POLSKA AKADEMIA NAUK
KOMITET STATYSTYKI I EKONOMETRII

PRZEGLĄD STATYSTYCZNY

STATISTICAL REVIEW

TOM 58

3-4

2011

WARSZAWA 2011

WYDAWCA

Komitet Statystyki i Ekonometrii Polskiej Akademii Nauk

Wydanie publikacji dofinansowane przez Ministra Nauki i Szkolnictwa Wyższego

Strona WWW „Przegląd Statystyczny”:
http://www.przegladstatystyczny.pan.pl

Nakład: 300 egz.

PAN Warszawska Drukarnia Naukowa
00-056 Warszawa, ul. Śniadeckich 8
Tel./fax:+48 22 628 76 14

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

MAGDALENA OSIŃSKA, MICHAŁ BERNARD PIETRZAK, MIROSŁAWA ŻUREK

OCENA WPŁYWU CZYNNIKÓW BEHAWIORALNYCH I RYNKOWYCH NA POSTAWY INWESTORÓW INDYWIDUALNYCH NA POLSKIM RYNKU KAPITAŁOWYM ZA POMOCĄ MODELU SEM

1. WSTĘP

W czasie globalizacji i liberalizacji przepływów kapitałowych znaczenia nabiera konieczność poznania motywów i czynników determinujących zachowania inwestorów na rynku papierów wartościowych. W wieloaspektowych analizach uwzględniać powinno się nie tylko czynniki fundamentalne, wpływające na decyzje podejmowane przez inwestorów, ale również czynniki behawioralne. Inklinacjami behawioralnymi inwestorów tłumaczone są obecnie wszelkie anomalie występujące na rynkach kapitałowych przeczące ich efektywności. W niniejszym artykule, podążając za rozwojem finansów behawioralnych, podjęta została próba empirycznego wyjaśnienia zależności skłonności do ryzyka i poziomu oczekiwanej, satysfakcjonującej inwestorów stopy zwrotu od czynników psychologicznych, leżących u podstaw decyzji inwestorów indywidualnych aktywnych na Giełdzie Papierów Wartościowych w Warszawie.

Podwaliny teoretyczne dla finansów behawioralnych pochodzą z teorii perspektywy [16]. Uchylenie bezwzględnego założenia o racjonalności inwestorów i dopuszczenie faktu występowania inklinacji behawioralnych zakłócających ich racjonalny proces decyzyjny, pozwoliło na opracowanie licznych modeli behawioralnych, do których należą modele: LSV [21], BSV [1], DHS [9] czy HS [16] oraz dało możliwość odniesienia się do anomalii występujących na rynku kapitałowym. Należy podkreślić, że modele te pozostają w zgodzie z nurtem rozwijanym nieco w innym zakresie, a mianowicie z modelami mikrostruktury rynków [10]. Teoria behawioralna nabiera szczególnego znaczenia w warunkach obecnego kryzysu ekonomicznego i finansowego. Wśród głównych prac z tego zakresu należy wymienić [8], [15], [32], [34], [37].

Do opisu zależności uwzględniających czynniki psychologiczne ma zastosowanie metodyka równań strukturalnych (*Structural Equation Model*, *SEM*), której założenia rozwinięte zostały między innymi przez Bollena [2], Kaplana [18] oraz Pearla [30]. W polskiej literaturze na temat *SEM* pisali między innymi Brzeziński [5], Gatnar [14], Osińska [28], Konarski [20].

Tematyka artykułu dotyczy identyfikacji czynników behawioralnych oraz ustalenia siły i kierunku ich oddziaływania na proces decyzyjny inwestorów indywidualnych

w Polsce. Badania takie ze względu na dużą złożoność tematyczną, interdyscyplinarny charakter, a także niepewność co do uzyskanych wyników, opartych na badaniach ankietowych i relatywnie wysokie koszty należą do rzadkości nie tylko w Polsce, ale także w krajach wysoko rozwiniętych. W Polsce, poza wstępnym opracowaniem autorów na mniejszej próbie [29], nikt do tej pory nie zastosował metodologii *SEM* dla badań własności rynku kapitałowego w aspekcie inklinacji behawioralnych. Podobne badanie chociaż w węższym zakresie tematycznym, przeprowadzili Wang, Shi i Fan [35] w odniesieniu do analizy psychologii inwestowania na chińskim rynku kapitałowym.

Celem badania była identyfikacja czynników charakteryzujących postawy inwestorów indywidualnych w postaci inklinacji behawioralnych tj. błędów w sferze opinii i błędów w sferze preferencji oraz skłonności do ryzyka. Wyodrębnione zostały także czynniki dotyczące umiejętności posługiwania się przez inwestorów analizą techniczną oraz postrzeganej przez nich jakości funkcjonowania rynku, co również powinno mieć wpływ na podejmowane przez nich decyzje inwestycyjne. Pozwoliło to na ustalenie siły zależności pomiędzy wyspecyfikowanymi czynnikami zarówno w całej badanej grupie, jak również w podziale na podgrupy. Dodatkowo na podstawie otrzymanych odpowiedzi w kwestionariuszu wyróżniono zmienną ilościową, opisującą oczekiwaną stopę zwrotu z inwestycji. Zmienna ta określa subiektywne oczekiwania co do satysfakcjonującej badanych inwestorów indywidualnych stopy zwrotu. W badaniu empirycznym postawione zostały następujące hipotezy:

1. Inklinacje inwestorów w sferze opinii i preferencji zwiększają ich skłonność do ryzyka, a także oczekiwaną przez nich stopę zwrotu z inwestycji.

2. Wpływ inklinacji behawioralnych na decyzje inwestycyjne jest mniejszy wśród inwestorów umiejących się posługiwać analizą techniczną we właściwy sposób, niż u pozostałych osób.

Pierwsza hipoteza wiąże się z tym, że popełniane błędy (reprezentowane skłonności) powodują więcej strat niż wynikałoby to z wartości przeciętnej, stąd też wyższa oczekiwana stopa zwrotu ma zrekompensować inwestorom błędne czy też obciążone decyzje.

Druga hipoteza wiąże się z panującym w literaturze poglądem, że inwestor racjonalny w swych decyzjach posługuje się sprawnie zarówno analizą techniczną jak i fundamentalną [3] lub co najmniej analizą fundamentalną [24]. Ponieważ w artykule ograniczono się wyłącznie do zbadania u inwestorów umiejętności właściwego wykorzystania analizy technicznej, przyjęto założenie, że wszyscy inwestorzy indywidualni uczestniczący w badaniu w podobnym stopniu posługują się analizą fundamentalną. Założono też, że inwestor posługujący się analizą techniczną w sposób właściwy jest jednostką bardziej świadomą mechanizmów i zasad działających na rynku kapitałowym, niż inwestor nieumiejący prawidłowo wykorzystywać tych narzędzi. Zgodnie z powyższym inwestorzy indywidualni podzieleni zostali na dwie grupy. Należy przypuszczać, że inwestorzy z pierwszej grupy w mniejszym stopniu podatni będą na określone skłonności, zarówno w sferze opinii jak i preferencji. Natomiast inwestorów z drugiej grupy określić można mianem *noise traders* na rynku kapitałowym [24].

2. ETAPY I METODY BADANIA

W celu przeprowadzenia analizy mechanizmów leżących u podstaw podejmowania decyzji na rynku kapitałowym przygotowany został kwestionariusz[1], zawierający pytania dotyczące skłonności inwestorów w sferze opinii oraz preferencji, skłonności do ryzyka, umiejętności posługiwania się analizą techniczną oraz jakości rynku kapitałowego w Polsce. Dodatkowo starano się uzyskać informację na temat oczekiwanej stopy zwrotu z dokonywanych inwestycji.

Ankieta przeprowadzona została w IV kwartale 2010 roku na próbie 315 respondentów, złożonej z inwestorów indywidualnych lokujących swoje środki na polskim rynku kapitałowym. Badanie wykonane zostało przez Stowarzyszenie Inwestorów Indywidualnych, co determinuje własności próby. Na podstawie raportu *Profil Inwestora Giełdowego w 2010 r.* [31] można założyć, iż około połowy inwestorów stosuje w podejmowaniu decyzji analizę techniczną, a nieco mniej (ok. 40%) analizę fundamentalną. Ponadto większość (66%) stanowią inwestorzy z niezbyt długim stażem inwestycyjnym (do 5 lat). Większość z nich, bo aż blisko 90% posiada w portfelu inwestycyjnym głównie akcje notowane na GPW w Warszawie. Z danych zebranych w badanej próbie inwestorów wynika, że blisko połowa (48%) spośród 315 osób inwestuje na giełdzie przez okres nie dłuższy niż 3 lata. Odpowiedzi na pytania zawierały warianty bazujące na pięciostopniowej skali Likerta, co umożliwiło ich wykorzystanie w modelach równań strukturalnych [22]. W załączniku 1 zaprezentowano poszczególne pytania kwestionariusza wraz z przypisanymi do nich zmiennymi obserwowalnymi.

Na podstawie otrzymanych wyników wykonano konfirmacyjną analizę czynnikową, w wyniku czego wyodrębniono pięć wymienionych wcześniej czynników, tj. błędy w sferze opinii, błędy w sferze preferencji, skłonność do ryzyka, umiejętność posługiwania się analizą techniczną oraz jakość rynku kapitałowego. W przypadku wszystkich czynników uzyskano wartości współczynnika Alfa-Cronbacha powyżej 0,7, co zapewnia rzetelność przyjętej skali i poprawny opis przez rozważane zmienne założonego czynnika [7].

Następnie w celu weryfikacji postawionych hipotez zbudowano model równań strukturalnych *SEM*, będący efektem połączenia konfirmacyjnej analizy czynnikowej oraz modelowania przyczynowo-skutkowego. Model równań strukturalnych został szczegółowo przedstawiony m.in. w pracach Bollena [2] i Kaplana [18]. Konstrukcja ta składa się z modelu opisującego powiązania pomiędzy zmiennymi ukrytymi, nazywanego modelem wewnętrznym oraz modelu pomiaru endogenicznych i egzogenicznych zmiennych nieobserwowalnych, określanego mianem modelu zewnętrznego. Model zewnętrzny jest reprezentacją wyników analizy czynnikowej pozwalającej na wyliczenie ładunków poszczególnych czynników kształtujących zmienną ukrytą. Model wewnętrzny przedstawia natomiast analizę ścieżkową, pozwalającą na określenie związków przyczynowo-skutkowych pomiędzy zmiennymi.

[1] Wykorzystane w badaniu pytania kwestionariusza zamieszczone zostały w załączniku 1 do niniejszego artykułu.

Model wewnętrzny, zwany modelem strukturalnym, ma postać:

$$\boldsymbol{\eta} = \boldsymbol{B}\boldsymbol{\eta} + \boldsymbol{\Gamma}\xi + \zeta, \tag{1}$$

gdzie: $\boldsymbol{\eta}_{m\times 1}$ – wektor endogenicznych zmiennych ukrytych, $\xi_{k\times 1}$ – wektor egzogenicznych zmiennych ukrytych, $\mathbf{B}_{m\times m}$ – macierz współczynników regresji przy zmiennych endogenicznych, $\boldsymbol{\Gamma}_{m\times k}$ – macierz współczynników przy zmiennych egzogenicznych, $\zeta_{m\times 1}$ – wektor składników losowych.

Model zewnętrzny, określany jako model pomiaru jest dany jako:

$$\mathbf{y} = \boldsymbol{\Pi}_{\mathbf{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon}, \tag{2}$$

$$\mathbf{x} = \boldsymbol{\Pi}_{\mathbf{x}}\xi + \boldsymbol{\delta}, \tag{3}$$

gdzie: $\mathbf{y}_{p\times 1}$ – wektor obserwowalnych zmiennych endogenicznych, $\mathbf{x}_{q\times 1}$ – wektor obserwowalnych zmiennych egzogenicznych, $\boldsymbol{\Pi}_y, \boldsymbol{\Pi}_x$ - macierze ładunków czynnikowych, $\varepsilon_{p\times 1}, \boldsymbol{\delta}_{q\times 1}$ – wektory błędów pomiaru.

