$

ad b) Wyeliminowano następujące zmienne: $X2, X3, X4, X15, X16, X18, X19, X20, X27, X28, X30, X31, X32, X52, X78$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z $X70$, jest nią $X43$ o współczynniku korelacji $r_h = 0,9681$. Końcowy model z estymacji KMNK z wykorzystaniem 16. obserwacji dla zmiennej zależnej $X70$ jest następujący:

$$X_{70}^{\wedge} = f(c, X5, X9, X10, X11, X12, X14, X23, X24, X25, X29, \mathbf{X36, X43, X76})$$

W tym modelu występuje jedna zmienna endogeniczna modelu Y .

Zmienna $X72$

ad b) Wyeliminowano następujące zmienne: $X3, X4, X14, X18, X19, X20, X27, X28, X29, X30, X31, X52, X78$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z $X72$, jest nią $X75$ o współczynniku korelacji

$r_h = 0,9818$. Po eliminacji *a posteriori* uzyskano model, w którym występują tylko zmienne egzogeniczne, jest on następujący:

$$X^{*72} = f(c, X_1, X_6, X_7, X_8, X_9, X_{10}, X_{11}, X_{12}, X_{13}, X_{15}, X_{16}, X_{56})$$

Zmienna X73

ad b) Wyeliminowano następujące zmienne: $X_4, X_{14}, X_{18}, X_{19}, X_{20}, X_{24}, X_{27}, X_{28}, X_{29}, X_{66}, X_{67}, X_{78}$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z X_{73} , jest nią X_1 o współczynniku korelacji $r_h = 0,9261$. W wyniku kolejnych iteracji uzyskano model, w którym występują trzy zmienne endogeniczne: X_{70}, X_{74}, X_{75} :

$$X^{*73} = f(c, X_{49}, X_{68}, X_{69}, X_{70}, X_{71}, X_{72}, X_{74}, X_{75}).$$

Zmienna X74

ad b) Wyeliminowano następujące zmienne: $X_3, X_4, X_{14}, X_{18}, X_{19}, X_{20}, X_{27}, X_{28}, X_{29}, X_{30}, X_{31}, X_{52}, X_{66}, X_{78}$.

ad c) Wyłoniono zmienną najsilniej skorelowaną z X_{74} , jest nią X_{49} o współczynniku korelacji $r_h = 0,9842$, którą jednak *GRET*L w drodze eliminacji zaproponował do usunięcia z modelu.

$$X^{*74} = f(c, X_1, X_2, X_{32})$$

Zatem w modelu pozostały tylko zmienne egzogeniczne.

7.3. Propozycja wstępna równań wraz z parametrami modelu ekonometrycznego

W tym podrozdziale dokonano zestawienia równań wraz z parametrami dla poszczególnych zmiennych, co stanowi wielorównaniowy model ekonometryczny wzajemnie zależnych zmiennych.

Tab. 7.8. Zestawienie równań modelu ekonometrycznego z uwzględnieniem zaokrąglonych wartości parametrów

Zmienna	Równanie
Y	$7013,05 - 25,89X_{21} + 0,13X_{36} - 24,62X_{47} + 7X_{57} + 7,96X_{60} - 48,24X_{61} - 1,24X_{62} - 3,06X_{63} - 0,04X_{70} + 36,65X_{72} - 67,61X_{73} + 109,45X_{74}$
X_{21}	$589,326 - 1,3609X_{44} - 0,38X_{46}$
X_{36}	$66764,8 + 3581,5X_6 + 7,5061X_{15} - 41,5674X_{16} - 124,367X_{17} + 98,2313X_{21} - 418,541X_{23} - 76,5408X_{45} + 401,587X_{47} - 46,8263X_{49} - 18,0817X_{57} - 2,7746X_{63} - 0,0206X_{65}$
X_{47}	$-354,736 - 8,4325X_6 + 0,4863X_7 + 0,4393X_{13} + 0,0479X_{15} - 0,5117X_{19} - 0,2818X_{21} + 1,5962X_{23} - 0,00004X_{24} + 0,0332X_{25} + 0,0022X_{36} - 0,0066X_{51} - 0,0081X_{52} - 1,7271X_{54}$
X_{57}	$-606,14 - 1,2105X_8 - 0,9305X_{12} - 1,2521X_{23} + 0,0006X_{24} + 3,9423X_{26} - 8,8576X_{34} + 0,027X_{35} + 0,0214X_{37} + 0,648X_{39} - 0,3237X_{46} - 0,1159X_{49}$
X_{60}	$-1799,38 + 0,1298X_2 - 1,2255X_8 - 1,2718X_{12} - 1,6075X_{17} + 0,0134X_{41} - 0,5566X_{46} - 0,5026X_{48} - 0,0255X_{50} - 0,1422X_{51} + 0,9077X_{59} - 0,3197X_{61} - 0,0088X_{62} - 0,0005X_{64}$
X_{61}	$-1517,33 + 0,064X_2 + 0,1608X_{31} + 0,5426X_{38} + 0,55X_{39} + 0,0003X_{64} + 0,0012X_{66} - 0,0019X_{67} + 0,0027X_{70} + 0,571X_{72} + 0,3317X_{73} + 0,9563X_{74} - 0,0605X_{76} + 1,2993X_{78}$
X_{62}	$-3049 - 0,0308X_{32} + 2,1664X_{40} - 0,0568X_{42} + 5,8495X_{46} + 2,0569X_{48} + 0,7354X_{49} + 0,1631X_{51} + 79,1398X_{54} - 0,463X_{60} - 0,7043X_{63} - 0,0018X_{64} - 0,0034X_{65} - 0,0312X_{70}$
X_{63}	$-2874,61 + 0,511X_8 - 3,6157X_{11} + 3,0316X_{12} - 0,1903X_{15} + 3,6365X_{17} - 3,6623X_{47} + 2,3678X_{48} + 0,128X_{50} + 0,4395X_{51} + 0,8447X_{52} + 3,5X_{53} - 0,3222X_{62} - 5,3146X_{74}$

Zmienna	Równanie
X_{70}	$61734 - 1,9437X_5 - 10,2337X_9 + 39,9036X_{10} - 18,0789X_{11} - 43,9417X_{12} + 111,227X_{14} + 36,2368X_{23} - 0,04X_{24} + 0,5808X_{25} - 9,8408X_{29} - 0,0508X_{36} + 2,5167X_{43} - 45,4584X_{76}$
X_{72}	$-530,918 + 0,0169X_1 - 3,17X_6 + 0,3028X_7 + 0,2173X_8 + 0,0651X_9 + 0,1489X_{10} - 0,2068X_{11} + 0,0979X_{12} + 0,0965X_{13} - 0,0214X_{15} - 0,055X_{16} - 0,0554X_{56}$
X_{73}	$286,369 + 0,1495X_{49} - 0,0006X_{68} + 0,0001X_{69} - 0,0022X_{70} - 0,0023X_{71} + 0,619X_{72} - 1,0668X_{74} - 2,5896X_{75}$
X_{74}	$456,627 - 0,01287X_1 + 0,006X_2 - 0,004X_{32}$

Źródło: Opracowanie własne.

7.4. Dalsza redukcja zmiennych

Podjęto próbę oszacowania parametrów równań wstępnego modelu ekonometrycznego w *GRET*L. W tym celu potraktowano je jako równania łącznie współzależne do rozwiązania których wykorzystywana jest między innymi podwójna metoda najmniejszych kwadratów (2MKN). Sformułowano następujące wyrażenia dla określenia równań (*equation*), przy czym litery zmiennych pisane są prosto:

```
equation Y const X21 X36 X47 X57 X60 X61 X62 X63 X70 X72 X73 X74
equation X21 const X44 X46
equation X57 const X8 X12 X23 X24 X26 X34 X35 X37 X39 X46 X49
equation X72 const X1 X6 X7 X8 X9 X10 X11 X12 X13 X15 X16 X56
equation X74 const X1 X2 X32
equation X36 const X6 X15 X16 X17 X21 X23 X45 X47 X49 X57 X63 X65
equation X47 const X6 X7 X13 X15 X19 X21 X23 X24 X25 X36 X51 X52 X54
equation X60 const X2 X8 X12 X17 X41 X46 X48 X50 X51 X59 X61 X62 X64
equation X61 const X2 X31 X58 X59 X64 X66 X67 X70 X72 X74 X76 X78
equation X62 const X32 X40 X42 X46 X48 X49 X51 X54 X60 X63 X64 X65 X70
equation X63 const X8 X11 X12 X15 X17 X47 X48 X50 X51 X52 X53 X62 X74
equation X70 const X5 X9 X10 X11 X12 X14 X23 X24 X25 X29 X36 X43 X76
equation X73 const X49 X68 X69 X70 X71 X72 X74 X75
endog Y X21 X36 X47 X57 X60 X61 X62 X63 X70 X72 X73 X74
instr X1 X2 X5 X6 X7 X8 X9 X10 X11 X12 X13 X14 X15 X16 X17 X19 X23 X24 X25 X26
X29 X31 X32 X34 X35 X37 X39 X40 X41 X42 X43 X44 X45 X46 X48 X49 X50 X51 X52 X53
X54 X56 X58 X59 X64 X65 X66 X67 X68 X69 X71 X75 X76 X78
```

Określenie *endog* oznacza zmienne endogeniczne (zaznaczone czcionką pogrubioną), natomiast *instr* - zmienne egzogeniczne równań. Okazało się jednak, że w wersji programu *GRET*L, pozyskanej z Internetu, dla tak obszernej ilości zmiennych nierealizowana jest procedura 2MKN. Z tego powodu pozostaje nam dokonanie zredukowania zmiennych w długich równaniach, w celu doprowadzenia do modelu równań o parametrach szacowanych KMKN. Rozpoczynamy od usunięcia w pierwszej kolejności zmiennych utrudniających estymację i postępujemy dalej według podpowiedzi *GRET*L-a.

Równanie zmiennej X_{47}

Usuamy z równania zmienne endogeniczne X_{21} i X_{36} i w rezultacie pracy z *GRET*L-em uzyskujemy raport modelu wynikowego zawierający tylko zmienne endogeniczne (tab. 7.9).

Tab. 7.9. Oszacowane parametry i statystyka opisowa zmiennej X_{47} .

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	-1112,89	70,5813	-15,77	7,31E-08	***
X_7	1,06515	0,235898	4,515	0,0015	***
X_{13}	1,02670	0,0906154	11,33	1,25E-06	***
X_{15}	0,101679	0,00770127	13,20	3,40E-07	***
X_{24}	-5,18801E-05	6,80481E-06	-7,624	3,24E-05	***
X_{25}	0,0707010	0,0120701	5,858	0,0002	***
X_{54}	4,56843	1,32070	3,459	0,0072	***

Średnia arytmetyczna zmiennej zależnej = 542,616

Odchylenie standardowe zmiennej zależnej = 236,879

Suma kwadratów reszt = 737,717

Błąd standardowy reszt = 9,05365

Wsp. determinacji $R^2 = 0,99912$

Skorygowany wsp. $R^2 = 0,99854$

Statystyka $F(6, 9) = 1709,88$ (wartość $p < 0,00001$)

Logarytm wiarygodności = -53,3508

Kryterium informacyjne Akaike'a (AIC) = 120,702

Źródło: Opracowanie własne w *GRET*L.

Równanie zmiennej X_{63}

Wyeliminowanie z modelu dla X_{63} zmiennych X_{74} oraz X_{62} , pozostawiając jednak X_{47} , którą można określić na podstawie wcześniejszego równania. W rezultacie pracy z *GRET*L-em uzyskujemy model zaprezentowany w tabeli 7.10.

Tab. 7.10. Oszacowanie parametrów i statystyka opisowa zmiennej X_{63}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	-8856,38	1163,65	-7,611	0,0003	***
X_{11}	-4,10857	0,676741	-6,071	0,0009	***
X_{12}	4,68431	0,685967	6,829	0,0005	***
X_{17}	5,12600	0,830251	6,174	0,0008	***
X_{47}	-6,13762	1,05329	-5,827	0,0011	***
X_{48}	3,03402	0,553192	5,485	0,0015	***
X_{50}	0,169401	0,0241502	7,014	0,0004	***
X_{51}	0,735094	0,115372	6,371	0,0007	***
X_{52}	1,30426	0,261111	4,995	0,0025	***
X_{53}	5,75515	0,646171	8,907	0,0001	***

Średnia arytmetyczna zmiennej zależnej = 484,669

Odchylenie standardowe zmiennej zależnej = 241,313

Suma kwadratów reszt = 6648,22

Błąd standardowy reszt = 33,2872

Wsp. determinacji $R^2 = 0,99239$

Skorygowany wsp. $R^2 = 0,98097$

Statystyka $F(9, 6) = 86,9235$ (wartość $p = 1,16e-005$)

Logarytm wiarygodności = -70,9391

Kryterium informacyjne Akaike'a (AIC) = 161,878

Źródło: Opracowanie własne w *GRET*L.

Równanie zmiennej X_{60}

Wyliminowanie z X_{60} zmiennych X_{62} , X_{61} i poprzez eliminację *a posteriori* w *GRET*L otrzymujemy raport modelu (tab. 7.11).

Tab. 7.11. Oszacowanie parametrów i statystyka opisowa zmiennej X_{60}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	676,25	100,504	16,68	7,57E-05	***
X_2	0,144977	0,0100261	14,46	0,0001	***
X_8	-1,24404	0,0615323	-20,22	3,53E-05	***
X_{12}	-1,30713	0,0692148	-18,89	4,63E-05	***
X_{17}	-1,70194	0,109336	-15,57	9,94E-05	***
X_{41}	0,0144662	0,000881244	16,42	8,06E-05	***
X_{46}	-0,554490	0,112667	-4,922	0,0079	***
X_{48}	-0,555539	0,0358288	-15,51	0,0001	***
X_{50}	-0,0277238	0,00268919	-10,31	0,0005	***
X_{51}	-0,141580	0,00936618	-15,12	0,0001	***
X_{59}	0,684444	0,123311	5,551	0,0052	***
X_{64}	-0,000713767	0,000205908	-3,466	0,0257	**

Średnia arytmetyczna zmiennej zależnej = 241,2

Odchylenie standardowe zmiennej zależnej = 137,227

Suma kwadratów reszt = 40,3279

Błąd standardowy reszt = 3,17521

Wsp. determinacji $R^2 = 0,99986$

Skorygowany wsp. $R^2 = 0,99946$

Statystyka $F(11, 4) = 2546,67$ (wartość $p < 0,00001$)

Logarytm wiarygodności = -30,0986

Kryterium informacyjne Akaike'a (AIC) = 84,1973

Źródło: Opracowanie własne w *GRET*L.

Po zredukowaniu powyższych zmiennych równania przyjmą postać zamieszczoną w tabeli 7.12.

Tab. 7.12. Postać równań po zredukowaniu zmiennych

Zmienna	Równanie
Y	$7013,05 - 25,89X_{21} + 0,13X_{36} - 24,62X_{47} + 7X_{57} + 7,96X_{60} - 48,24X_{61} - 1,24X_{62} - 3,06X_{63} - 0,04X_{70} + 36,65X_{72} - 67,61X_{73} + 109,45X_{74}$
X_{21}	$589,326 - 1,3609X_{44} - 0,38X_{46}$
X_{36}	$66764,8 + 3581,5X_6 + 7,5061X_{15} - 41,5674X_{16} - 124,367X_{17} + 98,2313X_{21} - 418,541X_{23} - 76,5408X_{45} + 401,587X_{47} - 46,8263X_{49} - 18,0817X_{57} - 2,7746X_{63} - 0,0206X_{65}$
X_{47}	$-1112,89 + 1,0652X_7 + 1,0267X_{13} + 0,1017X_{15} - 0,00005X_{24} + 0,07X_{25} + 4,5684X_{54}$
X_{57}	$-606,14 - 1,2105X_8 - 0,9305X_{12} - 1,2521X_{23} + 0,0006X_{24} + 3,9423X_{26} - 8,8576X_{34} + 0,027X_{35} + 0,0214X_{37} + 0,648X_{39} - 0,3237X_{46} - 0,1159X_{49}$
X_{60}	$1676,25 + 0,145X_2 - 1,244X_8 - 1,3071X_{12} - 1,7019X_{17} + 0,0145X_{41} - 0,5545X_{46} - 0,5555X_{48} - 0,0277X_{50} - 0,1416X_{51} + 0,6844X_{59} - 0,0007X_{64}$

Zmienna	Równanie
X_{61}	$-1517,33 + 0,064X_2 + 0,1608X_{31} + 0,5426X_{38} + 0,55X_{39} + 0,0003X_{64} + 0,0012X_{66} - 0,0019X_{67} + 0,0027X_{70} + 0,571X_{72} + 0,3317X_{73} + 0,9563X_{74} - 0,0605X_{76} + 1,2993X_{78}$
X_{62}	$-3049 - 0,0308X_{32} + 2,1664X_{40} - 0,0568X_{42} + 5,8495X_{46} + 2,0569X_{48} + 0,7354X_{49} + 0,1631X_{51} + 79,1398X_{54} - 0,463X_{60} - 0,7043X_{63} - 0,0018X_{64} - 0,0034X_{65} - 0,0312X_{70}$
X_{63}	$-8856,38 - 4,1086X_{11} + 4,6843X_{12} + 5,126X_{17} - 6,1376X_{47} + 3,034X_{48} + 0,1694X_{50} + 0,7351X_{51} + 1,3043X_{52} + 5,7552X_{53}$
X_{70}	$61734 - 1,9437X_5 - 10,2337X_9 + 39,9036X_{10} - 18,0789X_{11} - 43,9417X_{12} + 111,227X_{14} + 36,2368X_{23} - 0,04X_{24} + 0,5808X_{25} - 9,8408X_{29} - 0,0508X_{36} + 2,5167X_{43} - 45,4584X_{76}$
X_{72}	$-530,918 + 0,0169X_1 - 3,17X_6 + 0,3028X_7 + 0,2173X_8 + 0,0651X_9 + 0,1489X_{10} - 0,2068X_{11} + 0,0979X_{12} + 0,0965X_{13} - 0,0214X_{15} - 0,055X_{16} - 0,0554X_{56}$
X_{73}	$286,369 + 0,1495X_{49} - 0,0006X_{68} + 0,0001X_{69} - 0,0022X_{70} - 0,0023X_{71} + 0,619X_{72} - 1,0668X_{74} - 2,5896X_{75}$
X_{74}	$456,627 - 0,01287X_1 + 0,006X_2 - 0,004X_{32}$

Źródło: Opracowanie własne w Excelu.

7.5. Generacja zmiennych endogenicznych

Kolejność generacji zmiennych w modelu dla oszacowania parametrów klasyczna metoda najmniejszych kwadratów (KMNK) równania Y jest następująca: X_{21} , X_{57} , X_{60} , X_{72} , X_{74} , X_{47} , X_{63} , X_{36} , a następnie: X_{70} , X_{73} , X_{61} , X_{62} . Ustawiamy zakres próby na 1÷16 obserwacji i wpisujemy formułę:

$$\text{genrX21}=(589.326)-(1.3609)*X_{44}-(0.38)*X_{46}.$$

Program *GRET*L dokonuje generacji wartości zmiennej według podanej formuły. Wartości generuje dla obserwacji 1÷16, a obserwacje 17, 18 zarezerwowane są dla prognoz. Należy też wpisywać zamiast przecinka kropkę dziesiętną.

Przy generowaniu wartości dla zmiennej X_{57} *GRET*L przerywa za długą formułę w trakcie jej wpisywania. W tej sytuacji trzeba na nowo oszacować parametry modelu, eliminując dwie zmienne – X_{49} oraz X_{46} o najniższej statystyce t -Studenta i w efekcie pracy programu uzyskujemy raport z estymacji parametrów KMNK (zob. tabela 7.13).

Tab. 7.13. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej X_{57}

Zmienne	Współczynnik	Błąd stand.	t -Studenta	Wartość p	
const	-1824,72	356,720	-5,115	0,0022	***
X_8	-1,08563	0,0592977	-18,31	1,71E-06	***
X_{12}	-0,833096	0,0710941	-11,72	2,33E-05	***
X_{23}	-1,16805	0,101347	-11,53	2,56E-05	***
X_{24}	6,10644E-05	5,28539E-06	11,55	2,53E-05	***
X_{26}	4,24460	0,194740	21,80	6,09E-07	***
X_{34}	-7,49385	0,605864	-12,37	1,70E-05	***
X_{35}	0,0243131	0,00285575	8,514	0,0001	***
X_{37}	0,0229068	0,00322687	7,099	0,0004	***
X_{39}	0,734018	0,0292012	25,14	2,61E-07	***

Średnia arytmetyczna zmiennej zależnej = 1375,19

Odchylenie standardowe zmiennej zależnej = 539,294

Suma kwadratów reszt = 100,495
 Błąd standardowy reszt = 4,09258
 Wsp. determinacji $R^2 = 0,99998$
 Skorygowany wsp. $R^2 = 0,99994$
 Statystyka $F(9, 6) = 28939,7$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -37,4032
 Kryterium informacyjne Akaike'a (AIC) = 94,8064

Źródło: Opracowanie własne w programie *GRET*L.

Teraz dokonujemy aktualizacji wcześniej wygenerowanej zmiennej „genrX57” analogicznie do powyższej formuły korzystając z menu (*Zmienna/Edycja atrybutów*):

$$\text{genrX57} = (-1824,72) - (1,0856) * X8 - (0,833) * X12 - (1,168) * X23 + (0,00006) * X24 + (4,2446) * X26 - (7,4939) * X34 + (0,0243) * X35 + (0,0229) * X37 + (0,734) * X39.$$

Każdorazowo po wygenerowaniu sprawdzamy poprzez menu (*Dane/Pokaż wartości*).

Zmienna X60 zawiera 11 zmiennych objaśniających. Przed generacją dokonajmy w *GRET*L redukcji co najmniej 2-ch zmiennych z przyczyny podobnej jak dla zmiennej X57. Usunięto X64, X59, X46, X50, X51 i w efekcie KMNK otrzymano raport modelu podany w tabelicy 7.14.

Tab. 7.14. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej X60

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	907,102	190,641	4,758	0,0010 ***
X12	-0,890917	0,236097	-3,774	0,0044 ***
X17	-1,11315	0,293309	-3,795	0,0042 ***
X41	0,00717074	0,00124663	5,752	0,0003 ***
X48	-0,229530	0,0601886	-3,814	0,0041 ***
X2	0,0670694	0,0223263	3,004	0,0149 **
X8	-0,977296	0,224610	-4,351	0,0018 ***

Średnia arytmetyczna zmiennej zależnej = 241,2
 Odchylenie standardowe zmiennej zależnej = 137,227
 Suma kwadratów reszt = 3023,38
 Błąd standardowy reszt = 18,3284
 Wsp. determinacji $R^2 = 0,98930$
 Skorygowany wsp. $R^2 = 0,98216$
 Statystyka $F(6, 9) = 138,643$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -64,6354
 Kryterium informacyjne Akaike'a (AIC) = 143,271

Źródło: Opracowanie własne w programie *GRET*L.

Formuła do generacji nowej zmiennej genrX60:

$$\text{genrX60} = (907,102) - (0,8909) * X12 - (1,1132) * X17 + (0,0072) * X41 - (0,2295) * X48 + (0,0671) * X2 - (0,9773) * X8$$

Zmienna X72 opisywana jest 12 zmiennymi objaśniającymi. Redukujemy więc co najmniej trzy. Kolejno usunięto X56, X1, X13 oraz wg sugestii *GRET*L-a dalsze X7, X6, X10. W wyniku estymacji KMNK uzyskano raport modelu zaprezentowany w tabeli 7.15.

Tab. 7. 5. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej X_{72}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	80,1420	20,3807	3,932	0,0034 ***
X_8	0,0702757	0,0173395	4,053	0,0029 ***
X_9	0,0454568	0,0110396	4,118	0,0026 ***
X_{11}	-0,0395879	0,00665951	-5,945	0,0002 ***
X_{12}	0,0577609	0,0226252	2,553	0,0310 **
X_{15}	-0,0167265	0,00374487	-4,467	0,0016 ***
X_{16}	-0,0489996	0,0129811	-3,775	0,0044 ***

Średnia arytmetyczna zmiennej zależnej = 111,25

Odchylenie standardowe zmiennej zależnej = 11,5115

Suma kwadratów reszt = 40,4628

Błąd standardowy reszt = 2,12035

Wsp. determinacji $R^2 = 0,97964$

Skorygowany wsp. $R^2 = 0,96607$

Statystyka $F(6, 9) = 72,1869$ (wartość $p < 0,00001$)

Logarytm wiarygodności = -30,1254

Kryterium informacyjne Akaike'a (AIC) = 74,2508

Źródło: Opracowanie własne w programie *GRETL*.

Formuła do generacji nowych wartości nowej zmiennej $genX_{72}$:

$$genX_{72} = (80.142) + (0.07) * X_8 + (0.0455) * X_9 - (0.0396) * X_{11} + (0.0578) * X_{12} - (0.0167) * X_{15} - (0.049) * X_{16}$$

Zmienna X_{74} zawiera tylko 3 zmienne objaśniające, zatem możemy realizować bezpośrednio generowanie wartości tej zmiennej dla poszczególnych 16. Obserwacji według formuły:

$$genX_{74} = (456.627) - (0.0129) * X_1 + (0.006) * X_2 - (0.0004) * X_{32}.$$

Wygenerowane wartości są ujemne, co może wynikać z zaokrąglenia wartości parametrów równania, a dane rzeczywiste są dodatnie. W tej sytuacji ponownie przeprowadzono kontrolnie analizę doboru zmiennych w *GRETL* bazując na zmiennych wstępnych po etapie b) analizy współczynników korelacji. W wyniku procedury *a posteriori* wyeliminowano następujące zmienne: X_3 , X_4 , X_{14} , X_{18} , X_{19} , X_{20} , X_{27} , X_{28} , X_{29} , X_{30} , X_{31} , X_{52} , X_{66} , X_{78} . Wyłoniono zmienną najsilniej skorelowaną z X_{74} , jest nią X_{49} . Bez usuwania zmiennej X_{49} model pokazano w tabeli 7.16.

Tab. 7.16. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej X_{74}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	11,6438	2,06386	5,642	0,0005 ***
X_{49}	0,0838755	0,000928170	90,37	2,51E-013 ***
X_2	0,00139072	0,000208188	6,680	0,0002 ***
X_5	0,000543344	8,73826E-05	6,218	0,0003 ***
X_6	-2,05715	0,0659397	-31,20	1,21E-09 ***
X_7	0,231403	0,00811463	28,52	2,47E-09 ***
X_{12}	-0,0253246	0,00233207	-10,86	4,57E-06 ***
X_{15}	-0,00245324	0,000241769	-10,15	7,61E-06 ***

Średnia arytmetyczna zmiennej zależnej = 112,456
 Odchylenie standardowe zmiennej zależnej = 12,6207
 Suma kwadratów reszt = 0,37313
 Błąd standardowy reszt = 0,215966
 Wsp. determinacji $R^2 = 0,99984$
 Skorygowany wsp. $R^2 = 0,99971$
 Statystyka $F(7, 8) = 7316,84$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = 7,36432
 Kryterium informacyjne Akaike'a (AIC) = 1,27136

Źródło: Opracowanie własne w programie *GRET*L.

Następnie generujemy wartość zmiennej $genrX74$:

$$genrX74 = (11.6438) + (0.0839) * X49 + (0.0014) * X2 + (0.0005) * X5 - (2,0571) * X6 + (0,2314) * X7 - (0,0253) * X12 - (0,0024) * X15.$$

Teraz mamy wartości dodatnie przybliżone do wartości rzeczywistych.

W kolejnym etapie generujemy wartości teoretyczne zmiennej $X47$ obejmującej 6 zmiennych objaśniających:

$$genrX47 = (-1112.89) + (1.0652) * X7 + (1.0267) * X13 + (0.1017) * X15 - (0.00005) * X24 + (0.07) * X25 + (4.5684) * X54.$$

Zmienna $X63$ jest funkcją zmiennych objaśniających oraz jednej zmiennej endogenicznej $X47$. Generujemy więc wartości teoretyczne, przyjmując wartości zmiennej jako $genrX47$:

$$genrX63 = (-8856.38) - (4.1086) * X11 + (4.6843) * X12 + (5.126) * X17 - (6.1376) * genrX47 + (3.034) * X48 + (0.1694) * X50 + (0.7351) * X51 + (1.3043) * X52 + (5.7552) * X53.$$

Funkcja $genr$ sygnalizuje nadmierną ilość zmiennych i dlatego ponownie dokonujemy eliminacji zmiennych dla $X63$ korzystając z programu *GRET*L. W rezultacie otrzymano raport modelu zamieszczony w tabeli 7.17. Po ponownej eliminacji *a posteriori* nie występuje już zmienna endogeniczna $X47$ w równaniu zmiennej $X63$.

$$genrX63 = (-2268.72) - (0.9724) * X11 + (1.2675) * X12 + (0.0431) * X50 + (0.3049) * X51 + (3.9284) * X53.$$

Tab. 7.17. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej $X63$

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	-2268,72	652,443	-3,477	0,0059	***
X11	-0,972436	0,297423	-3,270	0,0084	***
X12	1,26747	0,611931	2,071	0,0651	*
X50	0,0431186	0,0154386	2,793	0,0190	**
X51	0,304886	0,126913	2,402	0,0372	**
X53	3,92842	0,895904	4,385	0,0014	***

Średnia arytmetyczna zmiennej zależnej = 484,669
 Odchylenie standardowe zmiennej zależnej = 241,313
 Suma kwadratów reszt = 51717,3
 Błąd standardowy reszt = 71,9148
 Współczynnik determinacji $R^2 = 0,94079$

Skorygowany współczynnik $R^2 = 0,91119$
 Statystyka $F(5, 10) = 31,7789$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -87,3507
 Kryterium informacyjne Akaike'a (AIC) = 186,701

Źródło: Opracowanie własne w programie GRETL.

Generowanie wartości teoretycznych zmiennej X_{36} rozpoczynamy od eliminacji nadmiernej ilości zmiennych objaśniających. Po estymacji KMNK z wykorzystaniem 16. obserwacji 1-16 użytkano model zmiennej zależnej X_{36} zawarty w tabeli 7.18.

Tab. 7.18. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej X_{36}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	-47622,3	16155,3	-2,948	0,0133	**
X_6	3121,59	953,847	3,273	0,0074	***
X_{21}	78,7148	18,0864	4,352	0,0012	***
X_{23}	-322,981	73,9680	-4,367	0,0011	***
X_{47}	382,024	12,2085	31,29	4,21E-012	***

Średnia arytmetyczna zmiennej zależnej = 165090
 Odchylenie standardowe zmiennej zależnej = 84695,9
 Suma kwadratów reszt = 2,2006e+008
 Błąd standardowy reszt = 4472,75
 Wsp. determinacji $R^2 = 0,99795$
 Skorygowany wsp. $R^2 = 0,99721$
 Statystyka $F(4, 11) = 1341,89$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -154,198
 Kryterium informacyjne Akaike'a (AIC) = 318,395

Źródło: Opracowanie własne w programie GRETL.

Pozostaje nam teraz wygenerowanie wartości teoretycznych nowej zmiennej $genrX_{36}$:

$$genrX_{36} = (-47622.3) + (3121.59) * X_6 + (78.7148) * X_{21} - (322.981) * X_{23} + (382.024) * (genrX_{47}).$$

W tym równaniu występuje zmienna endogeniczna X_{47} dla której wcześniej wygenerowaliśmy wartości teoretyczne.

Eliminujemy teraz nadmierną ilość zmiennych objaśniających w równaniu zmiennej X_{70} , kierując się ilością gwiazdek przy kolejnych modelach procedury iteracyjnej *a posteriori*. Kosztem jest jednak osłabienie współczynnika determinacji. W rezultacie po estymacji KMNK uzyskano model, a następnie wygenerowano wartości teoretyczne (modelowe).

Tab. 7.19. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej X_{70}

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p	
const	4595,61	1568,11	2,931	0,0126	**
X_5	-1,03810	0,364126	-2,851	0,0146	**
X_{10}	11,4281	3,88916	2,938	0,0124	**
X_{43}	2,41020	0,159050	15,15	3,46E-09	***

Średnia arytmetyczna zmiennej zależnej = 14021,2
 Odchylenie standardowe zmiennej zależnej = 2888,72

Suma kwadratów reszt = 4,51142e+006
 Błąd standardowy reszt = 613,149
 Wsp. determinacji $R^2 = 0,96396$
 Skorygowany wsp. $R^2 = 0,95495$
 Statystyka $F(3, 12) = 106,982$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -123,099
 Kryterium informacyjne Akaike'a (AIC) = 254,199

Źródło: Opracowanie własne w programie *GRET*L.

$$\text{genrX70} = (4595.61) - (1.0381) * X5 + (11.4281) * X10 + (2.4102) * X43.$$

Zmienna *X73* zawiera 8 zmiennych, w tym 3 endogeniczne *X70*, *X72*, *X74* dla których szeregi wartości teoretycznych zostały już wygenerowane w bazie *GRET*L-a o nazwie *Korelacja2.gdt*. Dokonajmy redukcji ilości zmiennych *eliminacją a posteriori*, a otrzymamy model ujęty w tabeli 7.20.

Tab. 7.20. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej *X73*

Zmienna	Współczynnik	Błąd standardowy	t-Student	wartość p
const	125,537	1,20875	103,9	2,29E-020 ***
X68	-0,000392519	0,000103145	-3,805	0,0022 ***
X71	-0,00167352	0,000504641	-3,316	0,0056 ***

Średnia arytmetyczna zmiennej zależnej = 110,4
 Odchylenie standardowe zmiennej zależnej = 8,21949
 Suma kwadratów reszt = 54,3983
 Błąd standardowy reszt = 2,0456
 Współczynnik. determinacji $R^2 = 0,94632$
 Skorygowany współczynnik $R^2 = 0,93806$
 Statystyka $F(2, 13) = 114,59$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -32,493
 Kryterium informacyjne Akaike'a (AIC) = 70,9859

Źródło: Opracowanie własne w programie *GRET*L.

Mając określone zmienne formuła do generacji wartości teoretycznych zmiennej *genrX73* jest następująca:

$$\text{genrX73} = (125.537) - (0.0004) * X68 - (0.0017) * X71.$$

Kolejne równanie zmiennej *X61* zawiera nadmierną ilość zmiennych objaśniających dla generacji modelu w *GRET*L. Dokonujemy więc dodatkowej eliminacji *a posteriori*, w wyniku czego uzyskujemy model zaprezentowany w tabeli 7.21.

Tab. 7.21. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej *X61*

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	-501,690	90,8163	-5,524	0,0003 ***
X2	0,0197422	0,00702897	2,809	0,0185 **
X31	0,0701244	0,0223826	3,133	0,0106 **
X58	0,301326	0,0444207	6,783	4,84E-05 ***
X59	0,686903	0,168683	4,072	0,0022 ***
X74	0,840382	0,365652	2,298	0,0444 **

Średnia arytmetyczna zmiennej zależnej = 82,5375
 Odchylenie standardowe zmiennej zależnej = 18,1787
 Suma kwadratów reszt = 107,973
 Błąd standardowy reszt = 3,28592
 Współczynnik determinacji $R^2 = 0,97822$
 Skorygowany współczynnik $R^2 = 0,96733$
 Statystyka $F(5, 10) = 89,8186$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -37,9773
 Kryterium informacyjne Akaike'a (AIC) = 87,9547

Źródło: Opracowanie własne w programie GRETL.

Generujemy wartości teoretyczne $X61$ przy jednej zmiennej endogenicznej $genrX74$ w postaci także wygenerowanej.

$$genrX61 = (-501.69) + (0.0197) * X2 + (0.07) * X31 + (0.3) * X58 + (0.6869) * X59 + (0.84) * (genrX74).$$

Dokonyjemy redukcji nadmiernej ilości zmiennych w równaniu $X62$ poprzez dodatkową eliminację *a posteriori*, uzyskując model podany w tabeli 7.22.

Tab. 7.22. Ponowne oszacowanie parametrów i statystyka opisowa zmiennej $X62$

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	4217,97	245,215	17,20	9,32E-09 ***
X32	-0,0342344	0,00243593	-14,05	6,53E-08 ***
X46	7,93617	1,48344	5,350	0,0003 ***
X48	2,11716	0,185095	11,44	4,58E-07 ***
X63	-0,551560	0,129318	-4,265	0,0016 ***
X65	-0,00543825	0,00107772	-5,046	0,0005 ***

Średnia arytmetyczna zmiennej zależnej = 3052,47
 Odchylenie standardowe zmiennej zależnej = 1029,09
 Suma kwadratów reszt = 47807,4
 Błąd standardowy reszt = 69,1429
 Współczynnik determinacji $R^2 = 0,99699$
 Skorygowany współczynnik $R^2 = 0,99549$
 Statystyka $F(5, 10) = 662,554$ (wartość $p < 0,00001$)
 Logarytm wiarygodności = -86,7218
 Kryterium informacyjne Akaike'a (AIC) = 185,444

Źródło: Opracowanie własne w programie GRETL.

Generujemy wartości teoretyczne $X62$ przy jednej zmiennej endogenicznej postaci $genrX63$.

$$genrX62 = (4217.97) - (0.0342) * X32 + (7.9362) * X46 + (2.1172) * X48 - (0.5516) * (genrX63) - (0.0054) * X65$$

Dotychczasowe formuły dla wygenerowanych wartości teoretycznych zmiennych w bazie danych *Korelacja2.gdt* zestawiono w tabeli 7.23.

Tab. 7.23. Formuły wygenerowanych wartości teoretycznych zmiennych

$\text{genrX21}=(589.326)-(1.3609)*X44-(0.38)*X46$
$\text{genrX57}=(-1824.7)-(1.0856)*X8-(0.833)*X12-(1.168)*X23+(0.00006)*X24+(4.2446)*X26-(7.4939)*X34+(0.0243)*X35+(0.0229)*X37+(0.734)*X39$
$\text{genrX60}=(907.102)-(0.8909)*X12-(1.1132)*X17+(0.0072)*X41-(0.2295)*X48+(0.0671)*X2-(0.9773)*X8$
$\text{genrX72}=(80.142)+(0.07)*X8+(0.0455)*X9-(0.0396)*X11+(0.0578)*X12-(0.0167)*X15-(0.049)*X16$
$\text{genrX74}=(11.6438)+(0.0839)*X49+(0.0014)*X2+(0.0005)*X5-(2.0571)*X6+(0.2314)*X7-(0.0253)*X12-(0.0024)*X15$
$\text{genrX47}=(-1112.89)+(1.0652)*X7+(1.0267)*X13+(0.1017)*X15-(0.00005)*X24+(0.07)*X25+(4.5684)*X54$
$\text{genrX63}=(-2268.72)-(0.9724)*X11+(1.2675)*X12+(0.0431)*X50+(0.3049)*X51+(3.9284)*X53$
$\text{genrX36}=(-47622.3)+(3121.59)*X6+(78.7148)*X21-(322.981)*X23+(382.024)*(\text{genrX47})$
$\text{genrX70}=(4595.61)-(1.0381)*X5+(11.4281)*X10+(2.4102)*X43$
$\text{genrX73}=(125.537)-(0.0004)*X68-(0.0017)*X71$
$\text{genrX61}=(-501.69)+(0.0197)*X2+(0.07)*X31+(0.3)*X58+(0.6869)*X59+(0.84)*(\text{genrX74})$
$\text{genrX62}=(4217.97)-(0.0342)*X32+(7.9362)*X46+(2.1172)*X48-(0.5516)*(\text{genrX63})-(0.0054)*X6$

Źródło: Opracowanie własne w Excelu.

Mając wygenerowane szeregi wartości dla powyższych zmiennych możemy dokonać korekty parametrów modelu zmiennej zależnej Y (PKB) stosując w *GRET* metodę *KMNC* i w efekcie uzyskać model z wykorzystaniem 16. obserwacji. Zwróćmy uwagę na zastosowanie teraz wygenerowanych wartości zmiennych nazwanych *genrX*.

Tab. 7.24. Oszacowanie parametrów oraz statystyka opisowa modelu zmiennej Y

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	66415,8	127856	0,5195	0,6393
genrX21	-36,6149	38,1353	-0,9601	0,4078
genrX36	0,154158	0,179112	0,8607	0,4527
genrX47	-21,8296	60,2922	-0,3621	0,7413
genrX57	0,122494	2,45106	0,04998	0,9633
genrX60	-9,78999	40,8451	-0,2397	0,8260
genrX61	-12,3220	64,4398	-0,1912	0,8606
genrX62	-1,86754	4,82064	-0,3874	0,7243
genrX63	-0,421850	7,61799	-0,05538	0,9593
genrX70	-0,745700	0,970868	-0,7681	0,4984
genrX72	-51,1900	220,885	-0,2317	0,8316
genrX73	-130,443	528,633	-0,2468	0,8210
genrX74	-92,0003	253,792	-0,3625	0,7410

Średnia arytmetyczna zmiennej zależnej = 16168,8

Odczylenie standardowe zmiennej zależnej = 8710,92

Suma kwadratów reszt = 6,27711e+006

Błąd standardowy reszt = 1446,5

Współczynnik determinacji = 0,99449

Skorygowany współczynnik determinacji = 0,97243

Statystyka $F(12, 3) = 45,0815$ (wartość $p = 0,00473$)

Logarytm wiarygodności = -125,742

Kryterium informacyjne Akaike'a (AIC) = 277,483

Wyluczając stałą, największa wartość p jest dla zmiennej 98 (*genrX57*)

Źródło: Opracowanie własne w *GRET*L.

Kierując się kolejnymi podpowiedziami programu *GRET*L na zakończeniu danego raportu wyeliminowano kolejno następujące wygenerowane zmienne o nazwie z przedrostkiem *genr*: *X57*, *X63*, *X61*, *X72*, *X73*, *X60*, *X47*, *X74*, *X62*. W efekcie końcowym stosując KMNK dla zmiennej zależnej uzyskano poniższy lepszy model biorąc pod uwagę kryterium informacyjne Akaike'e.

Tab. 7.25. Model wyników zmiennej Y

Zmienna	Współczynnik	Błąd stand.	t -Studenta	Wartość p	
const	17762,8	2397,01	7,410	8,16E-06	***
genrX21	-27,2193	3,60491	-7,551	6,76E-06	***
genrX36	0,0845154	0,00564217	14,98	3,95E-09	***
genrX70	-0,376227	0,130344	-2,886	0,0137	**

Średnia arytmetyczna zmiennej zależnej = 16168,8

Odchylenie standardowe zmiennej zależnej = 8710,92

Suma kwadratów reszt = 7,01451e+006

Błąd standardowy reszt = 764,554

Współczynnik determinacji = 0,99384

Skorygowany współczynnik determinacji = 0,99230

Statystyka $F(3, 12) = 645,056$ (wartość $p < 0,00001$)

Logarytm wiarygodności = -126,63

Kryterium informacyjne Akaike'a (AIC) = 261,26

Źródło: Opracowanie własne w *GRET*L.

Tak więc zmienna Produkt Krajowy Brutto (PKB - Y) jest funkcją bezpośrednio zmiennych endogenicznych:

X_{21} - zwolnieni w drodze wypowiedzenia przez zakład pracy w tys.,

X_{36} - wynagrodzenia osobowe w mln,

X_{70} - absolwenci szkół wyższych stacjonarnych (inżynieryjno-techniczne).

Przy czym z według tablicy 7.23 zmienna X_{21} zależy od:

X_{44} - zasiłki chorobowe w zł - rolnicy indywidualni,

X_{46} - odszkodowania wypadkowe w zł - rolnicy indywidualni.

Natomiast zmienna X_{36} od:

X_6 - osoby pracujące w tys. wykonujące pracę nakładczą,

X_{21} - zwolnienia z pracy w drodze wypowiedzenia przez zakład pracy w tys.,

X_{23} - przeniesieni na emeryturę, rentę z tytułu niezdolności do pracy, rehabilitację,

X_{24} - liczba dni roboczych ogółem nieprzepracowanych w związku z uczestnictwem w strajkach,

X_{47} - przeciętny miesięczny dochód rozporządzalny na 1 osobę w zł w gospodarstwie domowym.

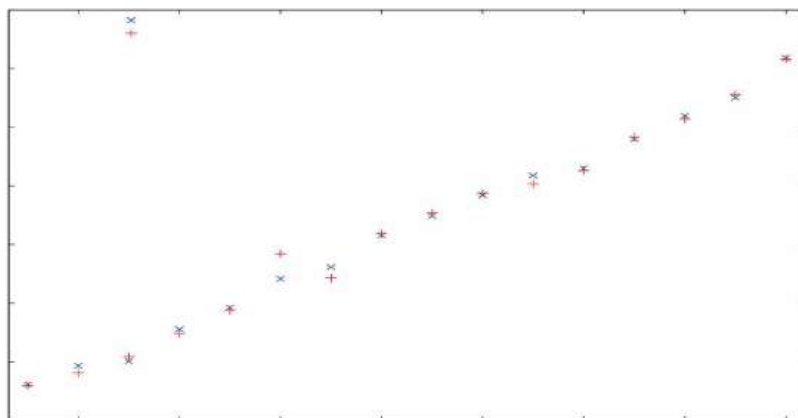
Zmienna X_{70} zależy od:

X_5 - pracodawcy i pracujący na własny rachunek w tys.,

X_{10} - przeciętne zatrudnienie w budownictwie w tys.,

X_{43} - zasiłki chorobowe w zł - poza rolnikami indywidualnymi.

Wykres wartości empirycznych oraz modelowych (wyrównanych) dla zmiennej zależnej Y uzyskamy po wywołaniu menu: (*Wykresy/wykres empirycznych i wyrównanych/względem obserwacji*).



Źródło: Opracowanie własne w *GRET*L.

Rys. 7.1. Wartości empiryczne i modelowe produktu krajowego brutto (zmiennej Y)

7.6. Określenie prognozy zmiennej Y na bazie wygenerowanych zmiennych

Wprowadzamy do nowej wersji bazy danych *Korelacja2.gdt* dodatkowe obserwacje 17. oraz 18. dla okresów przyszłych (prognozowanych) korzystając z menu: *Dane/dodaj obserwacje*. Określamy ilość dodawanych obserwacji - 2, a następnie wprowadzamy przykładowo jako planowane wartości z poniższej tabeli dla zmiennych egzogenicznych okresów 17. i 18. W tym względzie korzystamy z menu: *Dane/edycja wartości*. Dla okresu 16. nowych zmiennych z przedrostkiem *gener* dane wygenerował program *GRET*L. Natomiast dane dla zmiennych egzogenicznych są umownymi danymi wejściowymi.

Tab. 7.26. Wartości zmiennych dla okresów 16. 17. i 18.

Zmienna	Wartość zmiennej dla $T = 16$	F_{17}	F_{18}
genrX21	204,5637	199,5248	196,0430
X_{44}	264,8	268	270
X_{46}	64,2	66	68
genrX36	297272,0	305119,8	302882,4
X_6	3,7	3,8	4
genrX21	215,4	217	219
X_{23}	127,0	130	132
genrX47	935,5638		
genrX47	935,5638	954,4960	951,2839
X_7	10,9	11	12
X_{13}	780,3	785	787
X_{15}	7637,7	7640	7650
X_{24}	186213	1870	1880

Zmienna	Wartość zmiennej dla $T = 16$	F_{17}	F_{18}
X_{25}	3720,05	3730	3745
X_{54}	45,5	47	49
genrX70	17082,96	17134,95	17252,57
X_5	3498,6	3500	3520
X_{10}	576,8	580	590
X_{43}	3953,0	3960	3970

Źródło: Opracowanie własne w Excelu.

Program *GRET*L nie aktualizuje automatycznie wartości okresów 17. oraz 18. dla zmiennych generowanych. Z tego względu przyjmując poszerzoną próbę 18. obserwacji dokonujemy ponownie generowania następujących zmiennych rozszerzając nazwę o przykładowa literę *K*, w kolejności:

genrX21K, *genrX47K*, *genrX36K*, *genrX70K*.

W tym względzie korzystamy z menu:

Dodawanie zmiennej/definiowanie nowej zmiennej...

Wprowadzamy nazwę *genrX21K* oraz kopiujemy formułę z *genrX21* wywołując menu:

Zmienna/edycja atrybutów

i posługując się klawiszami *Ctrl+C* dla kopiowania, a następnie *Ctrl+V* dla wklejenia formuły.

Między nazwą a formułą wpisujemy znak równości. Podobnie postępujemy z dalszymi zmiennymi zaznaczonymi przedrostkiem *genr*. Teraz program *GRET*L automatycznie oblicza wartości także dla okresów prognozowanych 17. oraz 18.

Mając wygenerowane wartości zmiennych z literą *K* (na końcu nazwy) możemy przystąpić do określenia prognozy produktu krajowego brutto (PKB), czyli zmiennej *Y* na następne lata przykładowo 2008, 2009⁹³. Według „Rocznika statystycznego” roku 2008 PKB na rok 2007 wynosiło 30 834 zł w cenach bieżących w przeliczeniu na 1 mieszkańca.

Uzyskany ponownie w *GRET*L model PKB przy 16. obserwacjach nie ulega zmianie.

Tab. 7.27. Oszacowany ponownie model *Y* dla ustalenia prognoz

Zmienna	Współczynnik	Błąd stand.	<i>t</i> -Studenta	Wartość <i>p</i>	
const	17762,8	2397,01	7,410	8,16E-06	***
genrX21K	-27,2193	3,60491	-7,551	6,76E-06	***
genrX36K	0,0845154	0,00564217	14,98	3,95E-09	***
genrX70K	-0,376227	0,130344	-2,886	0,0137	**

Źródło: Opracowanie własne w *GRET*L.

Przy aktywnym powyższym modelu wywołujemy menu: *Analiza/prognoza...*

Teraz określamy zakres prognozy, czyli obserwację startową 17. i końcową 18. Zaznaczamy *Prognoza statyczna* oraz podajemy liczbę obserwacji przed prognozą = 16. Wówczas uzyskujemy z *GRET*L-a raport dla poziomu istotności $\alpha = 0,05$ oraz *iss* = 12 (patrz tabela 7.28).

⁹³ Okres zbierania danych statystycznych do niniejszej pracy to rok 2007, zatem lata 2008 i 2009 stanowiły prognozę. Procedura postępowania nie ulegnie zmianie jeśli szereg obserwacji będzie ruchomy.

Tab. 7.28. Wygaśnię i przyszłe prognozy zmiennej Y

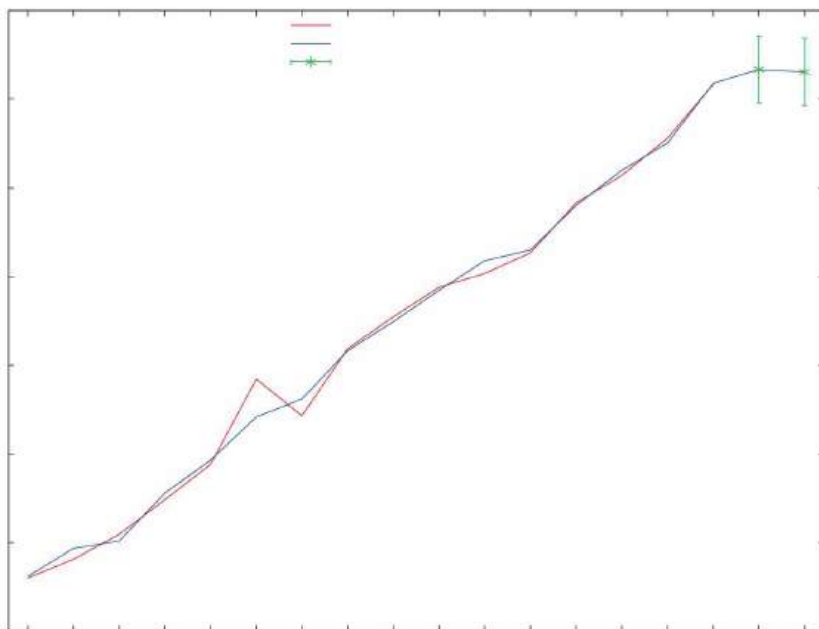
Obs.	Y	Prognoza	Błąd <i>ex ante</i>	95% przedział ufności
1	2996,00	3082,02		
2	4051,00	4663,38		
3	5459,00	5079,66		
4	7412,00	7782,64		
5	9395,00	9638,53		
6	14211,00	12076,49		
7	12144,00	13089,40		
8	15925,00	15816,09		
9	17723,00	17454,72		
10	19392,00	19217,76		
11	20169,00	20889,14		
12	21336,00	21494,20		
13	24153,00	24012,69		
14	25698,00	25987,26		
15	27803,00	27525,35		
16	30834,00	30891,68		
17		31672,53	871,058	29774,66 - 33570,40
18		31533,96	869,284	29639,95 - 33427,96

Źródło: Opracowanie własne w *GRET*L.