W celu estymacji parametrów model *SEM* musi być identyfikowalny. Do jego estymacji wykorzystuje się najczęściej metodę największej wiarygodności (MNW), uogólnioną metodę najmniejszych kwadratów (UMNK) oraz metodę asymptotycznie niewrażliwą na rozkład (ADF *Asymtotically Distribution-Free*). Wybór właściwej metody zależy od rodzaju danych, rozmiaru próby oraz rozkładów zmiennych. Metodę największej wiarygodności stosować można tylko dla wielowymiarowego rozkładu normalnego. W przypadku, gdy rozkład nie spełnia tego warunku stosować można metodę ADF, wymagającą próby liczącej od 200 do 500 obserwacji lub UMNK, dla której wymagana jest duża próba o liczebności powyżej 2500 [6].

Oszacowany model należy zweryfikować pod względem stopnia dopasowania oraz istotności parametrów. Stopień dopasowania modelu równań strukturalnych określa się najczęściej poprzez porównanie otrzymanego modelu z modelem nasyconym i niezależnym. W pierwszym z nich zakłada się, że wszystkie zmienne są ze sobą skorelowane, w drugim zaś, że korelacja nie występuje pomiędzy żadną spośród par zmiennych [17], [23].Wśród miar stopnia dopasowania modelu *SEM* za najważniejsze przyjmuje się statystykę *CMIN*/df_h, miary *IFI*, *TFI*, *RFI*, *NFI*, *CFI* porównujące estymowany model z modelem bazowym, średniokwadratowy błąd aproksymacji *RMSEA* oraz wartość krytyczną Hoeltera.

Pierwsza miara, oparta na statystyce χ^2 jest opisana wzorem:

$$CDMIN/df_h = \chi^2_{ML}/df_h, \tag{4}$$

$$\chi^2_{ML} = (n-1)[trace(S\hat{\mathrm{E}}^{-1}) - p + \ln(|\hat{\mathrm{E}}|) - \ln(|S|)], \tag{5}$$

gdzie df_h stanowi liczbę stopni swobody estymowanego modelu, n stanowi wielkość próby, p jest liczbą zmiennych obserwowalnych, S jest macierzą kowariancji dla próby, a $\hat{\mathrm{E}}$ jest macierzą odtworzenia S na podstawie oszacowanych parametrów.

Wartość statystyki mniejsza bądź równa dwa oznacza dobre dopasowanie modelu do danych [6]. Stopień dopasowania modelu *SEM* oceniany jest również przez szereg miar opartych na koncepcji porównywania modelu estymowanego z modelem bazowym. Przykładem takiego wskaźnika jest indeks *IFI* (*Incremental Fit Index*) określany wzorem:

$$IFI = \frac{T_b - T_h}{T_b - df_h}, \tag{6}$$

gdzie: T_h – statystyka chi-kwadrat estymowanego modelu, T_b – statystyka chi-kwadrat modelu niezależnego. Wartości wskaźnika *IFI* powinny zawierać się w przedziale od zera do jedności. Model uznaje się za dobrze dopasowany, jeśli wartość tego współczynnika jest większa od 0,95. W pracy [2] zalecane jest również wykorzystanie wskaźników *TFI, RFI, NFI* oraz*CFI,* których wyliczanie, jak i interpretacja opiera się na podobnej zasadzie co wskaźnik *IFI*.

Dla oceny modelu *SEM* wykorzystuje się powszechnie również wskaźnik *RMSEA* (*Root Mean Square Error of Approximation*). W przeciwieństwie do opisywanej miary *IFI*, podczas obliczania wskaźnika *RMSEA* nie następuje porównywanie modelu estymowanego z modelem bazowym. Wskaźnik ten oblicza się według wzoru:

$$RMSEA = \sqrt{\frac{T_h - df_h}{(n-1)df_h}}, \tag{7}$$

gdzie oznaczenia są analogiczne jak we wzorze (5). Im niższa wartość wyliczonego na podstawie modelu wskaźnika *RMSEA* tym lepszy stopień dopasowania modelu. Przyjmuje się, że dla wartości *RMSEA* mniejszej od 0,08 model jest dobrze dopasowany do danych.

Kolejną miarą służącą do oceny modelu jest wartość krytyczna Hoeltera obliczana dla danego poziomu istotności. Obliczana wartość krytyczna stanowi największą wielkość próby, dla której nie ma podstaw do odrzucenia hipotezy dotyczącej poprawności modelu przy danej wartości statystyki χ^2 oraz liczbie stopni swobody [6].

3. CZYNNIKI KSZTAŁTUJĄCE DECYZJE INWESTORÓW INDYWIDUALNYCH

Jednym z wyodrębnionych w analizie czynników wpływających na decyzje inwestorów była jakość rynku kapitałowego w Polsce. Przez jakość rynku rozumieć należy ogół cech rynku składających się na jego sprawne funkcjonowanie. Należą do nich: duża liczba kupujących i sprzedających, różnorodność potrzeb inwestycyjnych, obiektywność oraz dostępność opinii, informacji, łatwo dostępne miejsce, uzasadnione koszty transakcyjne, integralność rynku czy w efekcie uczciwość uczestników rynku [25]. W przeprowadzonym badaniu jakość rynku ograniczona została do subiektywnej oceny przez inwestorów indywidualnych informacji publicznych dostarczanych przez spółki pod względem ich kompletności, użyteczności, prawdziwości, porównywalności, jednoznaczności i dostępności, a także oceny sposobu zarządzania spółkami.

Oceniane elementy stanowią podstawę kształtowanych przez spółki relacji inwestorskich, stanowiących jedno z najważniejszych zagadnień w procesie budowy i funkcjonowania nowoczesnego, konkurencyjnego rynku finansowego. Zgodnie z definicją Amerykańskiego Instytutu Relacji Inwestorskich (NIRI), relacje inwestorskie stanowią element zarządzania strategicznego obejmujący finanse, komunikację, marketing oraz prawo papierów wartościowych, który umożliwia efektywną dwukierunkową komunikację między spółką i społecznością inwestorską, a w efekcie przyczynia się do rzetelnej wyceny papierów wartościowych przez rynek. Znaczenie relacji inwestorskich w dalszym ciągu rośnie, ze względu na następującą w procesie globalizacji deregulację i liberalizację przepływów kapitałowych, jak i całej gospodarki oraz rozwój nowych technologii informatycznych, telekomunikacyjnych oraz coraz szybszy i łatwiejszy dostęp do informacji [11], [12], [27].

Fundamentem relacji inwestorskich jest ujawnianie informacji, będące kluczowym czynnikiem dla funkcjonowania efektywnego rynku kapitałowego. Informacje publikowane przez spółki stanowią podstawę podejmowania decyzji przez inwestorów giełdowych, dlatego też firmom powinno zależeć na ich kompletności, aktualności i uczciwości. Ponadto rzetelne oraz przedstawione w profesjonalnej formie informacje przyczyniają się do wzrostu wiarygodności spółki dla inwestorów, kontrahentów czy analityków i obserwatorów rynku, co z kolei zwiększa zaufanie partnerów i instytucji finansowych. Dostępność informacji sprzyja pojawianiu się na rynku analiz, które przyczyniają się do wzrostu zainteresowania spółką.

Szczegółowy udział procentowy ocen udzielonych poszczególnym, analizowanym komponentom jakości rynku kapitałowego w Polsce zestawiono w tabeli 1. Oceniając jakość funkcjonowania rynku respondenci, w analizowanej grupie inwestorów indywidualnych, najczęściej przypisywali poszczególnym zmiennym tworzącym czynnik ocenę 3 (przeciętną) oraz 4 (powyżej przeciętnej). Najlepiej ocenione przez inwestorów zostały jakość zarządzania i prawdziwość informacji. W obu przypadkach odsetek nadanych ocen przeciętnych oraz wyższych był większy niż 90%. Najgorzej zaś inwestorzy indywidualni oceniają bezwłoczność i jednoznaczność informacji. W tym wypadku ponad 75% respondentów nadało ocenę przeciętną lub niższą. W pozostałych przypadkach ocenę przeciętną lub wyższą nadało 77-86% badanych.

Kolejny czynnik opisuje umiejętność posługiwania się analizą techniczną. Został on wyodrębniony na podstawie pytań dotyczących prognoz kształtowania się indeksów i cen akcji przy uwzględnieniu ich dotychczasowego trendu, formacji, linii oporu i wsparcia. Sama analiza techniczna powinna być narzędziem pomocniczym przy podejmowaniu decyzji inwestycyjnych, przydatnym w szczególności w krótkich okresach, a więc w przypadku inwestycji o charakterze spekulacyjnym [33]. Jednocześnie osoby posługujące się analizą techniczną w sposób prawidłowy powinny działać w sposób bardziej świadomy na rynku kapitałowym. Umiejętność posługiwania się narzędziami analizy technicznej posłużyła do podziału badanej zbiorowości na dwie grupy. Do pierwszej grupy zaliczeni zostali respondenci, dla których wartości tego czynnika kształtowały się powyżej średniej, natomiast do drugiej grupy zaliczono pozostałych.

Tabela 1.

Procentowy udział ocen poszczególnych zmiennych tworzących czynnik jakość funkcjonowania rynku

Ocena	1 – zła	2 – poniżej przeciętnej	3- przeciętna	4- powyżej przeciętnej	5- bardzo dobra
Kompletność informacji	2,67	15,67	44,33	31,67	5,67
Użyteczność informacji	1,00	13,00	39,00	38,00	9,00
Prawdziwość informacji	0,33	9,00	34,33	44,00	12,33
Bezwłoczność informacji	6,00	28,67	42,00	18,67	4,67
Porównywalność informacji	4,67	18,00	39,33	34,33	3,67
Jednoznaczność informacji	6,00	32,00	40,00	19,00	3,00
Dostępność informacji	5,00	16,00	28,33	37,33	13,33
Jakość zarządzania spółkami	1,00	6,67	51,33	39,33	1,67

Źródło: Opracowanie własne

W przypadku inklinacji w sferze opinii i preferencji, jak i ich wpływu na skłonność do ryzyka przełomową okazała się praca Longa, Shleifera, Summersa i Waldmanna [24], w której autorzy podjęli się wyjaśnienia przyczyn nieracjonalnego dyskonta w notowaniu jednostek zamkniętych funduszy inwestycyjnych. Dzieląc inwestorów na racjonalnych, podejmujących decyzje w oparciu o analizę fundamentalną i nieracjonalnych, kierujących się szumem informacyjnym pokazali oni, iż racjonalni uczestnicy rynku potrzebują dyskonta, które pozwalałoby im zrekompensować zwiększone ryzyko wahań notowań, wynikające ze złej wyceny jednostek przez inwestorów nieracjonalnych.

Inwestorzy racjonalni podejmują decyzje na podstawie właściwych sygnałów płynących z rynku, które następnie interpretują w prawidłowy sposób, dokonując wyceny aktywów i oceny ryzyka. Inwestorów bazujących zaś na szumie informacyjnym (*noise traders*) charakteryzują przede wszystkim błędne oczekiwania co do wariancji zwrotu, gdyż decyzje swoje podejmują oni w oparciu o szum informacyjny. Dokonując złej wyceny portfela, oceniają oni ryzyko inaczej, niż rynek. Wynika to z faktu, iż podejmując decyzje podlegają inklinacjom behawioralnym, przez co niepoprawnie interpretują zdarzenia podejmując decyzje błędne, bądź też opierające się na zdarzeniach nieistotnych z punktu widzenia konstruowania portfela. Powoduje to, że są oni w stanie podjąć większe ryzyko przy tym samym poziomie zasobów i awersji do ryzyka co inwestorzy racjonalni. Jednocześnie ponoszonemu, wyższemu ryzyku i popełnianym błędom towarzyszy większa częstotliwość zmian dokonywanych w portfelu.

Błędne wyobrażenia inwestorów *noise traders* dotyczące wariancji są wynikiem systematycznych błędów wnioskowania, które mają podłoże psychologiczne i wynikają z heurystyk stosowanych przez inwestorów. Mimo, iż heurystyki owe z jednej strony pomagają inwestorowi przebrnąć przez natłok informacji, to z drugiej zniekształcają one proces podejmowania decyzji poprzez nadmierną pewność siebie i własnej wiedzy,

złudzenie kontroli nad obserwowanymi zjawiskami, nadmierny optymizm, interpretowanie informacji w taki sposób, by potwierdzić wcześniejsze opinie (pułapka potwierdzenia), efekt myślenia wstecznego, dysonans poznawczy, selektywną percepcję oraz inne. Wymienione aspekty składają się na określone sentymenty czy też skłonności inwestorów zarówno w sferze opinii, jak i preferencji [36], [37]. W przeprowadzonym badaniu do określenia błędów w sferze opinii posłużyły przede wszystkim pytania dotyczące złudzenia kontroli oraz nadmiernego optymizmu inwestorów, zaś do identyfikacji inklinacji w sferze preferencji wykorzystane zostały efekty poprzednich inwestycji oraz perspektywa zysków i strat.

Wyodrębnione czynniki zestawiono w tabeli 2. Wartość każdej ze zmiennych obserwowalnych x_i jest przyporządkowana wybranej odpowiedzi na konkretne pytanie zawarte w kwestionariuszu.

Tabela 2.

Zmienne obserwowalne składające się na poszczególne czynniki

Czynnik (zmienna nieobserwowalna)	Oznaczenie	Zmienne obserwowalne
Umiejętność posługiwania się analizą techniczną	y_1	x_1, x_2, x_3, x_4, x_5, x_6
Jakość funkcjonowania rynku kapitałowego	y_2	x_7, x_8, x_9, x_{10},x_{11}
Inklinacje w sferze opinii	y_3	x_{12}, x_{13}, x_{14}
Inklinacje w sferze preferencji	y_4	x_{15}, x_{16}, x_{17}
Skłonność do ryzyka	y_5	x_{18}, x_{19}, x_{20}, x_{21}

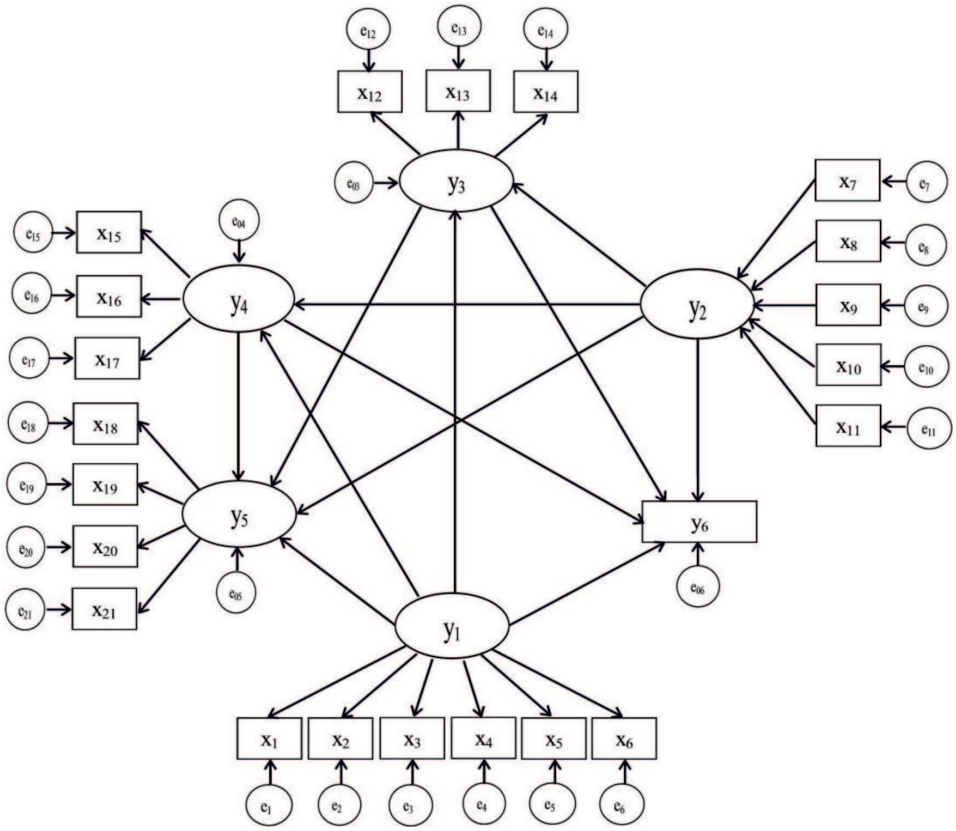
Źródło: Opracowanie własne

Zgodnie z teorią finansów behawioralnych, im większe błędy w sferze opinii i preferencji czynione przez inwestorów w procesie inwestowania, tym większa ich skłonność do ryzyka [24]. Jednocześnie popełniane błędy behawioralne powodują, że inwestorzy oczekują wyższej stopy zwrotu. Stąd w przeprowadzonym badaniu zakłada się, że popełniane błędy w sferze opinii i skłonności w sferze preferencji będą wpływać na wzrost oczekiwanej stopy zwrotu z dokonywanych inwestycji.

4. ANALIZA MECHANIZMÓW PODEJMOWANIA DECYZJI PRZEZ INWESTORÓW INDYWIDUALNYCH

Podstawą analizy mechanizmów podejmowania decyzji przez inwestorów indywidualnych jest model równań strukturalnych, którego schemat, dostosowany do celu niniejszego badania, zaprezentowany został na rysunku 1. Jest to model hipotetyczny, przyjęty na potrzeby ustalenia zależności strukturalnych pomiędzy czynnikami oraz wyjaśnienia postawionych w artykule hipotez. Opracowanie modelu, a następnie jego oszacowanie oraz weryfikacja dokonane zostały w pakiecie AMOS v. 16. Na rysunku 1 symbolami y_i {i=1,2,3,4,5} oznaczono czynniki nieobserwowalne, natomiast tworzące

je zmienne obserwowalne oznaczono symbolami x_i {i=1,2,...,21}. Zmienną opisującą oczekiwaną przez inwestorów stopę zwrotu określono symbolem y_6.



Rysunek 1. Schemat hipotetycznego modelu równań strukturalnych

Źródło: opracowanie własne w programie AMOS v.16

Podczas estymacji modelu *SEM* metodą największej wiarygodności czynnik y_1-umiejętność posługiwania się analizą techniczną oraz czynnik y_2-jakość funkcjonowania rynku kapitałowego okazały się statystycznie nieistotne i zostały usunięte z modelu. Ostatecznie w modelu pozostały czynnik y_3 charakteryzujący skłonność do ryzyka badanych inwestorów indywidualnych, czynniki y_4, y_5 opisujące inklinacje behawioralne oraz zmienna y_6 określająca satysfakcjonującą stopę zwrotu z inwestycji. Tabela 3 zawiera wyniki estymacji dla modelu zewnętrznego, a tabela 4 dla modelu wewnętrznego. W tabeli 5 zawarte zostały natomiast miary stopnia dopasowania modelu.

Wyniki uzyskane dla modelu zewnętrznego, zawarte w tabeli 3 wskazują, że wszystkie ładunki czynnikowe są statystycznie istotne. Tabela 4 zawiera wyniki estymacji modelu wewnętrznego. Istotne statystycznie oceny parametrów opisujących wpływ błędów w sferze opinii na poziomie 0,913 oraz błędów w sferze preferencji

Tabela 3.

Oszacowane parametry konfirmacyjnej analizy czynnikowej

Zależności	Parametry	Oceny parametrów	Wartość p
$x_{12} \leftarrow y_3$	α_1	0,284	–
$x_{13} \leftarrow y_3$	α_2	0,220	0,018
$x_{14} \leftarrow y_3$	α_3	0,694	0,004
$x_{15} \leftarrow y_4$	α_4	0,610	–
$x_{16} \leftarrow y_4$	α_5	0,539	0,000
$x_{17} \leftarrow y_4$	α_6	0,294	0,001
$x_{18} \leftarrow y_5$	α_7	0,556	–
$x_{19} \leftarrow y_5$	α_8	0,738	0,000
$x_{20} \leftarrow y_5$	α_9	0,268	0,000
$x_{21} \leftarrow y_5$	α_{10}	0,708	0,000

Źródło: opracowanie własne

Tabela 4.

Oszacowane parametry modelu wewnętrznego

Zależności	Parametry	Oceny parametrów	Oceny parametrów standaryzowanych	Wartość p
$y_5 \leftarrow y_3$	β_1	0,913	0,515	0,002
$y_5 \leftarrow y_4$	β_2	0,267	0,272	0,007
$y_6 \leftarrow y_3$	β_3	0,571	0,216	0,018
$y_6 \leftarrow y_4$	β_4	0,377	0,257	0,003

Źródło: opracowanie własne

Tabela 5.

Miary dopasowania modelu

Model	IFI	RMSEA	CMIN/DF	HOLTER
Estymowany	0,863	0,065	2,317	189
Nasycony	1	–	–	–
Niezależny	0,000	0,148	7,906	53

Źródło: opracowanie własne

na poziomie 0,267 wskazują na silne zwiększenie skłonności do ryzyka inwestorów indywidualnych w wyniku wymienionych błędów. Porównując przy tym wartości ocen parametrów standaryzowanych, równych odpowiednio 0,515 i 0,272, zauważalny jest znacznie silniejszy wpływ błędów popełnianych w sferze opinii na skłonność do ryzyka

inwestorów. Zgodnie z wcześniejszym założeniem ma miejsce równocześnie dodatni, statystycznie istotny wpływ tych czynników na oczekiwaną, satysfakcjonującą inwestorów stopę zwrotu. Wartości otrzymanych ocen wynoszą odpowiednio 0,571 dla wpływu błędów w sferze opinii oraz 0,377 dla błędów w sferze preferencji. Biorące pod uwagę oceny standaryzowane, można stwierdzić, że w tym przypadku wpływ obydwu czynników jest na zbliżonym poziomie.

Wartość wskaźnika *IFI* oszacowanego modelu *SEM* jest równa 0,863, a wartość *RMSEA* jest na poziomie 0,065, co pozwala stwierdzić, iż przedstawione miary świadczą o poprawnym dopasowaniu modelu do danych empirycznych.

Błędy popełnianie przez inwestorów oznaczają nieumiejętność wyciągania poprawnych wniosków z zaobserwowanych faktów. W związku z tym częściej podejmowane są działania ryzykowne, gdyż inwestorzy nie są w stanie oszacować w pełni stopnia ponoszonego ryzyka. Powyższe zależności są również wynikiem niższej jakości rynku, która jest cechą rynków wschodzących i rozwijających się, do których w dalszym ciągu zalicza się polski rynek kapitałowy[2]. Na tych rynkach inwestorzy są w stanie zaakceptować większe ryzyko inwestycji, ale oczekują przy tym większych zysków [26]. Na rynkach kapitałowych charakteryzujących się wyższą jakością ryzyko inwestycji jest mniejsze, przez co niższa powinna być również oczekiwana stopa zwrotu z inwestycji. Na wyższą jakość składają się między innymi kształtowane przez spółki relacje inwestorskie, których zadaniem jest nie tylko pozyskanie inwestorów, ale także umożliwienie im dokonanie właściwej i rzetelnej wyceny prowadzącej do zmniejszenia podejmowanego przez nich ryzyka inwestycyjnego.

W celu pogłębienia analizy, na podstawie wyodrębnionego i opisanego w punkcie poprzednim czynnika oznaczającego umiejętność posługiwania się analizą techniczną, podzielono badaną zbiorowość na dwie grupy. Pierwsza z nich, w której znaleźli się inwestorzy umiejący wykorzystywać analizę techniczną w poprawny sposób (tzn. otrzymana wartość czynnika była wyższa od średniej wynoszącej 2,17) składała się z 184 respondentów. W drugiej grupie, liczącej 131 inwestorów, znalazły się osoby, które posługiwały się narzędziami analizy technicznej w sposób niezadawalający, tj. wartość czynnika była na poziomie niższym lub równym średniej. Wyniki estymacji modelu wewnętrznego dla pierwszej i drugiej grupy zestawiono w tabelach 6 oraz 8, zaś ich miary dopasowania, odpowiednio w tabelach 7 i 9.