Raport zawiera numer okresu, realizację (wartość empiryczna obserwacji), prognozę, średni błąd prognozy *ex ante* (V_T), oraz 95% przedziału ufności prognozy. Dla okresów 17. i 18. względne błędy prognoz są następujące:

$$V_T^* = \frac{V_T}{Y_{T_p}} * 100 \quad V_{17}^* = \frac{V_{17}}{Y_{17p}} * 100 = \frac{871,058}{31672,53} * 100 = 2,75\% \quad V_{18}^* = \frac{869,284}{31533,96} * 100 = 2,76\%$$

W przedstawionym zadaniu dotyczącym *PKB* względne błędy *ex ante* nie przekraczają poziomu istotności 5%, czyli prognoza określona modelem z wygenerowanymi zmiennymi X_{21} , X_{36} , X_{70} może być praktycznie wykorzystywana. Dla porównania ułożenia się wartości empirycznych oraz wynikających opracowanego modelu sporządzono na podstawie tabeli 7.28 wykres punktowy 7.2 uwzględniający również okresy prognozowane.



Źródło: Opracowanie własne w *GRET*.

Rys. 7.2. Prezentacja graficzna wartości empirycznych i modelowych dla obserwacji 1-18 zmiennej Y

8. DOBÓR METOD I PARAMETRÓW



8.1. Wprowadzenie

Budowanie modeli ekonometrycznych jako relacji między zmienną objaśnianą a zmiennymi objaśniającymi jest procesem złożonym. Z tego względu coraz częściej w formalnym obrazowaniu zależności między zmiennymi wykorzystuje się ilościowe szeregi czasowe. Można je utworzyć na podstawie roczników statystycznych, czy też sprawozdawczości jednostek gospodarczych. Początkujący ma jednak pewne trudności w doborze odpowiedniej metody dającej właściwe równania dla określenia wygasłych już prognoz oraz predykcji zmiennej objaśnianej na okresy przyszłe.

Spotyka się w literaturze z różnorodnymi propozycjami metod oraz kuszony jest programami mającymi mu w tym pomóc. Nie zawsze ma w tym względzie odpowiednie przygotowanie teoretyczne i czas na ich rozpoznanie. Wychodząc naprzeciw potrzebie dokonano w tym rozdziale praktycznego rekonesansu poprzez podstawowe metody i sposoby doboru parametrów zawarte w komputerowym module do prognozowania objętym dostępnym w Internecie pakietem *WinQSB*. Dla przykładu bazowano na wykazującym tendencję rosnącą oraz sezonowość spożyciu piwa w ostatnich latach. Daje to nadzieję, że opracowanie w formie instrukcji pracy z systemem zachęci do badań nad spożyciem innych produktów, w celu odpowiedniego planowania produkcji oraz sprzedaży przez coraz większe już obiekty handlowe. W dobie przyspieszonej transformacji gospodarczej praktyczne opanowanie narzędzia komputerowego jakim jest moduł do prognozowania jest koniecznością, zwłaszcza gdy zależy nam na minimalizacji błędów prognoz.

Prognozowanie na podstawie szeregu czasowego wymaga doboru odpowiedniej metody, a następnie określenia najlepszych parametrów z punktu widzenia miary błędu *ex post*. Dla przedstawienia tego problemu oparto się na rocznych danych statystycznych Polski, przykładowo spożycia piwa otrzymanego ze słoju w litrach w przeliczeniu na jednego mieszkańca. Dane z lat 1989÷2005 podaje tabela 8.1.

Tab. 8.1. Dane dotyczące spożycia piwa

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Rok	1989	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005
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	77,5	80,7

Źródło: Roczniki Statystyczne Polski (lata 1989÷2006).

Dane statystyczne spożycia piwa obliczono bazując na produkcji powiększonej o import, pomniejszonej o eksport oraz o ubytki i straty u producentów i w obrocie, z uwzględnieniem zmiany stanu zapasów w magazynach handlowych. Wielkości spożycia piwa w litrach dotyczą lat gospodarczych, czyli obejmują okres od I.VII danego roku do 30.VI roku następnego.

Spożycie piwa wykazuje sezonowość w ciągu roku. Pobieżna obserwacja wskazuje na większe spożycie w kwartałach ciepłych niż chłodnych roku. Dla potrzeb niniejszego opracowania przyjęto następujące wagi dla obserwacji w poszczególnych kwartałach roku gospodarczego:

- kwartał III danego roku (lipiec, sierpień, wrzesień) - 0,35;
- kwartał IV danego roku (październik, listopad, grudzień) - 0,20;
- kwartał I roku następnego (styczeń, luty, marzec) - 0,15;
- kwartał II roku następnego (kwiecień, maj, czerwiec) - 0,30.

Dla poniższej prezentacji doboru metody prognozowania spożycia piwa na kwartał następny przyjęto 20 kwartałów 5-ciolecia (2001÷2005) i wyznaczono orientacyjnie na podstawie podanych wag wielkości spożycia w kwartałach. Suma roczna z kwartałów odpowiada spożyciu statystycznemu podanemu w tablicy 8.1.

Tab. 8.1. Orientacyjne spożycie piwa w kwartałach lat 2001-2005

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Kw.	III	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	IV	III	I	II
Litry	23	13,5	10	20	25	14	11	20,7	26	15	11	22,6	27	16	12	22,5	27,7	17	12	24

Źródło: Opracowanie własne w Excelu.

Błędy prognoz *ex post* (po fakcie) określone są w literaturze według różnych miar, a należą do nich⁹⁴:

- pierwiastek średniego kwadratu błędu - RMSE,
- średni absolutny błąd - MAE,
- średni procentowy absolutny błąd - MAPE,
- współczynnik Theila – THEIL oraz inne.

Dla wyboru metody określenia prognozy na okres 21, czyli kwartał III roku 2006 (nie występowało jeszcze w Roczniku statystycznym 2006 podczas pisania tej pracy) różnymi metodami, posłużono się miarą RMSE obliczaną z zależności:

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

Gdzie:

$S \div S+h$ - zakres próby dla której określamy błąd prognozy *ex post* zależny od okresów biorących udział w określonej metodzie,

Y - wartości rzeczywiste spożycia piwa w litrach,

$\hat{Y}$ - wartości teoretyczne obliczone przez daną metodę.

Równolegle przyjęto miarę procentową MAPE określaną ze wzoru:

$$MAPE = \sqrt{\frac{1}{1+h} \sum_{n=S}^{S+h} \left| \frac{Y_n - \hat{Y}_n}{Y_n} \right|} * 100$$

W przeglądzie metod posłużono się pakietem *WinQSB*⁹⁵. Po rozpakowaniu plików instalowane są na dysku mikrokomputera moduły programowe powyższego pakietu służące do wspomagania procesów decyzyjnych, a mianowicie:

- zagregowane planowanie,
- analiza decyzji,
- programowanie dynamiczne,
- alokacja środków,
- prognozowanie i regresja liniowa (*Forecasting and Linear Regression*),
- zarządzanie zapasami,
- planowanie pracy,
- procesy Markowa,
- gospodarka materiałowa,

⁹⁴ Lipiec - Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom II. *Ekonometria*, s. 25.

⁹⁵ Pakiet ten można było podczas pisania tego materiału zainstalować dla celów dydaktycznych pobierając 3 plik z Internetu (<http://oizet.p.lodz.pl>)

- modele sieciowe,
- programowanie nieliniowe,
- PERT-CPM,
- karty kontrolne,
- teoria kolejek,
- symulacja teorii kolejek,
- programowanie liniowe i całkowitoliczbowe,
- celowe programowanie,
- programowanie kwadratowe.

Dla doboru metody prognozowania przykładowego spożycia piwa interesuje nas moduł *Forecasting and Linear Regression* o symbolu *FC*.

8.2. Procedura pracy z modulem FC

Procedura dostępu do określonej metody w ramach pakietu *WinQSB* jest następująca:

- wywołanie modułu *WinQSB* z dysku komputera,
- wybór modułu *FC*,
- skorzystanie z opcji kolejno: *File* (zbiór), *New Problem* (nowy problem), *Problem Type* (typ problemu), a w ramach niego wybór: *Time Series Forecasting* (szereg czasowy), *Regression Forecasting* (prognozowanie według regresji zmiennych).

Wybieramy szereg czasowy i wprowadzamy informacje: *Problem Title* (tytuł problemu) - Spożycie piwa, *Time Unit* (jednostka czasu) - kwartał, *Number of Time Units (Periods)* - ilość okresów - 20, OK.

W kolejności wprowadzamy dane rzeczywiste spożycia piwa w litrach w poszczególnych kwartałach roku gospodarczego według tabeli 8.1. Zapisujemy plik na dysku przykładowo jako *PIWO.fc*. Następnie możemy wywołać funkcję *Solve and Analyze Perform Forecasting* określenia prognozy na okres 21. Udostępniana jest tabela z możliwością zaznaczenia metody prognozowania spośród:

- prosta średnia (*Simple average*) - SA,
- średnia ruchoma (*Moving average*) - MA,
- ważona średnia ruchoma (*Weighted moving average*) - WMA,
- średnia ruchoma z trendem (*Moving average with linear trend*) - MAT,
- proste wyrównywanie wykładnicze (*Single exponential smoothing*) - SES,
- proste wyrównywanie wykładnicze z trendem (*Single exponential smoothing with trend*) - SEST,
- podwójne wyrównywanie wykładnicze (*Double exponential smoothing*) - DES,
- podwójne wyrównywanie wykładnicze z trendem (*Double exponential smoothing with trend*) - DEST,
- adaptacyjne wyrównywanie wykładnicze (*Adaptive exponential smoothing*) - AES,
- trend liniowy (*Linear regression with time*) - LR,
- model Holta-Wintersa z sezonowością addytywną (*Holt-Winters Additive Algorithm*) - HWA,
- model Holta-Wintersa z sezonowością multiplikatywną (*Holt-Winters Multiplicative Algorithm*) - HWM.

Po wyborze metody prognozowania program umożliwia określenie sposobu wyznaczenia parametrów dla sformułowania modelu i wyznaczenia wygasłych już prognoz jak też prognozy *ex ante*, czyli na okres 21. poprzez zaznaczenie jednej z opcji:

- podania przez użytkownika wartości parametrów (*Assign values*),

- żądania wyszukania przez komputer drogą kolejnych iteracji wartości najlepszych dla modelu (*Search the best*) biorąc pod uwagę miarę błędu np. RMSE, przy czym:

$$MSE = \frac{1}{n} \sum (z_n^2)$$

n - liczba sumowanych reszt; $z_n = Y_n - \hat{Y}_n$ reszta.

Moduł *FC* w zależności od obranej metody sugeruje wybór kryterium obliczania błędów *ex post* spośród: MAD, CFE, MSE, MAPE. MSE stanowi sumę kwadratów reszt liczoną narastająco po każdej kolejnej obserwacji aż do okresu prognozowanego, zatem:

$$RMSE = \sqrt{MSE}$$

Wzór na miarę błędu *MAPE* podano powyżej. Miary *CFE*, *MAD* są rzadziej stosowane w literaturze przedmiotu i dlatego jak już wspomniano przyjęto w dalszych rozważaniach *RMSE* (*Root Mean Square Error*) i *MAPE* (*Mean Absolute Percentage Error*).

Metody prognozowania bazują na wzorach wypracowanych przez ekonometrię i w zależności od metody są to parametry α , β , γ . Ogólnie biorąc program umożliwia określenie wartości określonego parametru oraz wartości początkowej prognozy w sytuacji wyboru opcji *Assign values*. Gdy polecimy komputerowi dobór najlepszych parametrów (*Search the best*) to w opcji *Enter Search Domain* podajemy zakresy od do dla parametrów z krokiem iteracyjnym np. 0,01. Tu warto zauważyć, że program wymaga wpisywania wartości danych rzeczywistych i parametrów z kropką dziesiętna zamiast przecinka. Nie dla wszystkich metod prognozowania istnieje opcja wyboru najlepszych parametrów.

Bazując na orientacyjnych danych kwartalnych spożycia piwa w litrach na jednego mieszkańca Polski dokonamy teraz obliczeń poszczególnymi metodami dla określenia prognoz *ex post* i *ex ante* modulem programowym *FC*. Podano potencjalnemu użytkownikowi postępowanie dla otrzymania rozwiązania sumarycznego. Istnieje też możliwość przedstawienia rozwiązania w każdej z metod w formie graficznej.

8.3. Metody prognozowania

Prosta średnia (SA)

Wprowadzamy ilość okresów do prognozowania - 1. Metoda ta zakłada, że wartość przyszłej prognozy *F* będzie średnią arytmetyczną z wartości *Y* poprzednich okresów. Program liczy ze znaczną dokładnością. Dla prezentacji rezultatów i ograniczonych ram strony A4 niniejszego opracowania przyjęto dokładność dwóch miejsc po przecinku, z supresją zer nieznaczących. Mogą więc wystąpić nieznaczne różnice - gdybyśmy chcieli sprawdzać wyniki manualnie.

Tab. 8.2. Obliczenie prognoz jako średnich arytmetycznych

Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Kw.	II	IV	I	II	III	IV	I	II	III	IV	I	II	III	IV	I	II	IV	III	I	II	III
Y	2 3	13,5	10	20	25	14	11	20,7	26	15	11	22,6	27	16	12	22,5	27,71 7	17	12	24	
F		23	18,5	15,5	16,6 3	18,3	17,5 8	16,6 4	17,1 5	18,1 3	17,8 2	17,2 0	17,6 5	18,3 7	18,2	17,7 9	18,08	18,6 5	18,5 6	18,2 1	18,5 0
Y - F		-9,5	- 8,25	4,5	8,38	-4,3	- 6,58	4,06	8,85	- 3,13	- 6,82	5,4	9,35	- 2,37	- 6,20	4,71	9,62	- 1,65	- 6,56	5,79	
MSE		90,2 5	79,1 6	59,5 2	62,1 8	53,4 4	51,7 6	46,7 1	50,6 6	46,1 3	46,8 6	44,6 2	48,1 9	44,9 1	44,4 5	42,9 7	46,06	43,5 3	43,4 8	42,9 6	
MAP E %		70,3 7	76,4 4	58,4 6	52,2 2	47,9 2	49,9 1	45,5 8	44,1 3	41,5 5	43,6	41,8	41,2 1	39,1 8	40,0 7	38,7 9	38,54	36,8	37,8	37,1 3	

Źródło: Opracowanie własne w Excelu.

Program przyjmuje $F_2 = Y_1$, a błędy określa jako: $MSE = 42,96$ $MAPE = 37,11\%$

$$RMSE = \sqrt{MSE} = \sqrt{42,96} = 6,55$$

Średnia ruchoma (4-MA)

Wprowadzamy cztery okresy dla obliczania średnich ruchomych. Program dla okresu 5-tego oblicza średnią arytmetyczną na podstawie wartości obserwacji z okresów 1÷4.

$$(23 + 13,5 + 10 + 20)/4 = 16,63.$$

Dla okresu szóstego pobierane są przez program wartości z okresów 2÷5 i analogicznie dla kolejnych okresów prognozowania. Zatem prognoza na okres 21. wynosi:

$$(27,7 + 17 + 12 + 24)/4 = 20,17.$$

Dla kolejnych okresów prognozowania program oblicza narastająco błędy *ex post* według miar MSE oraz $MAPE$. Dla okresu ostatniego, czyli 20-tego są one następujące:

$$MSE = 35,71 \text{ przy czym } RMSE = 5,97 \quad MAPE = 32,15$$

W porównaniu z metodą SA występuje zmniejszenie błędów. Zatem metoda 4- MA jest lepsza od SA dla naszego szeregu czasowego spożycia piwa.

Średnia ruchoma z trendem (4-MAT)

Podajemy również cztery okresy średniej ruchomej dla określenia trendu liniowego segmentu (odcinka) postaci

$$Y^{\wedge} = a + b * t.$$

Program inicjuje prognozy wpisując dla okresu 5-tego wartość dla okresu 2-go, czyli 13,5. Obliczenia parametrów trendu liniowego realizowane są klasyczną metodą najmniejszych kwadratów ($KMNK$) na podstawie wartości z okresów 2÷5.

Korzystamy ze wzorów na określenie trendu odcinkowego, gdzie wartości z kreską nad zmienną to średnie arytmetyczne:

$$a_1 = \frac{\sum (y - \bar{y})(t - \bar{t})}{\sum (t - \bar{t})^2} \qquad a_0 = \bar{y} - a_1 * \bar{t}$$

Otrzymujemy w ten sposób model liniowy trendu i tak dla pierwszego segmentu czwartego okresowego ($m = 4$) jest w postaci:

$$Y^{\wedge} = a + b * t = 1,55 + 4,45 * t \qquad \text{zatem: } F_6 = 1,55 + 4,45 * 6 = 28,25$$

Prognoza *ex ante* na okres 21. Według modelu trendu dla okresów 17÷20 o parametrach trendu liniowego $a = 49,96$; $b = -1,61$ jest następująca:

$$F_{21} = 49,96 - 1,61 * 21 = 16,15$$

Błędy prognoz po okresie dwudziestym wynoszą:

$$MSE = 171,17 \quad RMSE = 13 \quad MAPE = 75,6\%$$

Tak więc spośród trzech powyższych metod z punktu widzenia miar błędów *ex post* nadal dominuje metoda 4- MA .

Ważona średnia ruchoma (WMA)

Program umożliwia tylko przypisanie wartości. Wprowadzamy ilość okresów do prognozowania (1), ilość okresów obliczania średniej ruchomej (4) oraz wagi dla wartości z naszych czterech okresów. Ponieważ jednak obserwacje już zostały wstępnie „zważone” wg spożycia rocznego przyjmijmy tylko dla przeglądu metody jednakowe wagi po 0,25, gdyż suma wag musi się równać jeden. Nic dziwnego, że dla tego przypadku system generuje identyczną tablicę jak dla metody 4- MA .

Proste wyrównywanie wykładnicze (SES)

Mamy tu dwie możliwości określenia parametrów tj. podając konkretne wartości, żądając określenia najlepszych parametrów przez program. Dla pierwszej opcji wpisujemy ilość okresów do prognozowania (1), przykładowo podajemy $\alpha = 0,2$ i nie inicjujemy prognozę zerową $F(0)$, chociaż możemy to uczynić gdy ją znamy. System przyjmuje $F(0) = F_2 = Y_1 = 23$. Prognoza dla kolejnych okresów obliczana jest wg wzoru:

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

$$\text{Np. } F_3 = 0,2 * 13,5 + (1 - 0,2) * 23 = 21,1 \quad F_{21} = 0,2 * 24 + (1 - 0,2) * 18,25 = 19,4.$$

Po zastosowaniu tej metody błędy są następujące:

$$MSE = 46,93 \quad RMSE = 6,85 \quad MAPE = 41,09\%$$

Nadal jednak okazuje się lepsza metoda 4-MA - $RMSE = 5,97$ a $MAPE = 32,1\%$.

Disponując komputerem możemy dokonywać symulacji doboru α i tak dla przykładowo dla $\alpha = 0,15$ i $F(0) = 23$ miary błędów są mniejsze i wynoszą:

$$MSE = 45,98 \quad RMSE = 6,78 \quad MAPE = 40,61\%.$$

Takie manualne doszukiwanie się α odpowiednio najlepszego jest pracochłonne. Skorzystajmy więc z automatycznego doboru przez program zaznaczając opcję *Search the best*, ilość okresów do prognozowania 1, nie podając wartości parametru α ani $F(0)$. Musimy jednak określić przedział dla parametru α : Start 0 End 1 z krokiem iteracyjnym 0,01. W efekcie pracy systemu otrzymaliśmy wyniki:

$$\alpha = 0,11; MSE = 40,84 \quad RMSE = 6,39 \quad MAPE = 40,84\%.$$

Tak więc metoda SES, z doбором automatycznym parametru wyrównującego α , nie jest jednak lepsza od 4-MA. Przykładowe prognozy wynoszą:

$$F_3 = 21,95 \quad F_{21} = 19,43.$$

Proste wyrównywanie wykładnicze z trendem (SEST)

Występują tu dwie opcje. Przyjmijmy od razu *Search the best* z kryterium MSE, prognozą na okres 21. bez podawania parametru wyrównywania α wartości zmiennej objaśnianej Y oraz parametru wyrównywania trendu β . W odpowiedzi otrzymujemy:

$$\alpha = 0,1; \beta = 0,02; F(0) = 23 = Y_1, T(0) = 0$$

oraz $MSE = 45,67 \quad RMSE = 6,76; MAPE = 40,48\%.$ $F_3 = 20,72$ a $F_{21} = 19,81.$

Z dotychczasowych metod wynika, że modele z trendem nie są adekwatne dla naszego szeregu czasowego z sezonowością kwartalną w okresie roku gospodarczego. Biorąc pod uwagę miarę $RMSE$ pozostajemy więc z dotychczas rozpatrzonych metod na średniej ruchomej 4-MA.

Podwójne wyrównywanie wykładnicze (DES)

W metodzie tej zwanej modelem Browna parametr α stosowany jest dla dwóch procedur wyrównywania. Oblicza się różnicę pomiędzy pojedynczo a podwójnie wygładzonymi wartościami zmiennej w celu estymacji trendu. Następnie wartość tej różnicy dodawana jest do pojedynczo wyrównywanej wartości zmiennej wraz ze skorygowaną wartością trendu z bieżącego okresu. Podobnie jak w metodzie SES dla opcji wyboru najlepszych parametrów, program oblicza kolejne wygasłe już prognozy $F_2 \div F_{20}$ oraz $F_{21} = 19,15$. Błędy prognoz są następujące:

$$MSE = 44,91 \quad RMSE = 6,7 \quad MAPE = 39,61\%$$

Widzimy więc, że podwójne wyrównywanie wykładnicze nie dało lepszego rezultatu niż 4-MA według przyjętych miar błędu $RMSE$ oraz $MAPE$. System zaproponował $\alpha = 0,25$ i przyjął inicjujące dla algorytmu wartości prognoz pierwszego i drugiego wyrównywania wykładniczego $F(0) = F^*(0) = Y_1$.

Podwójne wyrównywanie wykładnicze z trendem (DEST)

Dla przyspieszenia doboru α wybieramy opcję *Search the best*, podajemy ilość okresów prognozowania – 1, natomiast nie wprowadzamy wartości α , $F(0)$, $F'(0)$ podając jednak krok iteracji 0,01 i przedział 0÷1. W odpowiedzi otrzymujemy: $F_2 = 22,05$, dalsze wygasłe prognozy oraz prognozę $F_{21} = 18,84$.

Błąd prognozy $MSE = 45,68$ $RMSE = 6,76$, $MAPE = 40,39$ natomiast najlepsze α to 0,05. Model ten według $RMSE$ i $MAPE$ okazał się gorszy więc pozostajemy jako przy dotychczas najlepszym – (4-MA).

Adaptacyjne wyrównywanie wykładnicze (AES)

W adaptacyjnym wyrównywaniu wykładniczym parametr α zastąpiono zmiennym sygnałem adaptacyjnym TST_n . Istota tej metody polega na dostosowywaniu wartości parametru wyrównywania do każdego z okresów. Teoretyczne podstawy tej metody znajdujemy w rozdziale 2.11 publikacji „Wspomaganie procesów decyzyjnych”. Tom. II. *Ekonometria*⁹⁶. Program udostępnia tylko jako aktywną opcję przypisanie wartości parametrów. Podajemy więc ilość okresów prognozowania – 1, parametr $\alpha = 0,2$ oraz parametr β równy z założenia 0,2 (wg strony 75. publikacji podanej w rozdziale 2.11)⁹⁷.

Prognoza obliczana jest według wzoru:

$$F_n = F_{n-1} + TST_n (Y_{n-1} - F_{n-1}).$$

Dla zmniejszenia błędu *ex post* określonego miarami $RMSE$ i $MAPE$ możemy więc manewrować wartością parametru α . Przyjmujemy $F(0) = Y_1 = 23$. Moduł *FC* wyświetlił nam tabelę z wynikami wygasłych prognoz, w tym $F_3 = 19,2$, oraz prognozę przyszłą $F_{21} = 17,96$ a błędy wynoszą:

$$MSE = 43,39 \quad RMSE = 6,59 \quad MAPE = 39,55\%.$$

Możemy dla poprawy wyników dokonać jeszcze symulacji parametrem α , który przyjmijmy przykładowo kolejno równe 0,1; 0,05 i wówczas dla:

$$\alpha = 0,1 \quad MSE = 41,13 \quad RMSE = 6,41 \quad MAPE = 38,19$$

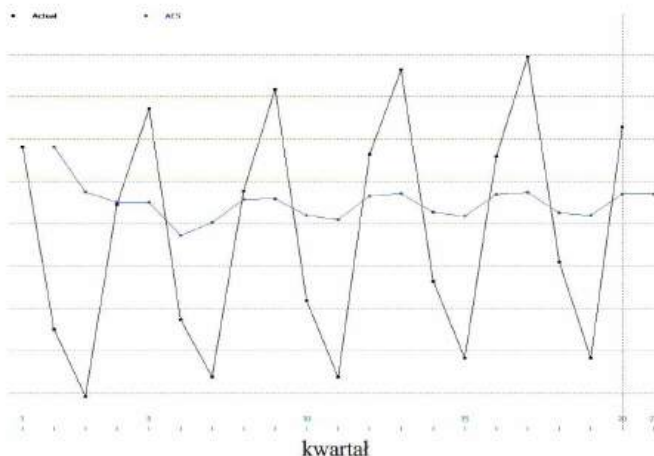
$$\alpha = 0,05 \quad MSE = 35,41 \quad RMSE = 5,95 \quad MAPE = 35,40$$

Tak więc metoda *AES* daje możliwości dopasowywania danych modelowych do empirycznych i z dotychczasowym wynikiem $RMSE = 5,95$ jest metodą najlepszą.

Jesteśmy chyba trochę zmęczeni liczbami; zobaczmy więc jak wygląda dołączone rozwiązanie w postaci graficznej otrzymane z modułu *FC*. Z wykresu wynika brak uwzględnienia sezonowości w kwartałach roku. Linia prognozy wykazuje pewną stacjonarność, podczas gdy łamana danych empirycznych niewielki, ale dodatni trend. Szukajmy więc dalszej bardziej odpowiedniej metody dla naszych obserwacji spożycia piwa w kwartałach pięciolecia 2001÷2005. Wykres z metody *AES* podano poniżej.

⁹⁶ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych* Tom II. *Ekonometria*, op. cit., rozdz. 2.11.

⁹⁷ Ibidem.



Źródło: Opracowanie własne programem WinQSB.

Rys. 8.1. Dane empiryczne i modelowe metody AES

Trend liniowy (LR)

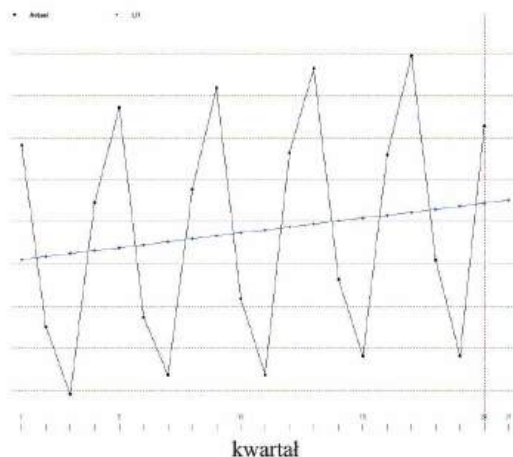
Dostępna jest tylko opcja *Assign values*. Możemy oprócz ilości okresów prognozowania wprowadzić parametry a i b trendu liniowego:

$$Y^{\wedge} = a + b * t$$

Pozostawmy komputerowi dobór parametrów a i b i po naciśnięciu OK uzyskujemy z modułu *FC* wygasłe prognozy a w tym: $F_1 = 17,03$; $F_2 = 17,19$; $F_3 = 17,34$ oraz przyszłą prognozę $F_{21} = 20,12$. Błędy są następujące:

$$MSE = 33,26 \quad RMSE = 5,77 \quad MAPE = 33,01\%$$

Trend liniowy jest zatem lepszy od *AES* gdzie $RMSE = 5,95$. System podaje też oszacowane parametry. Zwróćmy uwagę na poniższy rysunek 8.2 na którym współczynnik trendu liniowego jest dodatni.



Źródło: Opracowanie własne programem WinQSB.

Rys. 8.2. Dane empiryczne i modelowe metody LR

Model Holta-Wintersa z sezonowością addytywną (HWA)

W modelach Holta-Wintersa ze składową sezonową wprowadza się dodatkową procedurę dla wyrównywania sezonowości. Podstawy teoretyczne tych modeli znajdujemy w pracy podanej w rozdziale 2.14 publikacji „Wspomaganie procesów decyzyjnych”⁹⁸. Nasz szereg czasowy spożycia piwa charakteryzuje się trendem i wahaniami sezonowymi. Potwierdza to zamieszczony wykres uzyskany z metody trendu liniowego (LR). Z roku na rok oprócz tendencji wzrostowej spożycia piwa występują zmiany poziomu badanego zjawiska powtarzające się w tym samym kwartale roku gospodarczego. Jeśli amplituda (wysokość) wahań wokół trendu ma charakter stały budujemy model addytywny, natomiast gdy jest zmienna to model multiplikatywny. Z wykresu wynika, że mamy do czynienia z modelem multiplikatywnym, gdyż amplitudy mierzone prostopadłe do linii trendu są zmienne. Jednak dla przetestowania wszystkich metod zrealizowano również przebieg komputerowy dla modelu addytywnego.

Uzyskane wyniki tym modelem są dobre. Błędy prognoz uległy wyraźnemu obniżeniu w stosunku do wcześniejszego modelu AES wynoszą bowiem:

$$MSE = 0,54 \quad RMSE = 0,73 \quad MAPE = 2,96\%$$

Tak więc błąd względny jest mniejszy od 5%. Prognoza na III kwartał $F_{21} = 28,36$ jest adekwatna do wartości rzeczywistej przed rokiem gospodarczym $Y_{17} = 27,70$. System przyjął prognozę $F_5 = Y_1 = 23$ i policzył kolejne wygasłe prognozy do okresu 20. W rozwiązaniu program proponuje parametry:

$$\alpha = 0,1; \beta = 1; \gamma = 0,76.$$

Prognoza zerowa $F(0) = 16,625$ inicjująca obliczenia w modelu addytywnym stanowi średnią arytmetyczną z czterech pierwszych obserwacji ($Y_1 + Y_4$). Indeksy sezonowości wynoszą: $S(1) = 6,38$; $S(2) = 3,13$; $S(3) = 6,63$; $S(4) = 3,38$.

Model Holta-Wintersa z sezonowością multiplikatywną (HWM)

System umożliwia wybór spośród dwóch opcji. Wybieramy bezpośrednio *Search the best*, kryterium *MSE*. Następnie wpisujemy ilość okresów przyszłej prognozy – 1, długość cyklu sezonowości $c = 4$ (kwartały). Nie podajemy parametrów α, β, γ i zainicjowanej wartości dla czterech okresów $F(c)$ i $T(c)$ – niech system sam zaproponuje rozwiązanie najlepsze. Musimy jednak określić przedział okresów dla liczenia cykli ($1 \div 20$).

Program wymaga jeszcze podania indeksów sezonowości. Nasze wagi dla spożycia piwa w kwartałach były następujące: kw III - 0,35; kw IV - 0,20 danego roku kalendarzowego oraz kwartał I - 0,15 i kwartał II - 0,30 roku następnego stanowiące razem rok gospodarczy. Średnia z kwartałów stanowi 0,25. Orientacyjne zatem indeksy sezonowości, które powinny być w przybliżeniu pokazane w realizacji kolejnego modelu multiplikatywnego są następujące: $S(1) = 0,35/0,25 = 1,4$ $S(2) = 0,20/0,25 = 0,8$ $S(3) = 0,15/0,25 = 0,6$ $S(4) = 0,30/0,25 = 1,2$

W modelu addytywnym indeksy sezonowości były wyraźnie inne, gdyż wg wykresu z metody LR amplituda wahań okresowych nie jest stała. Z tego względu pozostawiamy zainicjowane przez system zerowe wartości indeksów sezonowości. Jako daną wejściową podajemy również krok iteracyjny 0,01 dla doboru parametrów α, β, γ z przedziału $0 \div 1$. Zaokrąglone wyniki realizacji prezentuje zamieszczona tabela 8.2 oraz rysunek 8.2.

⁹⁸ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych* Tom II. *Ekonometria*, op. cit., rozdz. 2.14.

Tab. 8.2. Prognozy według metody HWM

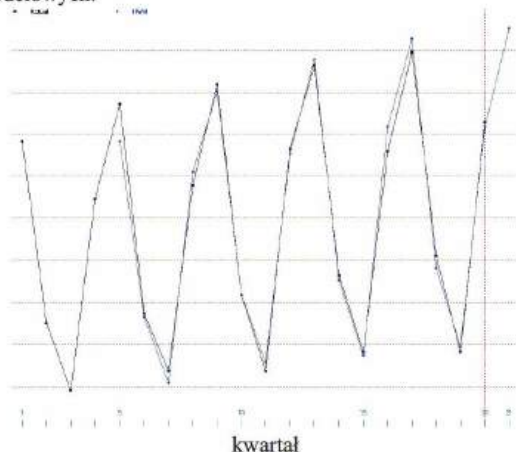
Lp.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
Kw.	II	IV	I	II	II	IV	I	II	III	IV	I	II	III	IV	I	II	IV	III	I	II	III
Y	2	13,	1	2	2	14	11	20,7	26	15	11	22,6	27	16	12	22,5	27,7	17	12	24	
	3	5	0	0	5																
F					2	13,8	10,4	21,4	25,6	14,9	11,4	22,3	27,3	15,7	11,8	23,7	28,4	16,3	12,2	23,7	28,9
					3	3	0	2	5	8	7	7	1	2	4	7	0	7	9	7	3
Y - F					2	0,17	0,6	-	0,35	0,02	-	0,23	-	0,28	0,16	-	-0,7	0,63	-	0,23	
								0,72			0,47	0,31	1,27					0,29			
MSE					4	2,01	1,46	1,23	1,01	0,84	0,75	0,66	0,6	0,55	0,5	0,59	0,59	0,57	0,54	0,51	
MAP					8	4,61	4,89	4,53	3,9	3,27	4,41	3,12	2,9	2,78	2,65	2,9	2,87	2,93	2,9	2,78	
E %																					
R ²					1	0,73	0,8	0,91	0,95	0,94	0,9	0,91	0,96	0,96	0,96	1	1	1	1	1	

Źródło: Opracowanie własne w WinQSB i w Excelu.

Nastąpiło dalsze zmniejszenie błędów ($RMSE = 0,71$; $MAPE = 2,78\%$) w stosunku do metody HWA. Współczynnik determinacji R^2 przy naszym zaokrągleniu jest w przybliżeniu równy jeden co świadczy o doskonałym dopasowaniu modelu multiplikatywnego do danych empirycznych. System zaproponował parametry: $\alpha = 0,14$; $\beta = 1$; $\gamma = 0,4$ a prognozę $F_{21} = 28,93$.

Okazuje się, że sprawdziły się nasze przewidywania i indeksy sezonowości są zbliżone do wstępnie założonych, bowiem: $S(1) = 1,38$; $S(2) = 0,81$; $S(3) = 0,6$; $S(4) = 1,2$.

Rezultat pracy programu dla metody HWM podaje rysunek 8.2. Widzimy dobre dopasowanie danych empirycznych do modelowych.



Źródło: Opracowanie własne programem WinQSB.

Rys. 8.2. Dane empiryczne i modelowe metody HWA

* * *

Spośród zaprezentowanych metod prognozowania z zastosowaniem modułu *Forecasting and Linear Regression (FC)* pakietu WinQSB (w których zastosowano również opcję najlepszego doboru parametrów modeli) wybrano model Holta-Wintersa z sezonowością multiplikatywną. Dla tego bowiem modelu przyjęte miary błędów $RMSE$ i $MAPE$ w odniesieniu do wygaśniętych już prognoz są najmniejsze.

Niech niniejsze opracowanie, pełniące rodzaj instrukcji pracy z powyższym modułem pakietu WinQSB, będzie nie tylko dla studentów, bodźcem do praktycznego opanowania, a następnie stosowania w obecnej nowej rzeczywistości metod prognozowania.

9. MODEL EKONOMETRYCZNY ZMIENNEJ „Liczba bezrobotnych”



9.1. Dane wejściowe

We wcześniejszym materiale przede wszystkim nacisk położono na sformułowanie modelu produktu krajowego brutto (PKB), w zależności od wybranych zmiennych z poszczególnych rynków⁹⁹. Jedną ze zmiennych rynku pracy podanej w *Rocznikach statystycznych* jest zmienna nazwana „Liczba bezrobotnych w tys.”. Teraz uwaga nasza skupiona będzie na określeniu modelu ekonometrycznego tej zmiennej w zależności od innych zmiennych objaśniających. W tym względzie posłużymy się programem *GRET*L oraz zastosujemy eliminację *a posteriori* dla selekcji zmiennych objaśniających. Wyjściem do opracowania modelu są zebrane na etapie opracowania książki „Modele ekonometryczne PKB obiektów struktury terytorialnej” dane statystyczne¹⁰⁰. Obejmują one okres szesnastu lat (1992-2007), z których wybrano 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 (X3));
- przeciętne zatrudnienie w tys.: przemysł (X9), budownictwo (X10), handel i naprawy (X11), transport, gospodarka magazynowa i łączność (X12), obsługa nieruchomości i firm (X13), działalność usługowa komunalna, społeczna i indywidualna, pozostałe (X14);
- koszty prac w zł/ogółem na jednego zatrudnionego przeciętnie miesięcznie (X25);
- rynek płacy/wynagrodzenia w mln/wskaźnik (przeciętne miesięczne wynagrodzenie nominalne brutto - rok poprzedni = 100 (X38));
- budżety gospodarstw domowych/przeciętne miesięczne wydatki na 1. gospodarstwo domowe w zł (X48);
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych): zużycie wody z wodociągów w ciągu roku w hm³ (X49), zużycie energii elektrycznej w ciągu roku w GWh (X50), zużycie gazu w sieci w ciągu roku w hm³ (X51);
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci – policealne w tys. (X59),
- rynek mieszkaniowy/wskaźnik zawartych małżeństw na 10000 ludności (X78).

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

ilość obserwacji > ilość parametrów (zmiennych + 1)

16 obserwacji > 14 zmiennych objaśniających + 1

Wartości obserwacji poszczególnych zmiennych podano w tabeli 9.1.

Tab. 9.1. Obserwacje zmiennych

Rok	t	X3	X9	X10	X11	X12	X13	X14	X25	X38	X48	X49	X50	X51	X59	X78
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

⁹⁹ Wornalkiewicz W., *Modele PKB obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski w Opolu, 2013

¹⁰⁰ Ibidem.

Rok	t	X3	X9	X10	X11	X12	X13	X14	X25	X38	X48	X49	X50	X51	X59	X78
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 *GRET*L-em dla określenia statystyki opisowej zarówno zmiennej objaśnianej *X3* jak i zmiennych objaśniających *X9-X14*, *X25*, *X38*, *X48-X51*, *X59*, *X78*.

Tab. 9.2. Statystyka opisowa

Zmienna	Średnia	Mediana	Min	Max	s	V	S	K
X3	2544,19	2494,50	1619,00	3431,00	585,608	0,230175	-0,00926730	-1,18380
X9	3064,93	3046,70	2639,10	3467,60	345,773	0,112816	-0,0272987	-1,77077
X10	632,781	670,350	453,100	823,000	108,517	0,171492	-0,210728	-1,05959
X11	1242,06	1295,50	996,100	1482,80	147,797	0,118993	-0,469248	-0,925767
X12	663,044	670,600	530,400	794,700	78,2167	0,117966	-0,0562422	-1,28409
X13	575,569	598,600	366,400	780,300	134,639	0,233924	-0,202479	-1,33192
X14	262,788	263,200	235,400	287,700	16,6576	0,0633880	0,0388373	-1,01035
X25	1968,13	2025,38	295,163	3720,05	1238,71	0,629383	-0,0405670	-1,65828
X38	116,444	111,800	102,600	138,900	12,9368	0,111099	0,531208	-1,28171
X48	1506,72	1793,32	404,700	2290,78	600,455	0,398517	-0,604703	-1,03547
X49	1455,58	1383,40	1219,40	1921,90	231,951	0,159353	0,780318	-0,678815
X50	21484,6	20918,5	18075,0	27713,0	3225,56	0,150134	0,904225	-0,362171
X51	4131,71	4052,80	3708,10	4775,10	347,356	0,0840706	0,407472	-1,20955
X59	75,0812	78,4000	40,1000	104,300	19,3223	0,257351	-0,628182	-0,582073
X78	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 *GRET*L.

Wyznaczenie procentowe współczynnika zmienności V_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.

Tab. 9.3. Współczynniki zmienności

Zmienna	X3	X9	X10	X11	X12	X13	X14	X25	X38	X48	X49	X50	X51	X59	X78
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 według danych z tabeli 9.2.

Z tabeli 9.3 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: *X14*, *X51*, *X78*. Zależność między zmiennymi modelu możemy przedstawić przy pomocy macierzy korelacji korzystając w *GRET*L z menu: *Widok/Macierz korelacji* w wyniku czego uzyskujemy raport przedstawiony w tabeli 9.4. Wartość krytyczna (przy dwustronnym 5% obszarze krytycznym) = 0,4973 dla $n = 16$.

Tab. 9.4. Współczynniki korelacji wzajemnej zmiennych

X3	X9	X10	X11	X12	X13	X14	X25	
1,0000	-0,6628	-0,5486	0,0927	-0,6162	0,3582	-0,0196	0,4105	X3
	1,0000	0,9060	-0,7265	0,9378	-0,9017	-0,2365	-0,9460	X9
		1,0000	-0,5920	0,8645	-0,7603	0,0197	-0,8578	X10
			1,0000	-0,7106	0,9170	0,5132	0,8694	X11
				1,0000	-0,8305	-0,0760	-0,9152	X12
					1,0000	0,4839	0,9600	X13
						1,0000	0,3139	X14
							1,0000	X25
X38	X48	X49	X50	X51	X59	X78		
-0,4713	0,2917	-0,3683	0,0815	-0,5123	0,2173	-0,6201	X3	
0,9094	-0,8529	0,8644	-0,7667	0,9032	-0,7209	0,0912	X9	
0,8072	-0,7515	0,8165	-0,7333	0,7610	-0,7093	0,1542	X10	
-0,9003	0,9562	-0,9063	0,8381	-0,8049	0,8931	0,2838	X11	
0,8936	-0,8571	0,8849	-0,7065	0,8526	-0,7210	0,1910	X12	
-0,9241	0,9309	-0,8925	0,8747	-0,8751	0,8199	0,1921	X13	
-0,3192	0,3729	-0,2782	0,3514	-0,3389	0,3678	0,4193	X14	
-0,9327	0,9440	-0,9243	0,8780	-0,9134	0,8063	0,1052	X25	
1,0000	-0,9691	0,9800	-0,7645	0,9229	-0,9056	0,0783	X38	
	1,0000	-0,9779	0,8202	-0,9054	0,9037	0,1080	X48	
		1,0000	-0,7807	0,8819	-0,9366	0,0463	X49	
			1,0000	-0,7086	0,7562	0,4510	X50	
				1,0000	-0,7789	0,0534	X51	
					1,0000	0,0116	X59	
						1,0000	X78	

Źródło: Opracowanie własne programem GRETL.

Dobór zmiennych objaśniających do modelu zmiennej X3 „Liczba bezrobotnych” możemy przeprowadzić stosując analizę współczynników korelacji. W tym celu możemy skorzystać z podanej powyżej wartości krytycznej współczynnika korelacji $r^* = 0,4973$. Jak już nadmieniono, postępowanie w tej analizie jest następujące¹⁰¹:

- a) Sprawdzamy, czy $|r_j| \leq r^*$ i eliminujemy zmienne X_j nieistotnie skorelowane z X3:

$$|r_j| \leq r^* = 0,4973$$

Kierując się tym kryterium i przeglądając w macierzy korelacji pary relacji zmiennej objaśnianej X3 względem odpowiadających zmiennych objaśniających X9, X10-X14, X25, X38, X48, X49-X51, X59, X78 możemy wyeliminować 9 zmiennych:

X11, X13, X14, X25, X38, X48, X49, X50, X59.

Zwróćmy uwagę, że spośród powyższych zmiennych według współczynnika zmienności eliminacji podlegała zmienna X14.

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

$$r_h = \max_j(r_j)$$

¹⁰¹ Dziechciarz J., *Ekometria Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003.

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 .

Dobór zmiennych objaśniających analizą współczynnika korelacji jest pracochłonne. Z tego względu łatwiej jest skorzystać z metody zwanej eliminacją *a posteriori* zawartej w *GRET*L.

9.2. Formułowanie modelu ekonometrycznego

Bazując na danych przekrojowych, zapisanych z uwzględnieniem zmiennych podanych w tabeli 9.1 w bazie danych *GRET*L-a pod nazwą *Korelacja.gdt*, możemy skorzystać z menu:

Model/Klasyczna metoda najmniejszych kwadratów.

Definiujemy zmienną zależną X_3 oraz zmienne niezależne (objaśniające): $X_9, X_{10}-X_{14}, X_{25}, X_{38}, X_{48}-X_{51}, X_{59}, X_{78}$, a rezultatem tej pracy programu jest raport modelu 1 (zob. tabela 9.5).

Tab. 9.5. Oszacowanie parametrów modelu zmiennej X_3

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 (X_{51}).

Źródło: Opracowanie własne programem *GRET*L.

W tabeli 9.5 szereg pojęć jest nam znanych ze statystyki. Na uwagę zasługuje jednak kryterium informacyjne Akaike'a (AIC), które jak już wspomniano w literaturze określane jest według wzoru¹⁰²:

¹⁰² Lipiec-Zajchowska (red.): *Wspomaganie procesów decyzyjnych*, T. II, *Ekonometria*. Wydawnictwo C.H. Beck,

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

Natomiast w programie *GRET*L obliczany jest wg formuły:

$$AIC = -2l(\hat{\theta}) + 2k$$

gdzie logarytm wiarygodności oznaczany jest symbolem:

$$l(\hat{\theta})$$

Kierując się podpowiedzią na końcu raportu budowano w *GRET*L modele eliminując kolejno zmienne: X_{51} , X_{49} , X_{25} , X_{11} , X_{14} , X_{38} , X_{78} , X_{12} uzyskując model 10 uznany przez program *GRET*L jako końcowy - bez sugestii dalszej eliminacji zmiennych (zob. tabela 9.6).

Tab. 9.6. Estymacja parametrów zmiennej X_3 (model 10)

Zmienna	Współczynnik	Błąd stand	t-Studenta	Wartość p
const	17753,3	964,731	18,40	1,89E-08 ***
X9	-4,33449	0,380482	-11,39	1,20E-06 ***
X10	3,19675	0,826869	3,866	0,0038 ***
X13	-2,29688	0,950987	-2,415	0,0389 **
X48	-0,887720	0,175447	-5,060	0,0007 ***
X50	-0,116366	0,0203389	-5,721	0,0003 ***
X59	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 programem *GRET*L.

Jak już wiemy, program *GRET*L 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 – brak statystycznej istotności ($\alpha > 0,1$).

Występuje zatem niski poziom istotności - popełnienia błędu w oszacowaniu parametrów oraz wysoki jest współczynnik determinacji $R^2 = 0,98024$ i nasz model jeszcze przed etapem weryfikacji jest następujący:

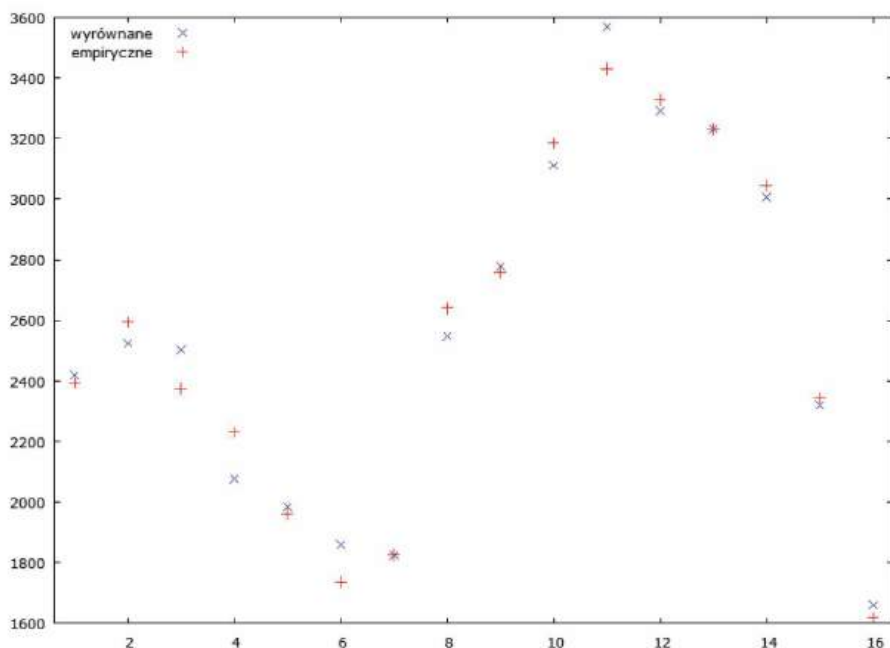
$$\hat{X}_3 = 17753,3 - 4,33449X_9 + 3,19675X_{10} - 2,29688X_{13} - 0,88772X_{48} - 0,116366X_{50} + 16,1492X_{59}$$

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 (X48);
- rynek usług komunalnych (w odniesieniu do gospodarstw domowych)/zużycie energii elektrycznej w ciągu roku w GWh (X50);
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach)/absolwenci - policealne w tys. (X59).

Program *GRET*L umożliwia pokazanie wartości z obserwacji i obliczonych na podstawie modelu na wykresie punktowym (zob. rys. 9.1).



Źródło: Opracowanie własne w *GRET*L.

Rys. 9.1. Wartości empiryczne i modelowe (wyrównane) zmiennej zależnej X3 „Liczba bezrobotnych”

9.3. Weryfikacja modelu

Etap weryfikacji modelu ekonometrycznego estymowanego klasyczną metodą najmniejszych kwadratów może obejmować wiele testów, którymi dysponuje program *GRET*L. W niniejszym temacie ograniczymy się jednak do sprawdzenia:

- normalności reszt,
- heteroskedastyczności (zmiennosć wariancji resztowej),
- współliniowość zmiennych objaśniających.

Po uzyskaniu na ekranie *Modelu* 10 wywołujemy menu:

Testy/Test na normalność rozkładu reszt

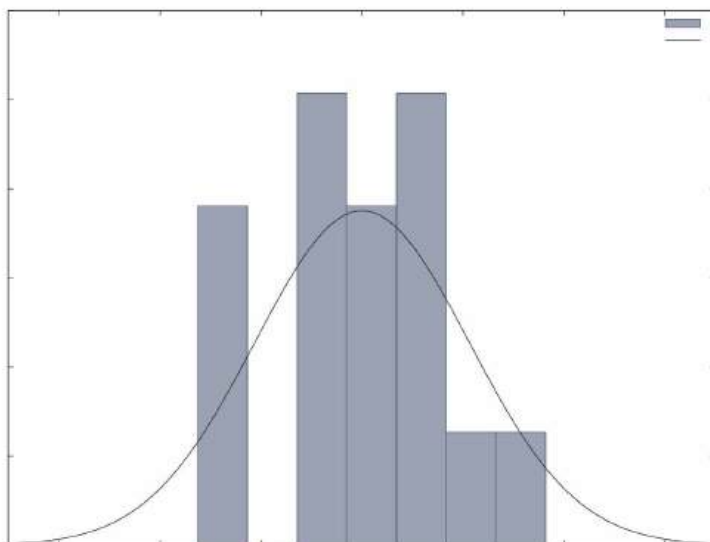
i w efekcie otrzymujemy szereg rozdzielczy reszt dla 16. obserwacji (tab. 9.7) oraz wykres funkcji gęstości i histogram częstości (rys. 9.2).

Tab. 9.7. 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 *GRET*L.



Źródło: Opracowanie własne w *GRET*L.

Rys. 9.2. Przedstawienie częstości występowania przedziałów reszt oraz funkcja gęstości

Hipoteza zerowa: dystrybuanta empiryczna posiada rozkład normalny. W *GRET*L zastosowano test Doornika-Hansena bazujący na współczynniku skośności i kurtozie. Z rysunku 9.2 odczytamy statystykę testu:

$$\chi^2_{(2)} = 0,346 \text{ z wartością } p = 0,84120$$

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. Następnie wywołujemy poniższe menu stosując test White'a (tab. 9.8).

Tab. 9.8. Raport testu heteroskedastyczności

Zmienna	Współczynnik	Błąd stand.	t-Studenta	Wartość p
const	7,01813E+06	4,36139E+06	1,609	0,2060
X9	-2059,16	1544,73	-1,333	0,2747
X10	-2002,50	1883,47	-1,063	0,3657
X13	-2670,96	1840,64	-1,451	0,2427
X48	276,069	189,069	1,460	0,2404
X50	-219,474	99,1065	-2,215	0,1136
X59	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 $R^2 = 0,86696$; Statystyka testu: $TR^2 = 13,871395$.

Hipoteza zerowa H_0 : heteroskedastyczność reszt nie występuje.

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

Raport podaje statystykę testu, która można obliczyć również według wzoru:

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

Test Withe'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.

W sytuacji występowania wysokich współczynników korelacji między zmiennymi objaśniającymi modelu T. Kufel zaleca zbadanie dodatkowo współliniowości zmiennych objaśniających¹⁰³. Sytuacja taka powoduje zaniżanie wartości statystyki t-Studenta w ocenie istotności parametru. Ocenę stopnia współliniowości zmiennych objaśniających w programie *GRET*L można dokonać posługując się miarą *VIF*_j określaną jako czynnik inflacji wariancji i korzystając z menu:

Testy/Ocena współliniowości VIF, czyli czynnika powiększania wariancji.

Zobacz tabela 9.9.

Tab. 9.9. Raport z oceny współliniowości

Minimalna możliwa wartość = 1.0
Wartości > 10.0 mogą wskazywać na problem współliniowości (rozdzęcia wariancji)
X9 22,992; X10 10,695; X13 21,778; X48 14,742; X50 5,717; X59 7,482;

Źródło: Opracowanie własne w *GRET*L.

Widzimy, że tylko dla zmiennych X50 oraz X59 miara *VIF*_j < 10. Miarę tę obliczamy według wzoru:

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

¹⁰³ Kufel T., *Ekometria Rozwiązanie problemów z wykorzystaniem programu GRET*L, Wydawnictwo Naukowe PWN, Warszawa 2007, rozdz. 4.3.6.

Gdzie: R_j^2 - współczynnik korelacji wielorakiej między zmienną x_j a pozostałymi zmiennymi modelem¹⁰⁴.

W tym modelu współczynniki korelacji między zmiennymi objaśniającymi podawane przez *GRET*L są wysokie (tab. 9.10).

Tab. 9.10. Współczynniki korelacji między zmiennymi objaśniającymi

X9	X10	X13	X48	X50	X59	
1,0000	0,9060	-0,9017	-0,8529	-0,7667	-0,7209	X9
	1,0000	-0,7603	-0,7515	-0,7333	-0,7093	X10
		1,0000	0,9309	0,8747	0,8199	X13
			1,0000	0,8202	0,9037	X48
				1,0000	0,7562	X50
					1,0000	X59

Źródło: Opracowanie własne w *GRET*L.

W zaistniałej sytuacji dokonamy korekty w zestawie zmiennych objaśniających dobierając intuicyjnie zmienne, które mogą mieć wpływ na zmienną zależną „Liczba bezrobotnych”:

- przeciętne zatrudnienie w tys./gospodarka magazynowa i łączność (X12);
- rynek edukacji i wychowania (szkoły, uczniowie i studenci w dziesiątkach): absolwenci - wyższe w tys. (X60), absolwenci - dla dorosłych w tys. (X61);
- rynek mieszkaniowy/wskaźnik rozwodów na 1000 ludności (X79).

Teraz rozpoczynamy procedurę eliminacji *a posteriori* od nowa realizując również każdorazowo test współliniowości zmiennych objaśniających. Zmienną zależną jest nadal X3. Raport *GRET*L-a bez podawania statystyki opisowej zaprezentowano w tabeli 9.11.

Tab. 9.11. Model po ponownej estymacji parametrów

Zmienna	Współczynnik	Błąd stand.	t-Student	Wartość p	
const	20051,9	3633,49	5,519	0,0027	***
X9	-4,77960	1,07768	-4,435	0,0068	***
X10	4,25611	1,71349	2,484	0,0556	*
X13	-2,12944	1,33769	-1,592	0,1723	
X48	-0,882948	0,273869	-3,224	0,0234	**
X50	-0,141497	0,0638607	-2,216	0,0775	*
X59	20,4479	13,4971	1,515	0,1902	
X12	-1,70175	1,79905	-0,9459	0,3876	
X60	-1,84050	1,98024	-0,9294	0,3953	
X61	-3,49460	9,19190	-0,3802	0,7194	
X79	303,742	667,829	0,4548	0,6683	

Wyluczając stałą, największa wartość *p* jest dla zmiennej X61.

Źródło: Opracowanie własne w *GRET*L.

Postępując dalej przy korzystaniu z podpowiedzi na końcu raportu oraz śledząc *VIF_j* w teście współliniowości wyeliminowano kolejno zmienne: X61, X79, X60, X13, otrzymując model końcowy wykazujący poprawę tej miary przy poszczególnych parametrach zmiennych (zob. tab. 9.12).

¹⁰⁴ Bliższe omówienie sposobu określania współczynnika korelacji wielorakiej: E. Nowak, *Zarys metod ekonometrii. Zbiór zadań*, Wydawnictwo Naukowe PWN, Warszawa 2002.

Tab. 9.12. Parametry modelu uznanego za końcowy według kryterium p

Zmienna	Współczynnik	Błąd stand.	t -Student	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 *GRET*L.

Tab. 9.13. Ponowna ocena współliniowości modelu VIF – czynnika powiększenia 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 *GRET*L.

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. 9.14).

Tab. 9.14. Skorygowany model według testu współliniowości

Zmienna	Współczynnik	Błąd stand.	t -Student	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 $R^2 = 0,84716$
 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 *GRET*L.

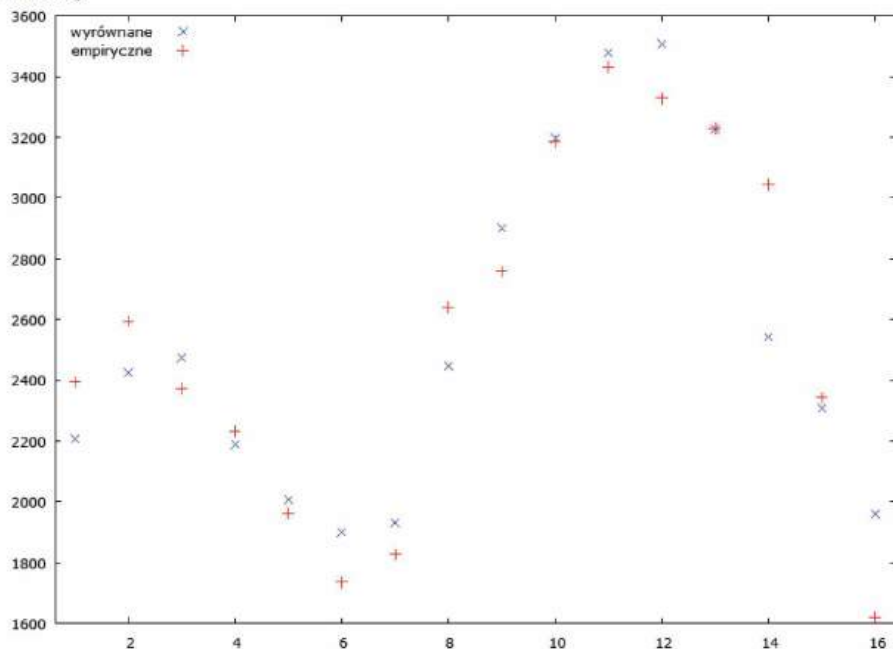
Tab. 9.15. Ocena współliniowości modelu podanego w tabeli 9.14.

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 *GRET*L.

Z tabeli 9.15 wynika, że dla zmiennych spełniony jest warunek wartości $VIF_j < 10$, co wskazuje na niewystępowanie problemu rozdęcia wariancji.

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* po każdym kolejnym kroku budowy modelu ekonometrycznego. Rezultatem pracy jest model stanowiący *consensus* w budowie modelu ekonometrycznego według dwóch miar jednocześnie: statystyki *t*-Studenta, miary *VIF* współliniowości danej zmiennej.



Źródło: Opracowanie własne w GRETL.

Rys. 9.3. Wartości empiryczne i modelowe zmiennej X_3

Tak więc w rezultacie końcowym zmienna zależna X_3 „Liczba bezrobotnych” zależy od zmiennych objaśniających: X_9 , X_{10} , X_{49} , X_{50} , X_{59} .

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

Zmienna te oznaczają:

- 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}).

Na rysunku 9.3 widzimy obraz danych empirycznych oraz modelowych na wykresie punktowym. Widzimy, jak możliwie dokładnie dla poszczególnych 16. obserwacji wartości określone przez model są przybliżone do wartości empirycznych (statystycznych) zmiennej X_3 „Liczba bezrobotnych” w tys.

10. WSPOMAGANIE KOMPUTEROWE BUDOWY MODELU PROCESU EKONOMICZNEGO



10.1. Wielomianowe modele trendu

Procesy ekonomiczne charakteryzują się niestacjonarnością średniej. Dla celów prognozowania stosowane są złożone modele trendu, sezonowości i autokorelacji danych rzeczywistych zmiennej objaśnianej. Zachodzi potrzeba określenia stopnia wielomianu trendu, amplitud sezonowości przykładowo w kwartałach roku oraz rzędu autokorelacji. W tym celu oraz dla ostatecznej estymacji parametrów modelu złożonego interesujące jest zastosowanie funkcji ogólnie dostępnego w Internecie programu *GRET*L.

Załóżmy, że mamy do czynienia z procesem niestacjonarnym typu trendowo-sezonowo-autoregresyjnym o podejściu określonym modelem:

$$Y_t = \sum_{j=0}^r a_j t^j + \sum_{i=1}^{m-1} d_i Q_{it} + \sum_p \rho_p Y_{t-p} + e'_t$$

Gdzie człon pierwszy P_t jest wielomianem j -tym trendu, drugi S_t dotyczy sezonowości w m kwartałach roku, trzeci AR_p autoregresji stopnia p , a e'_t jest resztą.

Według T. Kufla¹⁰⁵, przy założeniu stacjonarności wariancji, składnikowy model procesu ekonomicznego możemy zapisać również jako:

$$Y_t = P_t + S_t + C_t + \eta_t$$

Przy czym:

t - zmienna czasowa, $t = 1, 2, \dots, n$,

P_t oraz S_t jak wyżej,

C_t - składnik koniunkturalny zależny od rozwoju gospodarczego,

$\eta_t = e'_t$ - składnik przypadkowy, czyli reszta, obliczana jako różnica między wartością obserwacji a wynikiem z modelu dla danego t .

Trend jako wielomian zmiennej czasowej r - tego stopnia przed oszacowaniem parametrów określamy wzorem:

$$Y_t = \sum_{j=0}^r \alpha_j t^j + \eta_t$$

Dla określenia stopnia wielomianu r zabawimy na przykładzie liczby turystów w tysiącach korzystających z noclegów w pewnym mieście Polski porównując modele dla $r = 1, 2, 3$ dla zmiennej czasowej kwartałów lat 2001÷2005.

Tab. 10.1. Liczba turystów korzystających z noclegów

Rok/kw.	2001	2002	2003	2004	2005
I	133	137	134	150	188
II	238	257	258	302	336
III	262	267	273	333	393
IV	159	170	177	224	265

Źródło: Opracowanie własne wg przykładu 1.7¹⁰⁶.

¹⁰⁵ Kufla T., *Ekometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*L, Wydawnictwo Naukowe PWN, Warszawa 2007, rozdział 6.

¹⁰⁶ Praca zbiorowa, *Wybrane zagadnienia z prognozowania*, ECONOMICUS, Szczecin 2006.

Przedstawione dane charakteryzują się trendem rosnącym oraz sezonowością w poszczególnych kwartałach roku. Dla ułatwienia obliczeń skorzystamy z funkcji:

Model/Klasyczna metoda najmniejszych kwadratów ...

programu *GRET*L, który można zainstalować na własnym komputerze ze strony *kufel.torun.pl*. Program ten w wersji 1.6.5 wchodzi w skład bezpłatnego oprogramowania *Open Source*.

Przed dokonaniem estymacji danego modelu trendu stopnia r musimy dodać zmienne czasowe $time$, $time2$, $time3$ do bazy danych, która dla naszego przykładu obejmuje 20 obserwacji (5 lat x 4 kwartały) zmiennej objaśnianej Y w okresach t . Zmienne niezależne definiujemy jako:

$$time2=time^2$$

$$time3=time^3$$

Wpierw inicjujemy bazę danych składającą się z 20. obserwacji (2001:1÷2005.4) stosując funkcję:

Plik/Nowy zbiór danych

wprowadzając zmienną niezależną Y nazwaną *Nocleg*. Następnie dodajemy zmienną czasową:

Dodawanie zmiennych/time-zmienna czasowa t (1÷20).

W kolejności definiujemy zmienne $time2$ oraz $time3$ wprowadzając formuły jak wyżej. Rezultatem jest baza danych dla poszczególnych lat i kwartałów Q1÷Q4 o zawartości:

Obs.	Nocleg	time	time2	time3
2001Q1	133	1	1	1
2001Q2	238	2	4	8
2001Q3	262	3	9	27
2001Q4	159	4	16	64
2002Q1	137	5	25	125
2002Q2	257	6	36	216
2002Q3	267	7	49	343
2002Q4	170	8	64	512
2003Q1	134	9	81	729
2003Q2	258	10	100	1000
2003Q3	273	11	121	1331
2003Q4	177	12	144	1728
2004Q1	150	13	169	2197
2004Q2	302	14	196	2744
2004Q3	333	15	225	3375
2004Q4	224	16	256	4096
2005Q1	188	17	289	4913
2005Q2	336	18	324	5832
2005Q3	393	19	361	6859
2005Q4	265	20	400	8000

Teraz dokonujemy estymacji kolejnych trzech modeli trendu:

1. Liniowego:

$$\hat{Y} = a_0 + a_1 t$$

2. Kwadratowego:

$$\hat{Y} = a_0 + a_1 t + a_2 t^2$$

3. Wielomianu trzeciego stopnia.

$$\hat{Y} = a_0 + a_1 t + a_2 t^2 + a_3 t^3$$

Zastosujemy funkcję *GRET*L-a *Model/Klasyczna metoda najmniejszych kwadratów ...* traktując kolejno jako zmienne objaśniające:

$$t = \text{time}$$

$$t^2 = \text{time2}$$

$$t^3 = \text{time3}$$

Przykładowo dla zmiennej objaśnianej Y oraz zmiennej czasowej t uzyskamy statystykę, której fragment zamieszczono w tabeli 10.1. Jest to model 1. Estymacja KMNK z wykorzystaniem 20. obserwacji 2001:1-2005:4. Zmienna zależna: *Nocleg*.

Tab. 10.2. Parametry modelu liniowego

Zmienna	Współczynnik	Błąd stand.	Statystyka t
const	165,221	30,9422	5,3397
time	6,43609	2,583	2,4917

Średnia arytmetyczna zmiennej zależnej = 232,8

Odchylenie standardowe zmiennej zależnej = 75,1872

Suma kwadratów reszt = 79862,7

Błąd standardowy reszt $S_e = 66,6094$

Wsp. determinacji $R^2 = 0,256463$

Skorygowany współczynnik determinacji $R^{2*} = 0,215155$

Stopnie swobody = 18.

Źródło: Opracowanie własne w *GRET*L.

$$\hat{Y} = 165,221 + 6,43609 t$$

Model ten możemy zapisać jako ikonę sesji *GRET*L-a. Wykazuje on bardzo niskie dopasowanie do danych empirycznych $R^2 = 0,256463$. A teraz kolejne dwa modele (trendu kwadratowego oraz wielomianu trzeciego stopnia) również dla 20. obserwacji zmiennej zależnej *Nocleg*. Parametry modelu 2 pokazano w tabeli 10.3.

Tab. 10.3. Parametry modelu kwadratowego

Zmienna	Współczynnik	Błąd stand.	Statystyka t
const	198,844	49,9128	3,9838
time	-2,73377	10,9467	-0,2497
time2	0,43666	0,506334	0,8624

Średnia arytmetyczna zmiennej zależnej = 232,8

Odchylenie standardowe zmiennej zależnej = 75,1872

Suma kwadratów reszt = 76515,3

Błąd standardowy reszt = 67,0887

Wsp. determinacji $R^2 = 0,287628$

Skorygowany współczynnik determinacji $R^{2*} = 0,20382$

Stopnie swobody = 17.

Źródło: Opracowanie własne w *GRET*L.

$$\hat{Y} = 198,844 - 2,73377 t + 0,43666 t^2$$

A teraz parametry modułu 3.

Tab. 10.4. Parametry wielomianu trzeciego stopnia

Zmienna	Współczynnik	Błąd stand.	Statystyka t
const	180,809	75,2872	2,4016
time	6,46852	30,2928	0,2135
time2	-0,632584	3,30941	-0,1911
time3	0,0339442	0,103755	0,3272

Średnia arytmetyczna zmiennej zależnej = 232,8

Odchylenie standardowe zmiennej zależnej = 75,1872

Suma kwadratów reszt = 76006,8

Błąd standardowy reszt = 68,9233

Wsp. determinacji $R^2 = 0,292362$

Skorygowany współczynnik determinacji $R^2 = 0,15968$

Stopnie swobody = 16.

Źródło: Opracowanie własne w *GRET*L.

$$\hat{Y} = 180,809 + 6,46852t - 0,632584t^2 + 0,0339442t^3$$

Dla porównywania modeli wykorzystamy wariancję resztową (S_e^2) dla par modeli oraz zastosujemy test F :

Model trendu liniowego i kwadratowego:

$$F_{obl} = \frac{(S_{el}^2)}{(S_{ek}^2)} = \frac{66,6094^2}{67,0887^2} = 0,9858$$

Brak spadku wartości wariancji resztowej, czyli nie możemy odrzucić hipotezy zerowej na rzecz hipotezy alternatywnej

$$H_0 = \sigma^2_l = \sigma^2_k \quad H_1 = \sigma^2_l > \sigma^2_k$$

Dla $\alpha = 0,05$ oraz $r_1 = 18$, $r_2 = 17$ wartość krytyczna odczytana z tablicy rozkładu F :

$$F_\alpha = 2,26 \text{ zatem } F_{obl} < F_\alpha$$

Określenie wartości krytycznej F_α umożliwia również funkcja *GRET*L-a: *Narzędzia/Tablice statystyczne*.

$$F(18, 17)$$

prawostronne prawdopodobieństwo = 0,05

prawdopodobieństwo dopełnienia = 0,95

krytyczna wartość = 2,25667 $\approx$ 2,26

Model trendu kwadratowego i wielomianu trzeciego stopnia

$$F_{obl} = \frac{(S_{ek}^2)}{(S_{es}^2)} = \frac{67,0887^2}{68,9233^2} = 0,9475$$

Gdy $\alpha = 0,05$ oraz $r_1 = 17$, $r_2 = 16$ wartość krytyczna $F_\alpha = 2,30$ i też występuje sytuacja jak przy porównaniu modelu liniowego i kwadratowego.

Spróbujmy ocenić istotność parametrów strukturalnych poszczególnych modeli bazując na $n = 20$ obserwacjach, ilości zmiennych objaśniających m oraz współczynnika determinacji R^2 korzystając ze statystyki¹⁰⁷:

$$F_{obl} = \frac{n-m-1}{m} * \frac{R^2}{1-R^2}$$

¹⁰⁷ Dziechciarz J. (red.), *Ekometria. Metody, przykłady, zadania*; Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003.

Dla poszczególnych modeli trendu wartości F_{obl} oraz wartości krytyczne F_α wg tablicy rozkładu F dla $\alpha = 0,05$ raz $r_1 = m$, $r_2 = n - m - 1$ wynoszą:

- liniowy: $m = 1$, $R^2 = 0,256463$ $F_{obl} = 6,21$ $F_\alpha = 4,41$ zatem $F_{obl} > F_\alpha$

- kwadratowy: $m = 2$, $R^2 = 0,287628$ $F_{obl} = 3,43$ $F_\alpha = 3,59$ $F_{obl} > F_\alpha$

- wielomianu trzeciego stopnia: $m = 3$, $R^2 = 0,292362$ $F_{obl} = 2,20$ $F_\alpha = 3,24$ $F_{obl} < F_\alpha$

Ponieważ ocena parametrów dla wielomianu trzeciego stopnia jest nieistotna przyjmujemy trend kwadratowy, czyli stopień wielomianu zmiennej czasowej $r = 2$. Informację o potrzebie takiego wyboru podaje również *GRET*L w wyświetlonej pełnej statystyce opisowej dla modelu trzeciego.

Warto nadmienić, że poszczególne modele możemy zapisać w formie ikony, "przeciągnąć myszką" do ikony zbiorczej dla sesji: *Tabela model*, a następnie wyświetlić prezentacje wszystkich modeli których może być maksymalnie 12.

10.2. Modele ze zmiennymi sezonowymi

Dla naszego wieloskładnikowego modelu procesu ekonomicznego mamy już określony stopień wielomianu zmiennej czasowej. W modelu występuje jednak też człon S_t obejmujący zmienne Q_t dla czterech kwartałów roku. Dla estymacji modeli z trendem kwadratowym i zmiennymi sezonowymi umieszczamy dodatkowo w bazie danych zmienne o wartości zero lub jeden w zależności od odniesienia do określonego kwartału. Przyjmijmy, że nasz wstępny (z oszacowanymi parametrami) model jest bez ostatniej zmiennej sezonowej dla kwartału czwartego, a parametr (tzw. amplitudę) ostatniej zmiennej określamy z podanego warunku, że suma parametrów (amplitud) równa się zero.

$$Y_t = a_0 + a_1 t + a_2 t^2 + \sum_{k=1}^{m-1} d_{ok} Q_{kt} + e'_t \quad \sum_{k=1}^m \hat{d}_{ok} = 0 \quad \hat{d}_{o4} = -(\hat{d}_{o1} + \hat{d}_{o2} + \hat{d}_{o3})$$

Zastosujmy model Kleina¹⁰⁸ dla zbudowania powyższego modelu trendu liniowego z periodycznymi składnikami sezonowymi w kwartałach roku korzystając z rachunku macierzowego i wyrażenia na wektor parametrów.

$$\hat{\alpha} = [T^T T]^{-1} T^T Y$$

$$\hat{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ \alpha_2 \\ d_{01} \\ d_{02} \\ d_{03} \end{bmatrix} \quad T = \begin{bmatrix} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 4 & 0 & 1 & 0 \\ 1 & 3 & 9 & 0 & 0 & 1 \\ 1 & 4 & 16 & -1 & -1 & -1 \\ 1 & 5 & 25 & 1 & 0 & 0 \\ 1 & 6 & 36 & 0 & 1 & 0 \\ 1 & 7 & 49 & 0 & 0 & 1 \\ 1 & 8 & 64 & -1 & -1 & -1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & 20 & 400 & -1 & -1 & -1 \end{bmatrix} \quad \hat{Y} = \begin{bmatrix} 133 \\ 238 \\ 262 \\ 159 \\ 137 \\ 257 \\ 267 \\ 170 \\ \vdots \\ 265 \end{bmatrix}$$

Kolumny macierzy T dotyczą w kolejności zmiennych:

t_0 przy parametrze a_0 tożsamościowo równej 1,

t (time) w bazie danych *GRET*L'a,

t^2 (time2),

¹⁰⁸ Praca zbiorowa, *Wybrane zagadnienia z prognozowania*, *ECONOMICUS*, Szczecin 2006. Rozdział 1.21.

Q_1 dotyczącej I kwartału przyjmującej w zależności od obserwacji w kwartałach roku wartości 1 lub 0,

Q_2 dotyczącej II kwartału,

Q_3 dotyczącej III kw.

T^T - macierz transponowana (zamiana kolumn wierszami),

Y - wektor zmiennej objaśnianej.

Dla oszacowania modelu trendu i sezonowości musimy wprowadzić do bazy GRETLa dodatkowo zmienne Q_1, Q_2, Q_3 . W tym względzie skorzystamy wstępnie z funkcji:

Dodawanie zmiennych/periodyczne zmienne 0-1.

Program definiuje zmienne sezonowe i generuje je automatycznie jako $dg1, dg2, dg3, dg4$ dla danych kwartalnych, przy czym przykładowo wartość $dg1$ dla kwartału 1. jest jeden, a dla pozostałych kwartałów zero. Dla doprowadzenia do macierzy T metody Kleina dokonujemy korekty co czwartej zmiennej na (-1) stosując funkcję:

Dane/Edycja wartości.

Estymację parametrów modelu trendowo-sezonowego przeprowadzimy stosując klasyczną metodę najmniejszych kwadratów (KMNK) dla zmiennych:

- objaśnianej Y (Nocleg),

- zmiennych objaśniających z macierzy T (const, time, time2, $dg1, dg2, dg3$).

Fragment statystyki opisowej oraz informacje uzupełniające modelu z procesu estymacji KMNK dla 20. obserwacji 2001:1-2005:4 podano poniżej.

Zmienna	Współczynnik	Błąd stand.	Statystyka t
const	208,783	8,644	24,1535
time	-4,6083	1,88721	-2,4419
time2	0,504562	0,0871991	5,7863
$dg1$	-75,9233	4,52516	-16,7780
$dg2$	48,8983	4,47889	10,9175
$dg3$	70,3108	4,47889	15,6983

Średnia arytmetyczna zmiennej zależnej = 232,8

Odchylenie standardowe zmiennej zależnej = 75,1872

Suma kwadratów reszt = 1866,74

Błąd standardowy reszt = 11,5472

Wsp. determinacji $R^2 = 0,98262$

Skorygowany $R^2 = 0,976413$

Statystyka $F(5, 14) = 158,308$

Statystyka testu Durбина-Watsona = 1,79609

Autokorelacja reszt rzędu pierwszego = 0,0735706

Kryterium informacyjne Akaike'a = 159,482.

Na uwagę zasługuje bardzo dobre dopasowanie poniższego modelu do danych rzeczywistych ($R^2 = 0,98262$).

$$\hat{Y}_t = 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3$$

Obliczmy jeszcze brakujący parametr dla okresu czwartego:

$$\hat{a}_{04} = -(-75,9233 + 48,8983 + 70,3108) = -43,2858$$

Tak więc dla celów prognozowania wszystkich kwartałów możemy stosować model:

$$\hat{Y}_t = 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3 - 43,2858Q_4$$

Międzyprogramowa konwersja bazy danych jest następująca:

- ❖ GRETL: Dane/Zaznacz wszystko/Pokaz wartości; Prezentacja danych/Kopiuj
- ❖ Excel: Edycja/Wklej; usuwamy *time3*, zamieniamy *dg1*, *dg2*, *dg3* na zmienne Q_1 , Q_2 , Q_3 i dopisujemy Q_4 dla kwartału IV.
- ❖ Obliczenie wartości teoretycznych **Noclegt** dla zmiennej zależnej (objaśnianej) wg wyżej podanego modelu w którym nie występuje *time3*.

Tab. 10.5. Obliczenia pomocnicze zmiennych modelu

Rok kw.	Nocleg	time	time2	Q1	Q2	Q3	Q4	Noclegt	Noclegtr
20011	133	1	1	1	0	0	0	128,8	128,8
20012	238	2	4	0	1	0	0	250,5	250,5
20013	262	3	9	0	0	1	0	269,8	269,8
20014	159	4	16	0	0	0	1	155,2	155,1
20021	137	5	25	1	0	0	0	122,4	122,4
20022	257	6	36	0	1	0	0	248,2	248,2
20023	267	7	49	0	0	1	0	271,6	271,6
20024	170	8	64	0	0	0	1	161,0	160,9
20031	134	9	81	1	0	0	0	132,3	132,3
20032	258	10	100	0	1	0	0	262,1	262,1
20033	273	11	121	0	0	1	0	289,5	289,5
20034	177	12	144	0	0	0	1	182,9	182,9
20041	150	13	169	1	0	0	0	158,2	158,2
20042	302	14	196	0	1	0	0	292,1	292,1
20043	333	15	225	0	0	1	0	323,5	323,5
20044	224	16	256	0	0	0	1	221,0	220,9
20051	188	17	289	1	0	0	0	200,3	200,3
20052	336	18	324	0	1	0	0	338,2	338,2
20053	393	19	361	0	0	1	0	373,7	373,7
20054	265	20	400	0	0	0	1	275,2	275,2

-32	81,5966	60,1841	-64,6375	0,50456	-4,6083	197,497
a_6	a_5	a_4	a_3	a_2	a_1	a_0

Źródło: Opracowanie własne w Excelu.

Do estymacji parametrów modelu trendowo-sezonowego zastosowano alternatywnie (vide obliczenie *Noclegtr* oraz parametry podane w powyższej małej tabeli) również funkcję arkusza kalkulacyjnego Excel o składni:

$$=REGLINP(C3:C22;D3:I22;PRAWDA;PRAWDA)$$

Gdzie przykładowo:

C3:C22 - obszar zmiennej zależnej *Y* czyli *Nocleg*,

D3:I22 - obszar kolumn zmiennych objaśniających (*time*, *time2*, Q_1 , Q_2 , Q_3 , Q_4).

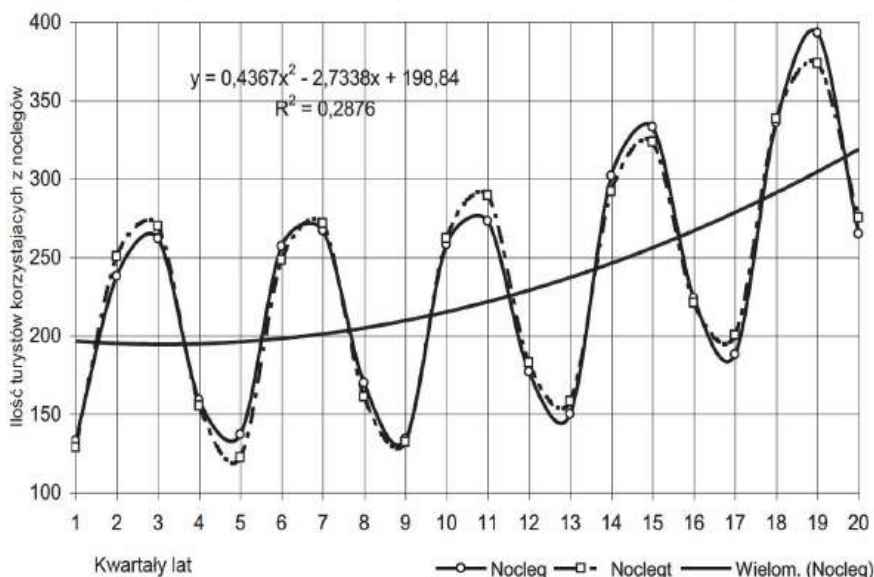
W obliczeniu wg *REGLINP* parametry stałej oraz przy zmiennych sezonowych różnią się od obliczenia wg modelu Kleina lecz wynik, czyli wartości w kolumnie *Noclegtr* są prawie identyczne. Excel umożliwia również sporządzenie wykresu zmiennej objaśnianej *Y* (*Nocleg*) oraz dla *Noclegt*. Na zamieszczonym wykresie podano również trend kwadratowy dla którego dopasowanie względem zmiennej *Nocleg* jest niskie - współczynnik determinacji $R^2 = 0,2876$.

Przedstawiony graficznie model jest zbliżony do wielorównaniowego modelu Holta-Wintersa z sezonowością addytywną stosowanego do określenia prognozy F na przyszły okres $n+m$ ¹⁰⁹. W modelu tym amplituda wahań zjawiska w czasie jest stała¹¹⁰.

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

Gdzie:

- S_n - wyrównana wykładniczo wartość zmiennej Y_n po usunięciu sezonowości,
- b_n - wyrównana wykładniczo wartość trendu na koniec okresu n ,
- m - horyzont prognozy,
- I_{n-L+1} - wyrównany wykładniczo indeks sezonowości w okresie $n - L + 1$,
- L - długość cyklu sezonowości (4 dla okresów kwartalnych),
- n - ilość obserwacji,
- Y_n - wartość rzeczywista zmiennej w okresie n ,
- α - parametr wyrównywania,
- β - parametr wyrównywania do oszacowania trendu b_n ,
- γ - parametr wyrównywania stosowany do obliczania indeksów sezonowości.



Źródło: opracowanie własne w Excelu.

Rys. 8.1. Dane empiryczne i teoretyczne zmiennej *Nocleg*

¹⁰⁹ Wg rozdziału 2.15 publikacji podanej w poniższym przypisie.

¹¹⁰ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*. Tom II. *Ekonometria*, Wydawnictwo C.H. Beck, Warszawa 2003.

10.3. Modele autoregresyjne

Dla procesów ekonomicznych trendowo-sezonowo-autoregresyjnych rząd modelu autoregresyjnego p określany jest wg reszt obliczonych wg modelu trendu i sezonowości¹¹¹, który określiliśmy jako:

$$\hat{Y}_t = 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3 - 43,2858Q_4$$

Model autoregresyjny procesu stacjonarnego o rzędzie opóźnienia p w którym wartość bieżąca określana jest na podstawie wartości z poprzednich okresów jest postaci:

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 Y_{t-2} + \dots + \alpha_p Y_{t-p} + \varepsilon_t$$

Warto zaznaczyć, że procesem stacjonarnym nazywamy taki proces dla którego szereg czasowy ma stałą średnią i wariancję, a kowariancja zależy od rzędu opóźnienia a nie od czasu¹¹². Rząd opóźnienia można określić w oparciu o funkcję autokorelacji cząstkowej, która wyraża zależność pomiędzy współczynnikiem autokorelacji cząstkowej a rzędem opóźnienia. Wartość współczynnika autokorelacji cząstkowej jest z przedziału $<-1;1>$. Przykładowo współczynnik autokorelacji cząstkowej dla $p = 3$ bada zależność między Y_t i Y_{t-3} , gdy pominiemy wpływ zmiennych Y_{t-1} oraz Y_{t-2} . Gdy funkcja autokorelacji zanika tj. wartości współczynników autokorelacji maleją wykładniczo i funkcja ta urywa się to rząd autoregresji jest równy p .

W GRETL dla oceny składnika autoregresyjnego procesu skorzystać można z funkcji

Wykresy/wykres reszt modelu/korelogram

przyjmując rząd opóźnienia 4 w menu okna wyników statystyki opisowej dla powyższego modelu trendowo-sezonowego. Maksymalna liczba opóźnienia p_{max} nie powinna przekraczać $15 \div 20\%$ długości szeregu i tak dla $n = 20$ $p_{max} = 3; 4$.

Podana jest tu funkcja autokorelacji (ACF) oraz autokorelacji cząstkowej (PACF) dla procesu resztowego. Fragment wyników podano poniżej przy czym nie występuje tu wyraźna tendencja wykładnicza zanikania oraz w końcu urywania się (przyjmowanie wartości bliskich zeru) dla autokorelacji cząstkowej PACF. Zaobserwowanie tego wymaga znacznie większej ilości obserwacji zmiennej zależnej.

Opóźnienia	ACF	PACF
1	0,0695	0,0695
2	-0,2720	-0,2782
3	-0,2226	-0,1950
4	0,0199	-0,0338

Współczynnik korelacji cząstkowej dla określonego rzędu opóźnienia przy poziomie istotności $\alpha = 0,05$ porównujemy z wartością krytyczną, która dla $n = 20$ obserwacji wynosi¹¹³:

$$PACF_\alpha = u_\alpha / \sqrt{n} = 1,96 / \sqrt{20} = 0,438$$

Żadna z wartości dla opóźnień 1,2,3,4 nie spełnia warunku, że $|IPACF| > PACF_\alpha$ i niska jest również wartość współczynnika autokorelacji. Wszystkie wartości oszacowanych współczynników autokorelacji cząstkowej są mniejsze od wartości granicznej, co świadczy, że proces ma cechy procesu losowego.

¹¹¹ Wg rozdziału 6.3 publikacji podanej w przypisie 1.

¹¹² Rozdział 3.3 publikacji podanej w przypisie 6.

¹¹³ Jak wyżej w przypisie 7

Dla dalszej prezentacji naszego przykładu wieloskładnikowego modelu procesu ekonomicznego niestacjonarnego typu trendowo-sezonalno-autoregresyjnego przyjmijmy, że występuje autokorelacja rzędu pierwszego.

$$Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \varepsilon_t$$

$$\hat{Y}_t = a_0 + a_1 t + a_2 t^2 + a_3 Q_1 + a_4 Q_2 + a_5 Q_3 + a_6 Q_4 + a_7 Y_{t-1}$$

W tej sytuacji model będzie postaci:

Dla uzupełnienia bazy danych o zmienną Y_{t-1} skorzystamy z funkcji:

$$\begin{aligned} \text{Dodawanie zmiennych/definiowanie nowej zmiennej} \\ \text{Noclego} = \text{Nocleg}(-1) \end{aligned}$$

Uzyskaliśmy wektor zmiennej objaśnianej przesunięty o 1 kwartał. Pozwoliło to nam na dokonanie estymacji programem *GRET* modelu złożonego z trzech członów klasyczną metodą najmniejszych kwadratów dla obserwacji 2÷20 (2001:2-2005:4) dla zmiennej zależnej *Nocleg*.

Zmienna	Współczynnik	Błąd stand.	Statystyka t
const	185,171	66,7923	2,7723
time	-3,33253	3,17657	-1,0491
time2	0,426265	0,201907	2,1112
dq1	-73,2031	13,382	-5,4703
dq2	56,3966	23,5836	2,3913
dq3	66,2029	15,2764	4,3337
Noclego	0,0922543	0,301296	0,3062

Średnia arytmetyczna zmiennej zależnej = 238,053

Odczylenie standardowe zmiennej zależnej = 73,3806

Suma kwadratów reszt = 1817,96

Błąd standardowy reszt = 12,3084

Współczynnik determinacji $R^2 = 0,981244$

Skorygowany $R^2 = 0,971865$

Statystyka $F(6, 12) = 104,63$

Ilość stopni swobody $iss = n - k = 20 - (m + 1) = 20 - 7 = 13$.

Ze względu na nieznaczne pogorszenie się współczynnika determinacji po wprowadzeniu członu z autoregresją pierwszego rzędu sprawdzimy istotność tego parametru przy pomocy testu *t*-Studenta; dla $\alpha = 0,05$ i $iss = 13$

$$|t_{\text{obs}}| > t_{\alpha} = 2,1604$$

Warunek ten nie spełniają parametry dla zmiennych *time* oraz oraz wyraźnie *Noclego*, więc zrezygnujemy ze zmiennej *Noclego* i nasz model pozostaje w postaci wcześniej już zdefiniowanej trendowo-sezonalnej.

$$\hat{Y}_t = a_0 + a_1 t + a_2 t^2 + a_3 Q_1 + a_4 Q_2 + a_5 Q_3 + a_6 Q_4$$

Model ten wykazuje przydatność do budowy prognoz. Obliczmy więc przykładowo prognozę na okres 21., czyli I kwartał kolejnego roku:

$$\begin{aligned} \hat{Y}_t &= 208,783 - 4,6083t + 0,504562t^2 - 75,9233Q_1 + 48,8983Q_2 + 70,3108Q_3 - 43,2858Q_4 = \\ &= 208,783 - 4,6083 \cdot 21 + 0,504562 \cdot 21^2 - 75,9233 \cdot 1 + 48,8983 \cdot 0 + 70,3108 \cdot 0 - 43,2858 \cdot 0 = 259 \end{aligned}$$

Zaprezentowano jedno z podejść do formułowania modelu trendowo-sezonowo-autoregresyjnego dla niestacjonarnego procesu ekonomicznego. Zarówno bowiem struktura modeli jak i etapy ich uzyskania mogą być różnorodne. Występuje jednak znaczna pracochłonność ich budowy i weryfikacji. Z tego względu już w trakcie studiowania ekonometrii jak i w praktyce gospodarczej powinniśmy intensywnie korzystać ze wspomagania komputerowego, a w tym z obszernych możliwości jakie daje dostępny w Internecie program *GRET*.

Próbkę ich w formie instrukcji korzystania z oprogramowania podano w niniejszym rozdziale. Niech to będzie kluczem do dalszego prześledzenia dogodności tego programu w celu poznania kolejnych metod i testów.

11. ESTYMACJA MODELI WIELORÓWNANIOWYCH W *GRETL*



11.1. Wstęp

Istnieje wiele modeli ekonometrycznych, a do szacowania ich parametrów literatura podaje różne sposoby. Istnieją obecnie też różnorodne programy komputerowe wspomagające rozwiązywanie modeli jedno i wielorównaniowych. Oprócz tradycyjnej już funkcjonalności *Excels* w zakresie rachunku macierzowego dla celów dydaktycznych proponuje się korzystanie z programu *GRETL*, który jest dostępny w Internecie. Obejmuje on liczne modele ekonometryczne oraz można skorzystać z komputerowej estymacji ich parametrów różnymi sposobami. W niniejszym rozdziale wymieniono te sposoby, dla części których spotykamy teoretyczne podstawy w dziele G. S. Maddali¹¹⁴. Ponadto zaprezentowano przykładowo podwójną metodę najmniejszych kwadratów w odniesieniu do modeli wielorównaniowych bazując na rozważaniu teoretycznym zawartym publikacji "Ekonometria. Metody, przykłady, zadania"¹¹⁵ oraz pracy Tadeusza Kufla¹¹⁶. Dokonano przetestowania tej metody na programie z Internetu, w wyniku czego otrzymano procedurę realizacji - jak się wydaje bardzo pomocną dla studiujących ekonometrię. Warto zaznaczyć, że program *GRETL* wersja 1.6.5 jest oprogramowaniem bezpłatnym należącym do grupy *Open Source*.

11.2. Model wielorównaniowy

Model strukturalny wielorównaniowy (bez zmiennych opóźnionych w czasie) w postaci macierzowej jest następujący:

$$BY + CX = \varepsilon$$

Przy czym:

B - macierz nieosobliwa parametrów przy zmiennych endogenicznych,

C - macierz parametrów przy zmiennych egzogenicznych,

Y - wektor kolumnowy zmiennych objaśnianych,

X - wektor kolumnowy zmiennych objaśniających,

ε - wektor kolumnowy składników losowych.

Macierzą nieosobliwą nazywamy macierz kwadratową, której wyznacznik jest różny od zera¹¹⁷.

$$B = \begin{bmatrix} 1 & -b_{12} & \dots & -b_{1m} \\ -b_{21} & 1 & \dots & -b_{2m} \\ \dots & \dots & \dots & \dots \\ -b_{m1} & -b_{m2} & \dots & 1 \end{bmatrix} \quad C = \begin{bmatrix} -c_{11} & -c_{12} & \dots & -c_{1k} \\ -c_{21} & -c_{22} & \dots & -c_{2k} \\ \dots & \dots & \dots & \dots \\ -c_{m1} & -c_{m2} & \dots & -c_{mk} \end{bmatrix} \quad Y = PX + \eta$$

¹¹⁴ Maddala G.S., *Ekonometria*. Wydawnictwo Naukowe PWN Warszawa 2006.

¹¹⁵ Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu 2003

¹¹⁶ Kufl T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN. Warszawa 2007.

¹¹⁷ Praca zbiorowa, *Leksykon naukowo-techniczny*. Warszawa 1984.

Po pomnożeniu równania $\mathbf{BY} + \mathbf{CX} = \varepsilon$ przez macierz odwrotną $\mathbf{B}^{-1}$ to otrzymamy postać zredukowaną modelu $\mathbf{Y} = \mathbf{PX} + \boldsymbol{\eta}$ w którym Y zależy tylko od zmiennych egzogenicznych X , przy czym:

$$\mathbf{P} = -\mathbf{B}^{-1}\mathbf{C} \text{ - parametry postaci zredukowanej,}$$

$$\boldsymbol{\eta} = \mathbf{B}^{-1}\boldsymbol{\varepsilon} \text{ - składnik losowy postaci zredukowanej.}$$

11.3. Model prosty

W tym modelu postać zredukowana jest identyczna ze strukturalną. Oznaczając przez $\mathbf{I}_m$ macierz jednostkową zmiennych objaśniających to:

$$\mathbf{B} = \mathbf{I}_m$$

W zbiorze modeli jednorównaniowych w żadnym z nich nie występuje zmienna endogeniczna. Parametry w poszczególnych równaniach możemy określić niezależnie od siebie klasyczną metodą najmniejszych kwadratów (*KMNK*), korzystając przykładowo z tej metody zawartej w programie *GRET*L. Alternatywnym rozwiązaniem jest zastosowanie funkcji *Excela* (*TRANSPONUI*; *ILO-CZYN.MACIERZ*; *MACIERZ.ODW*) dla oszacowania parametrów równań modelu prostego z zależności:

$$\mathbf{a} = [\mathbf{X}^T\mathbf{X}]^{-1}\mathbf{X}^T\mathbf{y}$$

Gdzie:

$\mathbf{X}$ - macierz zmiennych objaśniających wraz ze zmienną tożsamościowo równą 1 odpowiadającą stałej w danym równaniu,

$\mathbf{X}^T$ - macierz jak wyżej, lecz transponowana,

$\mathbf{y}$ - wektor określonej zmiennej objaśnianej.

11.4. Model rekurencyjny

Niech macierz $\mathbf{B}$ będzie macierzą trójkątną z zerami nad główną przekątną mającą jedynki lub daje się do takiej postaci sprowadzić. Poszczególne kolumny odpowiadają równaniom zmiennych w modelu wielorównaniowym. Przyjmijmy, że w pierwszym równaniu nie występują zmienne endogeniczne jako zmienne objaśniające, tak więc parametry tego równania można oszacować *KMNK*. W kolejnych równaniach mogą się znajdować zmienne endogeniczne pełniące funkcje zmiennych objaśniających. Estymację parametrów modelu rekurencyjnego przeprowadzamy wg procedury:

- ❖ obliczenie wartości teoretycznych $y^{\wedge}_1$ pierwszego równania dla poszczególnych obserwacji i , po wcześniejszym oszacowaniu parametrów *KMNK* o czym powiedziano wyżej,
- ❖ zastąpienie w macierzach obserwacji zmiennych równania drugiego wartości empirycznych wartościami obliczonymi - teoretycznymi zmiennej objaśnianej Y_1 ,
- ❖ oszacowanie parametrów równania drugiego również korzystając z *KMNK*,
- ❖ obliczenie wartości teoretycznych $y^{\wedge}_2$ kolejnych obserwacji zmiennej Y_2 ,
- ❖ postępujemy analogicznie z kolejnymi równaniami modelu rekurencyjnego.

11.5. Przykład zastosowania *GRET*L dla modelu rekurencyjnego

Jak już wspomniano program *GRET*L obejmuje między innymi metody ekonometryczne i sposoby estymacji parametrów równania (równań) określonego modelu. Polską wersję instalacyjną programu można pozyskać z Internetu pod adresem: [www.kufel.torun.pl]. Pomyślany jest jako

pomocny na zajęciach z ekonometrii.¹¹⁸ Dla konfrontacji obliczeń manualnych z realizacją komputerową bazowano na danych zaprezentowanych w przykładzie 8.1. publikacji podanej w przypisie 2. W zamieszczonej tabeli stanowiącej przedrędagowany w *Excelu* wydruk ze zbioru *GRET*L-a (z kropką dziesiętną zamiast przecinka) poszczególne zmienne oznaczają:

X_1 - liczba wydanych książek i broszur (mln. egz.),

X_2 - produkcja papieru (tys. t.),

X_3 - liczba bibliotek publicznych,

X_4 - liczba studentów (tys.),

X_5 - liczba uczniów w szkołach średnich (tys.),

X_6 - liczba uczniów w szkołach podstawowych (tys.),

X_7 - liczba abonentów telewizyjnych (mln.),

X_8 - liczba widzów w kinach (mln.),

X_9 - liczba chorych w szpitalach.

Naszym zadaniem jest oszacowanie parametrów modelu trzypięciowego, przy czym zmienne endogeniczne to X_1 , X_2 , X_3 , z których X_1 oraz X_2 stanowią jednocześnie zmienne objaśniające. Po przeniesieniu zmiennych endogenicznych na lewą stronę otrzymamy:

$$\hat{x}_1 = a_1 \hat{x}_2 + a_2 \hat{x}_4 + a_0$$

$$\hat{x}_2 = b_1 \hat{x}_4 + b_2 \hat{x}_7 + b_0$$

$$\hat{x}_3 = c_1 \hat{x}_1 + c_2 \hat{x}_5 + c_0$$

$$1\hat{x}_1 - a_1 \hat{x}_2 + 0\hat{x}_3 = a_2 \hat{x}_4 + a_0$$

$$0\hat{x}_1 + 1\hat{x}_2 + 0\hat{x}_3 = b_1 \hat{x}_4 + b_2 \hat{x}_7 + b_0$$

$$-c_1 \hat{x}_1 + 0\hat{x}_2 + 1\hat{x}_3 = +c_2 \hat{x}_5 + c_0$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 \\ -a_1 & 1 & 0 \\ 0 & -c_1 & 1 \end{bmatrix}$$

Źródło: opracowanie własne w *GRET*L bazujące na wartościach obserwacji zmiennych w tabeli 8.1 publikacji podanej w przypisie 2.

Macierz współczynników $\mathbf{B}$ przy zmiennych endogenicznych po doprowadzeniu do postaci macierzy trójkątnej dolnej (zamiana kolumny drugiej na pierwszą oraz wiersza pierwszego na drugi) wskazuje na istnienie modelu rekurencyjnego. Musimy zatem wpięrow oszacować parametry równania drugiego bazując na szeregach 24 obserwacji zmiennych X_4 , X_7 oraz X_0 (równę 1) oraz wektorze X_2 stanowiącym zmienną objaśnianą.

¹¹⁸ Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*L. Wydawnictwo Naukowe PWN Warszawa 2007 (Wstęp do wydania drugiego).

Tab. 11.1. Obliczenia pomocnicze w Excelu

Obserwacje zmiennych X_1 - X_9 oraz obliczone wartości X_1T , X_2T , X_3T wg równań modelu

Obsz.	Rok	X1	X2	X3	X4	X5	X6	X7	X8	X9	X2T	X1T	X3T
1	1970	112.2	764	8621	331	537	5389	4215	137.6	3263	858.359673	112.288	9097.0382
2	1971	131.1	839	8666	348	578	5186	4709	140.4	3414	878.660583	133.444	9169.0946
3	1972	137.6	899	8725	364	609	4978	5200	136.1	3487	898.770352	149.152	9177.9566
4	1973	135.8	940	8852	398	639	4779	5687	140.7	3610	920.088181	148.191	9139.1558
5	1974	139.5	969	8950	427	649	4306	6100	142.7	3717	938.179336	145.001	9151.4030
6	1975	143.9	981	8974	468	622	4448	6472	140.8	3848	955.600376	128.978	9202.6034
7	1976	159.5	1046	9060	491	582	4327	6820	144.2	3929	970.735608	143.122	9331.6370
8	1977	149.7	1081	9128	491	538	4257	7170	131.6	4075	984.207458	157.017	9315.6962
9	1978	143.7	1070	9199	485	491	4217	7474	116.1	4179	995.454714	155.692	9324.8342
10	1979	151.2	1010	9252	469	450	4217	7708	107.9	4224	1003.250920	139.984	9407.3762
11	1980	147.1	1033	9315	454	415	4260	7954	97.5	4356	1011.584686	156.720	9416.3366
12	1981	133.9	909	9355	426	393	4342	8188	100.1	4136	1018.472876	121.688	9359.7614
13	1982	178.1	965	9442	397	381	4465	8347	89.4	4344	1022.398573	158.623	9629.9006
14	1983	194.9	1026	9572	370	373	4327	8542	108.1	4495	1027.861282	196.529	9735.8174
15	1984	229.8	1042	9728	350	375	4771	8765	127.6	4588	1034.931415	213.021	9938.3078
16	1985	246.3	1071	9899	341	383	4928	9468	107.1	4571	1061.309576	229.097	10027.3958
17	1986	249.4	1100	10000	334	400	5058	9692	94.3	4605	1069.401884	244.159	10029.5354
18	1987	267.6	1158	10129	343	422	5165	9868	95.3	4611	1076.857312	262.622	10115.3906
19	1988	245.3	1220	10248	356	444	5213	10031	96.4	4622	1084.115029	280.645	9964.0778
20	1989	215.6	1164	10313	378	463	5257	10055	69.6	4520	1086.703509	247.259	9772.2566
21	1990	175.6	924	10269	404	494	5301	9919	32.8	4597	1083.436101	138.797	9508.8146
22	1991	125.5	946	9936	428	549	5323	9809	20.9	4680	1081.018123	135.363	9163.6190
23	1992	125.8	1031	9770	496	607	5324	10043	13.3	4742	1095.170441	134.632	9110.7398
24	1993	102.5	1070	9605	584	660	5288	10111	14.9	4934	1104.446613	105.499	8924.3690

$$\hat{x}_2 = b_1x_4 + b_2x_7 + b_0$$

Źródło: Opracowanie własne w Excelu.