Zgodnie z oczekiwaniami, wpływ czynnika oznaczającego inklinacje w sferze opinii na skłonność do ryzyka i satysfakcjonującą ich stopę zwrotu z inwestycji, w pierwszej grupie inwestorów okazał się statystycznie nieistotny. Podobnie nieistotny, choć w znacznie mniejszym stopniu (wartość p kształtująca się na poziomie 0,057) okazał się wpływ inklinacji w sferze preferencji na wybrane czynniki. Natomiast w drugiej grupie inwestorów zakładane zależności okazały się statystycznie istotne. Reprezentowane przez tych inwestorów indywidualnych inklinacje zarówno w sferze opinii, jak

[2] Na podstawie klasyfikacji krajów wg FTSE (por. FTSE Global Equity Index Series Country Classification – September 2008 – www.ftse.com)

Tabela 6.

Oszacowane parametry modelu wewnętrznego – grupa pierwsza

Zależności	Parametry	Oceny parametrów	Oceny parametrów standaryzowanych	Wartość p
$y_5 \leftarrow y_3$	β_1	1,634	0,485	0,181
$y_5 \leftarrow y_4$	β_2	0,248	0,251	0,057
$y_6 \leftarrow y_3$	β_3	0,699	0,139	0,298
$y_6 \leftarrow y_4$	β_4	0,310	0,211	0,061

Źródło: opracowanie własne

Tabela 7.

Miary dopasowania modelu – grupa pierwsza

Model	IFI	RMSEA	CMIN/DF	HOLTER
Estymowany	0,815	0,073	1,987	128
Nasycony	1	–	–	–
Niezależny	0,000	0,143	4,727	52

Źródło: opracowanie własne

Tabela 8.

Oszacowane parametry modelu wewnętrznego – grupa druga

Zależności	Parametry	Oceny parametrów	Oceny parametrów standaryzowanych	Wartość p
$y_5 \leftarrow y_3$	β_1	0,616	0,481	0,007
$y_5 \leftarrow y_4$	β_2	0,327	0,323	0,039
$y_6 \leftarrow y_3$	β_3	0,527	0,272	0,027
$y_6 \leftarrow y_4$	β_4	0,524	0,343	0,015

Źródło: opracowanie własne

Tabela 9.

Miary dopasowania modelu – grupa druga

Model	IFI	RMSEA	CMIN/DF	HOLTER 0,05
Estymowany	0,890	0,062	1,500	120
Nasycony	1	–	–	–
Niezależny	0,000	0,156	4,135	42

Źródło: opracowanie własne

i w sferze preferencji wpływają istotnie na ich skłonność do ryzyka oraz satysfakcjonującą stopę zwrotu. Również w tym przypadku wpływ błędów w sferze opinii na

skłonność do ryzyka okazał się znacznie silniejszy od wpływu skłonności w sferze preferencji. Natomiast wpływ obydwu inklinacji behawioralnych na satysfakcjonującą stopę zwrotu jest porównywalny. Przedstawione w tabelach 8, 10 miary jakości modelu *SEM* wskazują na poprawne własności statystyczne obydwu estymowanych modeli.

Inwestorzy indywidualni z grupy pierwszej są mniej podatni na oddziaływanie inklinacji behawioralnych, co potwierdza drugą hipotezę badawczą przyjętą w artykule. Oznacza to, iż inwestorów posługujących się prawidłowo narzędziami analizy technicznej uznać można za bardziej świadomych mechanizmów działających na rynku kapitałowym, w porównaniu z inwestorami z grupy drugiej.

5. ANALIZA BOOTSTRAP

Ze względu na niezbyt dużą liczebność próby przeprowadzono dodatkowo procedurę bootstrap, której wyniki potwierdzić miały poprawność statystyczną oszacowanego w artykule modelu *SEM*. W pracach [13] i [19] opisana została procedura bootstrap, którą autorzy wykorzystali do wnioskowania w zakresie modelu *SEM*. Procedura polega na losowaniu wielu podróbek z oryginalnych danych, co pozwala na sprawdzenie rozkładu parametrów w odniesieniu do wyjściowej próby.

Na podstawie posiadanych danych zastosowano procedurę bootstrap przy użyciu estymatora największej wiarygodności. Wykorzystano 5000 nowych oszacowań parametrów modelu obejmującego całą badaną zbiorowość. Wykonanie procedury pozwoliło na otrzymanie nowych ocen parametrów, które stanowią średnie z ocen dla wszystkich estymowanych modeli. Procedurę bootstrap wykonano trzykrotnie, najpierw dla całej próby inwestorów indywidualnych, a następnie dla grupy pierwszej oraz grupy drugiej. Uzyskane wyniki przedstawiono w tabeli 10, gdzie dla każdego z modeli pierwsze dwie kolumny zawierają ocenę parametru estymacji oraz średnie z ocen analizy bootstrap. Kolumna trzecia zawiera wartości p, świadczące o istotności albo nieistotności statystycznej parametrów wyznaczone na podstawie procedury bootstrap.

Tabela 10.

Wyniki analizy bootstrap

Zależności	Model dla całej próby			Model dla I grupy			Model dla II grupy		
	Ocena	Średnia	Wartość p	Ocena	Średnia	Wartość p	Ocena	Średnia	Wartość p
$y_5 \leftarrow y_3$	0,913	1,027	0,001	1,634	1,774	0,001	0,616	0,730	0,005
$y_5 \leftarrow y_4$	0,267	0,278	0,003	0,248	0,257	0,069	0,327	0,356	0,019
$y_6 \leftarrow y_3$	0,571	0,602	0,034	0,699	0,525	0,082	0,527	0,566	0,037
$y_6 \leftarrow y_4$	0,377	0,396	0,002	0,310	0,372	0,084	0,524	0,586	0,009

Źródło: opracowanie własne

W przypadku modelu szacowanego dla całej grupy respondentów potwierdzono istotność wszystkich zależności. Istotne okazały się również wszystkie zależności dla

drugiej grupy inwestorów. Natomiast w przypadku inwestorów posługujących się analizą techniczną w sposób zadowalający istotny okazał się jedynie wpływ błędów w sferze opinii na ich skłonność do ryzyka. Natomiast wpływ inklinacji w sferze preferencji na skłonność do ryzyka oraz wpływ inklinacji w sferze opinii oraz preferencji na oczekiwaną stopę zwrotu okazały się statystycznie nieistotne.

6. ZAKOŃCZENIE

W artykule przedstawiono wyniki badania empirycznego opisującego czynniki behawioralne determinujące decyzje podejmowane przez inwestorów indywidualnych na GPW w Warszawie. Na potrzeby analizy utworzony został hipotetyczny model *SEM*, w którym uwzględniono czynniki charakteryzujące błędy inwestorów indywidualnych w sferze opinii i skłonności w sferze preferencji, skłonność do ryzyka, umiejętność wykorzystania analizy technicznej oraz postrzeganą przez inwestorów jakość funkcjonowania rynku. W celu pozyskania danych opracowano kwestionariusz i przeprowadzono ankietę wśród inwestorów indywidualnych zrzeszonych w Stowarzyszeniu Inwestorów Indywidualnych. Uzyskane wyniki posłużyły do estymacji oraz weryfikacji modelu *SEM*, zarówno dla całej próby, jak i w dwóch podpróbach.

Zgodnie z postawionymi hipotezami, w modelu założono dodatni wpływ inklinacji behawioralnych na skłonność do ryzyka i oczekiwaną przez inwestorów stopę zwrotu oraz przyjęto, że wpływ ten będzie istotnie mniejszy w grupie inwestorów indywidualnych posługujących się analizą techniczną w sposób właściwy. Wyniki estymacji modelu pozwoliły na identyfikację założonych czynników w grupie badanych inwestorów indywidualnych oraz potwierdzenie wpływu inklinacji w sferze opinii i preferencji na wzrost skłonności do ryzyka. Wyniki przeprowadzonego badania okazały się zgodne z wnioskami zawartymi w pracy [24]. Ponadto wykazano wpływ inklinacji behawioralnych na wzrost oczekiwanej przez inwestorów stopy zwrotu, co rekompensować ma wzrost ryzyka z tytułu popełnianych przez nich błędów. Analiza przeprowadzona w grupie inwestorów indywidualnych, wyróżnionych ze względu na umiejętność właściwego posługiwania się narzędziami analizy technicznej, pozwoliła wykazać, że wpływ błędów w sferze opinii i skłonności w zakresie preferencji na proces decyzyjny jest u nich mniejszy, niż w grupie nieposiadającej tej umiejętności. Ze względu na relatywnie małą liczebność próby (n=315) przeprowadzono dodatkowo procedurę bootstrap, której wyniki potwierdziły statystyczną poprawność uzyskanego modelu i słuszność sformułowanych wniosków.

Katedra Ekonometrii i Statystyki,
Wydział Nauk Ekonomicznych i Zarządzania,
Uniwersytet Mikołaja Kopernika,
ul. Gagarina 13A, 87-100 Toruń
e-mail: emo@umk.pl

LITERATURA

[1] Barberis N., Shleifer A., Vishny R. (1998), A model of investor sentiment, *Journal of Financial Economics,* 49.

[2] Bollen K. A (1989), *Structural Equations with Latent Variables*, Wiley.

[3] Bollinger J. (1985), *The Capital Growth Letter*, Research and Market.

[4] Brown T.A. (2006) , *Confirmatory Factor Analysis for Applied Research*, Guilford Press.

[5] Brzeziński J. (1996), *Metodologia badań psychologicznych (Część V)*. PWN Warszawa.

[6] Byrne B.M. (2010), *Structural equation modeling with AMOS: Basic concepts, applications, and programming* (2nd ed.), Routledge/Taylor & Francis, New York.

[7] Cronbach L.J. (1951), Coefficient alpha and the internal structure of tests, *Psychometrika*, 16(3), 297-334.

[8] Czerwonka M., Gorlewski B. (2008), *Finanse behawioralne. Zachowania inwestorów i rynku*, SGH, Warszawa.

[9] Daniel K., Hirshleifer D., Subrahmanyam A. (1997), *A theory of overconfidence, self-attribution, and security market under- and over-reactions*. Unpublished working paper. University of Michigan.

[10] Doman M. (2011), *Mikrostruktura giełd papierów wartościowych*. Wydawnictwo Uniwersytetu Ekonomicznego w Poznaniu.

[11] Dziawgo D., Gajewska-Jedwabny A. (2006), Relacje inwestorskie – nowoczesna komunikacja spółek z rynkiem*, CEO Magazyn Top Menagerów*, kwiecień 2006.

[12] Dziawgo D. (2009), Idea zrównoważonego rozwoju w relacjach inwestorskich, w: Sidorczuk-Pietraszko E. (red.), *Funkcjonowanie przedsiębiorstw w warunkach zrównoważonego rozwoju i gospodarki opartej na wiedzy*, WSE, Białystok.

[13] Efron B. (1979), Bootsrap methods: Another look at the jackknife*, Annals of Statistic* no. 710.

[14] Gatnar E. (2003), *Statystyczne modele struktury przyczynowej zjawisk ekonomicznych*, Akademia Ekonomiczna, Katowice.

[15] Goldberg J., Von Nitzsch R. (2001), *Behavioral Finance*, New York.

[16] Hong H., Stein J.C (1999), A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets, *Journal of Finance,* 54, 2143-2184.

[17] Kahneman D., Tversky A. (1979), Prospect Theory: An Analysis of Decision under Risk, *Econometrica*, XLVII, 263-291.

[18] Kaplan D. (2000), *Structural Equation Modeling: Foundations and Extensions*, Sage Publications.

[19] Kotz S., Johnson N.I. (1992), *Breakthrough in statistics*, Springer-Verlag, New York.