W tym miejscu skorzystamy z funkcji estymacji parametrów równania *KMNK* występującej w programie *GRET*L, przy czym dla potrzeb języka skryptowego tego programu zmienne empiryczne zapisano symbolami: $X1$, $X2$, $X3$, $X4$, $X5$, $X6$, $X7$, $X8$, $X9$, a zmienne endogeniczne - obliczone - teoretyczne jako $X1T$, $X2T$, $X3T$. Początkowa procedura postępowania jest następująca:

- zainstalowanie pełnej wersji polskiej programu *GRET*L;
- w menu głównym korzystamy z opcji *Plik/Nowy zbiór danych*;
- mając powyższe dane dla kolejnych zmiennych, w oknie dialogowym wpisujemy liczbę obserwacji 24 (domyślnie 50), OK;
- w oknie "Struktura bazy danych" zaznaczamy opcję "Dane przekrojowe" i naciskamy przycisk "Naprzód";
- wpisujemy nazwę zmiennej *Rok* i wprowadzamy dane - lata 1970÷1993 akceptując strzałką w dół;
- po wprowadzeniu pełnego szeregu 24 lat naciskamy tylko klawisz "Zastosuj";
- w menu danego podokna wybieramy "Zmienna" oraz opcję "Dodaj";
- wpisujemy nową zmienną $X1$, wprowadzamy dane, naciskamy "Zastosuj", postępujemy podobnie z dalszymi zmiennymi $X2$ - $X9$, a na zakończenie zbioru danych naciskamy klawisz "Zamknij". Ponadto w menu "Plik" wybieramy opcję "Zapisz dane jako" i zapisujemy kolejno w 2-ch formatach:
 - standardowy plik danych *GRET*L'a np. *E4.gdt*,
 - bazy danych *E4.bin* z prowadzeniem nazwy bazy przykładowo "Obserwacje".

Całą pracę z *GRET*L'em, jako sesję, zapisujemy w formie pliku *E4.gretl* w podkatalogu *GRET*L-a dysku D.

Zaprezentowana poniżej struktura zbioru zapisana została na dysku jako plik tekstowy E4.txt. Widzimy tu dodatkowo wprowadzone zmienne wygenerowane w postaci szeregów odpowiadających określonym obserwacjom zmiennych wchodzących w trzech równaniach rozpatrywanego modelu rekurencyjnego. Zmienne wprowadzamy w nawiasach, parametry zmiennych zapisujemy z kropką dziesiętną. Całość poprzedzamy poleceniem *genr*, przykładowo dla X_{2T} - co widać na poniższej strukturze zbioru danych w postaci notacji komputerowej przepisanej z ekranu *GRET*L'a:

$$\text{genr}X_{2T}=(0.075668)*(X_4)+(0.038491)*(X_7)+(671.074)$$

Obserwacje 1 - 24 n = 24

Rok

X1

X2

X3

X4

X5

X6

X7

X8

X9

genrX2T (0.075668)*(X4)+(0.038491)*(X7)+(671.074)

genrX1T (0.397263)*(genrX2T)-(0.507356)*(X4)-(23.2019)

genrX3T (5.83453)*(genrX1T)-(0.948584)*(X5)+(8953.55)

Procedura oszacowania parametrów równań oraz wygenerowania wartości teoretycznych jest następująca:

❖ Mając zbiór zmiennych przystępujemy do oszacowania parametrów równań, lecz najpierw drugiego, gdyż zawiera tylko zmienne objaśniające X_4 oraz X_7 .

$$\hat{x}_2 = b_1 x_4 + b_2 x_7 + b_0$$

Z menu głównego wybieramy "Model" a następnie opcję "Klasyczna metoda najmniejszych kwadratów". Pojawia się okno dialogowe "Specyfikacja modelu". Wybieramy jako zmienną zależną X_2 (wg symbolizacji składni *GRET*L'a), a jako zmienne niezależne: const (stała), X_4 , X_7 . Program natychmiast wyświetla dane z estymacji podając dodatkowo statystykę opisową. Wyniki zapisujemy jako plik w formacie RTF(MS Word) na dysku D:\GRET\LE4.rtf. Dla równania drugiego uzyskaliśmy:

$$\hat{x}_2 = 0,075668x_4 + 0,038491x_7 + 671,074$$

❖ Generujemy szereg wartości teoretyczne dla x_2 w tym celu korzystamy z menu *Zmienna/Definiowanie nowej zmiennej* i wprowadzamy wcześniej już podaną formułę na wyliczenie zmiennej określonej jako X_{2T} .

❖ Szacujemy parametry dla równania pierwszego również *KM*NK w *GRET*L wstawiając zamiast x_2

$$\hat{x}_1 = 0,397263x_2 - 0,507356x_4 - 23,2019$$

wygenerowane wartości X_{2T}

Program numeruje kolejne wygenerowane modele i podaje dodatkowo statystykę opisową. Przykładowo dla powyższego równania zmiennej zależnej $X1$ ze zmienną teoretyczną $X2T$, zmienną objaśniającą $X4$ oraz stałą dla zbioru danych o 24. obserwacjach otrzymamy parametry zaprezentowane w tabeli 11.2.

Tab. 11.2. Parametry modelu

Zmienna	Współczynnik	Błąd standardowy	Statystyka t
const	-23,2019	84,0681	-0,2760
$X2T$	0,397263	0,080979	4,9058
$X4$	-0,507356	0,0880348	-5,7631

Statystyka opisowa:

Średnia arytmetyczna zmiennej zależnej = 168,4

Odchylenie standardowe zmiennej zależnej = 48,5454

Suma kwadratów reszt = 16207,7

Błąd standardowy reszt = 27,7812

Współczynnik determinacji $R^2 = 0,700981$

Skorygowany współczynnik determinacji = 0,672503.

Źródło: Opracowanie własne bazujące na danych z tabeli 11.1.

- ❖ Generujemy szereg wartości dla $X1$ wstawiając wartości z szeregów $X2T$ oraz $X4$.
- ❖ Szacowanie parametrów dla równania trzeciego wprowadzając zamiast $X1$ wartości szeregu dla $X1T$ oraz wartości obserwacji dla zmiennej objaśniającej $X5$.

$$\hat{x}_3 = 5,83453x_1 - 0,948584x_5 + 8953,55$$

- ❖ Możemy dodatkowo (vide wcześniejsza tabela 11.1) wygenerować również wartości szeregu $X3T$.

11.6. Model o równaniach łącznie współzależnych

W tego typu modelach wśród zmiennych objaśniających występują takie, których nie można zastąpić wartościami teoretycznymi z równań wcześniej oszacowanych KMNK przykładowo:

$$Y_1 = a_1 Y_2 + a_2 X_1 + a_0; \quad Y_2 = b_1 Y_1 + b_2 X_1 + b_0$$

Warunkiem estymacji jest identyfikowalność równań, czyli każde równanie musi być inne niż poprzednie. Stosujemy tu twierdzenie (wg publikacji podanej w przypisie 2 s. 196):

"Warunkiem identyfikowalności danego równania modelu m-równaniowego jest to, by macierz A_i utworzona ze współczynników stojących przy zmiennych występujących w innych równaniach modelu, a nie występujących w badanym (i-tym) równaniu była rzędu m-1". Ponadto dane równanie jest identyfikowalne jednoznacznie, gdy liczba zmiennych występujących w modelu, a nie występujących w badanym równaniu jest m-1".

"Rzędem macierzy nazywamy największy stopień niezerowego minora tej macierzy, czyli wyznacznika obliczonego z wybranych wierszy i kolumn tej macierzy"¹¹⁹.

Zastosowanie określonej metody estymacji zależy od badania identyfikowalności i tak gdy równanie jest identyfikowalne jednoznacznie stosujemy jedną z dwóch metod:

- ❖ podwójną metodę najmniejszych kwadratów,

¹¹⁹ Jurliewicz J., Skoczylas Z., *Algebra liniowa I. Przykłady i zadania*, Oficyna Wydawnicza GiS Wrocław 2003.

❖ pośrednią metodę najmniejszych kwadratów.

Rozpatrzmy model o $m = 3$ równaniach łącznie współzależnych (wg przypisu 2 s. 197) i niżej podanej macierzy (po przeniesieniu elementów na lewą stronę równań):

$$\begin{aligned}\hat{X}_1 &= a_1 X_2 + a_2 X_4 + a_0 \\ \hat{X}_2 &= b_1 X_3 + b_2 X_7 + b_0 \\ \hat{X}_7 &= c_1 X_1 + c_2 X_9 + c_0\end{aligned}\quad \mathbf{B} = \begin{bmatrix} 1 & -a_1 & 0 \\ 0 & 1 & -b_2 \\ -c_1 & 0 & 1 \end{bmatrix}$$

Zmienne endogeniczne X_1, X_2, X_7 są jednocześnie zmiennymi objaśniającymi. Sprawdźmy zatem identyfikowalność równań, poprzez badanie macierzy $\mathbf{A}_i$ obejmującej, przykładowo dla równania pierwszego zmienne egzogeniczne X_3, X_7, X_9 z równań drugiego i trzeciego, które nie występują w równaniu pierwszym.

$$\mathbf{A}_1 = \begin{bmatrix} -b_1 & -b_2 & 0 \\ 0 & 1 & -c_2 \end{bmatrix} \quad \mathbf{A}_2 = \begin{bmatrix} 1 & -a_2 & 0 \\ -c_1 & 0 & -c_2 \end{bmatrix} \quad \mathbf{A}_3 = \begin{bmatrix} -a_1 & 0 & -a_2 \\ 1 & -b_1 & 0 \end{bmatrix}$$

Rząd macierzy $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ jest 2, czyli równa się $m-1$ ¹²⁰. Zatem nasze trzy równania są identyfikowalne. Aby dana macierz była rzędu drugiego, to potrzeba aby chociaż jeden z wyznaczników - w naszym przykładzie "par kolumn" był różny od zera. Dla macierzy $\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3$ warunek ten jest spełniony. Niech k oznacza liczbę zmiennych które występują w modelu a nie występują w określonym równaniu. Dla $m = 3$ równania, w każdym z równań $k = 3$. Tak więc $m - 1 < k$, co oznacza, że równania są niejednoznacznie identyfikowalne i umożliwia to zastosowanie podwójnej metody najmniejszych kwadratów 2MKN do estymacji parametrów równań. Warto nadmienić, że metoda ta stosowana jest również dla równań identyfikowalnych jednoznacznie, czyli gdy występuje warunek $m - 1 = k$.

Procedura postępowania w metodzie 2MKN jest następująca:

❖ Szacowanie KMNK postaci zredukowanej modelu $\mathbf{Y} = \mathbf{P}\mathbf{X} + \eta$ co wyjaśniono na wstępie tego rozdziału. W wyniku otrzymujemy trzy równania w których występuje ta sama macierz zmiennych objaśniających $\mathbf{X}$. Możemy szacować każde zredukowane równanie osobno, ponieważ doprowadziliśmy do modelu prostego równań lub dokonać przeliczenia dla wszystkich równań jednocześnie (vide podstawy teoretyczne w publikacji wg przypisu 2 rozdział 8.6). Macierz transponowana $\mathbf{P}^T$ obliczamy z zależności:

$$\mathbf{P}^T = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{Y}$$

Gdzie:

$\mathbf{X}$ - macierz obserwacji zmiennych egzogenicznych łącznie ze zmienną X_0 równą 1 o wymiarze $n \times k$,

$\mathbf{Y}$ - macierz o wymiarze $n \times m$ obserwacji zmiennych endogenicznych.

$$\mathbf{X} = \begin{bmatrix} x_{11} & x_{12} & \dots & 1 \\ x_{21} & x_{22} & \dots & 1 \\ \dots & \dots & \dots & \dots \\ x_{n1} & x_{n2} & \dots & 1 \end{bmatrix} \quad \mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1m} \\ y_{21} & y_{22} & \dots & y_{2m} \\ \dots & \dots & \dots & \dots \\ y_{n1} & y_{n2} & \dots & y_{nm} \end{bmatrix} \quad \mathbf{P}^T = \begin{bmatrix} p_{11} & p_{12} & \dots & p_{m1} \\ p_{12} & p_{22} & \dots & p_{m2} \\ \dots & \dots & \dots & \dots \\ x_{1k} & x_{2k} & \dots & p_{mk} \end{bmatrix}$$

¹²⁰ Nowak E., *Zarys metod ekonometrii. Zbiór zadań*. Wydawnictwo Naukowe PWN. Warszawa 1990.

Kolumny macierzy $\mathbf{P}^T$ odpowiadają elementom i -tych równań postaci zredukowanej modelu.

- ❖ Obliczenie wartości teoretycznych kolejnych zmiennych objaśniających.

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{P}^T$$

Zasadniczo interesują nas te wartości teoretyczne zmiennych, które w kolejnych równaniach występują jako zmienne objaśniające. Jeśli stosujemy funkcję `ILOCZYN.MACIERZ(...;...)` Excela to mo-

$$\hat{\mathbf{Y}} = \begin{bmatrix} \hat{y}_{11} & \hat{y}_{12} & \dots & \hat{y}_{1m} \\ \hat{y}_{21} & \hat{y}_{22} & \dots & \hat{y}_{2m} \\ \dots & \dots & \dots & \dots \\ \hat{y}_{n1} & \hat{y}_{n2} & \dots & \hat{y}_{nm} \end{bmatrix}$$

żemy szybko obliczyć wszystkie wartości teoretyczne.

- ❖ Mając wartości teoretyczne zmiennych możemy dokonać estymacji parametrów równań postaci strukturalnej modelu. W tym względzie stosujemy ponownie *KMNK* budując macierz $\mathbf{X}_i$, zastępując jednak wartości obserwacji zmiennych endogenicznych ich wartościami teoretycznymi pobranymi z odpowiedniej kolumny macierzy $\hat{\mathbf{Y}}$. Korzystamy tu ze znanego nam wzoru na wektor ocen parametrów strukturalnych i -tego równania modelu bazując na wektorze obserwacji zmiennej objaśnianej y_i .

Ocenę parametrów przeprowadzamy oddzielnie dla każdego równania postaci strukturalnej, chociaż istnieje możliwość zredukowania tego do jednego etapu. Wyjaśnienie postępowania w tym drugim przypadku znajdujemy w rozdziale 8.7 pracy wymienionej w przypisie 2.

11.7. Przykład podwójnej metody najmniejszych kwadratów (2MKN)

$$\mathbf{b}_i = [\mathbf{X}_i^T \mathbf{X}_i]^{-1} \mathbf{X}_i^T \mathbf{y}_i$$

Jednym z powszechnie dostępnych sposobów estymacji parametrów wg 2MKN jest zastosowanie funkcji *Excela* w zakresie rachunku macierzowego. Przypomnijmy sobie jednak nasz model o równaniach łącznie współzależnych:

$$\hat{X}_1 = a_1 X_2 + a_2 X_4 + a_0$$

$$\hat{X}_2 = b_1 X_3 + b_2 X_7 + b_0$$

$$\hat{X}_7 = c_1 X_1 + c_2 X_9 + c_0$$

Widzimy, że zmiennymi endogenicznymi są X_1 , X_2 , X_7 a egzogenicznymi X_3 , X_4 , X_9 oraz $X_0 = 1$. Rozwiązanie bazujące na postępowaniu teoretycznym wg przykładu 8.3 w publikacji podanej w przypisie 2 realizujemy następująco:

- ❖ Określamy macierze $\mathbf{X}$ (X_3 , X_4 , X_9 , X_0) oraz $\mathbf{Y}$ (X_1 , X_2 , X_7) na podstawie wartości zmiennych zbioru danych z tabeli 11.1:

$$\mathbf{X} = \begin{bmatrix} 8621 & 331 & 3263 & 1 \\ 8666 & 348 & 3414 & 1 \\ \dots & \dots & \dots & \dots \\ 9605 & 584 & 4934 & 1 \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} 112,2 & 764 & 4215 \\ 131,1 & 839 & 4709 \\ \dots & \dots & \dots \\ 102,5 & 1070 & 10111 \end{bmatrix}$$

- ❖ Transponujemy macierz $\mathbf{X}$ korzystając z funkcji *Excels* – TRANSPONUJ(...) poprzedzonej znakiem równości,
- ❖ Szacujemy parametry postaci zredukowanej dla zmiennych X_3, X_4, X_9 korzystając ze wzoru:

$$\mathbf{P}^T = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{Y}$$

Dla obliczenia powyższego wektora używamy kolejno następujących funkcji *Excels*:

- MACIERZ.ILOCZYN(...;...) wstawiając zakresy dla macierzy $\mathbf{X}^T$ oraz $\mathbf{X}$,
- MACIERZ.ILOCZYN(...;...) wprowadzając zakres macierzy $\mathbf{X}^T$ i wektora $\mathbf{Y}$,
- MACIERZ.ODW(...) wpisując obszar macierzy uzyskanej z iloczynu $[\mathbf{X}^T \mathbf{X}]$,
- MACIERZ.ILOCZYN(...;...) wprowadzając adresy arkusza kalkulacyjnego dla macierzy odwrotnej $[\mathbf{X}^T \mathbf{X}]^{-1}$ oraz macierzy $[\mathbf{X}^T \mathbf{Y}]$.

Wynikiem powyższego działania uzyskujemy macierz $\mathbf{P}^T$, gdzie w kolumnach występują trzy równania modelu, a w wierszach parametry odpowiadające zmiennym X_3, X_4, X_9 .

$$\mathbf{P}^T = \begin{bmatrix} 0,026963 & 0,1105 & 1,76378 \\ -0,44326 & 0,27825 & 2,3231 \\ 0,032775 & 0,033545 & 2,0681 \\ -41,860 & -291,485 & -18380,367 \end{bmatrix}$$

$$\hat{x}_1 = 0,026963x_3 - 0,44326x_4 + 0,032775x_9 - 41,860$$

$$\hat{x}_2 = 0,1105x_3 - 0,27828x_4 + 0,033545x_9 - 291,485$$

$$\hat{x}_7 = 1,76378x_3 + 2,3231x_4 + 2,0681x_9 - 18380,367$$

Wartości z macierzy $\mathbf{P}^T$ możemy zapisać jako równania postaci zredukowanej modelu:

- ❖ Obliczamy wartości teoretyczne zmiennych endogenicznych X_1, X_2, X_7 dla wszystkich obserwacji korzystając ze wzoru $\mathbf{Y}^{\wedge} = \mathbf{X} \mathbf{P}^T$

Tab. 11.3. Wartości teoretyczne zmiennych endogenicznych

$X^{\wedge}_1$	$X^{\wedge}_2$	$X^{\wedge}_7$
150,8182	862,768	4342,299
149,4452	877,537	4773,444
146,3365	890,9582	5065,647
138,7214	918,5804	5623,007
132,0162	941,0697	6084,512
118,7833	958,5259	6493,01
113,562	978,1472	6865,641
120,1807	990,5593	7287,52
128,1632	1000,224	7613,893
138,1593	1003,138	7763,268
150,8333	1010,354	8112,529
157,1125	999,6023	7663,053
179,1301	1008,124	8179,297
199,5524	1020,041	8658,148
215,672	1034,834	9079,169
223,7149	1050,656	9324,71
230,6554	1061,01	9556,905
230,3410	19077,971	9817,749
228,1478	1095,108	10080,59

$X^{\wedge}_1$	$X^{\wedge}_2$	$X^{\wedge}_7$
216,8056	1104,992	10035,39
206,6182	1109,948	10177,43
189,7215	1082,612	9817,499
157,1360	1085,271	9810,903
119,9730	1097,967	10121,38

Źródło: Tabela 8.2 publikacji podanej w przypisie 2.

- ❖ Określenie macierzy X_i kolejno dla trzech równań postaci strukturalnej modelu zastępując dla określonego równania i zmiennej endogenicznej wartości empiryczne obserwacji wartościami teoretycznymi, obliczamy parametry równań postaci strukturalnej według *KMNC* bazując na wzorze:

$$b_i = [X_i^T X_i]^{-1} X_i^T y_i$$

przy czym wektorem y_i jest to szereg obserwacji y_i zmiennej objaśnianej.

W obliczeniach korzystamy z funkcji *Excelsa* w zakresie rachunku macierzowego otrzymując w rezultacie pierwsze równanie naszego 3-równaniowego modelu zmiennych łącznie współzależnych:

$$X_1 = \begin{bmatrix} 862,768 & 331 & 1 \\ 877,537 & 348 & 1 \\ \dots & \dots & \dots \\ 10917,967 & 584 & 1 \end{bmatrix} \quad \begin{aligned} \hat{x}_1 &= \hat{a}_1 x_2 + \hat{a}_2 x_4 + \hat{a}_0 \\ \hat{x}_1 &= 0,393x_2 - 0,507x_4 - 18,69 \end{aligned}$$

procedura określenia równania drugiego $\hat{x}_2 = b_1 x_3 + b_2 x_7 + b_0$ jest analogiczna jak pierwszego i uzyskujemy:

$$X_2 = \begin{bmatrix} 8621 & 4342,30 & 1 \\ 8666 & 4773,444 & 1 \\ \dots & \dots & \dots \\ 9605 & 10121,38 & 1 \end{bmatrix} \quad \begin{aligned} \hat{x}_2 &= \hat{b}_1 x_3 + \hat{b}_2 x_7 + \hat{b}_0 \\ \hat{x}_2 &= 0,00089x_3 + 0,039x_7 + 688,692 \end{aligned}$$

- ❖ określenie macierzy X_3 oraz oszacowanie parametrów dla równania trzeciego:

$$X_3 = \begin{bmatrix} 150,8182 & 3263 & 1 \\ 149,4452 & 3414 & 1 \\ \dots & \dots & \dots \\ 119,9730 & 4934 & 1 \end{bmatrix} \quad \hat{x}_7 = 5,770x_1 + 3,624x_9 - 8288,367$$

Powyższe trzy równania dla $\hat{x}_1$, $\hat{x}_2$, $\hat{x}_7$ z oszacowanymi parametrami stanowią określenie naszego modelu zmiennych łącznie współzależnych.

Innym sposobem estymacji parametrów równań zarówno postaci zredukowanej jak i strukturalnej jest zastosowanie funkcji *Excelsa* *REGLINP* o składni:

REGLINP(zakres wektora Y ; zakres macierzy X – bez X_0 ; PRAWDA; FAŁSZ)

Słowo PRAWDA oznacza podanie również stałej równania, a słowo FAŁSZ - nie generowanie statystyki opisowej.

Oczywiście możemy również skorzystać z programu *GRET* i opcji „Klasyczna metoda najmniejszych kwadratów”, którą szczegółowo przedstawiono na początku rozdziału. Program *GRET* dysponuje także funkcją obliczeń bezpośrednich dla podwójnej metody najmniejszych kwadratów. Procedura postępowania w tym zakresie przedstawiona zostanie w dalszej części na przykładzie

równań łącznie współzależnych. Warto jednak w tym miejscu wymienić modele jakimi dysponuje program *GRET* oczywiście poza klasyczną metodą najmniejszych kwadratów. Niżej wymienione można wywołać z menu *Model*:

- ❑ inne liniowe modele:
 - ważona metoda najmniejszych kwadratów (*UMNK*),
 - podwójna metoda najmniejszych kwadratów (*2MNK*),
 - *MNK* wysokiej precyzji,
- ❑ modele szeregów czasowych,
- ❑ modele panelowe,
- ❑ modele nieliniowe: logitowy, probitowy, tobitowy, Poissona, logistyczny, nieliniowa metoda najmniejszych kwadratów,
- ❑ odporne estymatory:
 - metoda najmniejszych wartości bezwzględnych,
 - korelacja rang,
- ❑ metoda największej wiarygodności,
- ❑ metoda momentów *GMM*,
- ❑ model równań współzależnych a w ramach niego w zależności od identyfikowalności równań następujące sposoby estymacji parametrów:
 - metoda pozornie niepowiązanych równań (*sur*),
 - potrójna metoda najmniejszych kwadratów (*3sls*),
 - metoda największej wiarygodności z pełną informacją (*fiml*),
 - metoda największej wiarygodności z ograniczoną informacją (*liml*),
 - klasyczna metoda najmniejszych kwadratów (*ols*),
 - podwójna metoda najmniejszych kwadratów (*tsls*),
 - ważona metoda najmniejszych kwadratów (*wls*).

Podstawy teoretyczne części z powyższych sposobów estymacji spotykamy w pracy Maddali wymienionej w przypisie 1., a także w publikacji Dziechciarza (przypis 2.) oraz w *Zarysie metod ekonometrii*¹²¹

11.8. Przykład zastosowania 2MNK w *GRET*

Bazowano na pliku danych rocznych (1962÷1985) GUS-u dla Polski dostępnym na stronie www.kufel.torun.pl. Plik obejmuje 24 obserwacje dotyczące przemysłu chemicznego scharakteryzowanego dziewięcioma zmiennymi z określeniem ich pól w strukturze zbioru danych:

- produkcja (produk);
- zatrudnienie (zatrud);
- inwestycje (inwest);
- zmienna czasowa *t* (*time1*);
- zmienna czasowa t^2 (*time2*) stanowiąca kwadrat zmiennej czasowej "*t*";
- zmienna czasowa t^3 (*time3*);
- *produk_1*=produk(-1), przesunięcie szeregu danych wielkości produkcji o 1 okres (rok);
- *zatrud_1*=zatrud(-1);
- *inwest_1*=inwest(-1).

¹²¹ Nowak E., *Zarys metod ekonometrii*, Wydawnictwo Naukowe PWN, Warszawa 1990.

Tab. 11.4. Wartości zmiennych dotyczących przemysłu chemicznego

Obserw. n	produkt P_t	zatrud Z_t	inwest I_t	time1 t	time2 t^2	time3 t^3	produkt1=produkt(-1) P_{t-1}	zatrud1=zatrud(-1) Z_{t-1}	inwest1=inwest(-1) I_{t-1}
1	176	301	52	1	1	1	163	280	51
2	185	317	52	2	4	8	176	301	52
3	188	342	53	3	9	27	185	317	52
4	198	356	53	4	16	64	188	342	53
5	199	373	56	5	25	125	198	356	53
6	201	381	59	6	36	216	199	373	56
7	205	407	62	7	49	343	201	381	59
8	211	432	66	8	64	512	205	407	62
9	216	436	70	9	81	729	211	432	66
10	214	438	70	10	100	1000	216	436	70
11	215	440	72	11	121	1331	214	438	70
12	220	444	73	12	144	1728	215	440	72
13	223	460	71	13	169	2197	220	444	73
14	224	480	65	14	196	2744	223	460	71
15	223	475	63	15	225	3375	224	480	65
16	224	483	57	16	256	4096	223	475	63
17	226	510	59	17	289	4913	224	483	57
18	225	501	60	18	324	5832	226	510	59
19	224	514	61	19	361	6859	225	501	60
20	226	520	61	20	400	8000	224	514	61
21	227	524	62	21	441	9261	226	520	61
22	230	538	65	22	484	10648	227	524	62
23	232	543	64	23	529	12167	230	538	65
24	235	560	63	24	576	13824	232	543	64

Źródło: Opracowanie własne bazujące na danych z w/w strony www.

Wartości danych z *GRET*L-a po edycji w *Excelu* podaje tabela 11.4. Ponieważ występują przesunięcia czasowe dla produkcji, zatrudnienia oraz inwestycji, dlatego dla tych informacji wprowadzono przykładowe umowne wartości dla obserwacji 1. W tym względzie skorzystano z ścieżki poprawy zbioru danych:

Edycja/Zastosuj/Zamknij.

Wydruk zbioru danych można otrzymać bezpośrednio z *GRET*L-a w postaci tekstowej poprzez wywołanie menu:

Dane/Zaznacz wszystko (wszystkie pola); Dane/Pokaż wartości/Drukuj.

Realizacja komputerowa w *GRET*L bazuje na modelu wielorównaniowym o równaniach łącznie współzależnych wykorzystujących¹²² niektóre ze zmiennych tabeli 11.4.

$$P_t = \beta_{12}Z_t + \alpha_{10} + \alpha_{11}P_{t-1} + \alpha_{13}I_{t-1} + \varepsilon_{1t}$$

$$Z_t = \beta_{21}P_t + \alpha_{20} + \alpha_{22}I_t + \alpha_{23}I_{t-1} + \varepsilon_{2t}$$

P_t i Z_t to zmienne endogeniczne; P_{t-1} , I_t , I_{t-1} są zmiennymi egzogenicznymi; ε_{1t} oraz ε_{2t} stanowią reszty poszczególnych równań. Po przekształceniu powyższy model dwurównaniowy **BY+CX** = ε można zapisać jako:

$$\begin{bmatrix} 1 & -\beta_{12} \\ -\beta_{21} & 1 \end{bmatrix} \begin{bmatrix} P_t \\ Z_t \end{bmatrix} + \begin{bmatrix} -\alpha_{10} & -\alpha_{11} & 0 & -\alpha_{13} & 0 \\ -\alpha_{20} & 0 & -\alpha_{22} & -\alpha_{23} & -\alpha_{24} \end{bmatrix} \begin{bmatrix} P_{t-1} \\ I_t \\ I_{t-1} \\ t \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \end{bmatrix}$$

Mając równanie w modelu równań łącznie współzależnych (chcąc dobrać odpowiednią metodę estymacji parametrów α przy poszczególnych zmiennych egzogenicznych) musimy wpięrować do końca sprawdzenia warunku identyfikowalności równań $G = 2$ wg macierzy **C**:

¹²² Publikacja jak w przypisie 3 rozdział 11.1.

P_t

- dwie zmienne nie występujące: I_t, t (ich parametry w powyższej macierzy C mają wartość zero);
- minimalna liczba powyższych zmiennych: $G - 1 = 2 - 1 = 1$.

Zatem jest to równanie identyfikowalne niejednoznacznie¹²³.

Z_t

- jedna zmienna P_{t-1} o zerowym parametrze.

Występuje równość ilości zmiennych i minimalnej ilości, tak więc równanie drugie jest jednoznacznie identyfikowalne. W tej sytuacji dla obu równań można zastosować podwójną metodę najmniejszych kwadratów (2MNK) o procedurze realizacji:

- określenie zmiennych współzależnych występujących również jako zmienne objaśniające (P_t, Z_t);

- zastąpienie powyższych zmiennych wartościami teoretycznymi:

$$\hat{P}_t; \hat{Z}_t$$

- szacowanie MNK parametrów postaci zredukowanej w skład której wchodzi cztery zmienne objaśniające (P_{t-1}, I_t, I_{t-1}, t) oraz stała *const*, czyli zmienne z góry ustalone.

W odniesieniu do dwóch naszych równań postać zredukowana jest następująca:

$$\hat{P}_t = \pi_{10} + \pi_{11}P_{t-1} + \pi_{12}I_t + \pi_{13}I_{t-1} + \pi_{14}t$$

$$\hat{Z}_t = \pi_{20} + \pi_{21}P_{t-1} + \pi_{22}I_t + \pi_{23}I_{t-1} + \pi_{24}t$$

Wymienione zmienne oraz stałe są podstawą określenia wartości teoretycznych zmiennych endogenicznych P_t, Z_t , a więc posłużą nam w kolejnej estymacji do określenia postaci strukturalnej rozpatrywanego modelu. Tyle odniesień teoretycznych, a teraz zastosowanie metody 2MNK w programie GRETL do naszego modelu dwóch równań współzależnych co można przeprowadzić dwoma sposobami:

- a) osobno każde równanie,
- b) jednocześnie system równań.

Bazując na opracowaniu T. Kufla¹²⁴ przedstawione zostaną na fragmentach wydruków komputerowych oba sposoby przetestowane przez autora niniejszego opracowania.

Ad a).

Wybieramy ścieżkę z menu:

Model/Inne liniowe modele/Podwójna metoda najmniejszych kwadratów (2MNK...)

- Specyfikacja pojedynczego równania dla "produkt":

- zmienna zależna - *produkt*,
- zmienne niezależne (X): domyślnie *const, zatrud, produkt1, inwest1*,
- instrumenty (zmienne objaśniające postaci zredukowanej): *const, produkt1, inwest, inwest1, time1*,

- OK (wyświetlane są wyniki estymacji dla zmiennej zależnej - endogenicznej "produkt" postaci zredukowanej pod określeniem "Model 1"). Zakres informacji zamieszczono w tabeli 10.5 oraz zamieszczoną pod nią fragment statystyki opisowej.

¹²³ Ibidem, s. 151.

¹²⁴ Kufel T., *Ekometria. Rozwiązywanie problemów*, op. cit.

Tab. 11.5. Parametry oraz statystyka opisowa *Modelu 1*

Zmienna	Współczynnik (parametr)	Błąd standardowy	Statystyka <i>t</i>
const	64,9576	9,07634	7,157
zatrud	0,089998	0,0230224	3,909
produkt	0,452118	0,10279	4,398
inwest	0,219223	0,0848503	2,584

Średnia arytmetyczna zmiennej zależnej - 214,458

Odchylenie standardowe zmiennej zależnej - 15,8772

Suma kwadratów reszt - 62,2475

Błąd standardowy reszt - 1,76419

Współczynnik determinacji R^2 - 0,989264

Skorygowany współczynnik determinacji - 0,987654.

Źródło: opracowanie własne w *GRET*L.

□ Specyfikacja równania dla "zatrud":

- zmienna zależna - *zatrud*,

- zmienne niezależne: *const*, *produkt*, *inwest*, *inwest1*, *time1*,

- instrumenty: *const*, *produkt1*, *inwest*, *inwest1*, *time1*,

- OK (pojawia się oszacowanie parametrów oraz statystyka opisowa pod nazwą "Model 2" (patrz tabela 10.6).

Tab. 11.6. Parametry oraz statystyka opisowa *Modelu 2*

Zmienna	Współczynnik (parametr)	Błąd standardowy	Statystyka <i>t</i>
const	-147,971	67,5382	-2,191
produkt	2,62417	0,451839	5,808
inwest	1,07858	0,596393	1,809
inwest1	-1,56254	0,610357	-2,560
time1	5,07182	0,8482234	5,979

Średnia arytmetyczna zmiennej zależnej - 448,958

Odchylenie standardowe zmiennej zależnej - 73,8927

Suma kwadratów reszt - 886,435

Błąd standardowy reszt - 6,83041

Współczynnik determinacji R^2 - 0,992943

Skorygowany współczynnik determinacji - 0,991457.

Źródło: Opracowanie własne w *GRET*L.

Po wywołaniu *Plik/Drukuj/OK* otrzymujemy wydruk dla *Modelu 1* oraz *Modelu 2*. Program *GRET*L umożliwia także sporządzenie wykresów rozrzutu danych z obserwacji przykładowo dla wskazanych kolumn bazy danych tj. *produkt*, *zatrud*, *inwest* w zależności od okresów czasu "*time1*". Mając oszacowane parametry równań dla zmiennych zależnych *produkt*, czy też *zatrud* możemy w *GRET*L sporządzić wykresy:

- empirycznej i wyrównywanej wartości danej zmiennej,

- reszty regresji dla zmiennej *produkt* oraz *zatrud*.

Gdy skorzystamy ze ścieżki: *Narzędzia/Pokaż dziennik poleceń* to program *GRET*L udostępni nam wykaz zarejestrowanych komend języka skryptowego tworzonych podczas pracy z menu, przykładowo:

open "C:\Documents and Settings\Dane aplikacji\GRET\PMNK.gdt"

```

gnuplot produk zatrud inwest time1
tsls produk const zatrud produk1 inwest1; const produk1 inwest\inwest1 time1
tsls zatrud const produk inwest inwest1 time1; const produk1 inwest\inwest1 time1

```

Przy czym symbol *tsls* oznacza metodę *2MNK*, a składnia tego polecenia zawiera:

- zmienną endogeniczną,
- stałą oraz zmienne objaśniające danego równania,
- po średniku (ustalone z góry zmienne modelu).

Program *GRET* umożliwia wybór podzbioru poleceń (komend) dotyczących określonego tematu np. estymacji, a w ramach niego opis powyższej komendy *tsls*.

Ad b)

Dla estymacji systemu równań, wywołujemy menu:

Model/Model równań współzależnych.

W specyfikacji wprowadzamy trzy polecenia:

equation - wskazanie zmiennych równania (najpierw zmiennych endogenicznych); powtarzamy to polecenie dla każdego równania,
endog - wskazanie nazw zmiennych endogenicznych,
instr - nazwy zmiennych z góry ustalonych (instrumentów) tworzących postać zredukowaną modelu.

Dla naszego przykładu specyfikacja poleceń jest następująca:

```

equation produk const zatrud produk1 inwest1
equation zatrud const produk inwest inwest1 time1
endog produk zatrud
instr produk1 inwest inwest1 time1

```

Okno specyfikacji modelu równań współzależnych pozwala wybrać metodę estymacji spośród siedmiu już wcześniej wymienionych, a wśród nich *tsls*. Wybór metody zależy od wcześniej przeprowadzonej identyfikacji równań. Dla naszego przykładu jak już wcześniej wykazano odpowiednia jest metoda *tsls*. Przy realizacji estymacji systemu równań *GRET* korzysta z pełnej bazy danych tj. 24 obserwacji i umożliwia wydrukowanie pełnego zestawu parametrów oraz statystyki opisowej dla wszystkich równań jednocześnie. Po opanowaniu składni poleceń - komend możemy dokonać estymacji modelu pisząc instrukcję dla programu i nadając tej instrukcji nazwę przykładowo "*Program.int*". Polecenia języka skryptowego tworzymy w oknie wywołanym menu:

Plik/Pliki poleceń skryptowych/Nowy plik skryptowy.

System sygnalizuje błędy w składni komend. Po naciśnięciu ikony "*Wykonaj*" następuje realizacja, przy czym program w naszym przykładzie pobiera obserwacje 2÷24 pomijając pierwszą, gdyż kieruje się przesunięciem jednookresowym zmiennych *produkt1*, *inwest1*. Zaprezentowanie wyników estymacji parametrów jest analogiczne do otrzymywanych z systemu równań.

Zaprezentowano tylko element oprogramowania *GRET* w zastosowaniu do modeli wielorównaniowych. Szerszą panoramę jego możliwości spotykamy w pracy Tadeusza Kufla "Ekonometria Rozwiązywanie problemów z wykorzystaniem programu *GRET*". Opanowanie tego narzędzia wymaga jednak bezpośrednich testowań na komputerze poszczególnych opcji rozbudowanego menu.

Wykonana przez autora niniejszej pracy próba prześledzenia procedur w kontekście założeń teoretycznych wynikających z literatury przedmiotu dla metody *2MNK*, może być zachętą do szerszego poznania dalszych metod estymacji z zastosowaniem tego programu.

12. PRZYKŁADY WPŁYWU NAKŁADÓW INWESTYCYJNYCH



12.1. Wybór zmiennych do określenia regresji jednoczynnikowych

Obliczono moduły otrzymanych z *GRET*L-a współczynników korelacji, a następnie posortowano malejąco oraz utworzono grupy przedziałowe wartości tych współczynników. Dane zaczerpnięto z publikacji, *Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*¹²⁵, gdzie:

X_1 - dynamika nakładów inwestycyjnych (ceny stałe),

Y_n - dynamiki zmiennych źródłowych,

Z_m - dynamiki zmiennych wynikowych.

Tab. 12.1. Przedziały współczynników korelacji względem zmiennej X_1

Z	LX11	Z	LX11	Z	LX11	Z	LX11	Z	LX11
Y42	1,0000	Y57	0,8994	Y147	0,7925	Y127	0,6818	Y30	0,5904
Y75	1,0000	Y74	0,8987	Y148	0,7904	Y78	0,6770	Y132	0,5775
Y76	1,0000	Z17	0,8962	Y149	0,7880	Y104	0,6682	Y94	0,5646
Y77	1,0000	Y87	0,8955	Y65	0,7875	Z2	0,6606	Y48	0,5167
Y40	0,9979	Y14	0,8947	Y63	0,7812	Y11	0,6542	4	
Y8	0,9952	Y5	0,8944	Z3	0,7636	Z6	0,6087		
Y80	0,9859	Y20	0,8942	Z18	0,7583	Y93	0,6028		
Y19	0,9858	Y37	0,8925	Z4	0,7528	7			
Y129	0,9791	Y101	0,8915	Y32	0,7487				
Y47	0,9790	Y6	0,8905	Z11	0,7454				
Y7	0,9782	Y69	0,8901	Z14	0,7438				
Y113	0,9781	Y137	0,8898	Z12	0,7337				
Y96	0,9779	Y38	0,8877	Y55	0,7315				
Y98	0,9770	Y67	0,8874	Y28	0,7295				
Y39	0,9768	Y24	0,8873	Z7	0,7217				
Y141	0,9742	Y70	0,8865	Y159	0,7131				
Y18	0,9734	Y135	0,8855	16					
Y71	0,9670	Y89	0,8847						
Z1	0,9662	Y86	0,8844						
Z10	0,9658	Y34	0,8841						
Y140	0,9640	Y16	0,8839						
Y100	0,9617	Y152	0,8817						
Y59	0,9597	Y136	0,8804						
Y116	0,9561	Y125	0,8795						
Y79	0,9503	Z13	0,8775						
Y133	0,9498	Y53	0,8758						

¹²⁵ Wornalkiewicz W., *Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*, tabela 3.1., Skrypty Wyższej Szkoły Zarządzania i Administracji w Opolu, Opole 2013.

Z	LX11	Z	LX11	Z	LX11	Z	LX11	Z	LX11
Y15	0,9482	Y154	0,8744						
Y58	0,9460	Y153	0,8736						
Y134	0,9441	Y130	0,8729						
Y155	0,9406	Y138	0,8668						
Y156	0,9399	Y2	0,8638						
Y81	0,9388	Y84	0,8634						
Y4	0,9376	Y83	0,8631						
Y36	0,9362	Y109	0,8607						
Y22	0,9328	Y43	0,8604						
Y72	0,9315	Y118	0,8563						
Y41	0,9297	Y62	0,8555						
Y115	0,9229	Y51	0,8547						
Y56	0,9213	Y158	0,8546						
Y126	0,9154	Y157	0,8541						
Y108	0,9133	Y139	0,8536						
Y111	0,9113	Y66	0,8506						
Y110	0,9109	Z16	0,8472						
Y3	0,9093	Y82	0,8471						
Y102	0,9087	Y33	0,8460						
Y112	0,9069	Y52	0,8444						
Y35	0,9055	Y117	0,8439						
Y99	0,9041	Y54	0,8420						
Y128	0,9002	Y50	0,8389						
49		Y88	0,8377						
		Y68	0,8352						
		Y131	0,8341						
		Z9	0,8315						
		Y25	0,8298						
		Y106	0,8266						
		Y151	0,8247						
		Y97	0,8229						
		Y61	0,8224						
		Y105	0,8175						
		Y150	0,8075						
		Y29	0,8063						
		Y64	0,8051						
		Y31	0,8023	63					

Źródło: Opracowanie własne w Excelu.

Z tabeli 12.1 wynika, że ilości zmiennych w poszczególnych przedziałach są następujące:

1-0,9	0,9-0,8	0,8-0,7	0,7-0,6	0,6-0,5
49	63	16	7	4

Dla zbadania regresji jednoczynnikowych między zmienną X_1 a pozostałymi zmiennymi zajmijmy się przykładowo przedziałem 1-0,9. Przypomnijmy sobie jednak te zmienne, które za-czerpnięto z tabeli 3.1¹²⁶.

Tab. 12.2. Zmienne do badania regresji jednoczynnikowych względem X_1

Poz.	Symbol	Zmienna	Nazwa
4	dPKB	Y3	Dynamika produktu krajowego brutto (PKB)
5	dPK	Y4	Dynamika popytu krajowego (ceny stałe)
8	dPKA	Y7	Dynamika akumulacji
9	dPKAS	Y8	Dynamika nakładów brutto na środki trwałe (w akumulacji)
16	dIP	Y15	Dynamika nakładów inwestycyjnych w przemyśle (ceny stałe)
19	dPRB	Y18	Dynamika produkcji budowlano-montażowej
20	dIB	Y19	Dynamika inwestycji w budownictwie
36	dPKBM	Y35	Dynamika produktu krajowego brutto na 1. mieszkańca
37	dPKM	Y36	Dynamika popytu krajowego na 1. mieszkańca
40	dAKM	Y39	Dynamika akumulacji na 1. mieszkańca
41	dAKSM	Y40	Dynamika udziału akumulacji na środki trwałe na 1. mieszkańca
42	dSHM	Y41	Dynamika sprzedaży detalicznej na 1. mieszkańca
43	dIKM	Y42	Dynamika nakładów inwestycyjnych na 1. mieszkańca
48	dPZH	Y47	Dynamika przeciętnego zatrudnienia w handlu
57	dDGO	Y56	Dynamika dochodu na 1. gospodarstwo domowe
59	dNBO	Y58	Dynamika nakładów na działalność badawczą
60	dNBS	Y59	Dynamika nakładów inwestycyjnych w działalności badawczej na maszyny
72	dZPB	Y71	Dynamika zobowiązań przedsiębiorstw budowlanych
73	dZPH	Y72	Dynamika zobowiązań przedsiębiorstw handlu
76	dNIO95	Y75	Dynamika nakładów inwestycyjnych ogółem
77	dNIM95	Y76	Dynamika nakładów inwestycyjnych na 1. mieszkańca
78	dNIS95	Y77	Dynamika nakładów inwestycyjnych na środki trwałe
80	dNIP	Y79	Dynamika nakładów inwestycyjnych w przemyśle
81	dNIB	Y80	Dynamika nakładów inwestycyjnych w budownictwie
82	dNIH	Y81	Dynamika nakładów inwestycyjnych w handlu
97	dSIO	Y96	Dynamika wartości środków trwałych uzyskanych z dział. inw. ogółem
99	dSIP	Y98	Dynamika wartości środków trwałych uzyskanych z dział. inw. w przemyśle
100	dSIB	Y99	Dynamika wartości środków trwałych uzyskanych z dział. inw. w budownictwie
101	dSIH	Y100	Dynamika wartości środków trwałych uzyskanych z dział. inw. w handlu
103	dSLO	Y102	Dynamika wartości zlikwidowanych środków trwałych ogółem
109	dIT95	Y108	Dynamika rachunku produktów – import towarów i usług
111	dPKB95	Y110	Dynamika rachunku produktów – produkt krajowy brutto
112	dPKBM	Y111	Dynamika rachunku produktów – PKB na 1. mieszkańca
113	dSM95	Y112	Dynamika rachunku produktów – spożycie na 1. mieszkańca
114	dAM95	Y113	Dynamika rachunku produktów – akumulacja na 1. mieszkańca
116	dPKBP95	Y115	Dynamika produktu krajowego brutto w przemyśle

¹²⁶ Ibidem.

Poz.	Symbol	Zmienna	Nazwa
117	dPKBB95	Y116	Dynamika produktu krajowego brutto w budownictwie
127	dPKBZ1	Y126	Dynamika kosztów związanych z zatrudnieniem (ceny bieżące)
129	dPKBDP	Y128	Dynamika wartości dodanej brutto w przemyśle
130	dPKBDB	Y129	Dynamika wartości dodanej brutto w budownictwie
134	dPKBDZP	Y133	Dynamika wartości dodanej brutto (koszty zatrudnienia) w przemyśle
135	dPKBDZB	Y134	Dynamika wartości dodanej brutto (koszty zatrudnienia) w budownictwie
141	dPKA	Y140	Dynamika popytu krajowego akumulacja (ceny bieżące)
142	dPKAT	Y141	Dynamika PK akumulacja – nakłady brutto na środki trwałe (ceny bieżące)
156	dDGR95	Y155	Dynamika dochodów realnych ogółem w gospodarstwach domowych
157	dDG-DRM95	Y156	Dynamika dochodów realnych na 1. mieszkańca w gospodarstwach dom.
1	dwiK	Z1	Dynamika wydajności inwestycji
10	dwiP	Z10	Dynamika wydajności inwestycji w przemyśle

Źródło: Opracowanie własne.

Do prezentacji graficznej wybrano intuicyjnie 16 zmiennych, które zaznaczono grubą czcionką, a są nimi zmienne:

Y3, Y4, Y8, Y18, Y36, Y41, Y47, Y56, Y58, Y96, Y108, Y112, Y126, Y155, Z1, Z10.

Teraz dokonajmy importu danych do programu *GRET* korzystając ze zbioru własnego *DANE4.xls*. Utworzony zostaje przez *GRET* zbiór *Dane4.gdt*. Fragment tego zbioru dla okresów 1-15 zawierającego zmienne X1, Y3-Z10 zamieszczono w tabeli 12.3.

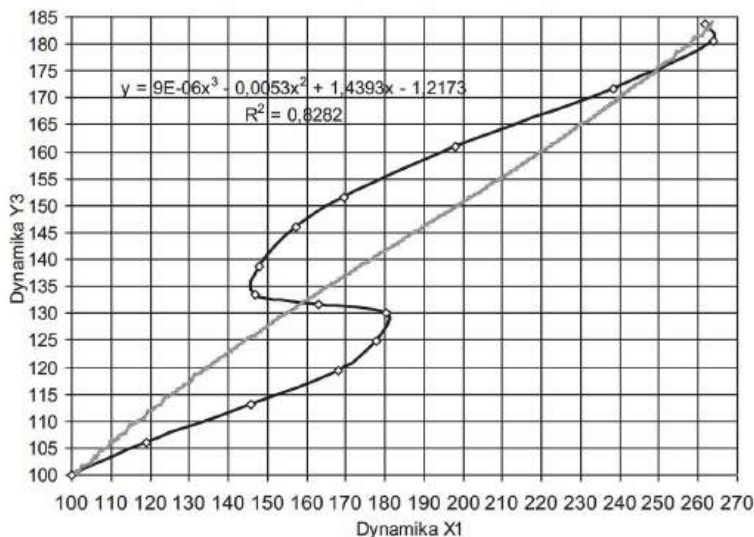
Tab. 12.3. Obserwacje zmiennych X1 oraz Y3-Z10

Obs.	X1	Y3	Y4	Y8	Y18	Y36	Y41	Y47	Y56	Y58	Y96	Y108	Y112	Y126	Y155	Z1	Z10
1	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0	100,0
2	119,2	106,0	109,7	119,7	103,0	107,8	103,4	103,8	127,3	129,5	129,5	128,0	107,1	128,2	104,8	117,0	120,8
3	145,6	113,2	119,4	135,6	120,0	115,2	110,5	115,4	157,0	157,6	182,6	155,4	113,5	159,0	112,2	138,9	137,3
4	167,9	119,4	127,4	166,2	134,9	121,3	113,5	122,5	166,2	187,8	219,5	184,1	118,2	186,7	117,4	156,6	154,2
5	177,9	124,8	134,1	177,1	143,2	129,2	119,0	122,2	181,5	215,2	250,4	186,0	123,5	206,1	119,7	170,6	158,6
6	180,4	130,2	138,2	181,9	144,7	133,2	120,2	122,8	199,4	224,9	269,5	215,0	126,6	223,3	121,4	177,0	148,6
7	163,2	131,7	136,4	164,2	135,4	136,2	120,4	120,1	210,0	227,8	252,6	203,5	128,8	238,5	123,0	161,1	148,1
8	146,9	133,6	137,8	153,9	135,1	141,0	122,7	119,7	216,2	212,0	232,0	208,8	132,4	236,0	124,0	170,0	143,7
9	147,8	138,8	141,7	153,8	136,2	144,1	127,3	121,3	211,0	213,7	237,5	228,2	135,6	240,1	124,6	172,1	154,9
10	157,4	146,1	150,4	163,6	126,7	151,0	130,5	120,1	217,9	241,7	264,7	263,1	143,7	249,4	124,0	183,9	168,3
11	169,5	151,4	154,2	174,2	128,6	154,2	127,4	127,9	225,6	261,4	280,9	275,5	147,5	262,6	129,0	196,2	179,4
12	198,0	160,9	165,5	200,2	146,3	162,1	136,7	131,3	247,1	276,3	320,0	323,4	156,4	280,8	134,1	226,5	203,0
13	238,4	171,6	179,9	235,4	171,7	170,2	147,2	139,9	275,0	312,9	398,6	370,4	164,8	311,8	140,1	263,4	234,5
14	263,9	180,6	190,0	258,0	188,2	179,9	154,5	151,4	309,6	361,3	455,4	400,5	175,2	355,0	147,0	280,1	247,2
15	261,8	183,7	188,1	255,2	199,3	183,5	157,0	149,4	330,0	425,3	433,3	350,8	178,2	367,7	152,5	280,3	244,5

Źródło: Opracowanie własne.

12.2. Przykłady prezentacji graficznej wpływu nakładów inwestycyjnych na wybrane zmienne

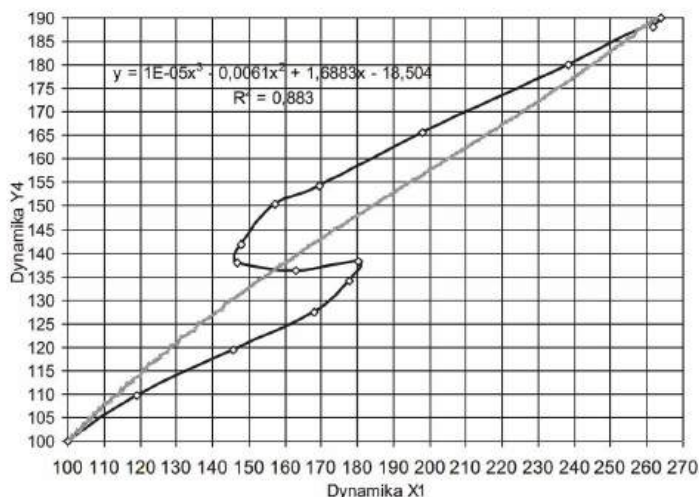
W zaprezentowanych wykresach wpływu zmiennej $X1$ od innych wybranych zmiennych przytoczono fragmenty nazw zmiennych ich pozycję oraz symbol występujące w tabelach 12.2 i 12.3.



1	dIK	$X1$	Dynamika nakładów inwestycyjnych (ceny stałe)
4	dPKB	$Y3$	Dynamika produktu krajowego brutto(PKB)

Źródło: Opracowanie własne w Excelu.

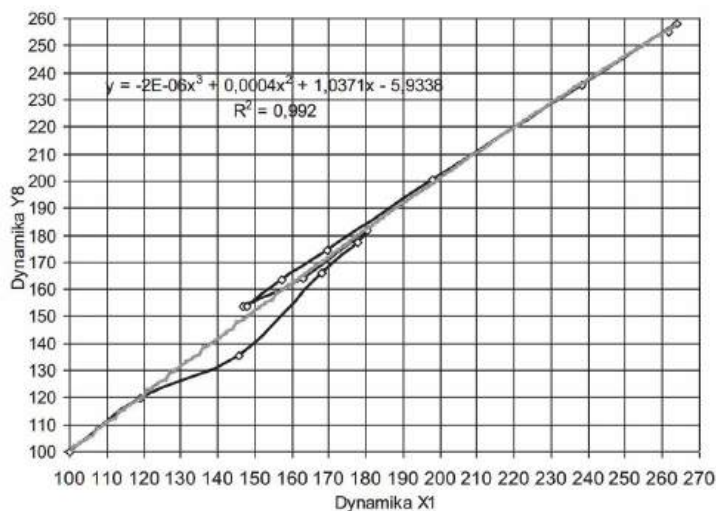
Rys. 12.1. Wpływ zmiennej $X1$ na zmienną $Y3$



1	dIK	$X1$	Dynamika nakładów inwestycyjnych (ceny stałe)
5	dPK	$Y4$	Dynamika popytu krajowego (ceny stałe)

Źródło: Opracowanie własne w Excelu.

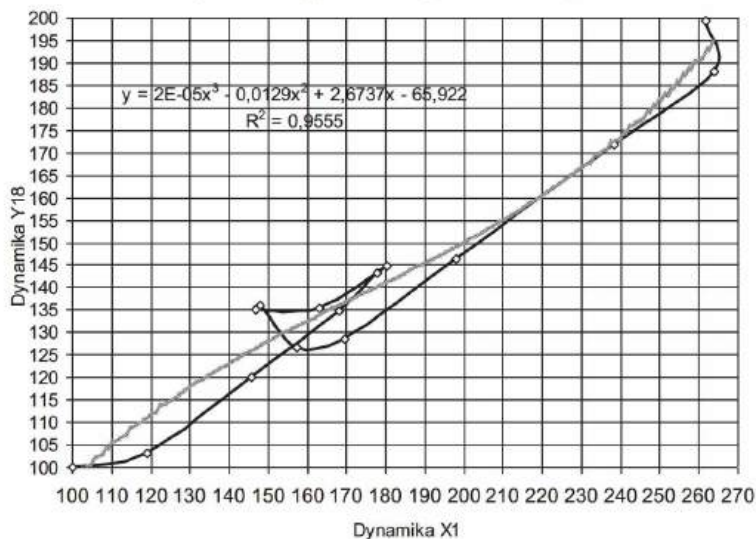
Rys. 12.2. Wpływ zmiennej $X1$ na zmienną $Y4$



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
9	dPKAS	Y8	Dynamika nakładów brutto na środki trwałe (w akumulacji)

Źródło: Opracowanie własne w Excelu.

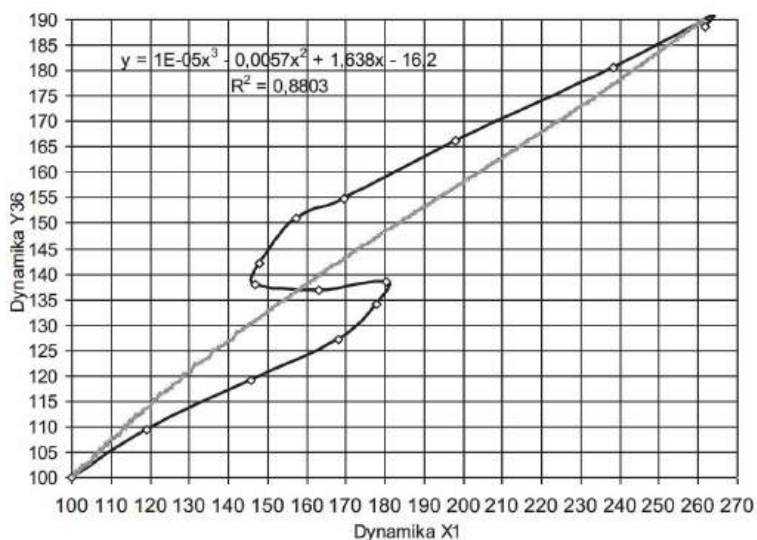
Rys. 12.3. Wpływ zmiennej X1 na zmienną Y8



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
19	dPRB	Y18	Dynamika produkcji budowlano-montażowej

Źródło: Opracowanie własne w Excelu.

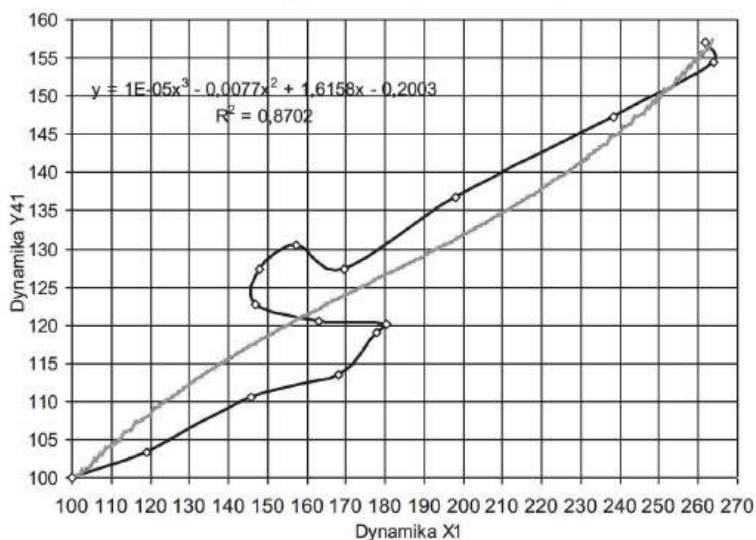
Rys. 12.4. Wpływ zmiennej X1 na zmienną Y18



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
37	dPKM	Y36	Dynamika popytu krajowego na 1. mieszkańca

Źródło: Opracowanie własne w Excelu.

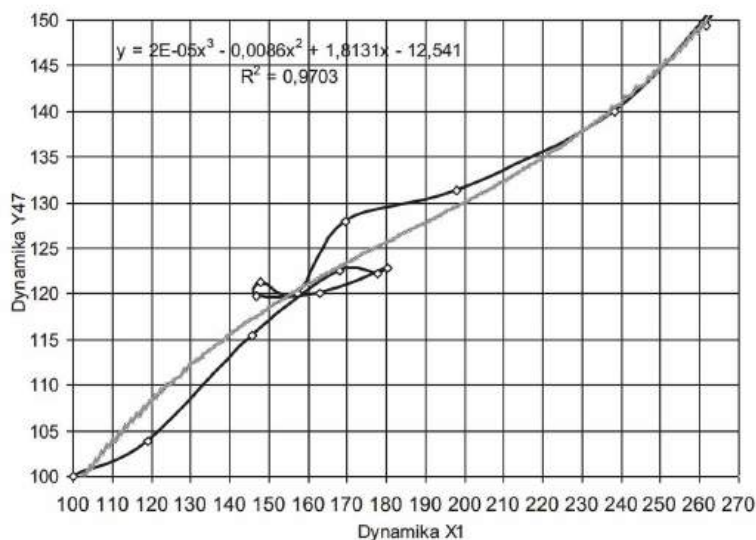
Rys. 12.5. Wpływ zmiennej X1 na zmienną Y36



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
42	dSHM	Y41	Dynamika sprzedaży detalicznej na 1. mieszkańca

Źródło: Opracowanie własne w Excelu.

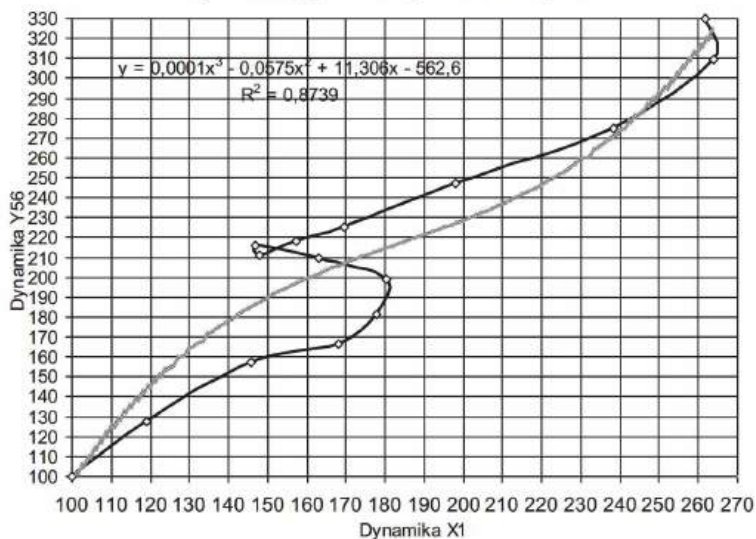
Rys. 12.6. Wpływ zmiennej X1 na zmienną Y41



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
48	dPZH	Y47	Dynamika przeciętnego zatrudnienia w handlu

Źródło: Opracowanie własne w Excelu.

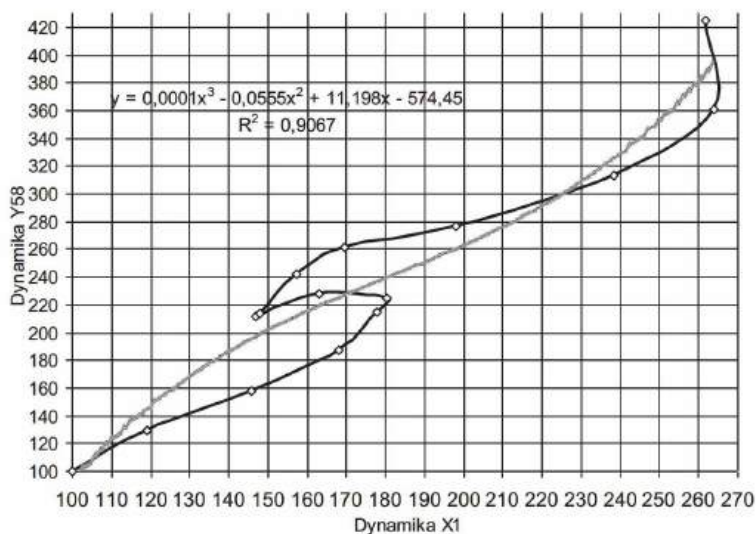
Rys. 12.7. Wpływ zmiennej X1 na zmienną Y47



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
57	dDGO	Y56	Dynamika dochodu na 1. gospodarstwo domowe

Źródło: Opracowanie własne w Excelu.

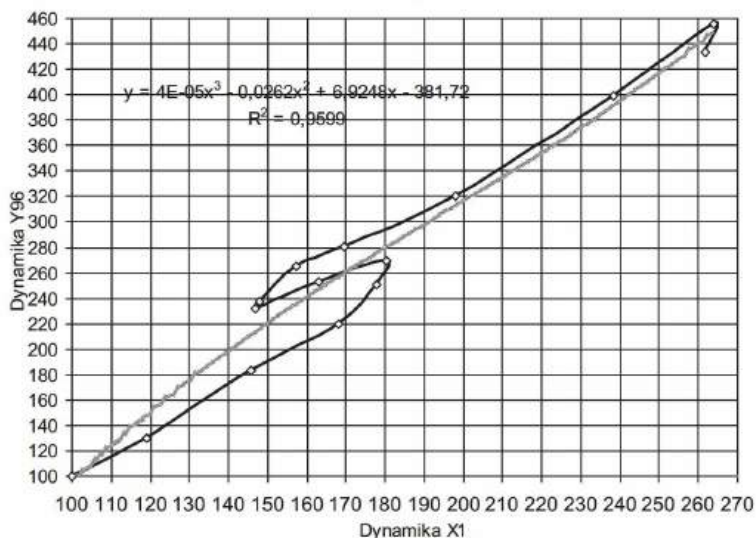
Rys. 12.8. Wpływ zmiennej X1 na zmienną Y56



I	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
59	dNBO	Y58	Dynamika nakładów na działalność badawczą

Źródło: Opracowanie własne w Excelu.

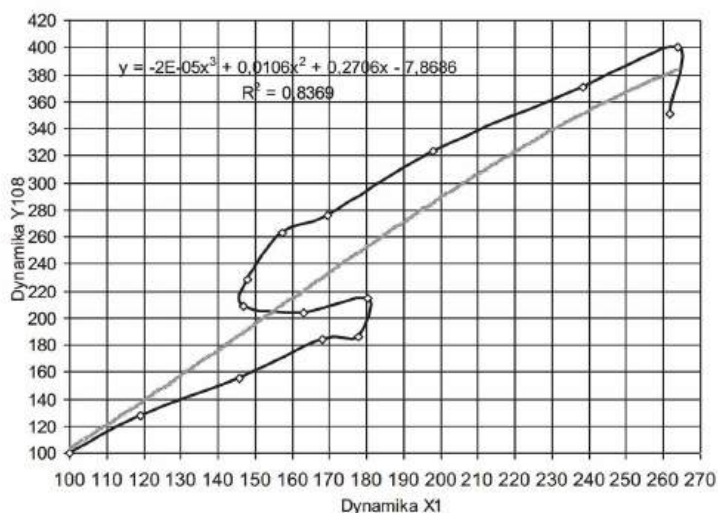
Rys. 12.9. Wpływ zmiennej X1 na zmienną Y58



I	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
97	dSIO	Y96	Dynamika wartości środków trwałych uzyskanych z dział. inv. ogółem

Źródło: Opracowanie własne w Excelu.

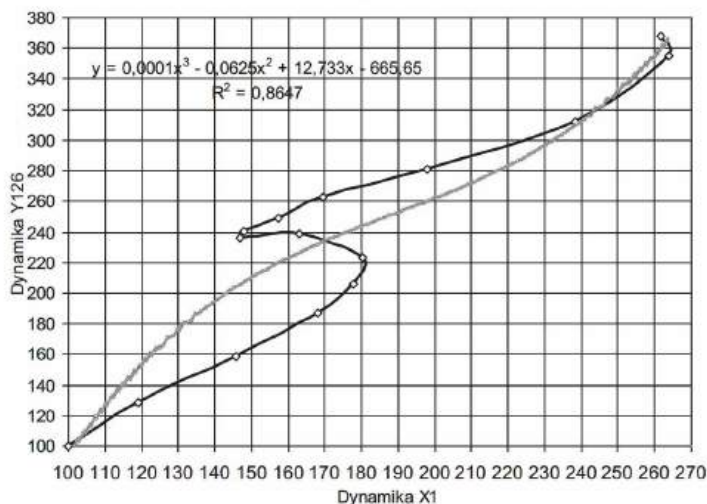
Rys. 12.10. Wpływ zmiennej X1 na zmienną Y96



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
97	dIT95	Y108	Dynamika rachunku produktów – import towarów i usług

Źródło: Opracowanie własne w Excelu.

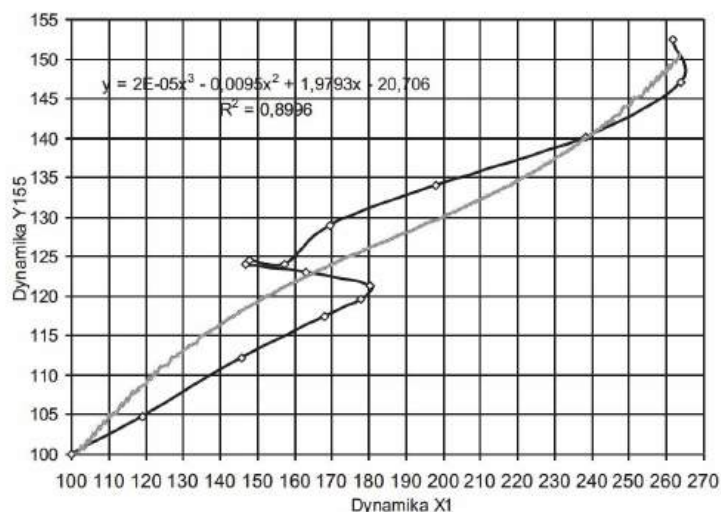
Rys. 12.11. Wpływ zmiennej X1 na zmienną Y108



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
127	dPKBZ1	Y126	Dynamika kosztów związanych z zatrudnieniem (ceny bieżące)

Źródło: Opracowanie własne w Excelu.

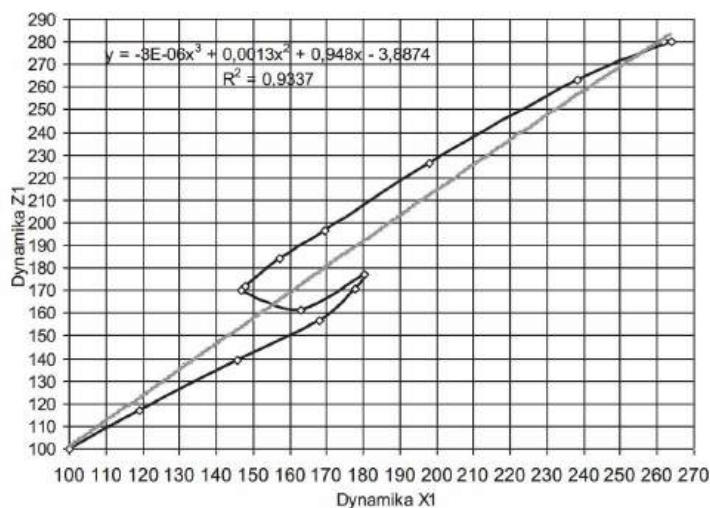
Rys. 12.13. Wpływ zmiennej X1 na zmienną Y126



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
156	dDGDR95	Y155	Dynamika dochodów realnych ogółem w gospodarstwach domowych

Źródło: Opracowanie własne w Excelu.

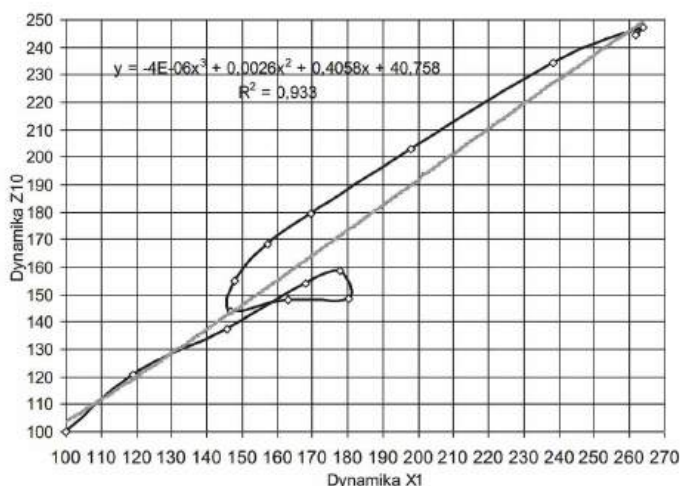
Rys. 12.14. Wpływ zmiennej X1 na zmienną Y155



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
1	dwiK	Z1	Dynamika wydajności inwestycji

Źródło: Opracowanie własne w Excelu.

Rys. 12.15. Wpływ zmiennej X1 na zmienną Z1



1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
10	dwiP	Z10	Dynamika wydajności inwestycji w przemyśle

Źródło: Opracowanie własne w Excelu.

Rys. 12.16. Wpływ zmiennej $X1$ na zmienną $Z10$

12.3. Propozycja modelu ekonometrycznego produktu krajowego brutto

Dotychczasowe rozważania doprowadziły nas do pokazania istotnych wpływów inwestycji na wybrane cechy statystyczne w oparciu o zmienną $X1$. Utwórzmy teraz na podstawie pliku *DANE4.xls* wariant bazy danych w *GRET*L o nazwie *Dane5.gdt* zawierającym zmienne wyszczególnione w tabeli 12.2. Załóżmy, że naszą zmienną objaśnianą jest:

4 | dPKB | Y3 | Dynamika produktu krajowego brutto (PKB)

a zmiennymi objaśniającymi $X1$ oraz pozostałe zmienne z tabeli 12.2. Posłużymy się teraz metodą *a posteriori* dla sformułowania modelu zmiennej $X3$ zawierającego również $X1$. Jest to metoda iteracyjna.

W kolejnych krokach biorąc pod uwagę test *t*-Studenta z początkowych 48-miu zmiennych eliminowane są zmienne statystycznie nieistotne, aż dochodzimy do zbudowania modelu ekonometrycznego. *GRET*L wymaga, aby liczba obserwacji > ilości parametrów przy zmiennych objaśniających modelu (łącznie z $Y0=1$ stojącym przy parametrze a_0). Z tego względu dla $n = 15$ obserwacji, liczba zmiennych nie może przekraczać 13. Dla sformułowania modelu ekonometrycznego wielowymiarowego korzystamy z menu: *Model/Klasyczna metoda najmniejszych kwadratów*.

Zmienną zależną jest $Y3$, a pierwszą porcję stanowią szeregi 13 zmiennych obejmujące $X1$, $Y4$ - $Y42$. Parametry *Modelu 1* po estymacji KMNK z wykorzystaniem 15. obserwacji (1-15) oraz statystykę opisową podano w tabeli 12.4.

Tab. 12.4. Parametry oraz statystyka opisowa *Modelu 1*

Zmienna	Współczynnik	Błąd stand.	Statystyka <i>t</i>	wartość <i>p</i>
const	1,42788	2,77487	0,5146	0,69745
X1	0,248466	1,11121	0,2236	0,85996
Y4	-1,64144	4,15644	-0,3949	0,76056
Y7	0,186169	0,499094	0,3730	0,77271
Y8	-0,0263754	0,0663266	-0,3977	0,75905
Y15	0,0172704	0,0479987	0,3598	0,78012
Y18	-0,00737973	0,0401969	-0,1836	0,88441
Y19	-0,001641	0,00782133	-0,2098	0,86834
Y35	1,02505	0,17321	5,9180	0,10657
Y36	1,52764	4,18343	0,3652	0,77711
Y39	-0,198173	0,436005	-0,4545	0,72841
Y40	0,118143	0,264742	0,4463	0,73279
Y41	0,0420354	0,105878	0,3970	0,75940
Y42	-0,303461	0,976195	-0,3109	0,80813

Średnia arytmetyczna zmiennej zależnej = 139,467

Odchylenie standardowe zmiennej zależnej = 26,0402

Suma kwadratów reszt = 0,0496978

Błąd standardowy reszt = 0,22293

Współczynnik determinacji $R^2 = 0,999995$; skorygowany $R^2 = 0,999927$

Statystyka $F(13, 1) = 14693,8$ (wartość $p = 0,00646$)

Logarytm wiarygodności = 21,5398

Kryterium informacyjne Akaike'a = -15,0795

Wyłączając stałą największa wartość p (popętnienia błędu) jest dla zmiennej Y18.

Źródło: Opracowanie własne w *GRET*L.

Program proponuje wyeliminowanie Y18, a w ostatniej kolumnie podaje informacyjnie ilością gwiazdek poziom istotności α a mianowicie:

*** - $\alpha = 0,01$; ** - $\alpha = 0,05$; * - $\alpha = 0,1$; puste brak statystycznej istotności ($\alpha > 0,1$).

Będziemy teraz dalej korzystać z podpowiedzi programu *GRET*L usuwając wskazaną zmienną i dodając jednocześnie kolejną. Wykonano kroki: (Y18, Y47); (Y19, Y56); (Y47, Y58); (Y15, Y59); (Y59, Y71). Pojawiły się pozycje z (**) i *GRET*L nie zaproponował usunięcia żadnej zmiennej. Użytkownicy *Model 5* podano w tabeli 12.5.

Tab. 12.5. Parametry oraz statystyka opisowa *Modelu 5*

Zmienna	Współczynnik	Błąd stand.	Statystyka <i>t</i>	Wartość <i>p</i>	
const	-11,307	0,78359	-14,4298	0,04405	**
X1	4,36531	0,233482	18,6965	0,03402	**
Y4	-20,6256	1,31106	-15,7320	0,04041	**
Y7	-4,25101	0,295666	-14,3777	0,04421	**
Y8	-0,645129	0,0406749	-15,8606	0,04009	**
Y35	-0,719043	0,118126	-6,0871	0,10366	
Y36	22,7985	1,46157	15,5986	0,04076	**
Y39	3,62419	0,253863	14,2762	0,04452	**
Y40	0,775593	0,0470906	16,4702	0,03861	**

Zmienna	Współczynnik	Błąd stand.	Statystyka <i>t</i>	Wartość <i>p</i>	
Y41	0,133385	0,00778997	17,1226	0,03714	***
Y42	-4,14269	0,216883	-19,1010	0,03330	***
Y56	-0,211626	0,0133047	-15,9061	0,03997	***
Y58	0,00959178	0,000863599	11,1067	0,05716	*
Y71	0,00159128	0,000346712	4,5896	0,13657	

Średnia arytmetyczna zmiennej zależnej = 139,467

Odchylenie standardowe zmiennej zależnej = 26,0402

Suma kwadratów reszt = 0,000184939

Błąd standardowy reszt = 0,0135992

Współczynnik determinacji $R^2 = 1$; skorygowany $R^2 = 1$

Statystyka $F(13, 1) = 3,94862e+006$ (wartość $p = 0,000394$)

Logarytm wiarygodności = 63,4924; Kryterium informacyjne Akaike'a = -98,9849.

Źródło: Opracowanie własne w *GRET*.

Jednak dla sprawdzenia wszystkich zmiennych wejściowych usunięto Y71 o największym „*p*”. Postępując analogicznie jak wcześniej wykonano kroki: (Y71, Y72); (Y72, Y75); (Y58, Y76); (Y35, Y77); (Y76, Y79); (Y79, Y80); (Y80, Y81); (Y81, Y96); (Y77, Y98); (Y98, Y99); (Y99, Y100); (Y100, Y102); (Y102, Y108). Pojawiły się pozycje z trzema gwiazdkami i *GRET* nie zaproponował usunięcia żadnej ze zmiennych. Otrzymany model zaprezentowano w tabeli 12.6.

Tab. 12.6. Parametry oraz statystyka opisowa *Modelu 19*

Zmienna	Współczynnik	Błąd stand.	Statystyka <i>t</i>	Wartość <i>p</i>	
const	-9,794	0,0575361	-170,2236	0,00374	***
X1	6,61303	0,104716	63,1519	0,01008	**
Y4	-10,7369	0,0603155	-178,0118	0,00358	***
Y7	-2,38529	0,00523392	-455,7362	0,00140	***
Y8	-0,396104	0,00065029	-609,1182	0,00105	***
Y36	12,0747	0,0596948	202,2740	0,00315	***
Y39	2,00895	0,00472209	425,4362	0,00150	***
Y40	0,393337	0,00247653	158,8258	0,00401	***
Y41	0,0651785	0,000706593	92,2433	0,00690	***
Y42	-2,72631	0,00737963	-369,4375	0,00172	***
Y56	-0,12351	0,00038678	-319,3289	0,00199	***
Y75	-3,69421	0,10408	-35,4939	0,01793	**
Y96	0,00381397	0,000337248	11,3091	0,05615	*
Y108	0,00123085	0,000299697	4,1070	0,15205	

Średnia arytmetyczna zmiennej zależnej = 139,467

Odchylenie standardowe zmiennej zależnej = 26,0402

Suma kwadratów reszt = 9,8289e-006

Błąd standardowy reszt = 0,00313511

Współczynnik determinacji $R^2 = 1$, skorygowany $R^2 = 1$

Statystyka $F(13, 1) = 7,42966e+007$ (wartość $p = 9,08e-005$)

Logarytm wiarygodności = 85,5027; Kryterium informacyjne Akaike'a = -143,005.

Źródło: Opracowanie własne w *GRET*.

Usuujemy jednak Y108 o największym p i wykonujemy kolejne kroki związane z usuwaniem i dodawaniem zmiennych: (Y108, Y110); (Y110, Y111); (Y111, Y112); (Y112, Y113); (Y113, Y115); (Y115, Y116); (Y126, Y128). Mimo, że nie ma propozycji usunięcia zmiennej w kolejnym kroku eliminujemy zmienną Y128 o największym p i uzyskujemy kolejny już 27. model (patrz tab. 12.7).

Tab. 12.7. Parametry oraz statystyka opisowa *Modelu 27*

Zmienna	Współczynnik	Błąd stand.	Statystyka t	Wartość p	
const	-10,3664	0,0220647	-469,8190	0,00136	***
X1	6,59081	0,0159214	413,9590	0,00154	***
Y4	-10,8872	0,00538235	-2022,7533	0,00031	***
Y7	-2,39098	0,000896357	-2667,4381	0,00024	***
Y8	-0,399755	0,000146838	-2722,4317	0,00023	***
Y36	12,2339	0,00508278	2406,9287	0,00026	***
Y39	2,01352	0,00079944	2518,6568	0,00025	***
Y40	0,398332	0,000319083	1248,3670	0,00051	***
Y41	0,0658437	0,000114801	573,5442	0,00111	***
Y42	-2,74996	0,0013551	-2029,3430	0,00031	***
Y56	-0,124702	7,21092e-05	-1729,3442	0,00037	***
Y75	-3,64858	0,0151154	-241,3810	0,00264	***
Y96	0,00355871	5,06107e-05	70,3153	0,00905	***
Y128	-0,0011417	4,35577e-05	-26,2113	0,02428	**

Średnia arytmetyczna zmiennej zależnej = 139,467

Odchylenie standardowe zmiennej zależnej = 26,0402

Suma kwadratów reszt = 2,55244e-007

Błąd standardowy reszt = 0,000505217

Współczynnik determinacji $R^2 = 1$; skorygowany $R^2 = 1$

Statystyka $F(13, 1) = 2,861e+009$ (wartość $p = 1,46e-005$)

Logarytm wiarygodności = 112,884; Kryterium informacyjne Akaike'a = -197,768

Źródło: Opracowanie własne w *GRET*.

Kolejne kroki usuwania i dodawania zmiennych były następujące: (Y128, Y129); (Y129, Y133); (Y133, Y134); (Y134, Y140); (Y140, Y141); (Y141, Y145); (Y145, Z1); (Z1, Z10); (Z10). Ostatni model po usunięciu jeszcze zmiennej Y96 z jedną gwiazdką zamieszczono w tabeli 12.8.

Tab. 12.8. Parametry oraz statystyka opisowa *Modelu 37*

Zmienna	Współczynnik	Błąd stand.	Statystyka t	wartość p	
const	-9,89463	0,369927	-26,7475	0,00011	***
X1	6,09128	0,401948	15,1544	0,00062	***
Y4	-10,9607	0,194493	-56,3552	0,00001	***
Y7	-2,34316	0,0238225	-98,3592	<0,00001	***
Y8	-0,393382	0,00342025	-115,0155	<0,00001	***
Y36	12,2961	0,188301	65,3001	<0,00001	***
Y39	1,97246	0,0223516	88,2468	<0,00001	***
Y40	0,390203	0,0114997	33,9316	0,00006	***
Y41	0,0631562	0,00442157	14,2837	0,00074	***
Y42	-2,70711	0,0451341	-59,9792	0,00001	***
Y56	-0,1202	0,00124638	-96,4392	<0,00001	***
Y75	-3,18973	0,391682	-8,1437	0,00387	***

Średnia arytmetyczna zmiennej zależnej = 139,467

Odchylenie standardowe zmiennej zależnej = 26,0402

Suma kwadratów reszt = 0,00127674

Błąd standardowy reszt = 0,0206296

Wsp. determinacji $R^2 = 1$

Skorygowany $R^2 = 0,999999$

Statystyka $F(11, 3) = 2,02789e+006$ (wartość $p < 0,00001$)

Logarytm wiarygodności = 49,0021; Kryterium informacyjne Akaike'a = -74,0043

Źródło: Opracowanie własne w *GRET*L.

Porównajmy jeszcze wartości empiryczne i teoretyczne naszego modelu zmiennej X_3 . Wykres w *GRET*L uzyskujemy korzystając z menu:

Wykresy/wartości empirycznych i wyrównanych/względem obserwacji.

Klikając na wykres dokonujemy kopiowania do schowka jako wykres czarno-biały, a następnie poprzez menu Word-a: *Edycja/wklej specjalnie/rysunek (EMF)* zapisania w dokumencie, przykładowo *Wpływy.doc*. Równanie modelu po zaokrągleniu parametrów, zwanych w *GRET*L współczynnikami jest następujące:

$$\hat{X}_3 = -9,89 + 6,09X_1 - 10,96X_4 - 2,34X_7 - 0,39X_8 + 12,3X_{36} + 1,97X_{39} + 0,39X_{40} + 0,06Y_{41} - 2,7Y_{42} - 0,12Y_{56} - 3,19Y_{75}$$

Gdzie:

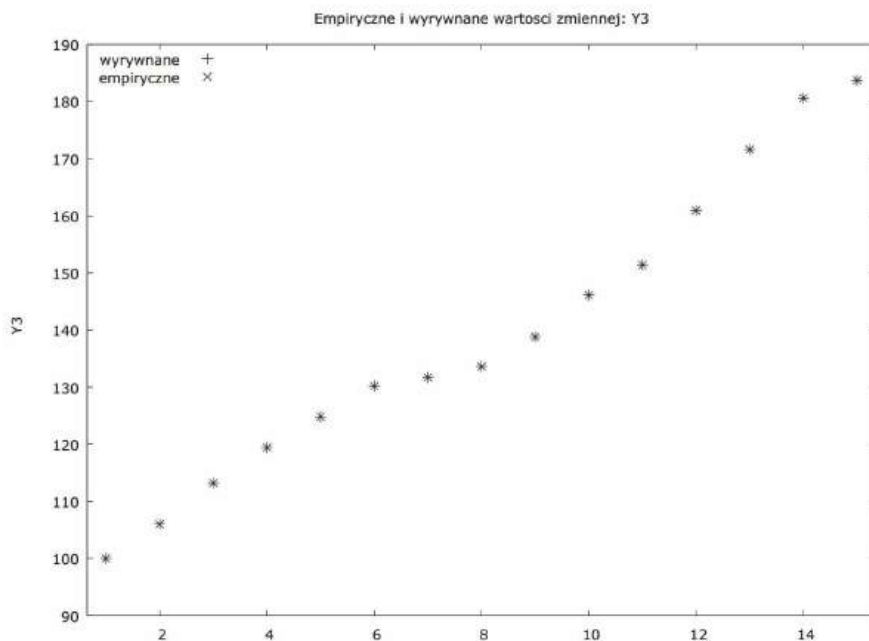
$$4 \mid \text{dPKB} \mid Y_3 \mid \text{Dynamika produktu krajowego brutto(PKB)}$$

a opisy pozostałe zmiennych występują w tabeli 12.9.

Tab. 12.9. Zmienne występujące w modelu X_3

Poz.	Symbol	Zmienna	Nazwa
1	dIK	X1	Dynamika nakładów inwestycyjnych (ceny stałe)
5	dPK	Y4	Dynamika popytu krajowego (ceny stałe)
8	dPKA	Y7	Dynamika akumulacji
9	dPKAS	Y8	Dynamika nakładów brutto na środki trwałe (w akumulacji)
37	dPKM	Y36	Dynamika popytu krajowego na 1. mieszkańca
40	dAKM	Y39	Dynamika akumulacji na 1. mieszkańca
41	dAKSM	Y40	Dynamika udziału akumulacji na środki trwałe na 1. mieszkańca
42	dSHM	Y41	Dynamika sprzedaży detalicznej na 1. mieszkańca
43	dIKM	Y42	Dynamika nakładów inwestycyjnych na 1. mieszkańca
57	dDGO	Y56	Dynamika dochodu na 1. gospodarstwo domowe
76	dNIO95	Y75	Dynamika nakładów inwestycyjnych ogółem

Źródło: Opracowanie własne (fragment tabeli 12.2)



Źródło: Opracowanie własne w programie *GRET*L.

Rys. 12.17. Graficzna prezentacja empirycznych i modelowych wartości zmiennej X_3

Widzimy bardzo dobre dopasowanie uzyskanego modelu do danych empirycznych. Współczynnik determinacji jest bliski jedności.

13. WERYFIKACJA MODELI EKONOMETRYCZNYCH WSPOMAGANA KOMPUTEROWO



13.1. Wstęp

Weryfikacja modeli ekonometrycznych obejmuje szereg pracochłonnych etapów. Zachodzi więc konieczność - zwłaszcza przy modelach o wielu zmiennych posługiwanie się wspomaganie komputerowym. Zaprezentowany materiał należy traktować jako rodzaj pomocy dla studiujących ekonometrię w zakresie poznania procedur zastosowania funkcjonalności arkusza kalkulacyjnego wraz jego dodatkami oraz pakietu *GRETL*, który jest bezpłatnie dostępny w Internecie.

13.2. Modele jednorównaniowe

Liniowy model ekonometryczny ma postać:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \alpha_j X_j + \dots + \alpha_m X_m + \varepsilon$$

Gdzie:

$\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m$ - parametry;

X_j - zmienna objaśniająca (niezależna) przy czym $j = 1, 2, \dots, m$;

Y - zmienna objaśniana (zależna);

ε - składnik losowy.

Po oszacowaniu, czyli estymacji parametrów otrzymujemy wartości teoretyczne zmiennej objaśnianej i wówczas parametry α zastępujemy oszacowanymi:

$$\hat{Y} = a_0 + a_1 X_1 + a_2 X_2 + a_j X_j + \dots + a_m X_m$$

Kolejne obserwacje zmiennej Y zapisujemy jako y_i ($i = 1, 2, \dots, n$), natomiast zmiennych X_j jako x_{ij} . Obserwacje zmiennej Y stanowią wektor kolumnowy, a wartości zmiennej X_j są macierzą o wymiarach ($n \times m$). Jedną z metod określenia wektora parametrów $\mathbf{a}$ przy pomocy klasycznej metody najmniejszych kwadratów (*KMNK*) jest posłużenie się równaniem:

$$\mathbf{a} = [\mathbf{X}^T \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{y}$$

Przy czym $\mathbf{X}^T$ - transpozycja macierzy $\mathbf{X}$. Podane w dalszej części tego tematu wzory bazują na rozdziale 4. publikacji wymienionej w przypisie¹²⁷.

Zależność między Y a X_j mierzona jest współczynnikiem korelacji wielorakiej przyjmującym wartości z przedziału $[0,1]$. Współczynnik ten wyznaczamy na podstawie wyznacznika macierzy korelacji R , obejmującej współczynniki korelacji między zmiennymi objaśniającymi X_j oraz macierzy rozszerzonej R^* :

$$R_w = \sqrt{1 - \frac{\det R^*}{\det R}} \qquad R^* = \begin{bmatrix} 1 & R_o^T \\ R_o & R \end{bmatrix}$$

Wektor korelacji R_o obejmuje współczynniki między zmienną Y a zmiennymi X_j . Im wartość R_w jest bliższa jednoci tym natężenie zależności jest większe. Dla obliczania R_w można skorzystać z funkcji Excela:

WYZNACZNIK.MACIERZY(...).

¹²⁷ Jajuga K., *Ekonometria. Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu 2002.

Modele jednorównaniowe obejmują:

- regresję jednej zmiennej z dopasowaniem funkcji liniowej lub nieliniowej (wykładniczej, potęgowej, logarytmicznej, logistycznej);
- trend liniowy lub nieliniowy (po wcześniejszej transformacji na postać liniową), gdy zmienną objaśniającą jest zmienna czasowa t ;
- regresję wielu zmiennych X_j , czyli tzw. regresję wieloraką.

Weryfikacja liniowych lub sprowadzalnych do liniowych modeli jednorównaniowych bazuje na resztach e_i stanowiących różnicę między wartościami empirycznymi, a modelowymi zmiennej objaśnianej obliczoną dla poszczególnych obserwacji. Pozwala to na określenie wariancji resztowej:

$$S_e^2 = \frac{\sum_{i=1}^n e_i^2}{n-k}$$

Gdzie: n - liczba ilość obserwacji, $k = m + 1$ - liczba zmiennych, w tym tożsamościowo równa 1 przy parametrze a_0 .

Standardowy błąd oceny poszczególnych parametrów a_j określamy z zależności:

$$S(a_j) = \sqrt{V(a_j)}$$

Wariancje parametrów $V(a_j)$ znajdują się na głównej przekątnej macierzy kowariancji wyznaczonej

$$D^2(a) = S_e^2 [X^T X]^{-1}$$

według:

Występuje tu zatem iloczyn wariancji resztowej i macierzy odwrotnej.

Istnieje szereg mierników dopasowania modelu do danych rzeczywistych (empirycznych), a należą do nich:

Współczynnik determinacji stanowiący również kwadrat współczynnika korelacji wielorakiej, przyjmujący wartości z przedziału $[0,1]$ obliczany ze wzoru:

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

Gdzie: $\bar{y}$ - średnia arytmetyczna zmiennej Y . Współczynnik R^2 jest miarą stopnia wyjaśnienia przez model zmiennej zależnej Y .

Alternatywnie wobec R^2 stosowany *współczynnik zbieżności*, który wskazuje jaką część zmiennej objaśnianej stanowi zmienność składnika losowego ε . Współczynnik zbieżności $\varphi^2 = 1 - R^2$, który można obliczyć bezpośrednio ze wzoru:

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

Wymienione wcześniej *odchylenie standardowe* S_e , określające przeciętne różnice między y_i a $\hat{y}_i$.

Współczynnik wyrazistości, wskazujący jaką część wartości średniej zmiennej Y stanowi odchylenie standardowe reszt:

$$W = \frac{S_e}{\bar{y}} 100$$

Wstępnie określony liniowy model jednorównaniowy może mieć część zmiennych objaśnianych nieistotnych statystycznie w relacji do zmiennej Y . Z tego względu wpięrow badamy istotność całego zestawu zmiennych X_j posługując się testem parametrycznym F , dla którego hipoteza zerowa zakłada nieistotność wektora parametrów bez wyrazu wolnego:

$$F = \frac{(n-k)R^2}{m(1-R^2)}$$

Jeśli $F > F^*$ to hipotezę zerową odrzucamy. Wartość F^* określamy na podstawie rozkładu F -Fishera dla poziomu istotności α i stopni swobody $n_1 = m$ oraz $n_2 = n - k = n - m - 1$.

Parametryczny test t -Studenta odnoszący się do określonej zmiennej X_j , stosowany przeważnie dla $n < 30$ obserwacji przy poziomie istotności $\alpha = 0,05$, umożliwia badanie statystycznej istotności danej zmiennej objaśniającej na zmienną objaśnianą.

$$t_j = \frac{Ia_j I}{S(a_j)}$$

Hipoteza zerowa zakłada brak statystycznej istotności danej zmiennej objaśniającej na zmienną Y . Gdy $t_j > t^*$ to odrzucamy hipotezę zerową. Wartość krytyczną t^* wyznaczamy według tablicy rozkładu t -Studenta.

Dalsze badania dotyczące składnika losowego (reszt) możemy prowadzić zarówno analitycznie jak i poprzez analizę graficzną, a mianowicie określając:

- rozkład reszt (częstość występowania reszty danej wartości), co pozwala na sprawdzenie założenia o normalności rozkładu reszt, czyli średnia reszt równa się zero, a wynika to z *KMNK*;
- symetrię reszt, czyli równość częstości występowania reszt dodatnich i ujemnych (bez uwzględnienia reszt o wartości zero);
- losowość (nie może być zbyt długich lub krótkich serii reszt dodatnich lub ujemnych); założenie, że reszty jako obserwacje pojawiają się kolejno w momentach lub okresach czasu, pozwala wyodrębnić kolejne grupy reszt o tendencji wzrostowej lub spadkowej;
- stałość/niestałość wariancji resztowej dla kolejnych przedziałów czasu; jeśli wariancja zwiększa się w czasie to dla „przebudowy” modelu należy stosować ważoną metodę najmniejszych kwadratów;
- brak zależności obserwacji od czasu; nie spełnienie tego założenia może oznaczać wzajemny liniowy wpływ zmiennych objaśnianych.

Jeśli mamy do czynienia z nieliniowym modelem ekonometrycznym z wieloma zmiennymi objaśniającymi, to jak już wspomniano przeprowadzamy transformację liniową, aby zastosować *KMNK*. W tym celu stosuje się podstawienia, odwrotności lub obustronne logarytmowanie danego równania nieliniowego. Dla modeli z tak oszacowanymi parametrami możemy stosować miary weryfikacji jak dla modeli liniowych, przy czym według Jajugi (vide rozdział „Analiza reszt”)¹²⁸ dla modeli nieliniowych stosujemy dodatkowo:

$$p = \frac{1}{n} \sum_{i=1}^n \frac{Ie_i I}{\hat{y}_i}$$

Gdzie: p - wskaźnik średniego względnego poziomu reszt, będący średnim poziomem wartości bezwzględnych reszt do wartości modelowej określonej po estymacji parametrów.

¹²⁸ Ibidem, rozdział „Analiza reszt”.

13.3. Zastosowanie funkcji *Excels*

Ogólnie dostępnym dla studiujących jak i praktyków narzędziem do wspomagania komputerowego obliczeń z zakresu estymacji jak i weryfikacji jednorównaniowych modeli ekonometrycznych jest program *Excel* z dodatkami programowymi. Pokażmy zastosowanie jego funkcji na przykładzie fragmentu bazy danych dotyczącej stopy bezrobocia w 50. powiatach, pozyskanej z Internetu na stronie www.kufel.torun.pl¹²⁹. Zbiór „powiaty.gdt” obejmuje 380 rekordów (obserwacji) dla 42. zmiennych objaśniających i zmiennej objaśnianej „bezrobot”. Procedura wyłonienia podzbioru z zastosowaniem menu programu *GRET*L jest następująca:

- menu *Próba/Zakres próby*, startowa obserwacja: 1 końcowa obserwacja: 50, ok.;
- zaznaczenie w strukturze rekordu informacji kopiowanych przez wybór z wciśnięciem klawisza *Ctrl*;
- menu *Dane/Pokaż wartości* (pojawia się okno prezentacji danych);
- kopiowanie do bufora;
- wklejenie do arkusza kalkulacyjnego, zastąpienie kropek dziesiętnych przecinkami, wykonanie obramowań i sformatowanie komórek (ustawienie czcionki, umiejscowienia);
- menu *Excels: Edycja/Kopiuj*;
- wstawienie do Worda *Edycja/Wklej specjalnie* jako tekst o formacie RTF.

Zamieszczona tabela 13.1 oprócz liczby porządkowej zawiera skróty 50. powiatów oraz informacje:

bezrobot - stopa bezrobocia w powiatach,
regon - liczba jednostek regon na 1 tys. mieszkańców,
lud_prod - udział ludności w wieku produkcyjnym,
noclegi - liczba miejsc noclegowych na 1 tys. mieszkańców.

a ponadto kolumnę z obliczonymi wartościami teoretycznymi zmiennej objaśnianej „bezrobot” i kolumnę „reszta” stanowiąca różnicę wartości empirycznej i modelowej.

Do estymacji parametrów oraz krótkiej statystyki opisowej *Excel* ma wbudowaną funkcję *REGLINP*, której przykładową składnię podano poniżej

=REGLINP(C3:C52;D3:F52;PRAWDA;PRAWDA)

Gdzie: C3:C52 - obszar zmiennej „bezrobot”; D3:F52 - obszar zmiennych: regon, lud_prod, noclegi; PRAWDA - odpowiednio w kolejności potwierdzenie występowania stałej, żądania statystyki opisowej.