[20] Konarski R. (2010), *Modele równań strukturalnych. Teoria i praktyka.*, PWN, Warszawa.

[21] Lakonishok J, Shleifer A, Vishny R.W. (1994), Contrarian Investment, Extrapolation, and Risk, *Journal of Finance*, 4.

[22] Likert R.(1932), Technique for the Measurement of Attitudes, *Archives of Psychology*, 140, 55.

[23] Loehlin J.C. (1987), *Latent variable models: An introduction to factor, path and structural analysis*, Erlbaum.

[24] Long J. B., Shleifer A., Summers L.H., Waldmann R.J. (1990), Noise Trader Risk in Financial Markets, *Journal of Political Economy*, University of Chicago Press, 98(4), 703-738.

[25] Maginn J.L., Tuttle D.L., McLeavey D.W., Pinto J.E. (2007), *Managing investment portfolios workbook*, John Wiley & Sons, New Jersey.

[26] Merrill Lynch, *The Whitepapers*, 2008.

[27] Niedziółka D.A. (2008), *Relacje inwestorskie*, PWN, Warszawa.

[28] Osińska M. (2008), *Ekonometryczna analiza zależności przyczynowych*, Uniwersytet Mikołaja Kopernika, Toruń.

[29] Osińska M., Pietrzak M., Żurek. (2011), Wykorzystanie modeli równań strukturalnych do opisu psychologicznych mechanizmów podejmowania decyzji na rynku kapitałowym, *Acta Universitatis Nicolai Copernici – Ekonomia*, w druku

[30] Pearl J. (2000), *Causality. Models, reasoning and inference*, Cambrige.
[31] *Profil inwestora giełdowego w 2010 r*. Raport SII na temat Ogólnopolskiego Badania Inwestorów w 2010, http://www.sii.org.pl/static/img/004235/RaportOBI2010m.pdf
[32] Shleifer A. (2000), *Inefficient Markets: An Introduction to Behavioral Finance and the Psychology of Investing. Financial Management Association*, Harvard Business School Press, Boston.
[33] Tarczyński W. (1997), *Rynki kapitałowe. Metody ilościowe*, tom I i II, Placet, Warszawa.
[34] Tyszka T. (2004), *Psychologia Ekonomiczna*, GWP, 2004.
[35] Wang X. L., Shi K., Fan H. X. (2006), *Chinese Stock Markets, psychological mechanisms, investment behaviors, risk perception, individual investors*, „Journal of Economic Psychology", 27(6), 762-780.
[36] Zaleśkiewicz T. (2003), *Psychologia inwestora giełdowego. Wprowadzenie do behawioralnych finansów*, Gdańskie Wydawnictwo Psychologiczne, Gdańsk.
[37] Zielonka P. (2006), *Behawioralne aspekty inwestowania na rynku papierów wartościowych*, CeDeWu, Warszawa

Załącznik 1. Pytania zawarte w kwestionariuszu, które posłużyły do wyodrębnienia zmiennych ukrytych

Czynnik Y_1: umiejętność posługiwania się analizą techniczną

13. Opierając się na sygnałach (informacjach) wskazanych poniżej, odpowiedz jak, Twoim zdaniem, w najbliższej przyszłości zachowa się indeks WIG?

(–2 oznacza silny spadek WIG, 0 oznacza brak wpływu sygnału na indeks WIG, 2 oznacza silny wzrost WIG)
– Przełamywanie kolejnych poziomów wsparcia

☐ -2 ☐ -1 ☐ 0 ☐ 1 ☐ 2 (x_1)

– Utworzenie przez wykres WIG formacji głowy przez spadający indeks WIG i ramion po dużych wzrostach

☐ -2 ☐ -1 ☐ 0 ☐ 1 ☐ 2 (x_2)

– Załamanie głównej linii trendu wzrostowego indeksu WIG

☐ -2 ☐ -1 ☐ 0 ☐ 1 ☐ 2 (x_3)

– Po spadkach cen akcji – duża przewaga popytu

☐ -2 ☐ -1 ☐ 0 ☐ 1 ☐ 2 (x_4)

– Rosnący indeks WIG kolejno potwierdza główną linię trendu

☐ -2 ☐ -1 ☐ 0 ☐ 1 ☐ 2 (x_5)

– Istotne zwiększenie się wartości zlecenia kupna

☐ -2 ☐ -1 ☐ 0 ☐ 1 ☐ 2 (x_6)

Czynnik Y_2: jakość funkcjonowania rynku kapitałowego

1. Jak oceniasz jakość publicznych informacji na temat spółek dostępnych na rynku kapitałowym?

(Wybór środkowego kwadratu oznacza brak przewagi jednej z dwóch, wykluczających się ocen; d*la każdego wiersza wybierz dokładnie jedną odpowiedź)*

Informacje są niekompletne	☐ ☐ ☐ ☐ ☐	Informacje są kompletne	(x_7)
Informacje są nieprawdziwe	☐ ☐ ☐ ☐ ☐	Informacje są prawdziwe	(x_8)
Informacje są podawane z opóźnieniem	☐ ☐ ☐ ☐ ☐	Informacje są podawane bezzwłocznie	(x_9)
Informacje są nieporównywalne	☐ ☐ ☐ ☐ ☐	Informacje są porównywalne	(x_{10})
Informacje są niejednoznaczne	☐ ☐ ☐ ☐ ☐	Informacje są jednoznaczne	(x_{11})

Czynnik Y_3: błędy w sferze opinii

6. Załóżmy, że na giełdzie od 6 miesięcy trwa hossa oraz panuje korzystna sytuacja makroekonomiczna. Jakie jest, według Ciebie, prawdopodobieństwo, że Tobie uda się wypracować ponadprzeciętny zysk?

☐ (0%-50%) ☐ (50%-60%) ☐ (60%-70%) ☐ (70%-80%) ☐ (80%-100%) (x_{12})

15. Załóżmy, że złożony przez Ciebie wniosek na rozpoczęcie działalności gospodarczej otrzymał dotację w sumie 50 000 złotych. Jakie jest, według Ciebie, prawdopodobieństwo, że Twój biznes się powiedzie?

☐ (0%-20%) ☐ (20%-40%) ☐ (40%-60%) ☐ (60%-80%) ☐ (80%-90%) (x_{13})

18. Jak często dokonujesz zmian w portfelu w wybranym przeze Ciebie horyzoncie inwestycyjnym?

(1 – nie dokonuję żadnych zmian, 5 –staram się jak najczęściej zmieniać skład portfela)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 (x_{14})

Czynnik Y_4: błędy w sferze preferencji

10. Twoja inwestycja w akcje przyniosła Ci 30% zysku. Na rynku od 6 miesięcy panuje hossa i nie ma wyraźnych sygnałów zmiany trendu. Jaki procent oszczędności byłbyś skłonny dodatkowo zainwestować?

☐ 10% ☐ 25% ☐ 50% ☐ 75% ☐ 100% (x_{15})

14. Załóżmy, że posiadasz portfel akcyjny o wartości 10 000 złotych, – przynoszący zysk 10% miesięcznie. Z informacji dochodzących z rynku wynika, że inwestycja w spółkę X zapewni co najmniej 60% zysku w okresie dwóch miesięcy. Jaka byłaby twoja decyzja odnośnie zmian w posiadanym portfelu?

☐ Brak zmian w posiadanym portfelu akcyjnym ☐ Zakup akcji spółki X o wartości 3 000 złotych

☐ Zakup akcji spółki X o wartości 6 000 złotych ☐ Zakup akcji spółki X o wartości 10 000 złotych

☐ Zakup akcji spółki X o wartości 20 000 złotych, posługując się kredytem na sumę 10 000 złotych (x_{16})

16. Załóżmy, że inwestycja 20000 złotych w akcje przyniosła Ci 5 000 złotych zysku. Analitycy rynku zapowiadają dalsze wzrosty i hossę trwającą co najmniej rok. Jak zamierzasz inwestować dalej?

☐ Robisz rewizję portfela inwestując 20 000 złotych i odkładasz zysk na inne cele.

☐ Inwestujesz 20 000 złotych wraz z uzyskanym zyskiem równym 5 000 zł.

☐ Inwestujesz 25 000, inwestując dodatkowo część swoich dodatkowych przychodów (np. 5 000 zł)

☐ Inwestujesz 25 000, inwestując dodatkowo całość swoich dodatkowych przychodów (np. 10 000 zł)

☐ Inwestujesz 25 000, całość swoich dodatkowych przychodów (np. 10 000 zł) i jeszcze dodatkowe 10 000 zł korzystając z kredytu. (x_{17})

Czynnik Y_5: skłonność do ryzyka

4. Czy zgadzasz się ze stwierdzeniem, że inwestujesz bardziej ryzykownie od innych?

(1 – całkowicie się nie zgadzam, 5 – całkowicie się zgadzam)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 (x_{18})

8. Czy inwestując na co dzień zdarza Ci się podejmować decyzje będące na krawędzi ryzyka?

(1 – nigdy , 5 – tak, zawsze)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 (x_{19})

12. Czy zgadzasz się ze stwierdzeniem: Jeżeli ktoś zainwestuje ryzykownie i zarobi, to mam ochotę postąpić podobnie?

(1 – całkowicie nie zgadzam się, 5 –całkowicie się zgadzam)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 (x_{20})

17. Czy podczas inwestowania zdarza się, że podejmujesz ryzyko pomimo, że nie jest to konieczne?

(1 – nie, nigdy , 5 – tak, zdarzało mi się to niejednokrotnie)

☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 (x_{21})

OCENA WPŁYWU CZYNNIKÓW BEHAWIORALNYCH I RYNKOWYCH NA POSTAWY INWESTORÓW INDYWIDUALNYCH NA POLSKIM RYNKU KAPITAŁOWYM ZA POMOCĄ MODELU SEM

S t r e s z c z e n i e

W artykule podjęta została próba empirycznego wyjaśnienia zależności skłonności do ryzyka i poziomu oczekiwanej stopy zwrotu od czynników psychologicznych, leżących u podstaw decyzji inwestorów indywidualnych aktywnych na Giełdzie Papierów Wartościowych w Warszawie. Celem artykułu jest identyfikacja i opis współzależności między nieobserwowalnymi czynnikami takimi jak inklinacje behawioralne,

skłonność do ryzyka, umiejętność posługiwania się przez inwestorów analizą techniczną, a także jakość rynku za pomocą modelu równań strukturalnych. Przyjęto hipotezę, że inklinacje behawioralne tj. błędy popełniane w sferze opinii i skłonności w zakresie preferencji powodują wzrost skłonności do ryzyka inwestorów indywidualnych i wzrost oczekiwanej przez nich stopy zwrotu. Założono także, że wpływ inklinacji behawioralnych na decyzje inwestycyjne jest mniejszy wśród inwestorów indywidualnych umiejących się posługiwać analizą techniczną w sposób właściwy, niż u pozostałych inwestorów. Wyniki estymacji modelu pozwoliły na identyfikację założonych czynników w grupie badanych inwestorów indywidualnych oraz potwierdzenie wpływu inklinacji w sferze opinii i preferencji na wzrost skłonności do ryzyka.

Słowa kluczowe: modele równań strukturalnych (*SEM*), finanse behawioralne, inklinacje behawioralne, skłonność do ryzyka, analiza bootstrap

EVALUATION OF IMPACT OF BEHAVIORAL AND MARKET FACTORS ON ATTITUDES OF INDIVIDUAL INVESTORS IN POLISH CAPITAL MARKET USING SEM MODEL

A b s t r a c t

In the paper we attempt to explain the impact of psychological factors on propensity for risk and the level of expected return exhibited by individual investors active on the Stock Exchange in Warsaw. The article aims at identification and description of the relationship between unobservable factors such as behavioral inclinations, propensity for risk, ability to use technical analysis as well as market quality with application of Structural Equation Model. The hypothesis was put that behavioral inclinations such as errors in the opinions and preferences cause an increase in individual propensity for risk as well as the increase in the level of expected and satisfactory return.
It was also assumed that investors that properly use technical analysis in their decision-making process are less sensitive for behavioral inclinations than the others. The results of estimation of the model allowed for identification of the unobservable psychological factors and confirmed the impact of the defined factors for the increase of individual propensity for risk.