Tab. 13.1. Wybrane dane dotyczące 50. powiatów

lp.	obs	bezrobot	regon	lud_prod	noclegi	bezrobot_t	reszta
1	aleksand	0,23731146	88,77903924	0,61042606	3,88940192	0,177008855	0,060302605
2	augustow	0,1932204	72,64343743	0,57643522	6,21548128	0,1790033	0,0142171
3	bartoszy	0,31262608	54,58067822	0,6120593	6,64126293	0,278419895	0,034206185
4	belchato	0,20068024	78,64737232	0,63906123	2,99502763	0,226923911	-0,02624367
5	bedzinsk	0,17313645	97,30628446	0,64825293	1,87554627	0,198274027	-0,02513758
6	bialski	0,15619965	46,96730826	0,55321287	7,48601029	0,193017156	-0,03681751
7	Biała Po	0,17700602	83,87472787	0,6397655	0,78166515	0,177962109	-0,00095609
8	bialobrz	0,18572871	83,89281683	0,57444596	6,53555938	0,166993567	0,018735143
9	bialogar	0,31052263	85,1998515	0,61458978	5,69648971	0,225101535	0,085421095
10	bialosto	0,13364342	64,52498231	0,58589608	6,35716246	0,210257678	-0,07661426

¹²⁹ Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*L. Wydawnictwo Naukowe PWN, Warszawa 2007.

lp.	obs	bezrobot	regon	lud_prod	noclegi	bezrobot_t	reszta
11	Bialysto	0,1159479	117,2178271	0,65325462	0,26840109	0,150446916	-0,03449902
12	bielski	0,10642451	93,37142468	0,61672618	2,25011053	0,150799068	-0,04437456
13	bielski	0,09712773	68,21023235	0,55115362	8,60218546	0,183633707	-0,08650598
14	Bielsko	0,10666292	139,5591745	0,64588516	0,58683139	0,113854908	-0,00719199
15	bierunsk	0,10678335	72,31736085	0,62652837	1,75211136	0,186819713	-0,08003636
16	bieszcz	0,24676314	97,12826486	0,6095605	8,17615698	0,248928781	-0,00216564
17	bilgoraj	0,16447494	59,60548885	0,5809129	4,910425	0,17867209	-0,01419715
18	bochensk	0,11816367	68,91857311	0,58917038	3,04723886	0,145681592	-0,02751792
19	boleslaw	0,19472202	69,72202035	0,6255337	3,34075218	0,219538689	-0,02481667
20	braniew	0,29777688	64,96623819	0,60874442	7,91663863	0,283802414	0,013974466
21	brodnick	0,22093662	65,73995956	0,59297202	4,84110723	0,192308658	0,028627962
22	brzeski	0,15175769	62,16931217	0,58637118	2,86409291	0,145472663	0,006285027
23	brzeski	0,23634522	89,93619865	0,61858214	3,16635772	0,17679099	0,05955423
24	brzezins	0,19058114	87,57657278	0,6204583	4,10332869	0,201840138	-0,011259
25	brzozow	0,25033179	49,42565558	0,57306302	3,59861486	0,151085051	0,099246739
26	buski	0,12205082	65,61599231	0,58515266	7,88272659	0,237452906	-0,11540209
27	bydgoski	0,16613259	83,56451339	0,62552228	3,65509823	0,207827136	-0,04169455
28	Bydgoszc	0,11421966	124,3337543	0,64257658	0,33686881	0,122308407	-0,00808875
29	Bytom	0,18391673	78,9212288	0,64397239	0,30109574	0,182899337	0,001017393
30	bytowski	0,31625195	74,80928226	0,60377308	7,79512987	0,259251651	0,057000299
31	Chelm	0,17973059	84,77036521	0,64336933	0,64013238	0,180870911	-0,00114032
32	chelmi	0,22705399	63,06622246	0,6110624	4,34274359	0,220341623	0,006712367
33	chelm	0,19664897	37,77596513	0,56698715	7,73838974	0,236034024	-0,03938505
34	chodzies	0,20626996	90,10649924	0,62630981	3,7285448	0,202321484	0,003948476
35	chojnick	0,25012773	74,66068374	0,60446054	3,40982296	0,174464524	0,075663206
36	Chorzow	0,18373607	83,68835418	0,62929007	0,43198937	0,151410822	0,032325248
37	choszcze	0,28483623	67,9078436	0,61360933	7,93877389	0,289686646	-0,00485042
38	chrzanow	0,1449684	79,43983145	0,62816024	1,75517412	0,180784517	-0,03581612
39	ciechano	0,1929599	77,26678291	0,607051	3,84699215	0,184624974	0,008334926
40	cieszyn	0,12057551	104,863689	0,62279219	2,1682075	0,145877732	-0,02530222
41	czarnkow	0,22369401	72,7488593	0,61686377	3,50591303	0,20240002	0,02129399
42	Czestoch	0,14663569	107,3946464	0,64624578	0,50329429	0,154431067	-0,00779538
43	czestoch	0,14817528	64,41460971	0,6023592	3,81314755	0,19163701	-0,04346173
44	czluchow	0,31031876	76,28742726	0,61792992	7,62171001	0,280840003	0,029478757
45	Dabrowa	0,18608169	95,18910741	0,6784947	0,69213313	0,23520656	-0,04912487
46	dabrowsk	0,1968168	43,99595676	0,58537923	4,60175778	0,201245068	-0,00442827
47	debicki	0,18682969	54,57289786	0,59657772	2,61297614	0,169741581	0,017088109
48	drawski	0,29538973	91,06803368	0,61712385	6,45539109	0,237272872	0,058116858
49	dzialdow	0,28138653	56,85883481	0,60242963	4,69332765	0,218846243	0,062540287
50	dzierz	0,22868148	93,47332205	0,61790746	1,99138004	0,147822057	0,080859423

Źródło: Opracowanie własne bazujące na pliku *powiaty.gdt*.

Zastosowanie *REGLINP* dało następujące wyniki podane w tabeli 13.2.

Tab. 13.2. Oszacowanie parametrów modelu

a_3	a_2	a_1	a_0	
0,0196761	1,9008	-0,00129131	-0,945176	parametry
0,00373272	0,425549	0,000474643	0,251316	błąd standardowy oszacowania parametrów
0,450304	0,047332			R^2, S_e
12,5609	46			F, iss
0,08442122	0,103055			ESS, RSS

Źródło: Opracowanie własne w Excelu.

Gdzie: R^2 - współczynnik determinacji, S_e - odchylenie standardowe reszt, F - statystyka Fishera, iss - ilość stopni swobody ($n - m - 1$), ESS - wyjaśniona przez model suma kwadratów, RSS - resztowa suma kwadratów. Tak więc model liniowy jest następujący:

$$\hat{Y} = -0,945176 - 0,00129131 * (\text{regon}) + 1,9008 * (\text{lud_prod}) + 0,0196761 * (\text{noclegi})$$

Opcja Excela:

Narzędzia/Analiza danych/Listy narzędzi statystycznych

umożliwia określenie między innymi: korelacji, kowariancji; statystyki opisowej, histogramu, regresji.

Korelacja

Tab. 13.3. Współczynniki korelacji między zmiennymi

	bezrobot	regon	lud_prod	noclegi
bezrobot	1			
regon	-0,273055	1		
lud_prod	-0,042493	0,701364	1	
noclegi	0,458887	-0,52811	-0,69885	1

Źródło: Opracowanie własne w Excelu.

Kowariancja

Tab. 13.4. Współczynniki kowariancji między zmiennymi

	bezrobot	regon	lud_prod	noclegi
bezrobot	0,003826			
regon	-0,338496	401,658397		
lud_prod	-0,000070	0,373030	0,000704	
noclegi	0,072086	-26,879798	-0,047101	6,449715

Źródło: Opracowanie własne w Excelu.

Statystyka opisowa

Przykład dotyczy zmiennej objaśnianej „bezrobot”. Wprowadzamy do okna dialogowego następujące informacje:

- zakres wejściowy Y: \$C\$2:\$C\$52; tytuły w pierwszym wierszu;
- statystyka podsumowująca; zapis w nowym arkuszu;
- poziom ufności 95%;
- największa oraz najmniejsza wartość zmiennej.

Tab. 13.5. Statystyka opisowa zmiennej *bezrobot*

<i>bezrobot</i>	
Średnia arytmetyczna	0,195567
Błąd standardowy	0,008748
Mediana	0,188705
Odchylenie standardowe	0,061855
Wariancja próbki	0,003826
Kurtoza	-0,654921
Skośność	0,386007
Zakres	0,219124
Suma wartości obserwacji	9,778363
Licznik (liczba obserwacji)	50
Największa wartość	0,316252
Najmniejsza wartość	0,097128
Poziom ufności (95,0%)	0,017579

Źródło: Opracowanie własne w *Excelu*.

W tabeli 13.5 na szczególną uwagę zasługuje *Błąd standardowy* jaki popełniamy przy prognozowaniu zmiennej *Y*, czyli stopy bezrobocia. Odchylenie standardowe jest pierwiastkiem z wariancji. Mediana jest wartością środkową stopy bezrobocia z 50-ciu obserwacji. Kurtoza jest miarą płaskości rozkładu wartości zmiennej *Y* i określana jest według wzoru podanego w *Excelu* jako:

$$K = \left\{ \frac{n(n+1)}{(n-1)(n-2)(n-3)} \sum \left(\frac{x_j - \bar{x}}{S} \right)^4 \right\} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

Skośność wyraża odchylenie rozkładu względem pionu określonego przez średnią, przy czym w *Excelu* podano wzór:

$$S_k = \frac{n(n+1)}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{S} \right)^3$$

Natomiast w rozdziale 1.4 publikacji Tadeusza Kufła¹³⁰ klasyczny współczynnik asymetrii (skośności) określony jest wg zależności:

$$S_k = \frac{1/n * \sum (x_i - \bar{x})}{S^3}$$

Z tabeli 13.5 wynika, że procent popełnienia błędu jest mniejszy od poziomu istotności $\alpha = 5\%$, gdyż wynosi 1,76 %.

Histogram

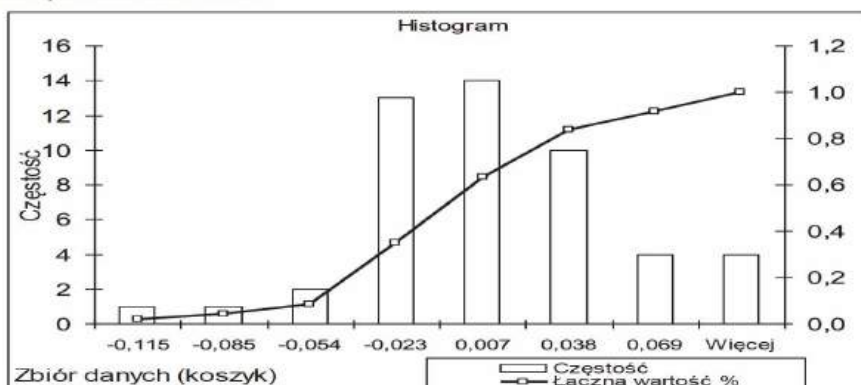
Skorzystanie z opcji *Histogram* umożliwia przykładowe wstępne zorientowanie się co do rozkładu normalnego reszt. Obliczone zostają częstości (liczności) występowania wartości *Y* w określonych przedziałach oraz podany jest wykres Pareto.

¹³⁰ Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, op. cit.

Tab. 13.6. Liczność występowania wartości zmiennych w przedziałach

Zbiór danych (koszyk)	Częstość - liczność	Łączna wartość %
-0,115	1	2,04%
-0,085	1	4,08%
-0,054	2	8,16%
-0,023	13	34,69%
0,007	14	63,27%
0,038	10	83,67%
0,069	4	91,84%
Więcej	4	100,00%

Źródło: Opracowanie własne w Excelu.



Źródło: Opracowanie własne w Excelu.

Rys. 13.1. Diagram oraz histogram zmiennej *bezrobot*

Regresja

Do okna dialogowego wprowadzamy przykładowo następujące informacje:

- zakres wejściowy Y: \$C\$3:\$C\$52,
- zakres wejściowy X: \$D\$3:\$F\$52,
- poziom ufności - 95%,
- wyjście na nowy arkusz,
- składniki resztowe, a w tym standaryzowane składniki resztowe, rozkład reszt.

W wyniku otrzymujemy raport podsumowujący o składnikach niżej zamieszczonych w tabelach 13.7÷ 13.10.

Tab. 13.7. Statystyki otrzymane z menu Regresja

Statystyki regresji	
Wielokrotność R	0,671047
R^2	0,450304
Dopasowany R^2	0,414454
Błąd standardowy reszt	0,0473321
Obserwacje (n)	50

Źródło: Opracowanie własne w Excelu.

W powyższej tabeli „Wielokrotność R ” to omówiony wcześniej współczynnik korelacji wielorakiej R_w . Dopasowany R^2 jest skorygowany o liczbę stopni swobody współczynnikiem determinacji i wyznaczany jest wg wzoru:

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

Tab. 13.8. Analiza wariancji

	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Istotność F</i>
Regresja	3	0,0844	0,0281	12,5609	0,000004
Resztkowy	46	0,1031	0,0022		
Razem	49	0,1875			

Źródło: Opracowanie własne w *Excelu*.

Gdzie:

$df = 3$, ilość stopni swobody w odniesieniu do obliczania MS ($m = 3$ – ilość zmiennych objaśniających);

$df = 46$, $50 - m - 1$, SS – suma kwadratów, przy czym dla:

Regresja

ESS w *REGLINP* – estymowana suma kwadratów, RSS – resztowa suma kwadratów

$$ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$$

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

MS – średnie z sum kwadratów obliczane jako: $MS = SS/df$, F – statyka Fishera określona dla łącznej oceny parametrów modelu, $Istotność F$ – prawdopodobieństwo błędu oszacowania.

Tab. 13.9. Wnioskowanie statystyczne

	<i>Współczynniki</i>	<i>Błąd standardowy</i>	<i>t Stat</i>	<i>Wartość-p</i>
Przecięcie	-0,945176	0,251316	-3,7609	0,000477
Zmienna X1	-0,00129131	0,000474643	-2,7206	0,009167
Zmienna X2	1,9008	0,425549	4,4667	0,00005
Zmienna X3	0,0196761	0,00373272	5,2712	0,000004

Źródło: Opracowanie własne w *Excelu*.

Ponieważ nie podano w programie nazw zmiennych tylko obszary ich wartości system nadał nazwy umowne: *przecięcie*, *zmienna1*, *zmienna2*, *zmienna3*. W tabeli tej współczynniki odpowiadają parametrom modelu, a ich wartości oraz błędy standardowe oszacowania a_0 , a_1 , a_2 , a_3 są identyczne jak otrzymane funkcją *REGLINP* (patrz tabela 9.2).

Statystyki „*t Stat*”, odpowiadają wcześniej omówionym otrzymanym z testu *t*-Studenta stosowanym przy badaniu istotności statystycznej określonej zmiennej X_j w modelu ekonometrycznym z prawdopodobieństwem popełnienia błędu określonym w kolumnie „*Wartość-p*”.

Kolejna tablica 13.10 stanowiąca opracowanie własne w *Excelu* dotyczy reszt, które program oblicza automatycznie na podstawie oszacowanych współczynników (parametrów) modelu. W zamieszczonej tabeli wprowadzono dodatkowo kolumnę „Standaryzowane składniki resztowe”. Standaryzację reszt przeprowadzono według:

$$z_i = \frac{(e_i - \bar{e})}{S_{ei}}$$

Gdzie: e_i - kolejne wartości reszt dla obserwacji 1÷50, $\bar{e}$ - średnia arytmetyczna reszt (w przybliżeniu równa zero), S_{ei} - odchylenie standardowe reszt.

Mając określony współczynnik determinacji $R^2 = 0,450304$ oraz wartość średnią zmiennej objaśnianej Y (patrz tabela 9.5), która wynosi 0,195567 możemy obliczyć dalsze pomocnicze miary dopasowania modelu do danych rzeczywistych:

- współczynnik zbieżności: $\varphi^2 = 1 - R^2 = 1 - 0,450304 = 0,549696$;
- współczynnik wyrazistości $W = S_{e/y} = 0,0473321/0,195567 = 0,242031$.

Tab. 13.10. Składniki resztowe

Obs.	$\hat{Y}$	Reszta	Standaryzacja reszt
1	0,177011406	0,060300054	1,328205901
2	0,179005575	0,014214825	0,313100216
3	0,27842227	0,03420381	0,753391652
4	0,226926608	-0,026246368	-0,578126502
5	0,198276844	-0,025140394	-0,553765538
6	0,193019235	-0,036819585	-0,811019609
7	0,177964923	-0,000958903	-0,021127004
8	0,166995835	0,018732875	0,412617956
9	0,225104011	0,085418619	1,88148508
10	0,210259973	-0,076616553	-1,687615642
11	0,15044986	-0,03450196	-0,759969986
12	0,150801731	-0,044377221	-0,977489418
13	0,183635756	-0,086508026	-1,905492184
14	0,113857839	-0,007194919	-0,158485886
15	0,186822408	-0,080039058	-1,763002146
16	0,248931136	-0,002167996	-0,047759329
17	0,178674426	-0,014199486	-0,312773397
18	0,145684066	-0,027520396	-0,606189127
19	0,2195413	-0,02481928	-0,546692459
20	0,28380473	0,01397215	0,30775489
21	0,192311057	0,028625563	0,630521267
22	0,145475123	0,006282567	0,138378705
23	0,176793611	0,059551609	1,311720114
24	0,201842719	-0,011261579	-0,248060982
25	0,151087402	0,099244388	2,1860212
26	0,237455126	-0,115404306	-2,541981987
27	0,207829753	-0,041697163	-0,918456588
28	0,122311315	-0,008091655	-0,178238008
29	0,182902184	0,001014546	0,02234158
30	0,259253967	0,056997983	1,255472163
31	0,180873748	-0,001143158	-0,025185528
32	0,220341623	0,006712367	0,147845785
33	0,236034024	-0,039385054	-0,867528432
34	0,202321484	0,003948476	0,086966392

Obs.	Y^	Reszta	Standaryzacja reszt
35	0,174464524	0,075663206	1,666605483
36	0,151410822	0,032325248	0,712013123
37	0,289688985	-0,004852755	-0,106895727
38	0,18078723	-0,03581883	-0,788976274
39	0,184627495	0,008332405	0,183529881
40	0,145880441	-0,025304931	-0,557389737
41	0,202402592	0,021291418	0,46897423
42	0,154433956	-0,007798266	-0,171775619
43	0,191639495	-0,043464215	-0,957378911
44	0,280842387	0,029476373	0,649261797
45	0,235209556	-0,049127866	-1,082130474
46	0,201247414	-0,004430614	-0,097597345
47	0,169744085	0,017085605	0,376334037
48	0,237275332	0,058114398	1,280063109
49	0,218848675	0,062537855	1,377497285
50	0,147824737	0,080856743	1,781001992

Źródło: Opracowanie własne w Excelu.

Średnia arytmetyczna reszt wynosi 0,000000251, a odchylenie standardowe populacji (50-ciu obserwacji) 0,045399439.

13.4. Dalsze badania modelu ekonometrycznego

Autokorelacja składnika resztowego. Stosujemy statystykę Durбина-Watsona¹³¹:

$$DW = \frac{\sum_{t=1}^{n-1} (e_{t+1} - e_t)^2}{\sum_{t=1}^n e_t^2}$$

Przy czym e_t - reszta dla obserwacji zmiennej objaśnianej w okresie t . Procedura postępowania jest następująca:

- wyznaczamy dla poszczególnych obserwacji wartości DW,
- dla przyjętego poziomu istotności np. $\alpha = 0,05$ przy danym n i ilości zmiennych $k = m + 1$ odczytujemy z tablicy rozkładu statystyki Durбина-Watsona wartości krytyczne d_L oraz d_U ,
- sprawdzamy czy:

$DW < d_L$, jeśli tak to odrzucamy hipotezę zerową H_0 o braku autokorelacji na rzecz hipotezy alternatywnej H_1 o dodatniej autokorelacji;

$DW > d_U$ - to brak jest podstaw do odrzucenia H_0 ;

$d_L \leq DW \leq d_U$ - nie podejmujemy żadnej decyzji.

Heteroscedastyczność. Stosowanie *KMNK* zakłada stałość wariancji składnika losowego (reszty) w czasie. Jednakże w modelach praktycznie stosowanych wariancje mają tendencję wzrostową lub malejącą. Dla zbadania tego zjawiska służy test Goldfelda i Quandta, a postępowanie w

¹³¹ Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003.

tym zakresie jest przykładowo następujące:

- dzielimy n obserwacji intuicyjnie według układu wektora reszt otrzymanego KMNK na dwie grupy: n_1, n_2 ,

- szacujemy wektory parametrów KMNK oddzielnie dla każdej grupy określonej jako A i B,
- określamy wariancje dla grup oraz statystykę korzystając ze wzorów:

$$s_1^2 = \frac{\sum_{i \in A} e_i^2}{n_1 - m - 1} \quad s_2^2 = \frac{\sum_{i \in B} e_i^2}{n_2 - m - 1} \quad F_e = \frac{s_1^2}{s_2^2}$$

- odczytujemy F_α z tablicy rozkładu F dla ilości stopni $n_1 - m - 1$ w liczniku oraz $n_2 - m - 1$ w mianowniku na poziomie istotności α np. 0,05,
- sprawdzamy, czy $F_e > F_\alpha$, jeśli tak to odrzucamy hipotezę zerową H_0 równości wariancji każdej z grup na rzecz hipotezy alternatywnej H_1 :

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad H_1 : \sigma_1^2 > \sigma_2^2$$

Gdy H_1 jest postaci jak niżej to stosujemy statystykę "odwrotną"

$$H_1 : \sigma_1^2 < \sigma_2^2 \quad F_e = \frac{s_2^2}{s_1^2}$$

i H_0 odrzucamy jeśli $F_e > F_{\alpha(n_2-m-1, n_1-m-1)}$.

Normalność rozkładu reszt. Jeśli próba jest mała $n \leq 30$ to przeprowadzamy test Hellwiga w sposób następujący:

- szacujemy odchylenie standardowe S dla ciągu e_t reszt kolejnych obserwacji i dokonujemy ich standaryzacji:

$$S = \sqrt{\frac{1}{n} \sum_{t=1}^n e_t^2} \quad e_t' = \frac{e_t}{S}$$

We wzorze na e_t' przyjmuje się wartość średniej reszt równą zero oraz traktuje się ciąg reszt jako populację dla obliczenia odchylenia standardowego. Jest to zbieżne z obliczeniem wartości standaryzowanych reszt zamieszczonych we wcześniejszej tabeli 13.10.

- sortujemy narastająco zestandaryzowane reszty i określamy dla nich na podstawie tablicy rozkładu normalnego wartości dystrybucyjny;
- dzielimy odcinek możliwych wartości dystrybucyjny [0; 1] na cele, czyli równe części i określamy liczbę K cel w których nie ma żadnej wartości;
- odczytujemy z tablicy testu Hellwiga dla przykładowo poziomu istotności $\alpha = 0,05$ wartości krytyczne K_1 oraz K_2 ,
- sprawdzamy hipotezę zerową (H_0 :) reszty mają rozkład normalny; gdy $K_1 \leq K \leq K_2$ to nie ma podstaw na odrzucenie hipotezy zerowej.

Dla dużych prób zmienna losowa K ma asymptotycznie rozkład Poissona z parametrem λ i według propozycji podanej w publikacji Dziechciarza¹³² stosujemy test Kołomogorowa o statystyce testowej dla maksymalnej różnicy:

$$D_n = \sup_x |F_n(x) - F_0(x)|$$

Przy czym: $x = e_t$, $F_0(x)$ - dystrybuanta rozkładu normalnego $N(0, S)$, $F_n(x)$ - dystrybuanta empiryczna z próby.

Procedura postępowania z przeprowadzeniem tego testu, w której możemy korzystać z wyrażenia *Excela* jest następująca:

¹³² Ibidem.

- obliczamy S dla próby n obserwacji na podstawie wzoru Hellwiga, następnie dokonujemy standaryzacji reszt;

- sortujemy narastająco reszty;
- określamy na podstawie tablicy dystrybucyjności rozkładu normalnego;
- obliczamy dystrybucyjności empiryczne;
- obliczamy różnice:

$$IF_n(e_i) - F(e_i)I$$

i odczytujemy różnicę maksymalną,

- określamy wartości obliczeniowe:

$$\lambda = \sqrt{n}D_n$$

- na podstawie tablicy rozkładu Kołmogorowa-Smirnowa określamy wartość krytyczną $\lambda_0 = Q(\lambda)$, przy czym:

$$Q(\lambda) = P(\sqrt{n}D_n < \lambda)$$

- sprawdzamy, czy dla danego poziomu istotności np. $\alpha = 0,05$

$$\lambda = \sqrt{n}D_n \geq \lambda_0$$

co oznacza niewystępowanie normalności rozkładu reszt.

Losowość. W zależności od ilości obserwacji w próbie do badania losowości reszt modelu liniowego stosujemy¹³³:

- test serii ($n \leq 30$),
- twierdzenie von Misesa ($n > 30$).

Procedura testu serii jest następująca:

- przyporządkowujemy symbole A resztom ($e_i > 0$) a wartościom ujemnym symbol B (w celu wyeliminowania reszt ujemnych możemy zwiększyć w *Excelu* dokładność obliczeń),
- obliczamy empiryczną liczbę serii K_e (ciągów o jednakowym symbolu),
- odczytujemy z tablicy liczby serii przy teście dwustronnym K_1 oraz K_2 ,
- sprawdzamy, czy $K_1 \leq K_e \leq K_2$, co oznacza potwierdzenie hipotezy zerowej

$$H_0 : \varepsilon_i \text{ losowy}$$

W przypadku dużej próby $n > 30$ nieuporządkowanych narastająco określamy:

- maksymalną długość serii r z liczby serii N_n o długości r w n doświadczeniach,
- obliczamy:

$$\mu = \frac{1 - 0,5^r}{0,5^{r+1}}$$

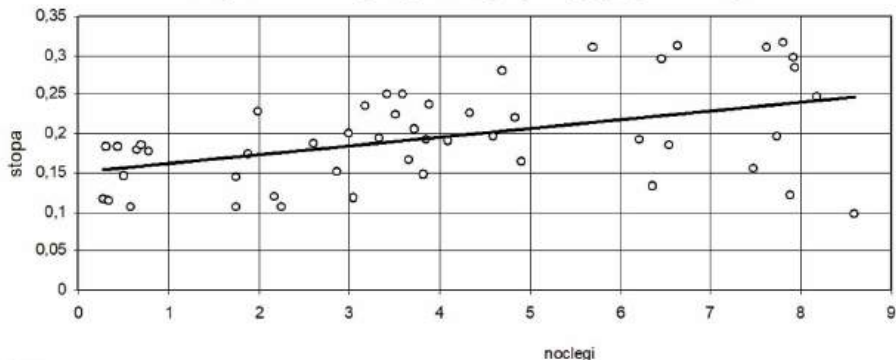
$$\sigma = \sqrt{2^{2r+2} - (2r+1)2^{r+1} - 2}$$

$$P(N_n \geq 1) = 1 - \Phi \left(\frac{(1 - \frac{n}{N_n})\mu^{3/2}}{\sigma\sqrt{n}} \right)$$

- szacujemy prawdopodobieństwo pojawienia się serii o długości r ;
- sprawdzamy, czy $P(N_n \geq 1) < 0,05$ (poziom istotności) jeśli tak to brak podstaw do odrzucenia hipotezy o losowości reszt.

¹³³ Dziechciarz J. (red.), *Ekometria. Metody, przykłady, zadania*, op. cit., rozdział 6.6.

Obserwacje odróżniające się od pozostałych (nietypowe, wpływowe). Interesują nas skutki oddziaływania określonych obserwacji na model ekonometryczny. Nietypowe cechują duże reszty, a po ich wyeliminowaniu następuje poprawa współczynnika determinacji R^2 , natomiast wpływowych wyraźna zmiana parametrów modelu. Do wykrywania obserwacji wpływowych stosujemy analizę wykresu rozrzutu obserwacji dla modelu z jedną zmienną objaśniającą, np. z "noclegi".



Źródło: Opracowanie własne w Excelu.

Rys. 13.2. Wykres rozrzutu obserwacji modelu $\text{bezrobot} = f(\text{noclegi})$

Wpływowe są obserwacje najbardziej oddalone od linii regresji $\text{bezrobot} = f(\text{noclegi})$.

Dla wielu zmiennych objaśniających m modelu ekonometrycznego mamy następującą liczbę wykresów¹³⁴:

$$\binom{m}{2} = \frac{m(m-1)}{2}$$

W dalszej części bazujemy na przykładzie modelu liniowego stopy bezrobocia "bezrobot" w powiatach o trzech zmiennych objaśniających:

regon - liczba jednostek regon na 1 tys. mieszkańców,

lud_prod - udział ludności w wieku produkcyjnym,

noclegi - liczba miejsc noclegowych na 1 tys. mieszkańców.

Analizę graficzną poszczególnych zmiennych stosuje się biorąc pod uwagę znaczne odbieganie obserwacji od pozostałych. Mogą jednak być obserwacja której wpływowość możemy zaobserwować dopiero w przestrzeni wielowymiarowej.

Obserwacje wpływowe wykrywane są stosując klasyfikacje wielowymiarową, przykładowo metodą taksonomii wrocławskiej. Wówczas obserwacje dzielone są na klasy, przy czym obserwacje wpływowe przydzielane są do osobnych klas biorąc znaczne oddalenie od pozostałych obserwacji¹³⁵. Jednak podstawowym narzędziem do wykrywania obserwacji wpływowych jest macierz rzutowania, która zależy tylko od macierzy zmiennych objaśniających:

$$H = X[X^T X]^{-1} X^T$$

¹³⁴ Według rozdziału 6.7 publikacji: Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, op. cit.

¹³⁵ Bliższe omówienie tego występuje w podrozdziale "Taksonomia wrocławska" publikacji: Jajuga K., *Ekonometria. Metody i analiza problemów ekonomicznych*, op. cit.

$$\hat{y} = Hy = \sum_{j=1}^n h_{ij} y_j$$

Ponieważ dla modelu liniowego $y^{\wedge} = Xa = X[X^T X]^{-1} X^T y$, tak więc:

Wielkość i -tego elementu h_i głównej przekątnej macierzy H określa wpływ i -tej obserwacji na parametry modelu. Korzystamy z poniższej formuły, aby zaobserwować jak zmienia się wektor parametrów a pod wpływem usunięcia z danych i -tej obserwacji o (y_i, x_i) ¹³⁶:

Gdzie e_i i -ta reszta dla obserwacji n określona dla obserwacji modelu oszacowanego KMNK, gdzie: $e_i/(1-h_i)$ - współczynnik proporcjonalności.

$$a(i) = a - [X^T X]^{-1} x_i^T \frac{e_i}{1-h_i}$$

Usunięcie obserwacji dla której h_i jest duże ma wpływ na estymację. Dla modeli z jedną zmienną objaśniającą elementy głównej macierzy H wskazują jak określona obserwacja (y_i, x_i) różni się od średniej:

$$h_i = \frac{1}{n} + \frac{(x_i - \bar{x})^2}{\sum_{j=1}^n (x_j - \bar{x})^2}$$

W metodzie macierz rzutowania wprowadzono pojęcie kwadrat odległości Mahalanobisa:

$$MD_i^2 = (n-1)(h_i - \frac{1}{n})$$

Najmniejsza wartość $h_i = 1/n$, występuje gdy obserwacja równa się średniej, natomiast gdy wartości h_i są bliskie jedności to obserwacja jest bardzo oddalona od średniej i jednocześnie od wszystkich innych obserwacji. Zatem wartości diagonalne macierzy rzutowania pozwalają wyłonić obserwacje wpływowe¹³⁷. Skoncentrujemy się jednak na naszym przykładzie bazującym na pliku o 50. obserwacjach stopy bezrobocia w powiatach. Dla takiej bazy oraz modelu liniowego o trzech zmiennych objaśniających parametry określone wcześniej funkcją *REGLINP* są następujące:

Tab. 13.11. Parametry modelu zmiennej *Stopa bezrobocia*

a_3	a_2	a_1	a_0
0,0196761	1,9008	-0,00129131	-0,945176

Źródło: Opracowanie własne w Excelu.

Przypomnijmy sobie model liniowy naszego przykładu:

$$Y^{\wedge} = -0,945176 - 0,00129131 * (\text{regon}) + 1,9008 * (\text{lud_prod}) + 0,0196761 * (\text{noclegi})$$

Postępowanie dla analitycznego wykrycia obserwacji wpływowych metodą macierzy rzutowania jest następujące:

- określenie h^* , czyli krytycznego poziomu h_i :

- wartość przeciętna $h_i = m/n = 3/50 = 0,06$, przy czym $h^* = 2(m/n) = 0,12$ lub

$h^* = 3(m/n) = 0,18$; przyjmijmy $h^* = 0,18$. Alternatywnym sposobem jest podzielenie przedziału $[0,1]$ dla h_i na trzy części informujące o „statusie odległościowym” obserwacji: $h_i \in [1/n; 0,2]$ bezpieczne, $h_i \in [1/n; 0,2]$ ryzykowne, $h_i \in [1/n; 0,2]$ niedopuszczalne.

¹³⁶ Wzór 6.14 publikacji podanej w: Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, op. cit.

¹³⁷ Przykład 6.15 realizacji graficznej oraz analitycznej wyłonienia obserwacji wpływowych dla modelu z jedną zmienną objaśniającą znajduje się w rozdziale 6.7 publikacji podanej w: Dziechciarz J. (red.), *Ekonometria. Metody, przykłady, zadania*, op. cit.

- o wyznaczenie macierzy rzutowania H o wymiarach (50×50) , po wcześniejszym określeniu poniższych macierzy z uwzględnieniem zmiennej dla stałej tożsamościowo równej 1:

X - o wymiarach (50×4) ,

X^T , czyli macierzy transponowanej o wymiarach (4×50) .

Prace te wykonano w *Excelu* korzystając z funkcji: TRANSPONUJ(...), MACIERZ.ILO-CZYN(...), MACIERZ.ODW(...).

Obszary macierzy X oraz X^T są duże i dlatego poniżej zamieszczono tylko iloczyn i odwrotność macierzy $[X^T X]$, a dla innych działań podano w nawiasach obszary macierzy:

$$[X^T X] \quad (4 \times 50) * (50 \times 4) = (4 \times 4),$$

Tab. 13.12. Iloczyn macierzy $X^T X$

50	3909,075505	30,588462	200,360680
3909,075505	325298,687567	2409,730620	14347,390431
30,588462	2409,730620	18,747590	120,266571
200,360680	14347,390431	120,2665707	1118,924083

Źródło: Opracowanie własne w *Excelu*.

$$[X^T X]^{-1} \quad (4 \times 4),$$

Tab. 13.13. Odwrotność macierzy $(X^T X)^{-1}$

28,192225	0,022077	-47,136638	-0,264898
0,022077	0,000101	-0,049324	0,000059
-47,136638	-0,049324	80,832990	0,384740
-0,264898	0,000059	0,384740	0,006219

Źródło: Opracowanie własne w *Excelu*.

$$X [X^T X]^{-1} \quad (50 \times 4) * (4 \times 4) = (50 \times 4); \quad X [X^T X]^{-1} X^T \quad (50 \times 4) * (4 \times 50) = (50 \times 50)$$

Tak więc otrzymaliśmy macierz rzutowania o wymiarach (50×50) , z zaokrąglenia do dwóch miejsc po przecinku. Wyeliminowano obserwacje według wstępnego kryterium o $h_i < 0,10$. W wyniku tego działania pozostało 14-cie obserwacji, a w tym dwie o wartości $h_i > 0,18$ (patrz tab. 13.14).

Tab.13.14. Zweryfikowana macierz rzutowania

0,11	0,01	-0,04	-0,03	-0,02	-0,09	0,04	0,10	-0,01	-0,07	0,07	0,11	0,09	0,08	3	1
0,01	0,12	0,08	-0,05	0,12	-0,04	0,01	0,02	0,09	-0,03	0,09	0,01	0,00	-0,10	6	2
-0,04	0,08	0,13	0,02	0,14	0,09	0,08	0,00	0,04	0,06	0,03	0,00	0,01	-0,10	8	3
-0,03	-0,05	0,02	0,10	-0,02	0,14	0,03	-0,02	-0,03	0,11	-0,07	-0,02	0,00	0,06	11	4
-0,02	0,12	0,14	-0,02	0,18	0,05	0,09	0,02	0,06	0,03	0,06	0,01	0,02	-0,13	13	5
-0,09	-0,04	0,09	0,14	0,05	0,24	0,09	-0,05	-0,04	0,19	-0,10	-0,04	-0,01	0,00	14	6
0,04	0,01	0,08	0,03	0,09	0,09	0,17	0,09	-0,07	0,04	0,00	0,10	0,11	0,00	16	7
0,10	0,02	0,00	-0,02	0,02	-0,05	0,09	0,11	-0,04	-0,05	0,06	0,12	0,11	0,06	20	8
-0,01	0,09	0,04	-0,03	0,06	-0,04	-0,07	-0,04	0,13	-0,01	0,07	-0,05	-0,06	-0,06	25	9
-0,07	-0,03	0,06	0,11	0,03	0,19	0,04	-0,05	-0,01	0,15	-0,07	-0,04	-0,02	0,02	28	10
0,07	0,09	0,03	-0,07	0,06	-0,10	0,00	0,06	0,07	-0,07	0,11	0,05	0,03	-0,03	33	11
0,11	0,01	0,00	-0,02	0,01	-0,04	0,10	0,12	-0,05	-0,04	0,05	0,13	0,12	0,07	37	12
0,09	0,00	0,01	0,00	0,02	-0,01	0,11	0,11	-0,06	-0,02	0,03	0,12	0,12	0,07	44	13
0,08	-0,10	-0,10	0,06	-0,13	0,00	0,00	0,06	-0,06	0,02	-0,03	0,07	0,07	0,19	45	14
3	6	8	11	13	14	16	20	25	28	33	37	44	45		

Źródło: Opracowanie własne w *Excelu*.

- o w wyniku przeanalizowania elementów $h_i > 0,18$, głównej przekątnej macierzy H , usuwamy i -tą obserwację o największym h_i ; jest nią obserwacja czternasta;
- o ponowna estymacja parametrów modelu funkcją *REGLINP* i znowu badanie jak wyżej;
- o jeśli $h_i \leq h^*$ to wówczas kończymy proces iteracyjny eliminacji obserwacji wpływowych.

Po usunięciu obserwacji 14. z macierzy X i ponownym obliczeniu macierzy rzutowania o wymiarach (49x49) na głównej przekątnej wystąpiły dwie obserwacje (28 oraz 45) o $h_i = 0,19$. Ponieważ mają tę samą wartość to usuniemy je jednocześnie z macierzy X . Przystąpmy w kolejnej iteracji do ponownego obliczenia skorygowanej macierzy rzutowania o wymiarach (47x47). Pojawiły się teraz 2 obserwacje o $h_i > 0,18$: 13 ($h_i = 0,21$), 16 ($h_i = 0,19$).

Dla przyspieszenia obliczeń usuwamy zarówno obserwację 13. jak i 16. Uzyskujemy macierz rzutowania w której dla jednej obserwacji tj. ósmej wartość $h_i = 0,21$. Ze względu na łamy niniejszego materiału pozostawiamy Czytelnikowi dalsze badanie, aż do doprowadzenia do warunku $h_i \leq 0,18$. Potraktujmy jednak, że w przybliżeniu spełniony jest warunek bezpieczny $h_i \in [1/n; 0,2]$ i dokonujemy ponownej estymacji modelu liniowego 45. obserwacji po trzech iteracjach funkcją *REGLINP*.

Tab. 13.15. Parametry modelu o 45-ciu obserwacjach

a_3	a_2	a_1	a_0	
0,020603	1,724477	-0,000944	-0,864073	parametry
0,004015	0,505905	0,000623	0,291230	błąd standardowy
0,424877	0,047287			R^2 ; S_e
10,096372	41			F ; iss
0,067729	0,091680			ESS; RSS

Źródło: Opracowanie własne w *Excelu*.

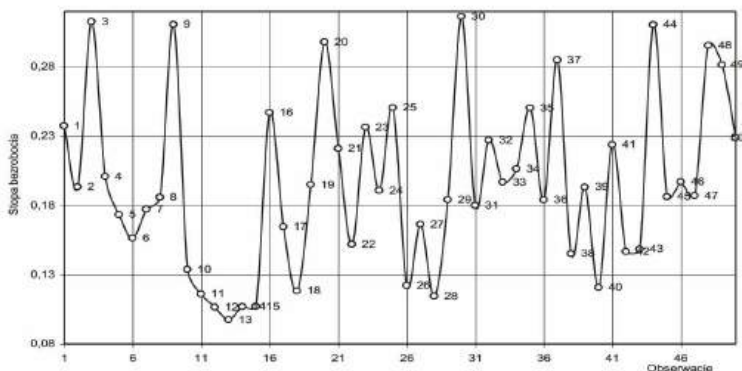
A teraz przypomnijmy sobie wcześniej określony model liniowy o następujących parametrach i statystyce opisowej.

Tab. 13.16. Parametry modelu o 50. obserwacjach

a_3	a_2	a_1	a_0	
0,0196761	1,9008	-0,00129131	-0,945176	parametry
0,00373272	0,425549	0,000474643	0,251316	błąd standardowy
0,450304	0,047332			R^2 ; S_e
12,5609	46			F ; iss
0,08442122	0,103055			ESS; RSS

Źródło: Tabela 13.11.

Według Dziechciarza (rozdział 6.7) „Obserwacje uważa się za wpływowe, jeśli w wyniku ... usunięcia jej z modelu znacznie zmieniają się oszacowane parametry modelu”¹³⁸.

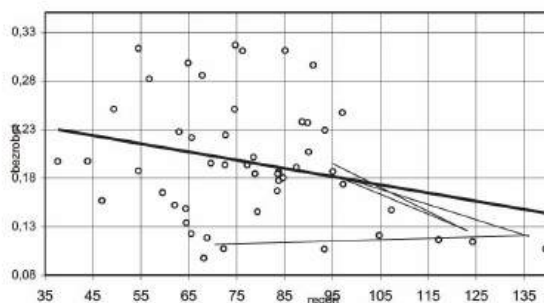


Źródło: Opracowanie własne w Excelu.

Rys. 13.3. Wartości empiryczne stopy bezrobocia (*bezrobot*) w 50. powiatach.

Porównanie nieznacznej zmiany współczynnika determinacji R^2 oraz wyraźnej zmiany parametrów a_0 , a_1 , a_2 przekonuje nas, że osunięte obserwacje 14, 28, 45, 13 oraz 16 były obserwacjami wpływowymi. Wykres punktowy wartości zmiennej objaśnianej w kolejnych obserwacjach (powiatach) podano na rysunku 13.3. Wykres jest tylko orientacją umiejscowienia wartości stopy bezrobocia

Zobaczmy jeszcze umiejscowienie obserwacji uznanych za wpływowe na wykresie punktowym względem średniej - 0,20. Wpływ zmiennych objaśniających dla obserwacji uznanych za wpływowe należy bowiem jak już wspomniano rozpatrywać w przestrzeni trójwymiarowej, przykładowej zmiennej *regon*.



Źródło: Opracowanie własne w Excelu.

Rys. 13.4. Wartości empiryczne zmiennej *regon* w 50. powiatach

Przykład wskazuje na niewystępowanie wyraźnej obserwacji nietypowej, która była by znacznie oddalona od innych obserwacji, lecz w przybliżeniu leżała wzdłuż w linii regresji. Natomiast z punktu widzenia jednej zmiennej objaśniającej występują obserwacje wpływowe (np. 13. o wartości

¹³⁸ Dziechciarz J. (red.), *Ekometria. Metody, przykłady, zadania*, op. cit.

regon $\approx 68,22$) znacznie oddalone równolegle od linii regresji. Otrzymane wcześniej według iteracji macierzy rzutowania obserwacje wyeliminowane (13, 14, 16, 28, 45) połączono linią ciągłą. Wzrostowa analiza ich wpływu na strukturę modelu budzi wątpliwości. Pamiętać należy jednak, że w przeszerzeniu trzywymiarowej wpływ na model ekonometryczny jest inny.

13.5. Zastosowanie *GRET*

Jak już wspomniano na wstępie pakiet *GRET* jest bezpłatnym dostępnym w Internecie oprogramowaniem do estymacji i weryfikacji modeli ekonometrii. Jego zastosowanie zaprezentowane zostanie na przykładzie niektórych miar oceny weryfikacji otrzymanego modelu ekonometrycznego. Ocena istotności statystycznej parametrów modelu o procedurze:

- określenie zmiennych niezależnych X_j (objaśniających) oraz zmiennej objaśnianej Y ,
- zastosowanie testu *t*-Studenta istotności parametru α_j . Hipoteza zerowa zakłada brak istotności statystycznej danego parametru $H_0: \alpha_j = 0$, natomiast hipoteza alternatywna $H_1: \alpha_j \neq 0$. Przypomnijmy sobie statystykę, która określona jest wzorem:

$$t_j = \frac{a_j}{S(a_j)}$$

- oszacowanie parametrów liniowego modelu ekonometrycznego na bazie danych dotyczących stopy bezrobocia w 50. powiatach, lecz rozszerzonej o pięć dalszych zmiennych objaśniających¹³⁹:

zatr_przem - udział zatrudnionych w przemyśle (sekcja C+D+E+F),

sklep_zat - liczba zatrudnionych na 1 sklep,

mieszk_new - liczba mieszkań oddanych do użytku na 1 tys. mieszkańców,

ha_urol - udział użytkowników rolnych w powierzchni powiatu,

lud_250tys - zmienna 0-1, przy czym 1- miasto o liczbie ludności powyżej 250 tys., 0 - pozostałe.

Zmienne objaśniające *regon*, *lud_prod*, *noclegi* już były w naszej wcześniej omawianej bazie stosowanej do obliczeń w *Excelu*. Po zdefiniowaniu struktury bazy w programie *GRET* korzystamy z menu: *Model/Klasyczna metoda najmniejszych kwadratów* i rezultatem jest *Model 1*, estymacji wraz ze statystyką opisową, dla zmiennej zależnej *bezrobot*, czego fragment podano poniżej.

Tab. 13.17. Parametry i statystyka opisowa *Modelu 1*

Zmienna	Współczynnik	Błąd stand.	Statystyka <i>t</i>	Wartość <i>p</i>	
const	-0,763574	0,285219	-2,6772	0,01063	**
regon	-0,00112881	0,000533592	-2,1155	0,04051	**
zatr_przem	-0,0900188	0,064573	-1,3941	0,17081	
sklep_zat	-0,0512426	0,0330824	-1,5489	0,12908	
mieszk_new	-0,00520729	0,00518114	-1,0050	0,32077	
lud_prod	1,91159	0,451284	4,2359	0,00013	***
ha_urol	-0,0299996	0,0495286	-0,6057	0,54805	
noclegi	0,0126064	0,00492128	2,5616	0,01419	**
lud_250tys	-0,0255601	0,0345797	-0,7392	0,46402	

Średnia arytmetyczna zmiennej zależnej = 0,195567

Odchylenie standardowe zmiennej zależnej = 0,061855

¹³⁹ Według zbioru powiaty2002.gdt dostępnego na stronie www.kufel.torun.pl.

Suma kwadratów reszt = 0,0909375

Błąd standardowy reszt = 0,0470955

Współczynnik determinacji $R^2 = 0,514938$

Skorygowany $R^2 = 0,420292$

Statystyka $F(8, 41) = 5,44066$ (wartość $p = 0,000107$)

Wyłączając stałą, największa wartość p jest dla zmiennej *ha_urol*.

W tabeli dla szybkiej orientacji symbolami gwiazdek program oznacza parametry istotnie różniące się od zera przy poziomach istotności: 1% - ***, 5% - **, 10% - *. Na zakończenie raportu program proponuje zmienną do usunięcia, która charakteryzuje się wysokim prawdopodobieństwem popełnienia błędu p w konstrukcji modelu z tą zmienną oraz niską statystyką t .

Źródło: Opracowanie własne programem GRETL.

W następnym kroku dokonujemy usunięcia zmiennej *ha_urol* ze struktury modelu 1 i dokonanie ponownej estymacji otrzymując kolejny Model 2.

Tab.13.18. Parametry i statystyka opisowa Modelu 2

Zmienna	Współczynnik	Błąd stand.	Statystyka t	Wartość p	
const	-0,837147	0,256108	-3,2687	0,00216	***
regon	-0,00106249	0,000518285	-2,0500	0,04664	**
zatr_przem	-0,0911435	0,064058	-1,4228	0,16217	
sklep_zat	-0,0462816	0,0318099	-1,4549	0,15312	
mieszk_new	-0,00578826	0,00505306	-1,1455	0,25849	
lud_prod	1,97947	0,433837	4,5627	0,00004	***
noclegi	0,0130892	0,00481956	2,7158	0,00956	***
lud_250tys	-0,0241506	0,0342403	-0,7053	0,48450	

Średnia arytmetyczna zmiennej zależnej = 0,195567

Odchylenie standardowe zmiennej zależnej = 0,061855

Suma kwadratów reszt = 0,0917513

Błąd standardowy reszt = 0,0467392

Współczynnik determinacji $R^2 = 0,510598$

Skorygowany $R^2 = 0,429031$

Statystyka $F(7, 42) = 6,25986$ (wartość $p = 4,69e-005$)

Wyłączając stałą, największa wartość p jest dla zmiennej *lud_250tys*.

Źródło: Opracowanie własne w GRETL.

Postępując analogicznie jak dla Modelu 1 eliminujemy teraz zmienną *lud_250tys* i ponownie szacujemy parametry. Program podpowiada nam w kolejnych iteracjach modeli wyeliminowanie zmiennych objaśniających: Model 3 - *mieszk_new*, Model 4 - *zatr_przem*, Model 5 - *sklep_zatr*. Model 6 zawiera wszystkie zmienne łącznie ze stałą oznaczone trzema gwiazdkami. Tak więc parametry tego modelu, a przez nie zmienne objaśniające są statystycznie istotne nawet na poziomie istotności 1% i nadają się do praktycznego wykorzystania, np. do celów prognozowania.

Tab. 13.19. Parametry oraz statystyka opisowa Modelu 6

Zmienna	Współczynnik	Błąd stand.	Statystyka t	Wartość p	
const	-0,945176	0,251316	-3,7609	0,00048	***
regon	-0,00129131	0,000474643	-2,7206	0,00917	***
lud_prod	1,9008	0,425549	4,4667	0,00005	***
noclegi	0,0196761	0,00373272	5,2712	<0,00001	***

Średnia arytmetyczna zmiennej zależnej = 0,195567
 Odchylenie standardowe zmiennej zależnej = 0,061855
 Suma kwadratów reszt = 0,103055
 Błąd standardowy reszt = 0,0473321
 Współczynnik determinacji $R^2 = 0,450304$
 Skorygowany $R^2 = 0,414454$
 Statystyka $F(3, 46) = 12,5609$ (wartość $p < 0,00001$)

Źródło: Opracowanie własne w *GRET*L.

Warto zauważyć, że w efekcie naszej analizy parametrów modelu uzyskaliśmy programem *GRET*L postać modelu identyczną jak wcześniej w *Excelu*.

$$\hat{Y} = -0,945176 - 0,00129131 * (\text{region}) + 1,9008 * (\text{lud_prod}) + 0,0196761 * (\text{noclegi})$$

Całościowa ocena przydatności modelu

Stosujemy statystykę F , dla której podany w "Modele jednorównaniowe" wzór możemy zapisać również w postaci:

$$F_{obl} = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)}$$

Gdzie: R^2 - współczynnik determinacji, m - ilość zmiennych objaśniających.

Hipoteza zerowa zakłada nieistotność statystyczną każdego z parametrów modelu:

$$H_0 : \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

Dla naszego *Modelu 6* przyjmijmy poziom istotności $\alpha = 0,05$, $m = 3$, $n - m - 1 = 50 - 3 - 1 = 46$. Statystyka Fishera, czyli F_{obl} podana w tabeli 13.19 przy bardzo małym prawdopodobieństwie błędu p wynosi:

$$F(3,46) = 12,5609$$

Tak więc kierując się tylko p uznajemy wszystkie zmienne *Modelu 6* jako istotne statystycznie. Innym sposobem jest porównanie F_{obl} z wartością krytyczną wyznaczoną z tablicy rozkładu F według $F(\alpha, m, n-m-1)$, a która wynosi $F^* = 2,81$. Ponieważ $F_{obl} > F^*$ to odrzucamy hipotezę zerową o nieistotności całego zestawu parametrów modelu.

Ocena stopnia dopasowania modelu

Na podstawie statystyki opisowej dołączonej do *Modelu 6* (patrz tabela 13.19) możemy wstępnie określić dopasowanie tego modelu do danych rzeczywistych stosując miary:

- błąd standardowy reszt $S_e = 0,0473321$, średnia arytmetyczna zmiennej zależnej $\bar{y} = 0,195567$;

- współczynnik zmienności resztowej; przy czym 10% to przyjęta wartość graniczna.

$$V_e = \frac{S_e}{\bar{y}} * 100 = \frac{0,0473321}{0,195567} * 100 = 24,2\% > V_e^* = 10\%$$

- współczynnik determinacji $R^2 = 0,450304$ oraz pomocniczo współczynnik zbieżności, czyli stopień nie wyjaśnienia zmienności Y przez oszacowany *Model 6*

$$\Phi^2 = 1 - R^2 = 1 - 0,450304 = 0,549696$$

- skorygowany $R^2 = 0,414454$, współczynnik ten stosujemy dla modeli o różnej liczbie obserwacji i oszacowanych parametrach; porównajmy zatem *Model 1* i *Model 6* o ośmiu oraz trzech zmiennych objaśniających: skorygowany $R^2_1 = 0,420292$; skorygowany $R^2_6 = 0,414454$.

Nastąpiło nieznaczne zmniejszenie, choć *Model 6* ma właściwie dobrzebrane zmienne objaśniające, a także *Modelu 6* o 50-ci obserwacjach i modelu z wyeliminowanymi pięcioma obserwacjami wpływowymi i $R^2 = 0,424877$ (patrz tabela 13.15) dla którego:

$$\bar{R}^2 = 1 - \frac{n-1}{n-m-1}(1-R^2) = 1 - \frac{45-1}{45-3-1}(1-0,424877) = 0,382794$$

W miarę więc eliminacji zmiennych nieistotnych oraz obserwacji wpływowych nastąpił nieznaczny spadek skorygowanego współczynnika determinacji. W tej sytuacji warto się zastanowić nad pobraniem ze zbioru 380 elementowego *stopy bezrobocia* w powiatach liczniejszej próby np. o 100 obserwacji. Pozostawia się jednak to badanie Czytelnikowi.

Ocena normalności składnika resztowego

Powróćmy do naszego *Model 6* i zastosujmy w *GRET*L test Jarque'a-Bery (*JB*) do oceny normalności reszt. Hipoteza zerowa zakłada normalność składnika resztowego. Statystykę *JB* określamy wg wzoru:

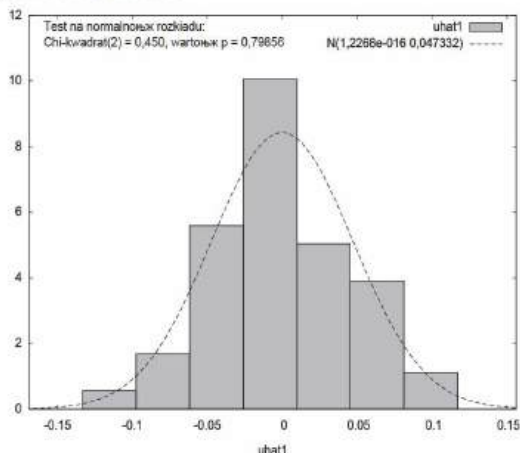
$$JB = n \left[\frac{S^2}{6} + \frac{(K-3)^2}{24} \right]$$

Gdzie: *S* - skośność, *K* - kurtოza.

Test oparty jest na rozkładzie χ^2 z dwoma stopniami swobody. Możemy też skorzystać z menu:

Testy/Test normalności rozkładu reszt

w oknie oszacowanego *Modelu 6*. Pojawi się nam wykres słupkowy testu normalności rozkładu. Kliknięcie na wykresie powoduje pojawienie się menu z którego wybieramy *Kopiuj do schowka*, aby potem wprowadzić do dokumentu Word-a. Wykres oraz raport z badania reszt podano na wykresie (zobacz rysunek 13.5) oraz w tabeli 13.20.



Rozkład częstości dla reszt (*uhat1*), obserwacje 1-50, liczba przedziałów = 7, średnia = 1,2268e-016, odchylenie standardowe = 0,0473321

Źródło: Opracowanie własne w *GRET*L.

Rys. 13.5. Diagram testu normalności rozkładu reszt.

Tab. 13.20. Raport z badania rozkładu reszt w przedziałach

Przedziały	Średnia	Liczba	Częstość	Skumulowana
< -0,097517	-0,11540	1	2,00%	2,00%
-0,097517 -0,061742	-0,079630	3	6,00%	8,00% **
-0,061742 -0,025967	-0,043855	10	20,00%	28,00% *****
-0,025967 0,0098074	-0,0080800	18	36,00%	64,00% *****
0,0098074 0,045582	0,027695	9	18,00%	82,00% *****
0,045582 0,081357	0,063470	7	14,00%	96,00% *****
≥ 0,081357	0,099244	2	4,00%	100,00% *

Test Jarque-Bera'y. H_0 : dystrybucja empiryczna posiada rozkład normalny: Chi-kwadrat(2) = 0,450 z wartością $p = 0,79856$

Źródło: Opracowanie własne w *GRET*L.

W końcowej uwadze program orzeka o normalności rozkładu reszt. Potwierdzeniem jest duże prawdopodobieństwo $p = 0,798556$ co pozwala nam przyjąć hipotezę zerową¹⁴⁰. Ponadto wartość krytyczna statystyki χ^2 dla $\alpha = 0,05$ wynosi 4,605, a określona w *GRET*L równa się 0,450, tak więc: $0,450 < 4,605$ i nie ma podstaw do odrzucenia hipotezy zerowej.

Ocena jednorodności wariancji reszt (heteroskedastyczności)

Program *GRET*L dla oceny jednorodności wariancji składnika losowego proponuje test White'a. W *Modelu 6* o 50-ciu obserwacjach i równaniu podstawowym:

$$\hat{Y} = -0,945176 - 0,00129131 * (\text{region}) + 1,9008 * (\text{lud_prod}) + 0,0196761 * (\text{noclegi})$$

występują, jak już wcześniej sprawdzono macierzą rzutowania, obserwacje wpływowe. Test White'a przyjmuje sprawdzenie istotności regresji wyznaczonej dla $m = 3$ zmiennych objaśniających modelu, ich kwadratami oraz iloczynami¹⁴¹. Powstanie równanie pomocnicze dla wariancji reszt o postaci ogólnej:

$$\sigma_i^2 = \alpha_0 + \alpha_1 x_{i1} + \alpha_2 x_{i2} + \alpha_3 x_{i3} + \alpha_4 x_{i1}^2 + \alpha_5 x_{i1} x_{i2} + \alpha_6 x_{i1} x_{i3} + \alpha_7 x_{i2}^2 + \alpha_8 x_{i2} x_{i3} + \alpha_9 x_{i3}^2 + v_i$$

a hipoteza zerowa

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_9$$

ma statystykę $T \cdot R^2$ o rozkładzie χ^2 z $J = 9$ stopniami swobody przy czym: T - liczba obserwacji, R^2 - współczynnik determinacji dla modelu pomocniczego.

Dla przeprowadzenia testu korzystamy z menu *Testy/test heteroskedastyczności* z okna estymacji *Modelu 6* po estymacji *KMNK*. W wynikach zamiast przecinka występuje kropka dziesiętna. Obraz testu White'a na heteroskedastyczność reszt (zmienność wariancji resztowej) po wprowadzeniu w formacie *RTF* do Word-a podano w tabeli 13.21.

Tab. 13.21. Estymacja parametrów modelu testu White'a (zmienna zależna: $uhat^2$ - kwadraty reszt)

Zmienna	Współczynnik	Błąd standardowy	Statystyka t	Wartość p
const	-0,232181	0,533736	-0,435	0,66589
region	-0,00110049	0,00159292	-0,691	0,49364
lud_prod	0,893611	1,78239	0,501	0,61887
noclegi	0,00681558	0,0134336	0,507	0,61469
sq_region	-1,94509E-06	1,84918E-06	-1,052	0,29917
region_lud_pr	0,00230064	0,00285027	0,807	0,42435

¹⁴⁰ Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, tom II, *Ekonometria*, rozdział 1.91.

¹⁴¹ Kufel T., *Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu GRET*L, rozdział 4.3.4.

regon_nocleg	5,57372E-06	2,06490E-05	0,270	0,78860
sq_lud_prod	-0,868303	1,50452	-0,577	0,56709
lud_pr_nocleg	-0,0114819	0,0221247	-0,519	0,60665
sq_noclegi	-1,55966E-05	0,000122828	-0,127	0,89959

przy czym: *sq* - kwadrat danej zmiennej, a np. **regon_lud_pr** iloczyn zmiennych.

Współczynnik determinacji $R^2 = 0,139724$, statystyka testu: $T \cdot R^2 = 6,986206$, z wartością $p = \chi^2(9) = 0,638556$. Prawdopodobieństwo popełnienia błędu (wartość $p = 0,638556$) wskazuje, że nie ma podstaw do odrzucenia hipotezy H_0 , gdyż $T \cdot R^2 = 50 \cdot 0,139724 = 6,9862 < \chi^2(5\%, 9) = 16,919$ (wg tabeli wartości krytycznych χ^2).

Źródło: Opracowanie własne w *GRET*L.

Występuje zatem jednorodność wariancji i wszystkie nietypowe obserwacje poprawnie opisuje *Model 6*.

*GRET*L ma wbudowane tablice statystyczne. W tym względzie stosujemy dla reszt menu *Narzędzia/Tablice statystyczne* dla otrzymania wartości krytycznej (np. dla $\alpha = 0,05$ oraz $J = 9$), która wynosi 16,919. Program posiada tablice rozkładu normalnego, *t*-Studenta, *F*-Fishera, Durbina-Watsona.

Ocena liniowości

Wykorzystanie liniowej analitycznej postaci naszego *Modelu 6* może stanowić o jego słabości. Test nieliniowości White'a dla postaci logarytmicznej sprawdza zasadność zbudowanego modelu liniowego względem potęgowego o zmiennych zlogarytmowanych.

$$H_0: \gamma_1 = \gamma_2 = \dots = \gamma_j = 0$$

Hipoteza zerowa H_0 : zależność liniowa, natomiast hipoteza alternatywna H_1 : nieliniowa (logarytmiczna).

Równanie pomocnicze odchylenia standardowego wobec równania podstawowego *Modelu 6* w postaci ogólnej jest następujące:

$$\sigma_i = \alpha_0 + \alpha_1 X_{i1} + \alpha_2 X_{i2} + \alpha_3 X_{i3} + \gamma_1 \ln X_{i1} + \gamma_2 \ln X_{i2} + \gamma_3 \ln X_{i3} + u_i$$

Korzystamy w programie *GRET*L z menu *Testy/test nieliniowości* (logarytmy) z poziomu oszacowania *KM*NK *Modelu 6*. Program dokonuje estymacji *KM*NK parametrów pomocniczego równania regresji z wykorzystaniem 50-ciu obserwacji zmiennej zależnej *reszty*. Zastosowany jest test White'a dla nieliniowości z użyciem logarytmów zmiennych. Wyniki *GRET*L-a prezentowane formie raportu zamieszczono w tabeli 13.22.

Tab. 13.22. Parametry równania pomocniczego

Zmienna	Współczynnik	Błąd standardowy	Statystyka <i>t</i>	Wartość <i>p</i>
const	21,8423	8,81987	2,476	0,01728 **
regon	0,000668302	0,00188353	0,355	0,72446
lud_prod	-23,8385	9,60029	-2,483	0,01700 **
noclegi	-0,00183896	0,00664546	-0,277	0,78332
l_regon	-0,0442363	0,145723	-0,304	0,76292
l_lud_prod	14,4515	5,80043	2,491	0,01666 **
l_noclegi	0,00260413	0,0201328	0,129	0,89769

Współczynnik determinacji $R^2 = 0,161275$, statystyka testu $T \cdot R^2$ przy trzech stopniach swobody dla logarytmów zmiennych z wartością $p = 0,0447127$.

Źródło: Opracowanie własne w *GRET*L.

Test nieliniowości dla logarytmów wskazał, że nie mamy podstaw na odrzucenie postaci liniowej, ponieważ iloczyn $T * R^2 = 50 * 0,161275 = 8,06375 < \chi^2(1\%,3) = 11,3449$ na poziomie istotności $\alpha = 0,01$. Wartość krytyczna $\chi^2(3)$ odczytana z tablicy statystycznej w *GRET*L wynosi 11,3449.

Całościowa ocena przydatności modelu

Stosujemy statystykę F , dla której podany w "Modele jednorównaniowe" wzór możemy zapisać również w postaci:

$$F_{obl} = \frac{R^2 / m}{(1 - R^2) / (n - m - 1)}$$

Gdzie: R^2 - współczynnik determinacji, m - ilość zmiennych objaśniających.

Hipoteza zerowa zakłada nieistotność statystyczną każdego z parametrów modelu:

$$H_0: \alpha_1 = \alpha_2 = \dots = \alpha_m = 0$$

Dla naszego Modelu 6 przyjmijmy poziom istotności $\alpha = 0,05$, $m = 3$, $n - m - 1 = 50 - 3 - 1 = 46$. Statystyka Fishera, czyli F_{obl} podana w tabeli 13.19 przy bardzo małym prawdopodobieństwie błędu p wynosi:

$$F(3,46) = 12,5609$$

Tak więc kierując się tylko p uznajemy wszystkie zmienne Modelu 6 jako istotne statystycznie. Innym sposobem jest porównanie F_{obl} z wartością krytyczną wyznaczoną z tablicy rozkładu F wg $F(\alpha, m, n - m - 1)$, a która wynosi $F^* = 2,81$. Ponieważ $F_{obl} > F^*$ to odrzucamy hipotezę zerową o nieistotności całego zestawu parametrów modelu.

Ocena stopnia dopasowania modelu

Na podstawie statystyki opisowej dołączonej do Modelu 6 (patrz tabela 13.19) możemy wstępnie określić dopasowanie modelu do danych rzeczywistych stosując miary:

- błąd standardowy reszt $S_e = 0,0473321$, średnia arytmetyczna zmiennej zależnej $\bar{y} = 0,195567$;
- współczynnik zmienności resztowej: przy czym 10% to przyjęta wartość graniczna.

$$V_e = \frac{S_e}{\bar{y}} * 100 = \frac{0,0473321}{0,195567} * 100 = 24,2\% > V_e^* = 10\%$$

- współczynnik determinacji $R^2 = 0,450304$ oraz pomocniczo współczynnik zbieżności, czyli stopień niewyjaśnienia zmienności Y przez oszacowany Model 6:

$$\Phi^2 = 1 - R^2 = 1 - 0,450304 = 0,549696$$

- skorygowany $R^2 = 0,414454$, współczynnik ten stosujemy dla modeli o różnej liczbie obserwacji i oszacowanych parametrach, porównajmy zatem Model 1 i Model 6 o ośmiu oraz trzech zmiennych objaśniających: skorygowany $R^2_1 = 0,420292$; skorygowany $R^2_6 = 0,414454$; nastąpiło nieznaczne zmniejszenie choć Model 6 ma właściwie dobrane zmienne objaśniające, a także Modelu 6 o 50-ci obserwacjach i modelu z wyeliminowanymi pięcioma obserwacjami wpływowymi i $R^2 = 0,424877$ (patrz tabela 13.15) dla którego:

$$\bar{R}^2 = 1 - \frac{n-1}{n-m-1}(1-R^2) = 1 - \frac{45-1}{45-3-1}(1-0,424877) = 0,382794$$

Zastosowanie komputera na poziomie studiujących ekonometrię może być znacznie szersze niż zaprezentowane powyżej jak i we wcześniejszych rozdziałach. W tym materiale zwrócono jednak

szczególną uwagę na zastosowanie arkusza kalkulacyjnego *Excel* do określenia macierzy rzutowania jako sprawnego narzędzia do wyłonienia obserwacji wpływowych.

Godnym uwagi jest również sposób usuwania z modelu liniowego nieistotnych zmiennych objaśniających z zastosowaniem programu *GRET*L. Warto zauważyć, że w efekcie naszej analizy parametrów modelu uzyskaliśmy programem *GRET*L postać modelu identyczną jak wcześniej w *Excelu*.

$$Y^{\wedge} = -0,945176 - 0,00129131 * (\text{region}) + 1,9008 * (\text{lud_prod}) + 0,0196761 * (\text{noclegi})$$

14. TABLICE STATYSTYCZNE¹⁴²



Nr tablicy	Nazwa
14.1	Dystrybuanta rozkładu normalnego
14.2	Wartości krytyczne $t(P, n)$ rozkładu t - Studenta
14.3	Wartości krytyczne $\chi^2(\alpha, r)$ rozkładu χ^2
14.4	Wartości krytyczne rozkładu Durбина-Watsona dla $\alpha = 0,05$ ¹⁴³
14.5	Wartości krytyczne $F(\alpha, r_1, r_2)$ rozkładu F
14.6	Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

Źródło: Zieliński R., *Tablice statystyczne*, PWN, Warszawa 1972.

¹⁴² Zamieszczone tablice statystyczne pochodzą z opracowania: Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych. Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski, Opole 2010.

¹⁴³ Fragment tabeli z publikacji Lipiec-Zajchowskiej (red.), *Wspomaganie procesów decyzyjnych*, tom II, *Ekonometria*, op. cit. s. 219.

Tab. 14.1. Dystrybuanta rozkładu normalnego

$$\Phi(u) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-u^2/2} du$$

u	0,0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	u
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	0,0
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	0,1
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	0,2
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	0,3
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	0,4
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	0,5
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	0,6
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	0,7
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	0,8
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	0,9
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	1,0
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	1,1
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	1,2
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	1,3
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	1,4
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	1,5
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	1,6
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	1,7
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	1,8
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	1,9
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	2,0
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	2,1
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	2,2
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	2,3
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	2,4
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	2,5
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	2,6

Tab. 14.1. cd. Dystrybuanta rozkładu normalnego

u	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09	u
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	2,7
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	2,8
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	2,9
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	3,0
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	3,1
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	3,2
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	3,3
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	3,4
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	3,5
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	3,6
3,7	0,9 ³ 8 922	0,9 ³ 8 964	0,9 ³ 9 039	0,9 ³ 9 426	0,9 ³ 9 799	0,9 ³ 1 158	0,9 ³ 1 504	0,9 ³ 1 858	0,9 ³ 2 159	0,9 ³ 2 468	3,7
3,8	0,9 ³ 2 765	0,9 ³ 3 052	0,9 ³ 3 327	0,9 ³ 3 593	0,9 ³ 3 848	0,9 ³ 4 094	0,9 ³ 4 331	0,9 ³ 4 558	0,9 ³ 4 777	0,9 ³ 4 988	3,8
3,9	0,9 ³ 5 190	0,9 ³ 5 385	0,9 ³ 5 573	0,9 ³ 5 753	0,9 ³ 5 926	0,9 ³ 6 092	0,9 ³ 6 253	0,9 ³ 6 406	0,9 ³ 6 554	0,9 ³ 6 696	3,9
4,0	0,9 ³ 6 833	0,9 ³ 6 964	0,9 ³ 7 090	0,9 ³ 7 211	0,9 ³ 7 327	0,9 ³ 7 439	0,9 ³ 7 546	0,9 ³ 7 649	0,9 ³ 7 748	0,9 ³ 7 843	4,0
4,1	0,9 ³ 7 934	0,9 ³ 8 022	0,9 ³ 8 106	0,9 ³ 8 186	0,9 ³ 8 263	0,9 ³ 8 338	0,9 ³ 8 409	0,9 ³ 8 477	0,9 ³ 8 542	0,9 ³ 8 605	4,1
4,2	0,9 ³ 8 665	0,9 ³ 8 723	0,9 ³ 8 778	0,9 ³ 8 832	0,9 ³ 8 882	0,9 ³ 8 931	0,9 ³ 8 978	0,9 ³ 9 026	0,9 ³ 9 071	0,9 ³ 9 116	4,2
4,3	0,9 ³ 1 460	0,9 ³ 1 837	0,9 ³ 2 199	0,9 ³ 2 545	0,9 ³ 2 876	0,9 ³ 3 193	0,9 ³ 3 497	0,9 ³ 3 788	0,9 ³ 4 066	0,9 ³ 4 332	4,3
4,4	0,9 ³ 4 587	0,9 ³ 4 831	0,9 ³ 5 065	0,9 ³ 5 288	0,9 ³ 5 502	0,9 ³ 5 706	0,9 ³ 5 902	0,9 ³ 6 089	0,9 ³ 6 268	0,9 ³ 6 439	4,4
4,5	0,9 ³ 6 602	0,9 ³ 6 759	0,9 ³ 6 908	0,9 ³ 7 051	0,9 ³ 7 187	0,9 ³ 7 318	0,9 ³ 7 442	0,9 ³ 7 561	0,9 ³ 7 675	0,9 ³ 7 784	4,5
4,6	0,9 ³ 7 888	0,9 ³ 7 987	0,9 ³ 8 081	0,9 ³ 8 172	0,9 ³ 8 258	0,9 ³ 8 340	0,9 ³ 8 419	0,9 ³ 8 494	0,9 ³ 8 566	0,9 ³ 8 634	4,6
4,7	0,9 ³ 8 699	0,9 ³ 8 761	0,9 ³ 8 821	0,9 ³ 8 877	0,9 ³ 8 931	0,9 ³ 8 983	0,9 ³ 9 030	0,9 ³ 9 078	0,9 ³ 9 125	0,9 ³ 9 161	4,7
4,8	0,9 ³ 2 067	0,9 ³ 2 453	0,9 ³ 2 822	0,9 ³ 3 173	0,9 ³ 3 508	0,9 ³ 3 827	0,9 ³ 4 131	0,9 ³ 4 420	0,9 ³ 4 696	0,9 ³ 4 958	4,8
4,9	0,9 ³ 5 208	0,9 ³ 5 446	0,9 ³ 5 673	0,9 ³ 5 889	0,9 ³ 6 094	0,9 ³ 6 289	0,9 ³ 6 475	0,9 ³ 6 652	0,9 ³ 6 821	0,9 ³ 6 981	4,9

Tab. 14.2. 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	P/n
1	1,0000	3,0777	6,3138	12,7062	31,8205	63,6567	127,3213	318,3088	636,6192	1
2	0,8165	1,8856	2,9200	4,3027	6,9646	9,9248	14,0890	22,3271	31,5991	2
3	0,7649	1,6377	2,3534	3,1824	4,5407	5,8409	7,4533	10,2145	12,9240	3
4	0,7407	1,5332	2,1318	2,7764	3,7469	4,6041	5,5976	7,1732	8,6103	4
5	0,7267	1,4759	2,0150	2,5706	3,3649	4,0321	4,7733	5,8934	6,8688	5
6	0,7176	1,5398	1,9432	2,4469	3,1427	3,7074	4,3168	5,2076	5,9588	6
7	0,7111	1,4149	1,8946	2,3646	2,9980	3,4995	4,0293	4,7853	5,4079	7
8	0,7064	1,3968	1,8595	2,3060	2,8965	3,3554	3,8325	4,5008	5,0413	8
9	0,7027	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2969	4,7809	9
10	0,6998	1,3732	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869	10
11	0,6974	1,3634	1,7959	2,2010	2,7181	3,1058	3,4966	4,0247	4,4370	11
12	0,6955	1,3562	1,7823	2,1788	2,6810	3,0545	3,4284	3,9296	4,3178	12
13	0,6938	1,3502	1,7709	2,1604	2,6503	3,0123	3,3725	3,8520	4,2208	13
14	0,6924	1,3450	1,7613	2,1448	2,6245	2,9768	3,3257	3,7884	4,1405	14
15	0,6912	1,3406	1,7530	2,1314	2,6025	2,9457	3,2860	3,7328	4,0728	15
16	0,6901	1,3368	1,7459	2,1199	2,5835	2,9208	3,2520	3,6862	4,0150	16
17	0,6892	1,3334	1,7396	2,1098	2,5669	2,8982	3,2224	3,6458	3,9651	17
18	0,6884	1,3304	1,7341	2,1009	2,5524	2,8784	3,1966	3,6105	3,9216	18
19	0,6876	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5794	3,8834	19
20	0,6870	1,3253	1,7247	2,0860	2,5280	2,8453	3,1534	3,5518	3,8495	20
21	0,6864	1,3232	1,7207	2,0796	2,5176	2,8314	3,1352	3,5272	3,8193	21
22	0,6858	1,3212	1,7171	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921	22
23	0,6853	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676	23
24	0,6848	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,4668	3,7455	24
25	0,6844	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251	25
26	0,6840	1,3150	1,7056	2,0555	2,4786	2,7787	3,0669	3,4350	3,7066	26
27	0,6837	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896	27
28	0,6834	1,3125	1,7911	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739	28
29	0,6830	1,3114	1,6991	2,0452	2,4620	2,7564	3,0380	3,3962	3,6594	29
30	0,6828	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460	30
32	0,6822	1,3086	1,6939	2,0369	2,4487	2,7385	3,0149	3,3653	3,6218	32
34	0,6818	1,3070	1,6909	2,0322	2,4411	2,7284	3,0020	3,3479	3,6007	34
36	0,6814	1,3055	1,6883	2,0281	2,4345	2,7195	2,9905	3,3326	3,5821	36
38	0,6810	1,3042	1,6890	2,0244	2,4286	2,7116	2,9803	3,3190	3,5657	38
40	0,6807	1,3031	1,6839	2,0211	2,4233	2,7045	2,9712	3,3069	3,5510	40
42	0,6804	1,3020	1,6820	2,0181	2,4185	2,6981	2,9630	3,2960	3,5377	42
44	0,6801	1,3011	1,6802	2,0154	2,4141	2,6923	2,9555	3,2861	3,5258	44
46	0,6799	1,3002	1,6787	2,0129	2,4102	2,6870	2,9488	3,2771	3,5150	46
48	0,6796	1,2994	1,6772	2,0106	2,4066	2,6822	2,9426	3,2689	3,5051	48
50	0,6794	1,2987	1,6759	2,0086	2,4033	2,6778	2,9370	3,2614	3,4960	50
55	0,6790	1,2971	1,6730	2,0040	2,3961	2,6682	2,9247	3,2561	3,4764	55
60	0,6786	1,2958	1,6706	2,0003	2,3901	2,6603	2,9146	3,2317	3,4602	60
65	0,6783	1,2947	1,6686	1,9971	2,3851	2,6535	2,9060	3,2204	3,4466	65
70	0,6780	1,2938	1,6669	1,9944	2,3808	2,6479	2,8987	3,2108	3,4350	70
80	0,6776	1,2922	1,6641	1,9901	2,3739	2,6387	2,8870	3,1953	3,4163	80
90	0,6772	1,2910	1,6620	1,9867	2,3685	2,6316	2,8779	3,1833	3,4019	90
100	0,6770	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,1737	3,3905	100
120	0,6765	1,2886	1,6577	1,9799	2,3578	2,6174	2,8599	3,1595	3,3735	120
150	0,6761	1,2872	1,6551	1,9759	2,3515	2,6090	2,8492	3,1455	3,3566	150
200	0,6757	1,2858	1,6525	1,9719	2,3451	2,6006	2,8385	3,1315	3,3398	200
250	0,6755	1,2849	1,6510	1,9695	2,3414	2,5956	2,8322	3,1232	3,3299	250
300	0,6753	1,2844	1,6499	1,9679	2,3388	2,5923	2,8279	3,1176	3,3233	300
400	0,6751	1,2837	1,6487	1,9659	2,3357	2,5882	2,8227	3,1107	3,3150	400
500	0,6750	1,2832	1,6479	1,9647	2,3338	2,5887	2,8195	3,1066	3,3101	500
∞	0,6745	1,2816	1,6449	1,9600	2,3263	2,5758	2,8070	3,0902	3,2905	∞

Tab. 14.3. 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	α/r
1	0,00393	0,005157	0,00393	0,0157	0,00982	0,00393	0,0158	0,0642	0,148	0,275	1
2	0,02100	0,02200	0,0100	0,0201	0,0506	0,103	0,211	0,446	0,713	1,022	2
3	0,0153	0,0243	0,0717	0,115	0,216	0,352	0,584	1,005	1,424	1,869	3
4	0,0639	0,0908	0,207	0,297	0,484	0,711	1,064	1,649	2,195	2,753	4
5	0,158	0,210	0,412	0,554	0,831	1,145	1,610	2,343	3,000	3,655	5
6	0,299	0,381	0,676	0,872	1,237	1,635	2,204	3,070	3,828	4,570	6
7	0,485	0,598	0,989	1,239	1,690	2,167	2,833	3,822	4,671	5,493	7
8	0,710	0,857	1,344	1,645	2,180	2,773	3,490	4,594	5,527	6,423	8
9	0,972	1,153	1,735	2,088	2,700	3,325	4,168	5,380	6,393	7,357	9
10	1,265	1,479	2,156	2,558	3,247	3,940	4,865	6,179	7,267	8,295	10
11	1,587	1,834	2,603	3,053	3,816	4,575	5,578	6,989	8,148	9,237	11
12	1,934	2,214	3,074	3,571	4,404	5,226	6,304	7,807	9,034	10,182	12
13	2,305	2,617	3,565	4,107	5,009	5,892	7,042	8,634	9,926	11,129	13
14	2,697	3,041	4,075	4,660	5,629	6,571	7,790	9,467	10,821	12,079	14
15	3,108	3,483	4,601	5,229	6,262	7,261	8,547	10,307	11,721	13,030	15
16	3,536	3,942	5,142	5,812	6,908	7,962	9,312	11,152	12,624	13,983	16
17	3,980	4,416	5,697	6,408	7,564	8,672	10,085	12,002	13,531	14,937	17
18	4,439	4,905	6,265	7,015	8,231	9,390	10,865	12,857	14,440	15,893	18
19	4,912	5,407	6,844	7,633	8,907	10,117	11,651	13,716	15,352	16,850	19
20	5,398	5,921	7,434	8,260	9,591	10,851	12,443	14,578	16,266	17,809	20
21	5,896	6,447	8,034	8,897	10,283	11,591	13,240	15,445	17,182	18,768	21
22	6,404	6,983	8,643	9,542	10,982	12,338	14,041	16,314	18,101	19,729	22
23	6,924	7,529	9,260	10,196	11,688	13,091	14,848	17,187	19,021	20,690	23
24	7,453	8,085	9,886	10,856	12,401	13,845	15,659	18,062	19,943	21,652	24
25	7,991	8,649	10,520	11,524	13,120	14,611	16,473	18,940	20,867	22,616	25
26	8,538	9,222	11,160	12,198	13,844	15,379	17,292	19,820	21,792	23,579	26
27	9,093	9,803	11,808	12,879	14,573	16,151	18,114	20,703	22,719	24,544	27
28	9,656	10,391	12,461	13,565	15,308	16,928	18,939	21,588	23,647	25,509	28
29	10,227	10,986	13,121	14,256	16,047	17,708	19,768	22,475	24,577	26,475	29
30	10,804	11,588	13,787	14,953	16,971	18,493	20,599	23,364	25,508	27,442	30
31	11,389	12,196	14,458	15,655	17,539	19,281	21,434	24,255	26,440	28,409	31
32	11,979	12,811	15,134	16,362	18,291	20,072	22,271	25,148	27,373	29,376	32
33	12,576	13,431	15,815	17,073	19,047	20,867	23,110	26,042	28,307	30,344	33
34	13,176	14,057	16,501	17,789	19,806	21,664	23,952	26,938	29,242	31,313	34
35	13,788	14,688	17,192	18,509	20,569	22,465	24,797	27,836	30,178	32,282	35
36	14,401	15,324	17,887	19,233	21,336	23,269	25,643	28,735	31,115	33,252	36
37	15,020	15,965	18,586	19,960	22,106	24,075	26,492	29,635	32,053	34,222	37
38	15,644	16,611	19,289	20,691	22,878	24,884	27,343	30,537	32,992	35,192	38
39	16,273	17,262	19,996	21,426	23,654	25,695	28,196	31,441	33,932	36,163	39
40	16,906	17,916	20,707	22,164	24,433	26,509	29,051	32,345	34,872	37,134	40
41	17,544	18,575	21,421	22,906	25,215	27,326	29,907	33,251	35,813	38,105	41
42	18,186	19,238	22,138	23,650	25,909	28,144	30,765	34,157	36,775	39,077	42
43	18,832	19,905	22,859	24,398	26,785	28,965	31,625	35,065	37,698	40,050	43
44	19,482	20,576	23,584	25,148	27,575	29,787	32,487	35,974	38,641	41,022	44
45	20,137	21,251	24,311	25,901	28,366	30,612	33,350	36,884	39,585	41,955	45
46	20,794	21,929	25,041	26,657	29,160	31,439	34,215	37,795	40,529	42,968	46
47	21,456	22,610	25,775	27,416	29,959	32,268	35,081	38,708	41,474	43,942	47
48	22,121	23,295	26,511	28,177	30,755	33,098	35,949	39,621	42,420	44,915	48
49	22,789	23,983	27,249	28,491	31,555	33,930	36,818	40,534	43,366	45,889	49
50	23,461	24,674	27,991	29,707	32,357	34,764	37,689	41,449	44,313	46,864	50
51	24,136	25,368	28,735	30,475	33,162	35,600	38,560	42,365	45,261	47,838	51
52	24,814	26,065	29,481	31,246	33,968	36,437	39,433	43,281	46,209	48,813	52
53	25,495	26,765	30,230	32,018	34,776	37,276	40,308	44,199	47,157	49,788	53
54	26,179	27,468	30,981	32,793	35,586	38,116	41,183	45,117	48,106	50,764	54
55	26,866	28,173	31,735	33,570	36,398	38,958	42,060	46,036	49,054	51,739	55

Tab. 14.3. cd. 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	cd_r
1	0,455	0,708	1,074	1,642	2,706	3,841	5,024	6,636	7,879	10,828	12,116	1
2	1,386	1,833	2,408	3,219	4,605	5,991	7,378	9,210	10,597	13,816	15,202	2
3	2,366	2,946	3,665	4,642	6,251	7,815	9,348	11,345	12,838	16,266	17,730	3
4	3,357	4,045	4,878	5,989	7,779	9,488	11,143	13,277	14,860	18,467	19,997	4
5	4,351	5,132	6,064	7,289	9,236	11,070	12,832	15,086	16,750	20,515	22,105	5
6	5,345	6,211	7,231	8,558	10,645	12,592	14,449	16,812	18,548	22,458	24,103	6
7	6,346	7,283	8,383	9,803	12,017	14,067	16,013	18,475	20,278	24,322	26,018	7
8	7,344	8,351	9,524	11,030	13,362	15,507	17,535	20,090	21,955	26,125	27,868	8
9	8,343	9,414	10,656	12,242	14,684	16,919	19,023	21,666	23,589	27,877	29,666	9
10	9,342	10,473	11,781	13,442	15,987	18,307	20,483	23,209	25,188	29,588	31,420	10
11	10,341	11,530	12,899	14,631	17,275	19,675	21,920	24,725	26,757	31,264	33,136	11
12	11,340	12,584	14,011	15,812	18,549	21,026	23,336	26,217	28,300	32,909	34,821	12
13	12,340	13,636	15,119	16,985	19,812	22,362	24,735	27,688	29,819	34,528	36,478	13
14	13,339	14,685	16,222	18,151	21,064	23,685	26,119	29,141	31,319	36,123	38,109	14
15	14,339	15,733	17,322	19,311	22,307	24,996	27,488	30,578	32,801	37,697	39,719	15
16	15,338	16,780	18,418	20,465	23,542	26,296	28,845	32,000	34,267	39,252	41,308	16
17	16,338	17,824	19,511	21,615	24,769	27,587	30,191	33,409	35,718	40,790	42,879	17
18	17,338	18,868	20,601	22,760	25,989	28,869	31,526	34,805	37,156	42,312	44,434	18
19	18,338	19,910	21,689	23,900	27,204	30,144	32,852	36,191	38,582	43,820	45,973	19
20	19,337	20,951	22,775	25,038	28,412	31,410	34,170	37,566	39,997	45,315	47,498	20
21	20,337	21,991	23,858	26,171	29,615	32,671	35,479	38,932	41,401	46,797	49,010	21
22	21,337	23,031	24,939	27,301	30,813	33,924	36,781	40,289	42,796	48,268	50,511	22
23	22,337	24,069	26,018	28,429	32,007	35,172	38,086	41,638	44,181	49,728	52,000	23
24	23,337	25,106	27,096	29,553	33,196	36,415	39,364	42,980	45,558	51,179	53,479	24
25	24,337	26,143	28,172	30,675	34,382	37,652	40,646	44,314	46,928	52,618	54,947	25
26	25,336	27,179	29,246	31,795	35,563	38,885	41,923	45,642	48,290	54,052	56,407	26
27	26,336	28,214	30,319	32,912	36,741	40,113	43,194	46,963	49,645	55,476	57,858	27
28	27,336	29,249	31,391	34,027	37,916	41,337	44,461	48,278	50,993	56,892	59,300	28
29	28,336	30,283	32,461	35,139	39,087	42,557	45,722	49,588	52,336	58,301	60,735	29
30	29,336	31,316	33,530	36,250	40,256	43,773	46,979	50,892	53,672	59,703	62,162	30
31	30,336	32,349	34,598	37,359	41,422	44,985	48,232	52,191	55,003	61,098	63,582	31
32	31,336	33,381	35,665	38,466	42,585	46,194	49,480	53,486	56,328	62,487	64,995	32
33	32,336	34,413	36,731	39,572	43,745	47,400	50,725	54,776	57,648	63,870	66,402	33
34	33,336	35,444	37,795	40,676	44,903	48,602	51,966	56,061	58,964	65,247	67,803	34
35	34,336	36,475	38,859	41,778	46,059	49,802	53,203	57,342	60,275	66,619	69,199	35
36	35,336	37,505	39,922	42,879	47,212	50,998	54,437	58,619	61,581	67,985	70,588	36
37	36,336	38,535	40,984	43,978	48,363	52,192	55,668	59,892	62,882	69,346	71,972	37
38	37,335	39,564	42,045	45,076	49,513	53,384	56,895	61,162	64,181	70,703	73,351	38
39	38,335	40,593	43,105	46,173	50,660	54,572	58,120	62,428	65,476	72,055	74,725	39
40	39,335	41,622	44,165	47,269	51,805	55,768	59,342	63,691	66,766	73,402	76,095	40
41	40,335	42,651	45,224	48,363	52,949	56,942	60,561	64,950	68,053	74,745	77,459	41
42	41,335	43,679	46,282	49,456	54,090	58,124	61,777	66,206	69,336	76,084	78,820	42
43	42,335	44,706	47,339	50,548	55,230	59,304	62,990	67,459	70,616	77,419	80,176	43
44	43,335	45,734	48,396	51,639	56,369	60,481	64,201	68,709	71,893	78,749	81,528	44
45	44,335	46,761	49,452	52,729	57,505	61,656	65,410	69,957	73,166	80,077	82,876	45
46	45,335	47,787	50,507	53,818	58,641	62,830	66,617	71,201	74,437	81,400	84,220	46
47	46,335	48,814	51,562	54,906	59,774	64,001	67,821	72,443	75,704	82,720	85,560	47
48	47,335	49,840	52,616	55,993	60,907	65,171	69,023	73,683	76,969	84,037	86,897	48
49	48,335	50,866	53,670	57,079	62,038	66,339	70,222	74,919	78,231	85,351	88,231	49
50	49,335	51,892	54,723	58,164	63,167	67,505	71,420	76,154	79,490	86,661	89,561	50
51	50,335	52,917	55,775	59,248	64,295	68,669	72,616	77,386	80,747	87,968	90,881	51
52	51,335	53,942	56,827	60,332	65,422	69,832	73,810	78,616	82,001	89,272	92,211	52
53	52,335	54,967	57,879	61,414	66,548	70,993	75,002	79,843	83,253	90,573	93,531	53
54	53,335	55,992	58,930	62,496	67,673	72,153	76,192	81,069	84,502	91,872	94,849	54
55	54,335	57,016	59,980	63,577	68,796	73,311	77,380	82,292	85,749	93,167	96,163	55

Tab. 14.3. cd. Wartości krytyczne χ^2 rozkładu χ^2

r/a	0,9995	0,999	0,995	0,99	0,975	0,95	0,90	0,80	0,70	0,60	α/r
56	27,555	28,881	32,490	34,350	37,212	39,801	42,937	46,955	50,005	52,715	56
57	28,248	29,591	33,248	35,131	38,027	40,646	43,816	47,876	50,956	53,691	57
58	28,943	30,305	34,008	35,913	38,844	41,492	44,696	48,797	51,909	54,667	58
59	29,640	31,020	34,770	36,698	39,662	42,339	45,577	49,718	52,857	55,643	59
60	30,340	31,738	35,535	37,485	40,482	43,188	46,459	50,641	53,809	56,620	60
61	31,043	32,459	36,301	38,273	41,303	44,038	47,342	51,564	54,761	57,597	61
62	31,748	33,181	37,068	39,063	42,126	44,889	48,226	52,487	55,714	58,574	62
63	32,455	33,906	37,838	39,855	42,950	45,741	49,111	53,412	56,666	59,551	63
64	33,165	34,633	38,610	40,649	43,776	46,595	49,996	54,336	57,619	60,528	64
65	33,877	35,362	39,383	41,444	44,603	47,450	50,883	55,262	58,573	61,506	65
66	34,591	36,093	40,158	42,240	45,431	48,305	51,770	56,188	59,527	62,484	66
67	35,307	36,826	40,935	43,038	46,261	49,162	52,659	57,115	60,481	63,461	67
68	36,025	37,561	41,713	43,838	47,092	50,020	53,548	58,042	61,436	64,440	68
69	36,745	38,298	42,494	44,639	47,924	50,879	54,438	58,970	62,391	65,418	69
70	37,467	39,036	43,275	45,442	48,758	51,739	55,329	59,898	63,346	66,396	70
71	38,192	39,777	44,058	46,246	49,592	52,600	56,221	60,827	64,302	67,375	71
72	38,918	40,520	44,843	47,051	50,428	53,462	57,113	61,756	65,258	68,353	72
73	39,646	41,262	45,629	47,858	51,265	54,325	58,006	62,686	66,214	69,332	73
74	40,376	42,010	46,417	48,666	52,103	55,189	58,900	63,616	67,170	70,311	74
75	41,107	42,757	47,206	49,475	52,942	56,054	59,795	64,547	68,127	71,290	75
76	41,841	43,507	47,997	50,286	53,782	56,920	60,690	65,478	69,084	72,270	76
77	42,576	44,258	48,788	51,097	54,623	57,786	61,586	66,409	70,042	73,249	77
78	43,313	45,010	49,582	51,910	55,466	58,654	62,483	67,341	70,999	74,229	78
79	44,051	45,764	50,376	52,725	56,309	59,522	63,308	68,274	71,957	75,208	79
80	44,791	46,520	51,172	53,540	57,153	60,391	64,278	69,207	72,915	76,188	80
81	45,533	47,277	51,969	54,357	57,998	61,261	65,176	70,140	73,874	77,168	81
82	46,276	48,036	52,767	55,174	58,845	62,132	66,076	71,074	74,833	78,148	82
83	47,021	48,796	53,567	55,993	59,692	63,004	66,976	72,008	75,792	79,128	83
84	47,767	49,557	54,368	56,813	60,540	63,876	67,876	72,943	76,751	80,108	84
85	48,515	50,320	55,170	57,634	61,389	64,749	68,777	73,878	77,710	81,089	85
86	49,264	51,085	55,973	58,456	62,239	65,623	69,679	74,813	78,670	82,069	86
87	50,015	51,850	56,777	59,279	63,089	66,498	70,581	75,749	79,630	83,050	87
88	50,767	52,617	57,582	60,103	63,941	67,373	71,484	76,685	80,590	84,031	88
89	51,521	53,386	58,389	60,928	64,793	68,249	72,387	77,622	81,550	85,012	89
90	52,276	54,155	59,196	61,754	65,647	69,126	73,291	78,558	82,511	85,993	90
91	53,032	54,926	60,005	62,581	66,501	70,003	74,196	79,496	83,472	86,974	91
92	53,790	55,698	60,815	63,409	67,356	70,882	75,101	80,433	84,433	87,955	92
93	54,549	56,472	61,625	64,238	68,211	71,760	76,006	81,371	85,394	88,936	93
94	55,309	57,246	62,437	65,068	69,068	72,640	76,912	82,309	86,356	89,917	94
95	56,070	58,022	63,250	65,898	69,925	73,520	77,818	83,248	87,317	90,899	95
96	56,833	58,799	64,063	66,730	70,783	74,400	78,725	84,187	88,279	91,881	96
97	57,597	59,577	64,878	67,562	71,642	75,282	79,633	85,126	89,241	92,862	97
98	58,362	60,356	65,694	68,396	72,501	76,164	80,541	86,065	90,204	93,844	98
99	59,128	61,136	66,510	69,230	73,361	77,046	81,449	87,005	91,166	94,826	99
100	59,896	61,918	67,328	70,075	74,222	77,929	82,358	87,945	92,129	95,808	100

Funkcje pomocnicze $p_1(\alpha)$ i $p_2(\alpha, \nu)$ ($\nu = 1/\sqrt{2r}$)

ν/α	0,9995	0,999	0,995	0,99	0,975	0,950	0,90	0,80	0,70	0,60	α/ν
0,00	6,652	5,700	3,757	2,941	1,894	1,137	0,428	-0,194	-0,483	-0,624	0,00
0,01	6,537	5,691	3,757	2,945	1,901	1,145	0,436	-0,189	-0,487	-0,622	0,01
0,02	6,522	5,681	3,758	2,949	1,908	1,153	0,443	-0,183	-0,491	-0,620	0,02
0,03	6,506	5,670	3,758	2,953	1,914	1,160	0,451	-0,177	-0,495	-0,618	0,03
0,04	6,489	5,659	3,758	2,956	1,921	1,168	0,459	-0,171	-0,499	-0,616	0,04
0,05	6,471	5,647	3,758	2,959	1,927	1,176	0,466	-0,165	-0,503	-0,614	0,05
0,06	6,452	5,635	3,757	2,962	1,933	1,183	0,474	-0,159	-0,507	-0,612	0,06
0,07	6,432	5,621	3,755	2,964	1,940	1,191	0,481	-0,153	-0,511	-0,609	0,07
0,08	6,411	5,607	3,754	2,967	1,946	1,198	0,489	-0,147	-0,516	-0,607	0,08
$p_1(\alpha)$	-3,2905 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	$p_1(\alpha)$

Tab. 14.3. cd. Wartości krytyczne χ^2 rozkładu χ^2

r/α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005	cd/r
56	55,335	58,040	61,031	64,658	69,918	74,468	78,567	83,513	86,994	94,460	97,475	56
57	56,335	59,064	62,080	65,737	71,040	75,624	79,752	84,733	88,236	95,751	98,784	57
58	57,335	60,088	63,129	66,816	72,160	76,778	80,936	85,950	89,477	97,039	100,090	58
59	58,335	61,111	64,178	67,894	73,279	77,931	82,117	87,166	90,715	98,324	101,394	59
60	59,335	62,135	65,226	68,972	74,397	79,082	83,298	88,379	91,952	99,607	102,695	60
61	60,335	63,158	66,274	70,049	75,514	80,232	84,476	89,591	93,186	100,888	103,993	61
62	61,335	64,181	67,322	71,125	76,630	81,381	85,654	90,802	94,419	102,166	105,289	62
63	62,335	65,204	68,369	72,201	77,745	82,529	86,830	92,010	95,649	103,442	106,583	63
64	63,335	66,226	69,416	73,276	78,860	83,675	88,004	93,217	96,878	104,716	107,875	64
65	64,335	67,249	70,462	74,351	79,971	84,821	89,177	94,422	98,105	105,988	109,164	65
66	65,335	68,271	71,508	75,425	81,086	85,965	90,349	95,626	99,330	107,258	110,451	66
67	66,335	69,293	72,554	76,498	82,197	87,108	91,519	96,828	100,554	108,526	111,736	67
68	67,334	70,315	73,600	77,571	83,308	88,250	92,688	98,028	101,776	109,791	113,018	68
69	68,334	71,337	74,645	78,643	84,418	89,391	93,856	99,227	102,996	111,055	114,299	69
70	69,334	72,358	75,689	79,715	85,527	90,531	95,023	100,425	104,215	112,317	115,578	70
71	70,334	73,380	76,734	80,786	86,635	91,670	96,189	101,621	105,432	113,577	116,854	71
72	71,334	74,401	77,778	81,857	87,743	92,808	97,353	102,816	106,648	114,835	118,129	72
73	72,334	75,422	78,822	82,927	88,850	93,945	98,516	104,010	107,862	116,092	119,402	73
74	73,334	76,443	79,865	83,997	89,956	95,081	99,678	105,202	109,074	117,346	120,673	74
75	74,334	77,464	80,908	85,066	91,061	96,217	100,839	106,393	110,286	118,599	121,942	75
76	75,334	78,485	81,951	86,135	92,166	97,351	101,999	107,582	111,495	119,850	123,209	76
77	76,334	79,505	82,994	87,203	93,270	98,484	103,158	108,771	112,704	121,100	124,475	77
78	77,334	80,526	84,036	88,271	94,374	99,617	104,316	109,958	113,911	122,348	125,739	78
79	78,334	81,546	85,078	89,338	95,476	100,749	105,473	111,144	115,117	123,594	127,001	79
80	79,334	82,566	86,120	90,405	96,578	101,879	106,629	112,329	116,321	124,839	128,261	80
81	80,334	83,586	87,161	91,472	97,608	103,010	107,783	113,512	117,524	126,082	129,520	81
82	81,334	84,606	88,202	92,538	98,780	104,139	108,937	114,695	118,726	127,324	130,778	82
83	82,334	85,626	89,243	93,604	99,880	105,267	110,090	115,876	119,927	128,565	132,033	83
84	83,334	86,646	90,284	94,669	100,980	106,395	111,242	117,057	121,126	129,804	133,288	84
85	84,334	87,665	91,325	95,734	102,079	107,522	112,393	118,236	122,325	131,041	134,540	85
86	85,334	88,685	92,365	96,799	103,177	108,648	113,544	119,414	123,522	132,277	135,792	86
87	86,334	89,704	93,405	97,863	104,275	109,773	114,693	120,591	124,718	133,512	137,041	87
88	87,334	90,723	94,445	98,927	105,372	110,898	115,841	121,767	125,913	134,745	138,290	88
89	88,334	91,742	95,484	99,991	106,469	112,022	116,989	122,942	127,106	135,978	139,537	89
90	89,334	92,761	96,524	101,054	107,565	113,145	118,136	124,116	128,299	137,208	140,782	90
91	90,334	93,780	97,563	102,117	108,661	114,268	119,282	125,289	129,491	138,438	142,027	91
92	91,334	94,799	98,602	103,179	109,756	115,390	120,427	126,462	130,681	139,666	143,269	92
93	92,334	95,818	99,641	104,241	110,850	116,511	121,571	127,633	131,871	140,893	144,511	93
94	93,334	96,836	100,679	105,303	111,944	117,632	122,715	128,803	133,059	142,119	145,751	94
95	94,334	97,855	101,717	106,364	113,038	118,752	123,858	129,973	134,247	143,344	146,990	95
96	95,334	98,873	102,755	107,425	114,131	119,871	125,000	131,141	135,433	144,567	148,228	96
97	96,334	99,892	103,793	108,486	115,223	120,990	126,141	132,309	136,619	145,789	149,465	97
98	97,334	100,910	104,831	109,547	116,315	122,108	127,282	133,476	137,803	147,010	150,700	98
99	98,334	101,928	105,868	110,607	117,407	123,225	128,422	134,642	138,987	148,230	151,934	99
100	99,334	102,946	106,906	111,667	118,498	124,342	129,561	135,807	140,169	149,449	153,167	100

Funkcje pomocnicze $p_1(\alpha)$ i $p_2(\alpha, \nu)$ ($\nu = 1/\sqrt{2r}$)

ν/α	0,50	0,40	0,30	0,20	0,10	0,05	0,025	0,01	0,005	0,001	0,0005	α/ν
0,00	-0,667	-0,624	-0,483	-0,194	0,428	1,137	1,894	2,941	3,757	5,700	6,552	0,00
0,01	-0,667	-0,626	-0,479	-0,200	0,421	1,129	1,887	2,937	3,755	5,708	6,565	0,01
0,02	-0,667	-0,628	-0,476	-0,206	0,413	1,121	1,880	2,933	3,754	5,716	6,578	0,02
0,03	-0,667	-0,630	-0,472	-0,212	0,405	1,113	1,873	2,928	3,752	5,723	6,590	0,03
0,04	-0,666	-0,631	-0,468	-0,218	0,398	1,105	1,866	2,923	3,750	5,730	6,601	0,04
0,05	-0,666	-0,633	-0,464	-0,224	0,390	1,097	1,859	2,918	3,748	5,736	6,612	0,05
0,06	-0,666	-0,635	-0,460	-0,229	0,382	1,089	1,851	2,913	3,745	5,741	6,622	0,06
0,07	-0,666	-0,637	-0,457	-0,235	0,375	1,081	1,844	2,908	3,742	5,747	6,631	0,07
0,08	-0,666	-0,639	-0,453	-0,241	0,367	1,073	1,836	2,902	3,739	5,741	6,640	0,08
$p_1(\alpha)$	0,0000	0,2533	0,5244	0,8416	1,2815	1,6448	1,9599	2,3263	2,5758	3,0902	3,2905	$p_1(\alpha)$
	0000	4710	0051	2123	5157	5363	6398	4787	2930	3231	2673	

Tab. 14.4. 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

Tab. 14.5. 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	r_2/r_1
1	161	200	216	225	230	234	237	239	241	242	244	245	246	247	248	1
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	2
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	3
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	4
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	5
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	6
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	7
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	8
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	9
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	10
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	11
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	12
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	13
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	14
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	15
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	16
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	17
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	18
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	19
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	20
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	21
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	22
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	23
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	24
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	25
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	26
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	27
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	28
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	29
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	30
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	32
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	34
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	36
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	38
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	40
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	42
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	44
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	46
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	48
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	50
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	60
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	80
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	100
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	125
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	150
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	200
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	300
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	500
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	1000
∞	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	∞