Keywords: structural equations model (*SEM*), behavioral finance, behavioral inclinations, propensity for risk, bootstrap analysis

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

EMIL PANEK

SYSTEM WALRASA I ZAPASY

1. WSTĘP

Przez system Walrasa w ekonomii matematycznej rozumiemy zazwyczaj układ równań różniczkowych dynamiki cen w n-produktowej gospodarce

$$\dot{p}(t) = \sigma f(p(t)), \tag{1}$$

gdzie:

t – zmienna czasu, $t \in R^1_+ = [0, +\infty)$,

n – liczba towarów (wytwarzanych i/lub zużywanych) w gospodarce,

$p(t) = (p_1(t), ..., p_n(t))$ – wektor cen towarów w momencie t,

$\dot{p}(t) = (\dot{p}_1(t), ..., \dot{p}_n(t))$ – pochodna funkcji (trajektorii) cen w momencie t,

$f(p(t)) = (f_1 p(t), ..., f_n p(t))$ – funkcja popytu nadwyżkowego(excess demand function),

$$f(p(t)) = f^d(p(t)) - f^s(p(t)),$$

$f^d(p(t)) = (f^d_1(p(t)), ..., f^d_n(p(t)))$ – funkcja zagregowanego popytu na towary w momencie t,

$f^s(p(t)) = (f^s_1(p(t)), ..., f^s_n(p(t)))$ – funkcja zagregowanej podaży towarów w gospodarce w momencie t,

σ – dodatni wskaźnik proporcjonalności (prędkości reakcji cen na zmianę popytu i podaży) [1].

Układ równań (1) opisuje prosty mechanizm dostosowywania cen do popytu i podaży, zgodnie z którym:

- jeżeli popyt na pewien towar przekracza jego podaż, a więc gdy w gospodarce mamy do czynienia z niedoborem towaru, jego cena zaczyna rosnąć,
- jeżeli popyt na towar jest niższy od jego podaży, tj. gdy w gospodarce mamy nadmiar pewnego towaru, wówczas jego cena zaczyna maleć,
- jeżeli popyt na towar jest równy jego podaży, cena towaru nie zmienia się, zob. np. [4], [5].

Opisany mechanizm zakłada nie tylko antropomorficzne ucieleśnienie rynku konkurencyjnego w postaci „niewidzialnej ręki", ale także bezwzględne jego podporządkowanie zasadzie, że transakcje kupna-sprzedaży towarów dochodzą do skutku dopiero

[1] Nie jest to najogólniejsza postać równania dynamiki cen w gospodarce Walrasa, zob. np. [2]. W celu zachowania prostoty dalszych wywodów świadomie pozostajemy przy równaniu (1).

po osiągnięciu przez gospodarkę równowagi. Nie można w żadnym razie układu (1) traktować jako opisu codziennych transakcji w gospodarce rynkowej, które są zawierane permanentnie – bez względu na to czy ceny $p(t)$, po których są sprzedawane i/lub kupowane towary, są w chwili t zawierania transakcji cenami równowagi walrasowskiej, czy nie. Dobrze oczywiście, jeżeli są, ale najczęściej tak nie jest i nikt się w praktyce specjalnie nad tym nie zastanawia. Wówczas w (1) $f(p(t)) \neq 0$, a to oznacza, że chwilowo ma miejsce niedobór jednych towarów ($f_i(p(t)) > 0$), prowadzący do wzrostu cen i w konsekwencji do wzrostu ich podaży w przyszłości oraz nadmiar innych ($f_j(p(t)) < 0$), prowadzący do powstawania zapasów, których w systemie Walrasa w ogóle się nie uwzględnia.

W modelu, który prezentujemy poniżej, zapasy nie mają wpływu na ceny towarów. Rosną, gdy na rynku ma miejsce nadwyżka podaży i maleją, gdy zachodzi sytuacja odwrotna. Za gromadzenie i dystrybucję zapasów odpowiada rząd lub powołana przez rząd agencja, która przyjmuje (nieodpłatnie) nadwyżki towarów od producentów po to, aby w sytuacji niedoboru móc je przekazać (również nieodpłatnie) finalnym odbiorcom zbiorowym, reprezentowanym np. przez takie instytucje pożytku publicznego jak opieka społeczna, zdrowotna czy pozarządowe organizacje charytatywne.

Pokażemy, że przy standardowych założeniach również w takiej nietypowej gospodarce istnieje (globalnie stabilny) stan równowagi konkurencyjnej.

2. PROSTY SYSTEM WALRASA Z ZAPASAMI

Niech $f^s(p(t)) = (f_1^s(p(t)), ..., f_n^s(p(t)))$ oznacza wektor podaży towarów w gospodarce w momencie t, który dalej utożsamiamy z ich produkcją w tym momencie. Przez $z(t) = (z_1(t), ..., z_n(t))$ oznaczamy wektor zapasów w momencie t, natomiast przez $\mu \in (0, 1)$ stopę naturalnego ubytku zapasów.

Jeżeli popyt $f_i^d(p(t))$ na towar i-ty przekracza podaż (produkcję) $f_i^s(p(t))$, jego zaspokojenie wymaga sięgnięcia po zapasy, które w rezultacie zmaleją.

Jeżeli $f_i^d(p(t)) = f_i^s(p(t))$, wtedy w momencie t nie ma potrzeby sięgania po zapasy, ale zmieniają się one z tytułu ich naturalnego ubytku (ze stopą μ).

Jeżeli $f_i^d(p(t)) = f_i^s(p(t)) - \mu z_i(t)$, wtedy zapasy pozostają na stałym poziomie.

Jeżeli $f_i^d(p(t)) < f_i^s(p(t)) - \mu z_i(t)$, zapasy towaru i-tego rosną.

W prezentowanym modelu, podobnie jak w jego wersji oryginalnej, ceny towarów na rynku kształtują się zgodnie z równaniem (1). Strumień dochodów konsumentów w momencie t jest równy wartości zrealizowanego popytu $\left\langle p(t), \min\left\{f^s(p(t)), f^d(p(t))\right\}\right\rangle$ czyli produkcji sprzedanej (a nie wytworzonej) przez producentów; symbolem

$$\min\left\{f^s(p(t)), f^d(p(t))\right\} = (\min\{f_1^s(p),\ f_1^d(p)\}, ..., \min\{f_n^s(p),\ f_n^d(p)\})$$

oznaczamy tutaj wektor popytu zrealizowanego przy cenach p, a symbolem $\langle x, y\rangle$ iloczyn skalarny wektorów $x, y \in R^n$: $\langle x, y\rangle = \sum_{i=1}^{n} x_i, y_i$. Wektor towarów, które nie

znalazły w momencie t nabywcy na rynku

$$\max\left\{f^s(p(t))-f^d(p(t)),0\right\}=\left(\max\left\{f_1^s(p(t))-f_1^d(p(t)),0\right\},...,\max\left\{f_n^s(p(t))-f_n^d(p(t)),0\right\}\right)$$

zostaje nieodpłatnie gromadzony w postaci zapasów. Tam gdzie zapasy towaru i są dodatnie, $z_i(t) > 0$, ich dynamikę opisuje równanie

$$\dot{z}_i(t) = f_i^s(p(t)) - f_i^d(p(t)) - \mu z_i(t). \tag{2a}$$

Natomiast tam gdzie dochodzi do wyczerpania zapasów, tj. gdy $z_i(t) = 0$, przyjmujemy:

$$\dot{z}_i(t) = \max\{f_i^s(p(t)) - f_i^d(p(t)) - \mu z_i(t),\ 0\}. \tag{2b}$$

Przez system Walrasa z zapasami rozumiemy układ $2n$ równań różniczkowych dynamiki cen towarów i ich zapasów w gospodarce konkurencyjnej:

$$\dot{p}(t) = \sigma[f^d(p(t)) - f^s(p(t))], \tag{3}$$

$$\dot{z}_i(t) = \begin{cases} f_i^s(p(t)) - f_i^d(p(t)) - \mu z_i(t), & gdy \quad z_i(t) > 0, \\ \max\{f_i^s(p(t)) - f_i^d(p(t)) - \mu z_i(t), \quad 0\}, & gdy \quad z_i(t) = 0, \end{cases} \tag{4}$$

$$i = 1, ..., n.$$

Weźmy dowolny wektor cen $p^0 > 0$ i dowolny wektor $z^0 \geqq 0$.

Δ **Definicja 1.** Określone na półosi czasu $R_+^1 = [0, +\infty)$ rozwiązanie $p(t), z(t)$ układu (3)-(4) z dodatnią funkcją $p(t)$ i nieujemną funkcją $z(t)$, spełniające warunek początkowy

$$p(0) = p^0 > 0, \quad z(0) = z^0 \geqq 0 \tag{5}$$

nazywamy (p^0, z^0, ∞) – dopuszczalną trajektorią cen i zapasów w gospodarce Walrasa.

▲

Obowiązuje standardowy układ założeń [2]:

(I) $f^d, f^s \in C^1(R_+^n \setminus \{0\} \to R^n)$,

(II) $p_i = 0 \Rightarrow f_i(p) > 0$,

(III) $\forall p \geqq 0\ \forall \lambda > 0 \left(f^d(\lambda p) = f^d(p)\ \ \&\ \ f^s(\lambda p) = f^s(p)\right)$,

(IV) $\forall p \geqq 0 \left(\langle p, f^d(p)\rangle = \langle p, f^s(p)\rangle\right)$ (prawo Walrasa) ,

(V) $\forall p \in P_+^n(1) = \{p \in R_+^n \left| \sum_{i=1}^n p_i = 1\right\} \quad \forall \lambda \in R^n \setminus (R_+^n \cup R_-^n)$

$$\lambda J(p) \lambda^T < 0$$

[2] Z ich interpretacją można zapoznać się np. w pracy [4] rozdz. 3

gdzie

$$J(p) = \left(\frac{\partial f_i}{\partial p_j}\right)_{(n,n)}, \quad f = f^d - f^s$$

(λ jest wektorem wierszowym, T jest znakiem transpozycji, R^n_+ oznacza nieujemny, a R^n_- niedodatni orthant przestrzeni R^n).

O własnościach (p^0, z^0, ∞)-dopuszczalnych trajektorii cen i zapasów mówi poniższe twierdzenie.

□ **Twierdzenie 1.** Przy założeniach (I) – (IV) :

(i) każde rozwiązanie układu (3) – (4) spełnia warunek:
$\forall t \geqslant 0\ (p(t) > 0\ \ \&\ \ z(t) \geqq 0)$,

(2i) trajektoria cen $p(t)$ leży na powierzchni kuli n-wymiarowej $K^n_r(0)$ o promieniu $r = \left\|p^0\right\|_E$ i środku w 0 (przez $\|x\|_E$ oznaczamy normę Euklidesa wektora $x \in R^n$).

Dowód. (i) Załóżmy, że $\exists(t' > 0)\exists i \in \{1, ..., n\}\quad (p_i(t') = 0)$. Wtedy $p_i(t') > 0$ (zgodnie z warunkiem (II)). Trajektoria cen jest funkcją ciągłą i różniczkowalną na obszarze określoności zatem $\exists\varepsilon > 0\forall t \in (t' - \varepsilon, t')\,(p_i(t) < 0)$ i ponieważ $p^0_i > 0$, więc $\exists\ t" > 0\ \ (p_i(t") = 0\ \ \&\ \ \dot{p}_i(t") \leqslant 0)$, co jest sprzeczne z warunkiem (II) i kończy dowód części (i).