Tab. 14.5. cd. 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	∞	r_2/r_1
1	249	249	249	250	250	251	251	251	252	252	252	253	254	254	254	1
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	2
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	3
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	4
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	5
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	6
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	7
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	8
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	9
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	10
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	11
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	12
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	13
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	14
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	15
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	16
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	17
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	18
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	19
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	20
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	21
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	22
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	23
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	24
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	25
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	26
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	27
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	28
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	29
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	30
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	32
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	34
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	36
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	38
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	40
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	42
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	44
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	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	48
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	50
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	60
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	80
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	100
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	125
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	150
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	200
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	300
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	500
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	1000
∞	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. 14.6. Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
2	2	-	-	-	-	4	4	4	4
	3	-	-	-	-	5	5	5	5
	4	-	-	-	-	5	5	5	5
	5	-	-	-	-	5	5	5	5
	6	-	-	-	-	5	5	5	5
	7	-	-	-	-	5	5	5	5
	8	-	-	-	2	5	5	5	5
	9	-	-	-	2	5	5	5	5
	10	-	-	-	2	5	5	5	5
	11	-	-	-	2	5	5	5	5
	12	-	-	2	2	5	5	5	5
	13	-	-	2	2	5	5	5	5
	14	-	-	2	2	5	5	5	5
	15	-	-	2	2	5	5	5	5
	16	-	-	2	2	5	5	5	5
	17	-	-	2	2	5	5	5	5
	18	-	-	2	2	5	5	5	5
	19	-	2	2	2	5	5	5	5
	20	-	2	2	2	5	5	5	5
3	3	-	-	-	-	6	6	6	6
	4	-	-	-	-	6	7	7	7
	5	-	-	-	2	7	7	7	7
	6	-	-	2	2	7	7	7	7
	7	-	-	2	2	7	7	7	7
	8	-	-	2	2	7	7	7	7
	9	-	2	2	2	7	7	7	7
	10	-	2	2	3	7	7	7	7
	11	-	2	2	3	7	7	7	7
	12	2	2	2	3	7	7	7	7
	13	2	2	2	3	7	7	7	7
	14	2	2	2	3	7	7	7	7
	15	2	2	3	3	7	7	7	7
	16	2	2	3	3	7	7	7	7
	17	2	2	3	3	7	7	7	7
	18	2	2	3	3	7	7	7	7
	19	2	2	3	3	7	7	7	7
	20	2	2	3	3	7	7	7	7
4	4	-	-	-	2	7	8	8	8
	5	-	-	2	2	8	8	8	9
	6	-	2	2	3	8	8	9	9
	7	-	2	2	3	8	9	9	9
	8	2	2	3	3	9	9	9	9
	9	2	2	3	3	9	9	9	9
	10	2	2	3	3	9	9	9	9
	11	2	2	3	3	9	9	9	9
	12	2	3	3	4	9	9	9	9
	13	2	3	3	4	9	9	9	9
	14	2	3	3	4	9	9	9	9
	15	3	3	3	4	9	9	9	9
	16	3	3	4	4	9	9	9	9
	17	3	3	4	4	9	9	9	9
	18	3	3	4	4	9	9	9	9
	19	3	3	4	4	9	9	9	9
	20	3	3	4	4	9	9	9	9

Tab. 14.6. cd. Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
5	5	-	2	2	3	8	9	9	10
	6	2	2	3	3	9	9	10	10
	7	2	2	3	3	9	10	10	11
	8	2	2	3	3	9	10	11	11
	9	2	3	3	4	10	11	11	11
	10	3	3	3	4	10	11	11	11
	11	3	3	4	4	11	11	11	11
	12	3	3	4	4	11	11	11	11
	13	3	3	4	4	11	11	11	11
	14	3	3	4	5	11	11	11	11
	15	3	4	4	5	11	11	11	11
	16	3	4	4	5	11	11	11	11
	17	3	4	4	5	11	11	11	11
	18	4	4	5	5	11	11	11	11
	19	4	4	5	5	11	11	11	11
	20	4	4	5	5	11	11	11	11
6	6	2	2	3	3	10	10	11	11
	7	2	3	3	4	10	11	11	12
	8	3	3	3	4	11	11	12	12
	9	3	3	3	4	11	11	12	12
	10	3	3	4	5	11	12	13	13
	11	3	4	4	5	12	12	13	13
	12	3	4	4	5	12	12	13	13
	13	3	4	5	5	12	13	13	13
	14	4	4	5	5	12	13	13	13
	15	4	4	5	6	13	13	13	13
	16	4	4	5	6	13	13	13	13
	17	4	5	5	6	13	13	13	13
	18	4	5	5	6	13	13	13	13
	19	4	5	6	6	13	13	13	13
	20	4	5	6	6	13	13	13	13
7	7	3	3	3	4	11	12	12	12
	8	3	3	4	4	12	12	13	13
	9	3	4	4	5	12	13	13	14
	10	3	4	5	5	12	13	14	14
	11	4	4	5	5	13	13	14	14
	12	4	4	5	6	13	13	14	15
	13	4	5	5	6	13	14	15	15
	14	4	5	5	6	13	14	15	15
	15	4	5	6	6	14	14	15	15
	16	5	5	6	6	14	15	15	15
	17	5	5	6	7	14	15	15	15
	18	5	5	6	7	14	15	15	15
	19	5	5	6	7	14	15	15	15
	20	5	6	6	7	14	15	15	15

Tab. 14.6. cd. Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
8	8	3	4	4	5	12	13	13	14
	9	3	4	5	5	13	13	14	14
	10	4	4	5	6	13	14	14	15
	11	4	5	5	6	14	14	15	15
	12	4	5	6	6	14	15	15	16
	13	5	5	6	6	14	15	16	16
	14	5	5	6	7	15	15	16	16
	15	5	5	6	7	15	15	16	17
	16	5	6	6	7	15	16	16	17
	17	5	6	7	7	15	16	17	17
	18	6	6	7	8	15	16	17	17
	19	6	6	7	8	15	16	17	17
9	9	4	4	5	6	13	14	15	15
	10	4	5	5	6	14	15	15	16
	11	5	5	6	6	14	15	16	16
	12	5	5	6	7	15	15	16	17
	13	5	6	6	7	15	16	17	17
	14	5	6	7	7	16	16	17	17
	15	5	6	7	8	16	17	17	18
	16	6	6	7	8	16	17	17	18
	17	6	7	7	8	16	17	18	18
	18	6	7	8	8	17	17	18	19
	19	6	7	8	8	17	17	18	19
	20	7	7	8	9	17	17	18	19
10	10	5	5	6	6	15	15	16	16
	11	5	5	6	7	15	16	17	17
	12	5	6	7	7	16	16	17	18
	13	5	6	7	8	16	17	18	18
	14	6	6	7	8	16	17	18	18
	15	6	7	7	8	17	17	18	19
	16	6	7	8	8	17	18	19	19
	17	7	7	8	9	17	18	19	19
	18	7	7	8	9	18	18	19	20
	19	7	8	8	9	18	19	19	20
	20	7	8	9	9	18	19	19	20
	21	8	8	9	10	18	19	20	20
11	11	5	6	7	7	16	16	17	18
	12	6	6	7	8	16	17	18	18
	13	6	6	7	8	17	18	18	19
	14	6	7	8	8	17	18	19	19
	15	7	7	8	9	18	18	19	20
	16	7	7	8	9	18	19	20	20
	17	7	8	9	9	18	19	20	21
	18	7	8	9	10	19	19	20	21
	19	8	8	9	10	19	20	21	21
	20	8	8	9	10	19	20	21	21
	21	8	9	10	10	20	20	21	22
	22	8	9	10	11	20	21	22	22
12	12	6	7	7	8	17	18	18	19
	13	6	7	7	8	17	18	19	20
	14	7	7	8	9	18	19	20	20
	15	7	8	8	9	18	19	20	21
	16	7	8	9	10	19	20	21	21
	17	8	8	9	10	19	20	21	21
	18	8	8	9	10	20	20	21	22
	19	8	9	10	10	20	21	22	22
	20	8	9	10	11	20	21	22	22
	21	8	9	10	11	20	21	22	22
	22	8	9	10	11	20	21	22	22
	23	8	9	10	11	20	21	22	22

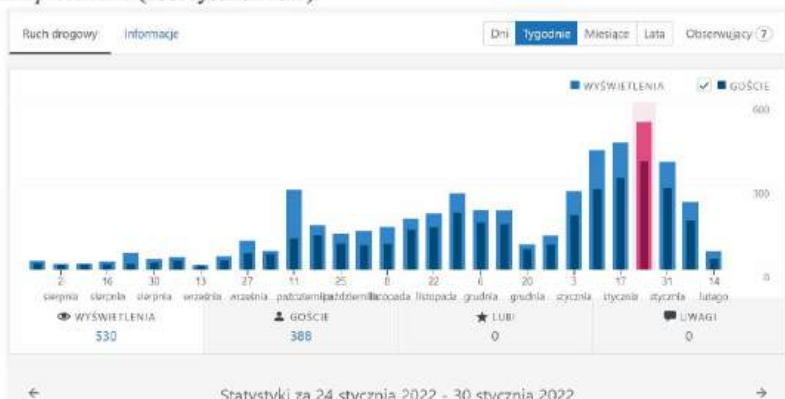
Tab. 14.6. cd. Test serii. Wartości krytyczne $r_{n_1, n_2}(\alpha)$

n_1	n_2	α							
		0,005	0,010	0,025	0,050	0,950	0,975	0,990	0,995
13	13	7	7	8	9	18	19	20	20
	14	7	7	9	9	19	19	20	21
	15	7	8	9	10	19	20	21	21
	16	8	8	9	10	20	20	21	22
	17	8	9	10	10	20	21	22	22
	18	8	9	10	11	20	21	22	23
	19	9	9	10	11	21	22	23	23
	20	9	10	10	11	21	22	23	23
14	14	7	8	9	10	19	20	21	22
	15	8	8	9	10	20	21	22	22
	16	8	9	10	11	20	21	22	23
	17	8	9	10	11	21	22	23	23
	18	9	9	10	11	21	22	23	24
	19	9	10	11	12	22	22	23	24
	20	9	10	11	12	22	23	24	24
	21	9	10	11	12	22	23	24	24
15	15	8	9	10	11	20	21	22	23
	16	9	9	10	11	21	22	23	23
	17	9	10	11	11	21	22	23	24
	18	9	10	11	12	22	23	24	24
	19	10	10	11	12	22	23	24	25
	20	10	11	12	12	23	24	25	25
	21	10	11	12	13	23	24	25	25
	22	10	11	12	13	23	24	25	26
16	16	9	10	11	11	22	22	23	24
	17	9	10	11	12	22	23	24	25
	18	10	10	11	12	23	24	25	25
	19	10	11	12	13	23	24	25	26
	20	10	11	12	13	24	24	25	26
	21	10	11	12	13	24	25	26	26
	22	11	11	12	13	24	25	26	27
	23	11	12	13	14	24	25	26	27
17	17	10	10	11	12	23	24	25	25
	18	10	11	12	13	23	24	25	26
	19	10	11	12	13	24	25	26	26
	20	11	11	13	13	24	25	26	27
	21	11	12	13	14	24	25	26	27
	22	11	12	13	14	25	26	27	28
	23	12	12	13	14	25	26	27	28
	24	12	13	14	15	26	27	28	29
18	18	11	11	12	13	24	25	26	26
	19	11	12	13	14	24	25	27	27
	20	11	12	13	14	25	26	27	28
	21	12	12	13	14	25	26	27	28
	22	12	13	14	15	26	27	28	29
	23	12	13	14	15	26	27	28	29
	24	13	13	14	15	26	27	28	29
	25	13	14	15	16	27	28	29	30
19	19	11	12	13	14	25	26	27	28
	20	12	12	13	14	26	26	28	28
	21	12	13	14	15	26	27	28	29
	22	13	13	14	15	27	27	29	29
	23	13	14	15	16	27	28	29	30
	24	14	14	15	16	28	28	30	30
	25	14	15	16	17	28	29	30	31
	26	15	15	16	17	29	29	31	31
20	20	12	13	14	15	26	27	28	29
	21	13	13	14	15	27	28	29	30
	22	13	14	15	16	27	28	30	30
	23	14	14	15	16	28	29	31	31
	24	14	15	16	17	28	29	31	32
	25	15	15	16	17	29	30	32	32
	26	15	16	17	18	29	30	32	33
	27	16	16	17	18	30	31	33	34

15. DOSTĘP DO PUBLIKACJI W LATACH 2008-2025 (Władysław Wornalkiewicz)



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 15.1).



Źródło: Opracowanie własne na podstawie blogu „Procesy informacyjne w teorii i praktyce”.

Rys. 15.1. 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 zmianami o moim dorobku naukowym na stronie WWW Akademii Nauk Stosowanych WSZiA w Opolu. 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 (2008-2024).

2008

„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

„Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera”, monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Marian Duczmał, Opole: Wydawnictwo Instytut Śląski, 2010, 597 s., ISBN 978-8388980-94-7, 978-83-62105-41-0.

Link: https://www.wsza.opole.pl/wp-content/uploads/2020/05/Wstep_do_ekonometrii_i_badan_operacyjnych.pdf.



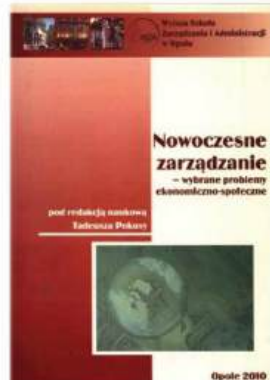
„Zarządzanie i polityka społeczna - wybrane problemy”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2010, 443 s., artykuł „Estymacja modeli wielorównaniowych w GRETL” strony: 255-276, ISBN 978-83-88980-90-9, 978-83-62105-23-6.

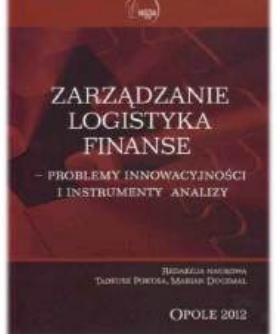
Link: <https://wordpress.com/media/wornalkiewicz.wordpress.com>.



„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.

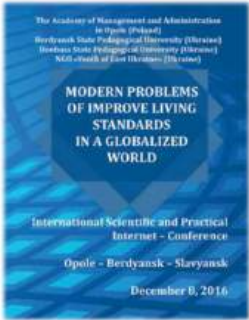
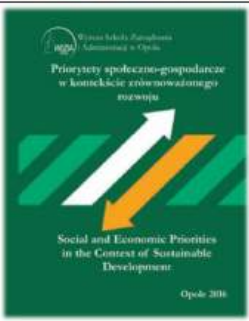
Link: <https://integro.bg.polsl.pl/172200374055/ksiazka/nawoczesne-zarzadzanie-wybrane-problemy-ekonomiczno-spoeczne?bibFile=17>.



2011 „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 „Wstęp do ekonometrii i badań operacyjnych Zbiór przykładów z zastosowaniem mikrokomputera”, wydanie II rozszerzone, część I, monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Marian Duczmał, 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 Duczmał, Opole: Wydawnictwo Instytut Śląski, 2012, 223 s., ISBN 978-83-62683-30-7, 978-83-7511-259-0.	
„Zarządzanie Logistyka Finanse - Problemy innowacyjności i instrumenty analizy”, monografia zbiorowa recenzowana, Opole: Wyższa Szkoła Zarządzania i Administracji w Opolu, 2012, 287 s., artykuły: „Model wielorównaniowy PKB”, strony: 223-249, „Etapy i procedury budowy modelu produktu brutto podregionów”, strony: 261-279, ISBN Link: https://w.bibliotecz.pl/1798765/Zarz%C4%85dzenie+logistyka+finanse.	
2013 „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.	

„<i>Modele ekonometryczne PKB obiektów struktury terytorialnej</i>”, 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/womalkiewicz-wladyslaw/modele-ekonometryczne-pkb-objektow-struktury-terytorialnej.	
„<i>Zarządzanie Logistyka - procesy, koncepcje, narzędzia</i>”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2013, 304 s., artykuł „<i>Programowanie wielokriterialne z zastosowaniem WinQSB</i>”, strony: 168-184, ISBN 978-83-62683-49-9, 978-83-7511-192-7. Link: https://integro.ciniba.edu.pl/integro/192504222555/ksika/zarządzanie-logistyka-koncepcje-procesy-narzędzia.	
„<i>Spoleczno-ekonomiczne uwarunkowania zarządzania i administracji - innowacyjność, komunikacja</i>”, monografia zbiorowa recenzowana, Opole: Wydawnictwo Instytut Śląski, 2013, 251 s., artykuł „<i>Wybór lokalizacji obiektu z zastosowaniem Expert Choice</i>”, strony: 124-140, ISBN 978-83-62683-53-6, 978-83-7511-103-4. Link: https://wordpress.com/media/wornalkiewicz.wordpress.com.	
2014 „<i>Formułowanie modeli ekonometrycznych na potrzeby zarządzania</i>” 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. „<i>Formułowanie modeli ekonometrycznych na potrzeby zarządzania</i>” 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 „<i>Wdrożenie zintegrowanego systemu informatycznego wspomagającego zarządzanie</i>”, 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.	
„<i>Ekonomiczno-społeczne uwarunkowania rozwoju gospodarczego - zarządzanie informacją i nowymi technologiami</i>”, monografia zbiorowa recenzowana, Opole: Wydawnictwo „Instytut Śląski”, 2015, 507 s., artykuły i strony: „<i>Przejawy wdrożenia systemów informatycznych</i>” (85-108), „<i>Modelowanie procesów zarządzania</i>” (124-146), współautorstwo: Ryszard Broszkiewicz, „<i>EDI w procesie logistycznym</i>” (278-301), „<i>Analityka biznesowa</i>” (356-375). Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/2_2021.pdf.	
„<i>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”</i>”, praca zbiorowa, Opole: Wydawnictwo Instytut Śląski, 102 s., artykuł „<i>Implementacja systemów klasy ERP w logistyce</i>”, ISBN 978-83-62683-75-8, 978-83-7511-235-1. Link: https://wordpress.com/media/wornalkiewicz.wordpress.com.	
2016 „<i>Wprowadzenie do projektowania systemów informatycznych zarządzania</i>” 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. „<i>Wprowadzenie do projektowania systemów informatycznych zarządzania</i>” 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.	
„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.	
„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_kontekście_zrownowazonego_rozwoju.pdf.	
„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_zarządzania_i_administracji_w_opolu_interdyscyplinarnosc_badan.pdf.	

„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 978-678-7291-44-1.

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



„Освіта і суспільство. Міжнародний збірник наукових праць”,
Бердянський державний педагогічний університет, 2016, 418 с.,
artykuł „Ranking metod ilościowych w Internecie”, strony: 275-285,
ISBN 978-617-7291-80-9 (електронне видання).

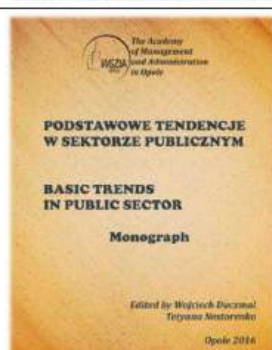
Link: https://bdpu.org.ua/wp-content/uploads/2019/10/Papers_Berdvansk_2016.pdf.

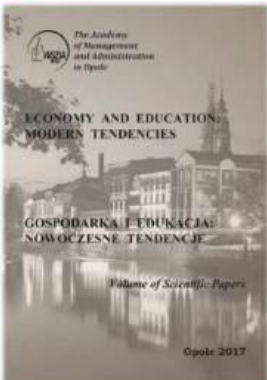


„Basic Trends in Public Sector” („Podstawowe tendencje w sektorze publicznym”), monografia zbiorowa recenzowana, artykuł: *Designing of managerial consoles* (Projektowanie pulpityw menadżerskich), strony: 58-74.

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.

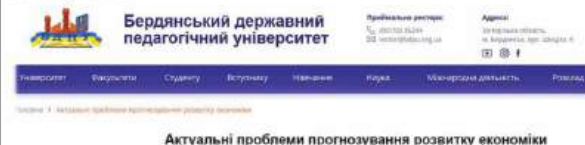


2017 „ECONOMY AND EDUCATION: MODERN TENDENCIES” - GOSPODARKA I EDUKACJA: NOWOCZESNE TENDENCJE, Volume of Scientific Papers, monografia zbiorowa recenzowana, Wyższa Szkoła Zarządzania i Administracji w Opolu, 2017, 362 s., artykuł „Prognostowanie z wykorzystaniem zasady postarzania informacji”, strony: 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.	
„Modern Problems of Management: Economics, Education, Health Care and Pharmacy”, Conference Proceedings of the 5th 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_economics_education_health_care_and_pharmacy.pdf.	
„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_internerice_wornalkiewicz.pdf.	

„Problems and Prospects of Territories' Socio-Economic Development”, Conference Proceedings of the 6th 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_prospects_of_territories_socio-economic_development.pdf.

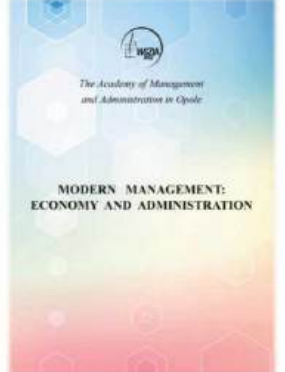


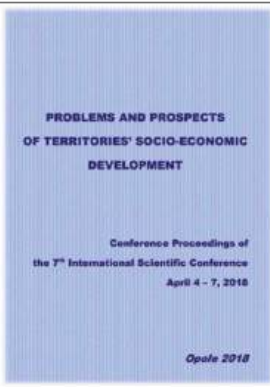
„Актуальні проблеми прогнозування розвитку економіки...” („Bieżące problemy prognozowania ekonomiki Ukrainy”), zbiór prac naukowych - monografia zbiorowa recenzowana, 370 s., КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ТАРАСА ШЕВЧЕНКА (Kijowski Narodowy Uniwersytet im. Tarasa Szewczenki) i inne, 2017, „Optymalizacja 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/>.



„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_nowoczesnym_spoleczenstwie_aspekty_ekonomiczne_nosferanko_pokusa.pdf.



2018 „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_improve_living_in_a_globalized_world.pdf.	
„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.	
„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_pharmacy.pdf.	

„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.	 The cover is blue with white text. It reads: 'PROBLEMS AND PROSPECTS OF TERRITORIES' SOCIO-ECONOMIC DEVELOPMENT', 'Conference Proceedings of the 7th International Scientific Conference', 'April 4 - 7, 2018', and 'Opole 2018'.
„Uwarunkowania demograficzne rozwoju społecznego i ekonomicznego miasta Nysa i powiatu nyskiego”, Opole-Nysa, Wyższa Szkoła Zarządzania i Administracji w Opolu, 2018, 260 s., artykuł: „Prognozowanie migracji ludności z uwzględnieniem wag harmonicznych”, strony: 212-233, ISBN 978-83-946765-0-6.	 The cover is grey with a central graphic of a map of Nysa and its district. Text includes: 'WYŻSZA SZKOŁA ZARZĄDZANIA I ADMINISTRACJI W OPOLU', 'FACULTY OF MANAGEMENT AND ADMINISTRATION IN OPOL', 'UWARUNKOWANIA DEMOGRAFICZNE ROZWOJU SPOŁECZNEGO I EKONOMICZNEGO MIASTA NYSA I POWIATU NYSKIEGO', and a list of authors: 'PAŹ DOMINIKA NAJEMSKA, BOŻENA BŁĘDZISZAK, TADEUSZ FURATY, TOMASZ CHROMIŁKA'. It also says 'Opole-Nysa 2018'.
2019 „Elementy inżynierii i analizy danych systemów zarządzania Wybrane aspekty logistyczne”, monografia zbiorowa recenzowana (podręcznik akademicki), współautorstwo - Wojciech Duczmał, 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_analazy_systemow_zarzadzania.pdf.	 The cover is dark grey with red and white text. It reads: 'Wydawnictwo Zarządzania i Administracji w Opolu', 'WŁADYSŁAW WORNALCZAK, WOJCIECH DUCZMAŁ', 'ELEMENTY INŻYNIERII I ANALIZY SYSTEMÓW ZARZĄDZANIA', 'Wybrane aspekty logistyczne', and 'Opole 2019'.

„*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, „*Prognosowanie 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.



„*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.



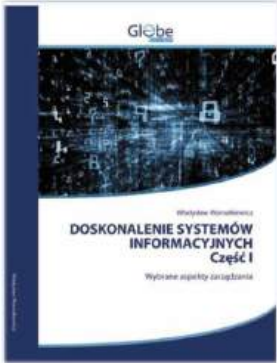
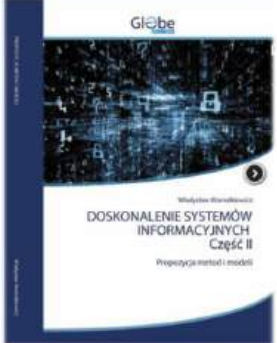
„*Modern Technologies in Education*”, Collective Scientific Monograph, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2019, 495 s., artykuł: „*Converting PDF to DOC*” (*Konwersja pliku formatu PDF na DOC*), strony: 184-194, ISBN 978-83- 946765-5-1.
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_modern_technologies_in_education_dauczan.pdf.

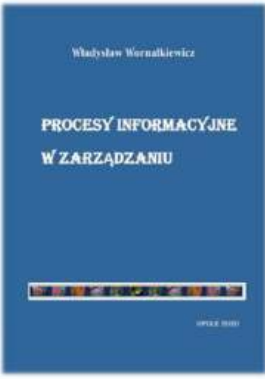


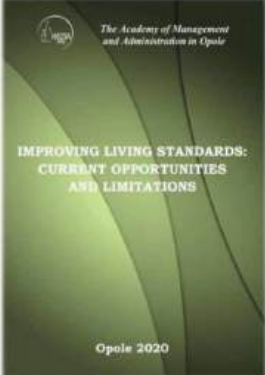
„Modern Technologies in Economy and Management”, Collective Scientific Monograph, monografia zbiorowa recenzowana, Opole: The Academy of Management and Administration in Opole, 2019, 493 s., artykuł: „Mortality modeling” (<i>Modelowanie umieralności</i>), strony: 148-160, ISBN 978-83-946765-4-4. Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_moderntechologies_in_ecenomy_and_management.pdf.	
„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” (<i>Techniki rozwiązań optymalizacyjnych zadania decyzyjnego</i>), 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.	
„Наука III 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” (<i>Konwersja pliku audio na tekstowy</i>), strony: 199-200, УДК 378:001. (063) Н 34.	

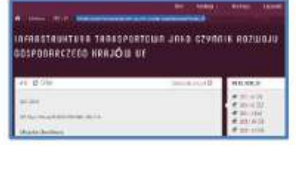
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/09/2019_%D0%9D%D0%B0%D1%83%D0%BA%D0%B0_%D0%86%D0%86%D0%86_%D1%82%D0%B8%D1%81%D1%8F%D1%87%D0%BE%D0%BB%D1%96%D1%82%D1%82%D1%8F.pdf.

„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 → magistralska droga 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. „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%B3wpraca-specjalizowanych-system%C3%B3w-informatycznych/isbn/978-613-9-42041-4.	
2020 „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/store/gb/book/narz%C4%99dzia-wdra%C5%BCania-system%C3%B3w-informatycznych/isbn/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.	

„Dokształcanie 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/dokskonalenie-system%C3%B3w-informacyjnych-cz%C4%99%C5%9B%C4%87-i/isbn/978-620-0-59233-0.	
„Dokształcanie 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/dokskonalenie-system%C3%B3w-informacyjnych-cz%C4%99%C5%9B%C4%87-ii/isbn/978-620-0-59542-3.	
„Освіта і суспільство V: Міжнародний збірник наукових праць”, Бердянський державний педагогічний університет, Ополь: видавництво Вищої школи управління і адміністрації в Ополь, 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.wsza.opole.pl/wp-content/uploads/2020/09/2020_OCBITA_I_SUPLYSTWO_V.pdf.	

„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://wormalkiewicz.files.wordpress.com/2020/02/implementacja-informatyczna-wybranych-metod-ilościowych.pdf.	
„Informatyka w wybranych obszarach zarządzania”, monografia indywidualna recenzowana, Berdiańsk: Wydawca „Svidler A.L.”, 2020, 450 s., ISBN 978-617-627-145-1. Link: https://wormalkiewicz.files.wordpress.com/2020/10/informatyka-w-wybranych-obszarach-zarzadzania.pdf.	
„Procesy informacyjne w zarządzaniu”, monografia indywidualna recenzowana, 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.	

„Journal of Modern Economic Research”, Bratysława, monografia zbiorowa recenzowana, artykuł „<i>Innovative logistic solutions</i>” (<i>Innowacyjne rozwiązania logistyczne</i>), strony: 53-63, współautorstwo: Maksym Kutsenko, 2020, ISSN 2644-4380 na-druk; 2644-6332 online. Link: https://denakyrpublishing.science/index.php/jmer/article/view/40.	
„<i>Improving living standards: current opportunities and limitations</i>”, monografia zbiorowa recenzowana, 594 s., Opole: The Academy of Management and Administration in Opole, 2020, konferencja, artykuł „<i>POS systems</i>” (<i>Systemy klasy POS</i>), s. 139-167, ISBN 978 - 83 - 66567 - 21 - 4. Link: https://www.wsia.opole.pl/wp-content/uploads/2020/05/Zbirnyk_Osvita-i-suspilstvo-VI_new.pdf.	
„<i>Organizational problems of life quality management in the conditions of globalization</i>”, Opole: The Academy of Management and Administration in Opole, 428 s., 2020, artykuł „<i>Existing and intended logistic projects</i>” (<i>Istniejące i zamierzone przedsięwzięcia logistyczne</i>), strony: 309-334, ISBN 978 - 83 - 66567 - 20 - 7. Link: https://www.wsia.opole.pl/wp-content/uploads/2020/05/2_2021.pdf.	

2021 „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	
„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.wsia.opole.pl/wp-content/uploads/2020/05/Systemy_informacyjne_w_logistyce_Wornalkiewicz.pdf.	
„Економічні науки”, monografia zbiorowa recenzowana, Вісник Хмельницького національного університету, 2021, № 5, Том I, artykuł: <i>THE OVERSIZED CARGO FORWARDING: ECONOMIC AND LEGAL ASPECTS</i>, UDC 330, strony: 175-180, ISSN 2307-5740, współautorstwo: YURII KRAVCHYK, ALLA KAPLUNOVSKA. Link: vknk-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: <i>THE OVERSIZED CARGO FORWARDING: ECONOMIC AND LEGAL ASPECTS</i>, UDC 330, strony: 175-180, ISSN 2307-5740, współautorstwo: YURII KRAVCHYK, ALLA KAPLUNOVSKA. Link: vknk-2021-n-5-298-175-180[1].pdf (khnu.km.ua).	
„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.	

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: Ievgen 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

„*SCIENTIFIC NOTES OF THE PEDAGOGICAL DEPARTMENT*”, 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-Tekst cтaтti-34489-1-10-20210721 (2).pdf; <https://periodicals.karazin.ua>.

„*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.

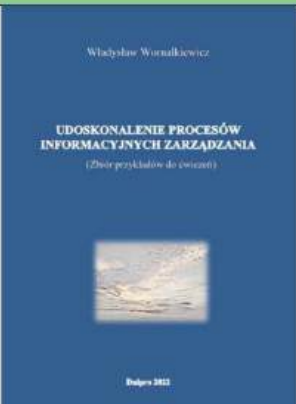
Link: https://www.wszia.opole.pl/wp-content/uploads/2020/05/7_2021.pdf.

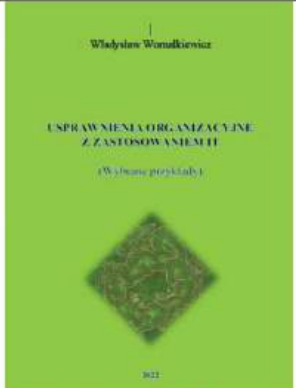
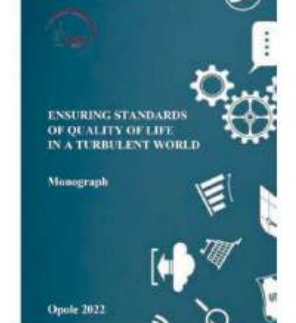


„*ОСВІТА І СУСПІЛЬСТВО VI*” Міжнародний збірник наукових праць, monografia zbiorowa recenzowana, Biediańsk: Wydawnictwo - Wyższa Szkoła Zarządzania i Administracji w Opolu, 2021, artykuł „*Opportunities to make milk reception logistics more modern*” („*Możliwości unowocześnienia logistyki odbioru mleka*”), 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.

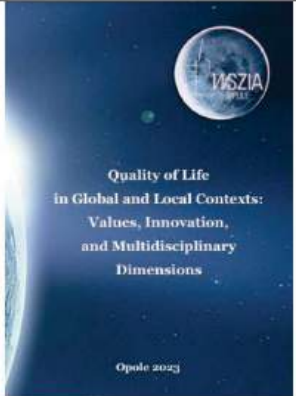
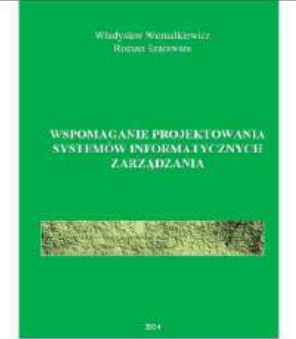


„SCIENTIFIC NOTES OF THE PEDAGOGICAL DEPARTMENT”, artykuł „ONLINE LEARNING AT UNIVERSITIES: POLISH-UKRAINIAN EXPERIENCE” „Nauczanie online na uniwersytetach: polsko-ukraińskie doświadczenia”, strony: 123-132, j. ang., УДК 378.147.31, 2021, współautorstwo: Olena Taranukha, Olena Fona-riuk. Link: 17560-Текст статті-34489-1-10-20210721 (2).pdf, https://periodicals.karazin.ua/pedagogy/article/view/17560.	
„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”, („Spoeczno-logistyczne aspekty przekopu Mierzei Wiślanej”), strony: 302-304, удк 371: 378. Link: https://www.wsia.opole.pl/wp-content/uploads/2020/05/Spoleczno_logistyczne_aspekty_przekopu.pdf.	
2022 „Udoskonalenie procesów informacyjnych zarządzania (Zbiór przykładów do ćwiczeń), monografia indywidualna recenzowana, Dnipro: Wydawca „Svidler A.L.”, 241 s., ISBN 978-617-627-170-3.	

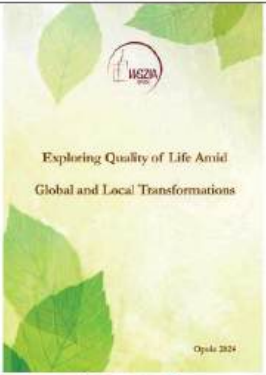
„Usprawnienia organizacyjne z zastosowaniem IT (Wybrane przykłady), monografia indywidualna recenzowana, Kijów: Wydawca „Majster Knyg”, s. 340, ISBN 978-617-7652-59-4. Link do strony internetowej Narodowego Uniwersytetu Gospodarki Miejskiej im. O.M. Beketova w Charkowie: http://eprints.kname.edu.ua/61802/1/Monografia_7.pdf	
„Zasilanie alternatywne pojazdów samochodowych”, Zeszyty Naukowe Wyższej Szkoły Technicznej w Katowicach, 2022, nr 15, s. 30, ISSN 2082-7016, eISSN 2450-5552.	
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”, s.: 15. Wydanie przez: ANS Opole, Wydział Pedagogiczny Uniwersytetu w Preszowie (Słowacja), Państwowy Uniwersytet Pedagogiczny w Bierańsku (Ukraina).	
„Projects in the field of 5G network construction” (Przedsięwzięcia w zakresie budowy sieci 5G), s. 277-293, Ensuring Standards of Quality of Life in a Turbulent World. Monograph. The Academy of Applied Sciences – Academy of Management and Administration in Opole.	

2023 Artykuły w ramach monografii zbiorowej „<i>Modern Approaches to sustainable development</i> „AŚ Katowice 2023: ○ <i>Flood Prevention (Przeciwdziałanie powodziom wiosennym)</i>, 1.12. s. 95-105, referat na konferencji. ○ <i>Organization of a rescue action in a situation earthquake (Organizacja akcji ratowniczej w sytuacji trzęsienia ziemi)</i>, 2.3. s. 285-294. Monografia pokonferencyjna 4th International Scientific Conference „<i>Role of Science and Education in Sustainable Development</i>” (Znaczenie nauki i edukacji w zrównoważonym rozwoju), konferencja międzynarodowa internetowa.	
Artykuł „<i>Procedure implementation of logistics services</i>” (<i>Implementacja procedury optymalizacji usług logistycznych</i>), Zeszyty Naukowe Wyższej Szkoły Technicznej w Katowicach, 2023, nr 16, s. 157-170. ISSN: 2082-7016; e-ISSN: 2450-5512; DOI: 10.54264/0070.	
Artykuł „<i>Cyber security</i>” (<i>Cybersecurity</i>), abstract, s. 5, VII International Scientific and Practice Conference „<i>Innovative potential and legal support of social and economic development of Ukraine: the challenge of the global world</i>”, konferencja internetowa, referat na platformie ZOOM, Poltawski Instytut Ekonomii i Prawa Uniwersytetu „Ukraina”, s. 16-21, Poltava, may, 17-18, 2023, УДК 330.111.66(477):34.	
Artykuł „<i>Iteracyjne korzystanie ze sztucznej inteligencji</i>” - <i>Iterative use of artificial intelligence</i>, s. 12-15. Zбірник матеріалів V Всеукраїнської науково-практичної конференції, м. Полтава, Полтавський інститут економіки і права, 5-6 грудня 2023 року 12 Полтавський інститут економіки і права, РОЗДІЛ 1 ПРАВОВЕ РЕГУЛЮВАННЯ СУСПІЛЬНИХ ВІДНОСИН В УКРАЇНІ В УМОВАХ ВОЄННОГО СТАНУ; prezentacja w Power Point. ISBN 978-617-8368-11-1 УДК 330.111.66(477):34	

Презентація нового видання – підручника, виданого в межах українопольської співпраці: Wornalkiewicz W., Szarawara R. „Techniki rozwiązań optymalizacyjnych. Przykłady zastosowań”, Połtawa: Połtawski Instytut Ekonomii i Prawa Uniwersytetu „Ukraina”, 2023, s. 243.	
Wornalkiewicz W., Szarawara R., podręcznik „Techniki rozwiązań optymalizacyjnych Przykłady zastosowań”, Połtawski Instytut Ekonomii i Prawa Uniwersytetu „Ukraina”, s. 244, UDK 338.24:001.82(075), ISBN 978-966-388-674-9. DOI: https://doi.org/10.36994/978-966-388-674-9-2023-243	
Wornalkiewicz W., Szarawara R., podręcznik „Techniki rozwiązań optymalizacyjnych Przykłady zastosowań”, Wydawnictwo Naukowe PWN S.A., Warszawa 2024, s. 246, ISBN 978-83-01-23503-1, DOI: https://doi.org/10.53271/2024.020.	
Wornalkiewicz W., „Modelowanie ekonometryczne podstawą prognozowania”, s. 340. Państwowy Uniwersytet Pedagogiczny w Bierańsku (Zaporoże), w druku.	

Artykuł „<i>Artificial intelligence applications</i>” (<i>Aplikacje sztucznej inteligencji</i>) - w monografii pokonferencyjnej, s. 12, <i>Quality of Life in Global and Local Contexts: Values, Innovation, and Multidisciplinary Dimensions</i> (<i>Jakość życia w kontekście globalnym i lokalnym: wartości, innowacje i wymiary multidyscyplinarne</i>), The Academy of Applied Sciences – Academy of Management and Administration in Opole, 2023, ISBN 978-83-66567-53-5.	
2024 Artykuł "<i>Metody ochrony przed cyberbezpieczeństwem</i>" s. 126-134, w Zeszytach Naukowych <i>Vzdelávanie a spoločnosť VIII</i>, Prešovská univerzita v Prešove Pedagogická fakulta, PREŠOV 2024, ISBN 978-80-555-3260-8, EAN 9788055532608.	
Womalkiewicz W., Szarawara R., „<i>Wspomaganie projektowania systemów informatycznych zarządzania</i>”, podręcznik, s. 408, Połtawski Instytut Ekonomii i Prawa. Uniwersytet „Ukraina”, ISBN 978-966-388-689-3, UDK 005.8:004.415(075.8),s.	

Artykuł „Niewykorzystane możliwości energii wodnej”, s. 213-217, referat, konferencja 19.04.2023, Kijów, National University of Live and Environmental Sciences of Ukraine, Katedra Technologii Środków Transportu Kompleksu Rolno-Przemysłowego, monografia tylko w wersji elektronicznej. ISBN978-617-8368-11-1, УДК 631.17+62-52-631.3.	
Womalkiewicz W., „Tendencje w zakresie wytwarzania energii ze źródeł odnawialnych. Wybrane sektory”, Globe Edit, ISBN 978-620-6-79987-0, s. 132.	
Artykuły: Konferencja „Digital Economy and Digital Society 6th International Scientific Conference, April 9-10, 2024, Akademia Śląska w Katowicach, Monografia „Exploring the Digital Landscape: Interdisciplinary Perspectives”, Edited by Olha Blaha and Iryna Ostapolets, The University of Technology in Katowice Press 2024, ISBN 978-83-969890-9-3, DOI: 10.54264/M036: - „Breaking social anxiety – green light for nuclear power plants” - „Przełamanie niepokoju społecznego – zielone światło dla elektrowni atomowych”, s. 457-477. - „Carbon-Nuclear Transformation” - „Transformacja węglowo-jądrowa”, s. 478-891, referat. - „Trend of application of AI in search engines” - „Tendencje zastosowania AI w wyszukiwarkach”, s. 892-906.	

Artykuł „<i>Współpraca mieszkańców wsi w ustalaniu priorytetu remontu dróg lokalnych</i>”, s.: 447-459, monografia po konferencji międzynarodowej z udziałem przedstawicieli Ukrainy pt. „Exploring Quality of Life Amid Global and Local Transformations”, 9th International Scientific and Practical – Conference „<i>Modern problems of Improve living standards in a globalized world</i>”, november 13-14, 2024. Prezentacja artykułu w Power Point na Konferencji.	
Artykuł „<i>IT technology in a small tailoring shop</i>” - „<i>Technika IT w małym zakładzie krawieckim</i>”, w ramach monografii EDUKACJA I SPOŁECZEŃSTWO IX, Zbiór prac naukowych, Katowice 2024, strony: 221-236, ISBN 978-83-972085-8-2, DOI 10.54264/M041.	
Artykuł „<i>Usprawnienie obiegu dokumentów w administracji publicznej</i>”, Zeszyty Naukowe WST nr 19, Akademia Śląska w Katowicach, w trakcie druku.	
Podręcznik „<i>Przedsięwzięcia w zakresie udoskonalenia pomocy humanitarnej i logistyki (Studium przypadku)</i>”, Uniwersytet Pedagogiczny w Berdiańsku (siedziba w Zaporozże), ISBN 978-617-627-189-5, Wydawca „Svidler A.A.”, Dnipro 2024, 478 s.	

Artykuł "Dobór pakietu integracji modułów wsparcia rolnictwa", Poltawski Instytut Ekonomii i Prawa, Uniwersytet „Ukraina”, wygłoszony referat w dniu 12.12.2024, monografia pokonferencyjna wydana w wersji elektronicznej. УДК 330.111.66(477):34	
Artykuł „Usprawnienie obiegu dokumentów w administracji publicznej”, Zeszyty Naukowe Wyższej Szkoły Technicznej w Katowicach”, ISSN 2082-7016; eISSN 2450-5552 2024, nr 19, DOI: 10.54264/xxxx, w druku.	

BIBLIOGRAFIA



- Bartosiewicz S. (red.), *Ekometria wspomagana mikrokomputerem*, Wydawnictwo Akademii Ekonomicznej, Wrocław 1991.
- Bartosiewicz S., *Ekometria. Technologia ekonometrycznego przetwarzania informacji*, PWE, Warszawa 1989.
- Bieчек P., *Przewodnik po pakiecie R*, Oficyna Wydawnicza GiS, Wrocław 2008.
- Cieślak M. (red.), *Prognozowanie gospodarcze. Metody i zastosowania*, Wydawnictwo Naukowe PWN, Warszawa 2005.
- Dobosz M., *Wspomagana komputerowo statystyczna analiza wyników badań*, Akademicka Oficyna Wydawnicza EXIT, Warszawa 2001.
- Drabska E., *Statystyczna analiza danych za pomocą komputerowych metod statystycznych*, Politechnika Białostocka, 1992 (wg strony www w Internecie: //brun.fuw.edu.pl/edu/STAT).
- Duczmal M., Wornalkiewicz W., *Wstęp do ekonometrii i badań operacyjnych. Zbiór przykładów z zastosowaniem mikrokomputera*, Wydawnictwo Instytut Śląski, Opole 2010.
- Dziechciarz J. (red.), *Ekometria. Metody, przykłady, zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2003.
- Frączak E. (red.), *Wielowymiarowa analiza statystyczna. Teoria - przykłady zastosowań z systemem SAS*, SGH, Warszawa 2009.
- Gajda J., *Ekometria praktyczna*, Wydawnictwo: Absolwent, Łódź 1998.
- Gajda J., *Ekometria. Wykład i łatwe obliczenia w programie komputerowym*, Wydawnictwo C.H. Beck, Warszawa 2004.
- Górnai J., Wachnicki J., *Pierwsze kroki w analizie danych SPSS PL for Windows*, SPSS Polska, Kraków 2000.
- http://krawczyk.hostil.pl/images/S_opisowa_part2_marco.pdf.
- <http://mk.po.opole.pl/treść/programy.htm#statystyczne>.
- <http://www.google.pl>, hasło: „standardized skewness”).
- Jajuga K. (red.), *Ekometria. Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2003.
- Jajuga K. (red.), *Ekometria. Metody i analiza problemów ekonomicznych*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, 2002.
- Jurlewicz T., Skoczylas J., *Algebra liniowa I. Przykłady i zadania*, Wydawnictwo GiS, Wrocław 2003.
- Koślacz P. (red.), *Nowoczesne pakiety komputerowe w podejmowaniu decyzji*, UW, Warszawa 1998.
- Kowerski M., *Ekometria. Przewodnik po wykładach*, Wyższa Szkoła Zarządzania i Administracji w Zamościu, 1998.
- Krokosz-Krynke Zofia, *Symulacja w nauczaniu zarządzania produkcją*, referat w ramach publikacji na konferencję "Symulacja systemów gospodarczych", Prace Naukowe Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej, 2005
- Kufel T., *Ekometria. Rozwiązywanie problemów z wykorzystaniem programu Gretl*, Wydawnictwo Naukowe PWN, Warszawa 2007.
- Kufel T., *Ekometria. Rozwiązywanie problemów z wykorzystaniem programu GRETL*, Wydawnictwo Naukowe PWN, Warszawa 2004.
- Lipiec-Zajchowska M. (red.), *Wspomaganie procesów decyzyjnych*, Tom I. *Statystyka*, Wydawnictwo C.H. Beck, Warszawa 2003.
- Lipiec-Zajchowska M. (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.
- Lipiec-Zajchowska, M. Sikora E., Zajchowski J., *Modele podejmowania decyzji w pakietach komputerowych*, UW, Warszawa 1995.
- Luszniewicz A., Słaby T., *Statystyka z pakietem komputerowym STATISTICA PL. Teoria i zastosowania*, Wydawnictwo C.H. Beck, Warszawa 2008.
- Luszniewicz A., Słaby T., *Statystyka stosowana*, PWE, 1996.
- Maddala G.S., *Ekonometria*, Wydawnictwo Naukowe PWN, Warszawa 2006.
- Microsoft, *Excel. Opis funkcji*, Microsoft Corporation, Ireland 1993.
- Nowak E., *Zarys metod ekonometrii. Zbiór zadań*, Wydawnictwo Naukowe PWN, Warszawa 2002.
- Nowak E., *Zarys metod ekonometrii. Zbiór Zadań*. Wydawnictwo Naukowe PWN, Warszawa 1990.
- Osińska M. (red.), *Ekonometria współczesna*, Towarzystwo Naukowe Organizacji i Kierownictwa, Toruń 2007.
- Pilatowska M., *Repetitorium ze statystyki*, Wydawnictwo Naukowe PWN, Warszawa 2014.
- Podgórski J., *Statystyka z komputerem*, Warszawa 1995.
- Praca zbiorowa, *Badania operacyjne w przykładach i zadaniach*, Wydawnictwo Naukowe PWN, Warszawa 2004.
- Praca zbiorowa, *Leksykon naukowo-techniczny*. Warszawa 1984.
- Praca zbiorowa, *Modelowanie ekonometryczne z Excelem*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004.
- Praca zbiorowa, *Pakiety statystyczne dostępne na WNE UW* (wg strony www w Internecie //win.wne.uw.edu.pl/panel/materialy/opis-programow.pdf).
- Praca zbiorowa, *Pakiety statystyczne dostępne na WNE UW*, fragment: Kuhl K., "Analiza danych panelowych za pomocą SAS".
- Praca zbiorowa, *Pakiety statystyczne dostępne na WNE UW*, fragment: Tomaszewski A. „Możliwości przeprowadzenia badań statystycznych na danych panelowych przy pomocy programu Stata”.
- Praca zbiorowa, *Przykłady i zadania z badań operacyjnych i ekonometrii*, Wydawnictwo Akademii Ekonomicznej w Poznaniu, 2003.
- Praca zbiorowa, *Statystyka Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006.
- Praca zbiorowa, *Statystyka Elementy teorii i zadania*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2006.
- Praca zbiorowa, *Statystyka w biznesie. Zastosowanie narzędzi Excela*. Wydawca Elipsa 2005.
- Praca zbiorowa, *Wybrane zagadnienia z prognozowania*, ECONOMICUS, Szczecin 2006.
- Radzikowska B. (red.), *Metody prognozowania. Zbiór zadań*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu, Wrocław 2004.
- Radzikowska B. (red.), *Metody prognozowania. Zbiór zadań*, Wydawnictwo Akademii Ekonomicznej we Wrocławiu. 2004.
- Skrzypek J. Szubra M., *Symulacyjna gra decyzyjna TESS-6 jako narzędzia e-learningu*, Symulacja systemów gospodarczych. Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Skrzypek J. Szubra M., *TESS-2. Symulacyjna gra decyzyjna*, Poltex, Warszawa 1996.
- Strzała K., Przechleński T., *Ekonometria Inaczej*, UG 1994.
- Symulacja systemów społeczno-gospodarczych*, Prace Naukowe Instytutu Organizacji i Zarządzania Politechniki Wrocławskiej. Seria: Studia i Materiały. Wrocław 2005.
- Szczurkowski L., *Komputerowa gra ekonomiczna. Symulacja systemów gospodarczych*, Oficyna Wydawnicza Politechniki Wrocławskiej, Wrocław.
- Trzaskalik T., *Badania operacyjne z komputerem*, Seria: *Metody badań operacyjnych*, Absolwent, Łódź 1997.

- Trzaskalik T., *Modelowanie optymalizacyjne*. Seria: *Metody badań operacyjnych*, Absolwent, Łódź 1997.
- Walesiak M., Gatnar E. (red.), *Statystyczna analiza danych z wykorzystaniem programu R*, Wydawnictwo Naukowe PWN, Warszawa 2009.
- Wornalkiewicz W., *Metoda badania przyczynowo-skutkowego związków między cechami statystycznymi*, Skrypty Wyższej Szkoły Zarządzania i Administracji w Opolu, Opole 2013.
- Wornalkiewicz W., *Modele ekonometryczne PKB obiektów struktury terytorialnej*, Wydawnictwo Instytut Śląski, Opole 2013.
- Zieliński R., *Tablice statystyczne*, PWN, Warszawa 1972.
- Żądło T., Wywiół J., *Prognozowanie szeregów czasowych za pomocą pakietu SPSS*.

Format 60x84/16
Konwencjonalnie zadrukowane arkusze 9,5
Wydanie 50 kopie
Zamówienie № 01/12052025

ОП "ШвидкоДрук"
36003, м. Полтава, вул. Чорновола 2б
Свідоцтво ФОП Іпполітова Т.С.
В02 №414529 від 27.11.2007 р.