(2i) Z prawa Walrasa otrzymujemy:

$$\forall t \geqslant 0\ \left(2\langle p(t), \dot{p}(t)\rangle = 2\sigma\langle p(t), f^d(p(t)) - f^s(p(t))\rangle = 0\right),$$

czyli

$$\forall t \geqslant 0\ \left(2\int_0^t \langle p(\theta), \dot{p}(\theta)\rangle d\theta = \langle p(t), p(t)\rangle - \langle p^0, p^0\rangle = \|p(t)\|^2_E - \left\|p^0\right\|^2_E = 0\right),$$

stąd

$$\forall t \geqslant 0\ \left(\sum_{i=1}^n p_i^2(t) = r^2\right), \text{ gdzie } r = \left\|p^0\right\|_E.$$

■

3. RÓWNOWAGA

Δ **Definicja 2.** Mówimy, że system Walrasa z zapasami jest w równowadze (długookresowej), jeżeli ustalą się takie ceny $p(t) \equiv \bar{p}$ oraz takie zapasy $z(t) \equiv \bar{z}$, przy których

$$f^d(\bar{p}) = f^s(\bar{p}). \tag{6}$$

▲

W równowadze spełnione są jednocześnie dwa układy równań:

$$f^d(\bar{p}) - f^s(\bar{p}) = 0 \tag{7a}$$

oraz

$$f^s(\bar{p}) - f^d(\bar{p}) - \mu\bar{z} = 0, \tag{7b}$$

a stąd wynika, że $\bar{z} = 0$. System Walrasa (3) – (4) w równowadze pozbywa się zapasów. Łatwo zauważyć też, że ceny $\bar{p}$ w równowadze są dodatnie. Rzeczywiście, zakładając $\bar{p}_i = 0$ dla pewnego i, z warunku (II) otrzymujemy $f_i(\bar{p}) = f_i^d(\bar{p}) - f_i^s(\bar{p}) > 0$, co przeczy (7a). Z dodatniej jednorodności stopnia 0 funkcji popytu $f^d(p)$ i podaży $f^s(p)$ wnioskujemy, że jeżeli $\bar{p} > 0$ jest wektorem cen równowagi, to jest nim także jego dowolna dodatnia wielokrotność.

Półprostą

$$P = \{\lambda\bar{p} \mid \lambda > 0\}$$

nazywamy tradycyjnie promieniem cen równowagi.

Pokażemy, że w systemie Walrasa z zapasami ceny równowagi istnieją i są określone jednoznacznie z dokładnością do struktury.

□ **Twierdzenie 2.** Przy założeniach (I) – (V) istnieje dokładnie jeden wektor cen równowagi rynkowej (spełniający warunek (6)) określony z dokładnością do struktury.

Dowód. Utwórzmy funkcję

$$w = (w_1, ..., w_n) : P_+^n(1) \to R^n,$$

$$w_i(p) = \max\{p_i + f_i(p), \frac{1}{2}\, p_i\}, \quad i = 1, ..., n. \tag{$*$}$$

Wówczas $w_i \in C^0(P_+^n(1) \to R^n)$ oraz

$$\forall p \in P_+^n(1)\ \ (w(p) > 0)\,.$$

Rzeczywiście, jeżeli $p_i = 0$, to zgodnie z warunkiem (II) $f_i(p) > 0$, czyli $w_i(p) \geqslant p_i + f_i(p) = f_i(p) > 0$. Jeżeli zaś $p_i > 0$, to $w_i(p) \geqslant \frac{1}{2}p_i > 0$.
Niech

$$\phi(p) = \frac{w(p)}{\|w(p)\|},$$

gdzie $\|x\| = \sum_i |x_i|$ (wszędzie dalej tak samo).

Wówczas $\phi \in C^0(P_+^n(1) \to R_+^n)$ oraz

$$\forall p \in P_+^n(1)\,(\phi(p) > 0\ \ \&\ \ \|\phi(p)\| = \left\|\frac{w(p)}{\|w(p)\|}\right\| = 1)\,,$$

zatem ϕ jest ciągłym odwzorowaniem simpleksu $P^n_+(1)$ w siebie. Z twierdzenia Brouwera (zob. np. [3], rozdz.1, §4) płynie wówczas wniosek, że $\exists \bar{p} \in P^n_+(1)(\phi(\bar{p}) = \bar{p} > 0)$. Zgodnie z definicją odwzorowania $w(p)$,

$$\forall p \in P^n_+(1)\left(w(p) \geqslant \frac{1}{2}p\right),$$

czyli w szczególności $w(\bar{p}) \geqslant \frac{1}{2}\bar{p}$, tzn. $\|w(\bar{p})\| \geqslant \frac{1}{2}\|\bar{p}\| = \frac{1}{2}$. Pokażemy, że $\|w(\bar{p})\| > \frac{1}{2}$. W tym celu załóżmy $\|w(\bar{p})\| = \frac{1}{2}$. Wtedy $w(\bar{p}) = \frac{1}{2}\bar{p}$, tzn. zgodnie z (*) $\frac{1}{2}\bar{p} = w(\bar{p}) \geqslant \bar{p} + f(\bar{p})$, czyli $f(\bar{p}) \leqslant -\frac{1}{2}\bar{p} < 0$. Oznacza to, że $\langle \bar{p}, f(\bar{p})\rangle < 0$, co jest sprzeczne z prawem Walrasa (IV). Zatem

$$\|w(\bar{p})\| > \frac{1}{2}, \text{ tzn. } \bar{p} = \phi(\bar{p}) = \frac{w(\bar{p})}{\|w(\bar{p})\|} < \frac{w(\bar{p})}{\frac{1}{2}}, \qquad (**)$$

czyli $w(\bar{p}) > \frac{1}{2}\bar{p}$ i wobec (*)

$$w(\bar{p}) = \bar{p} + f(\bar{p}),$$

co zgodnie z (**) prowadzi do warunku:

$$\bar{p} = \frac{\bar{p} + f(\bar{p})}{\|w(\bar{p})\|}. \qquad (***)$$

Mnożąc obustronnie (***) przez wektor $f(\bar{p})$ otrzymujemy znowu (zgodnie z prawem Walrasa):

$$\langle \bar{p}, f(\bar{p})\rangle = \frac{\langle \bar{p}, f(\bar{p})\rangle + \langle f(\bar{p}), f(\bar{p})\rangle}{\|w(\bar{p})\|} = 0,$$

czyli $\langle f(\bar{p}), f(\bar{p})\rangle = 0$, co oznacza że

$$f(\bar{p}) = f^d(\bar{p}) - f^s(\bar{p}) = 0.$$

Obie funkcje $f^d(p)$, $f^s(p)$ są dodatnio jednorodne stopnia zero (warunek (III)), więc również

$$\forall \lambda > 0\left(f(\lambda\bar{p}) = f^d(\lambda\bar{p}) - f^s(\lambda\bar{p}) = f^d(\bar{p}) - f^s(\bar{p}) = 0\right).$$

Załóżmy, że istnieją dwa wektory cen równowagi $\bar{p}^1, \bar{p}^2$ o różnej strukturze:
$\tilde{p}^1 = \frac{\bar{p}^1}{\|\bar{p}^1\|} \neq \tilde{p}^2 = \frac{\bar{p}^2}{\|\bar{p}^2\|}$ i przyjmijmy oznaczenie: $p(\tau) = \tilde{p}^1 + \tau(\tilde{p}^2 - \tilde{p}^1), \quad \tau \in [0, 1]$.
Wówczas

$$p(0) = \tilde{p}^1, \quad p(1) = \tilde{p}^2 \text{ oraz } \forall \tau \in [0, 1] \quad (p(\tau) \in P^n_+(1)).$$

Niech $\phi(\tau) = \langle \tilde{p}^2 - \tilde{p}^1, f(p,\tau)) - f(\tilde{p}^1)\rangle$. Wówczas $\phi(0) = 0$, $\phi \in C^1\left([0,1] \to R^1\right)$ oraz

$$\dot{\phi}(\tau) = \lambda J(p(\tau))\lambda^T$$

gdzie $\lambda = \tilde{p}^2 - \tilde{p}^1 \neq 0$, tzn. $\dot{\phi}(\tau) < 0$ na $[0,1]$ (zgodnie z warunkiem (V)), zatem

$$\phi(1) = \langle \tilde{p}^2 - \tilde{p}^1, f(\tilde{p}^2) - f(\tilde{p}^1)\rangle < 0,$$

co jest niemożliwe, gdyż $f(\tilde{p}^1) = f(\tilde{p}^2) = 0$.

Otrzymana sprzeczność kończy dowód całego twierdzenia.

■

4. STABILNOŚĆ

W standardowej wersji modelu Walrasa stabilność gospodarki oznacza zbieżność jej dowolnej dopuszczalnej trajektorii cen do pewnego wektora cen równowagi z promienia P. Teraz musimy uwzględnić zbieżność zarówno cen, jak i zapasów do ich poziomu w równowadze. W przypadku cen będzie to pewien wektor $\bar{p} \in P$, a w przypadku zapasów wektor $\bar{z} = 0$.

Δ **Definicja 3.** System Walrasa z zapasami nazywamy (globalnie) asymptotycznie stabilnym, jeżeli $\forall p^0 > 0 \forall z^0 \underset{=}{\geq} 0$ każda (p^0, z^0, ∞) – dopuszczalna trajektoria cen i zapasów spełnia warunek:

$$(i) \exists\, \bar{p} \in P\left(\lim_t p(t) = \bar{p}\right),$$

$$(2i) \lim_t z(t) = 0.$$

▲

□ **Twierdzenie 3.** Przy założeniach (I) – (V) system Walrasa z zapasami jest globalnie asymptotycznie stabilny.

Dowód. Pokażemy najpierw, że

$$\forall p^1 \in P \quad \forall p^2 \notin P\,((\,\langle p^1, f(p^2)\rangle > 0\,))\,.$$

Dowód wzorowany jest na końcowym fragmencie dowodu twierdzenia 2. Weźmy dowolną parę wektorów cen $p^1 \in P$ i $p^2 \notin P$ oraz wprowadźmy oznaczenia:

$$\tilde{p}^1 = \frac{p^1}{\left\|p^1\right\|}\,,\ \tilde{p}^2 = \frac{p^2}{\left\|p^2\right\|}, \quad p(\tau) = \tilde{p}^1 + \tau(\tilde{p}^2 - \tilde{p}^1), \quad \tau \in [0,1].$$

Mamy wówczas $p(0) = \tilde{p}^1$, $p(1) = \tilde{p}^2$. Niech

$$\phi(\tau) = \langle \tilde{p}^2 - \tilde{p}^1, f(p(\tau)) - f(\tilde{p}^1)\rangle.$$

Wówczas $\phi(0) = 0$, $\phi \in C^1([0,1] \to R^1)$ oraz zgodnie z warunkiem (V)

$$\forall \tau \in [0,1]\left(\frac{d\phi(\tau)}{d\tau} = \lambda J(p(\tau))\lambda^T < 0\right),$$

gdzie $\lambda = \tilde{p}^2 - \tilde{p}^1 \neq 0$. Zatem $\phi(1) = \langle \tilde{p}^2 - \tilde{p}^1, f(\tilde{p}^2) - f(\tilde{p}^1)\rangle < 0$.

Ponieważ $\tilde{p}^1 \in P$, więc $f(\tilde{p}^1) = 0$, czyli

$$\phi(1) = \langle \tilde{p}^2, f(\tilde{p}^2)\rangle - \langle \tilde{p}^1, f(\tilde{p}^2)\rangle < 0.$$

W myśl prawa Walrasa (warunek IV) $\langle \tilde{p}^2, f(\tilde{p}^2)\rangle = 0$, zatem $\langle \tilde{p}^1, f(\tilde{p}^2)\rangle > 0$, a stąd (wobec dodatniej jednorodności stopnia 0 funkcji popytu nadwyżkowego):

$$\langle p^1, f(p^2)\rangle > 0. \qquad (*)$$

Weźmy teraz dowolną (p^0, z^0, ∞) – dopuszczalną trajektorię cen $p(t)$ i zapasów $z(t)$ (rozwiązanie układu (3)-(4) z warunkiem początkowym (5)). Wówczas zgodnie z twierdzeniem 1 (2i) $\forall t \geqslant 0\,(p(t) \in K_r^n(0))$, gdzie $r = \left\|p^0\right\|_E$. Niech $\bar{p}$ będzie punktem, w którym promień cen równowagi P przebija kulę $K_r^n(0)$.

Jeżeli $p^0 = \bar{p}$, to $\forall t \geqslant 0\,(p(t) = \bar{p})$ oraz

$$\dot{z}_i(t) = \begin{cases} -\mu z_i(t) & , \quad gdy \quad z_i^0(t) > 0, \\ 0 & , \quad gdy \quad z_i^0(t) = 0, \end{cases}$$

tzn. $\lim\limits_t z(t) = 0$.

Załóżmy teraz, że $p^0 \neq \bar{p}$. Wówczas $\forall t \geqslant 0\,(p(t) \notin P)$. Utwórzmy funkcję

$$V(t) = \frac{1}{2}\langle p(t) - \bar{p}, p(t) - \bar{p}\rangle \geqslant 0.$$

Oczywiście $V \in C^1[0, +\infty)$ oraz $\forall t \geqslant 0$:

$$\dot{V}(t) = \langle p(t) - \bar{p}, \dot{p}(t)\rangle = \sigma\langle p(t), f^d(p(t)) - f^s(p(t))\rangle -$$

$$-\sigma\langle \bar{p}, f^d(p(t)) - f^s(p(t))\rangle = -\sigma\langle \bar{p}, f(p(t))\rangle < 0 \qquad (**)$$

(zgodnie z (*)). Aby pokazać, że $p(t) \underset{t}{\to} \bar{p}$ wystarczy udowodnić, że $V(t) \underset{t}{\to} 0$. Ponieważ $\dot{V}(t) < 0$ oraz $V(t) \geqslant 0$ na półosi $[0, +\infty)$, zatem $\lim\limits_t V(t) = \bar{V} \geqslant 0$. Załóżmy, że $\bar{V} > 0$. Wówczas $\exists \delta > 0 \forall t \geqslant 0\,(\|p(t) - \bar{p}\|_E \geqslant \delta)$.

Niech $F(p(t)) = -\sigma\langle \bar{p}, f(p(t))\rangle$ oraz $\Omega = \{p \in K_r^n(0) \mid \|p - \bar{p}\|_E \geqslant \delta\}$.

Zbiór Ω jest zwarty i zgodnie z (**)

$$\forall t \geqslant 0\left(\dot{V}(t) = F(p(t)) < 0 \;\;\&\;\; F(p) \in C^0(\Omega \to R^1)\right),$$

czyli istnieje

$$\max_{p\in\Omega} F(p) = -\varepsilon < 0.$$

Wówczas

$$\forall t \geqslant 0 \quad \left(\dot{V}(t) \leqslant -\varepsilon < 0\right),$$

tzn.

$$0 \leqslant V(t) \leqslant -\varepsilon t \underset{t}{\to} -\infty.$$

Otrzymana sprzeczność dowodzi, że $\lim V(t) = 0$, czyli $\lim_t p(t) = \bar{p}$.

Rozwiązaniem równania (4) jest funkcja zapasów i – tego towaru

$$z_i(t) = \max\{z_i^0 e^{-\mu t} + e^{-\mu t} \int_0^t e^{\mu\theta} f_i(\theta) d\theta,\ 0\}, \tag{8}$$

gdzie

$$z_i^0 e^{-\mu t} + e^{-\mu t} \int_0^t e^{\mu\theta} f_i(\theta) d\theta = e^{-\mu t}\left[z_i^0 - \sigma^{-1} \int_0^t e^{\mu\theta} \dot{p}_i(\theta)\, d\theta\right] =$$

$$= e^{-\mu t}\left[z_i^0 - (1-\mu)\sigma^{-1}\ p_i^0\right] - \sigma^{-1}(1-\mu)p_i(t) \underset{t}{\to}\ -\sigma^{-1}(1-\mu)\bar{p}_i < 0.$$

Zatem zgodnie z (4)

$$\forall i\ (\lim_t z_i(t) = 0),$$

co zamyka dowód twierdzenia.

■

Łatwo pokazać, że w szczególnym przypadku, gdy stopa ubytku zapasów $\mu = 0$, a początkowy zapas towarów $z^0 > 0$, wtedy dodatnim cenom $\bar{p}$ odpowiada w równowadze dodatni wektor zapasów $\bar{z}$. Również taki stan równowagi jest globalnie asymptotycznie stabilny.

Uniwersytet Ekonomiczny w Poznaniu
Wydział Informatyki i Gospodarki Elektronicznej
Katedra Ekonomii Matematycznej

LITERATURA

[1] McKenzie L.W. (2002), Classical General Equilibrium Theory, The MIT Press, Cambridge.
[2] Morishima M. (1996), Dynamic economic theory, Cambridge University Press.
[3] Nikaido H., (1968), Convex Structures and Economic Theory, Acad. Press, New York and London.
[4] Panek E., (2003), Ekonomia matematyczna, Wydawnictwo AEP, Poznań.
[5] Takayama A., (1997), Mathematical Economics, Cambridge University Press.

SYSTEM WALRASA I ZAPASY

Streszczenie

W literaturze z zakresu równowagi ogólnej L. Walrasa standardowo zakłada się, że o dynamice cen towarów w gospodarce konkurencyjnej decydują relacje między popytem i podażą, co jest oczywiście prawdą. Prawdą jest jednak również to, że różnice między popytem i podażą prowadzą do powstawania zapasów towarów, czego w standardowych modelach równowagi ogólnej najczęściej już się nie uwzględnia.

Artykuł jest skromną próbą wypełnienia tej luki. Przedstawiamy w nim nawiązujący do Walrasa model kształtowania cen, uzupełniony o układ równań dynamiki zapasów w gospodarce konkurencyjnej. Dowodzimy istnienia stanu równowagi konkurencyjnej w takiej gospodarce oraz jego globalnej stabilności.

Słowa kluczowe: system Walrasa, równowaga konkurencyjna, stabilność

WALRASIAN SYSTEM AND STOCKS

Abstract

In the literature concerning Walrasian general equilibrium it is normally assumed that in the competitive economy the decisive factor influencing the price dynamics is the relation between demand and supply, what is actually true. However it is also true that the differences between demand and supply lead to the creation of stocks of goods, what in standard models of the general equilibrium is usually not taken into consideration.

This article is a modest attempt to fill this gap. We present in it a model of price adjustment, which refers to Walras and is complemented by a system of equations of stocks dynamics in the competitive economy. We prove the existence of competitive equilibrium in this economy and its global stability.

Keywords: Walras System, competitive equilibrium, stability

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

JACEK KWIATKOWSKI

WYBRANE MODELE ZAWIERAJĄCE STOCHASTYCZNY PIERWIASTEK JEDNOSTKOWY W ANALIZIE KURSÓW WALUTOWYCH[1]

1. WSTĘP

Granger i Swanson [5] wykazali, że finansowe i makroekonomiczne procesy, które wymagają obliczenia pierwszych różnic nie zawsze są procesami zintegrowanymi rzędu pierwszego. Dowodzą oni, że mają one często pierwiastek jednostkowy, który nie jest stały lecz losowy. Zmienny parametr, odpowiedzialny za stopień zintegrowania traktowany jest tu jako odrębny proces stochastyczny, którego realizacje oscylują wokół jedynki. Tego typu proces określa się jako proces zawierający stochastyczny pierwiastek jednostkowy (ang. *Stochastic Unit Root, STUR*).

Celem artykułu jest przedstawienie wybranych modeli zawierających stochastyczny pierwiastek jednostkowy oraz opis badań empirycznych dokonanych na podstawie dziennych realizacji kursu złotego w stosunku do najważniejszych walut europejskich. Prezentowany artykuł bezpośrednio nawiązuje do publikacji dotyczących modeli STUR (zob. [5], [8], [10], [15], [16] i [17]), a także jest ich rozszerzeniem. Po pierwsze, zbadano nieznaną, jak dotąd, moc wyjaśniającą niektórych specyfikacji STUR w stosunku do standardowych modeli zmienności tj. SV (ang. *Stochastic Volatility*) i GARCH (ang. *Generalized Autoregressive Conditional Heteroskedasticity*). Po drugie, w modelach STUR przyjęto założenie, że losowy w czasie pierwiastek jest procesem stochastycznej zmienności. Jest to nowe podejście, które nie było dotąd jeszcze przedmiotem rozważań w literaturze przedmiotu. Po trzecie, opracowano metody numeryczne wykorzystywane w schemacie wnioskowania bayesowskiego w kontekście rozważanych tu specyfikacji.

Układ artykułu jest następujący. Część druga przedstawia model STUR w wersji Grangera i Swansona [5]. Część trzecia dotyczy modelu STUR w wersji Leybourne'a ([10] i [11]), natomiast część czwarta opisuje model dwuliniowy, zaproponowany przez Francqa, Makarovą i Zakoïana [3]. Ostatnia, piąta część artykułu zawiera wyniki badań empirycznych. Na końcu pracy zamieszczono najważniejsze wnioski.

[1] Praca naukowa finansowana ze środków na naukę w latach 2009-2011 jako projekt badawczy numer N N111 431737. Autor dziękuje anonimowemu Recenzentowi za bardzo wnikliwe uwagi i sugestie. Adres elektroniczny: *jkwiat@umk.pl*

2. MODEL STUR W WERSJI GRANGERA I SWANSONA

Oznaczmy przez $\{y_t,\ t = 1,\ ...,\ T\}$ proces stochastyczny. Granger i Swanson [5] rozważali następujący proces STUR:

$$y_t = a_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N\left(0,\ \sigma^2\right), \tag{1}$$

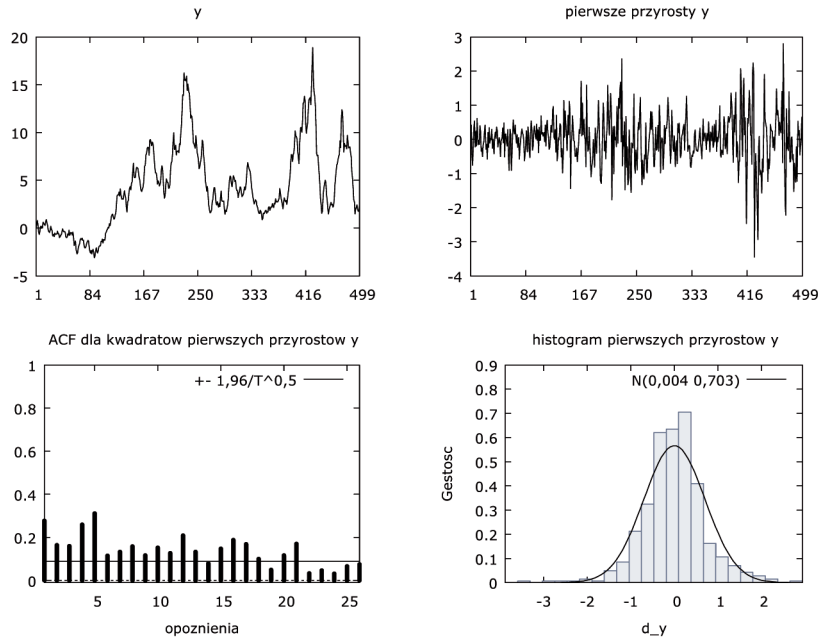
gdzie:

$$a_t = \exp\left(\delta_t\right), \quad E\left(a_t\right) = 1. \tag{2}$$

Przez $\{\delta_t,\ t = 1,\ ...,\ T\}$ oznaczono stacjonarny, nieobserwowalny proces autoregresyjny rzędu pierwszego:

$$\delta_t = \phi_0 + \phi_1 \delta_{t-1} + \eta_t, \quad \eta_t \overset{iid}{\sim} N\left(0,\ \omega^2\right), \tag{3}$$

natomiast przez $\varepsilon_t \overset{iid}{\sim} N\left(0,\ \sigma^2\right) \quad t = 1,\ ...,\ T$ oznaczono niezależne zmienne losowe o rozkładzie normalnym z zerową średnią i wariancją σ^2 (z ang. *independent and identically distributed; iid*). Dodatkowo zakłada się, że składniki losowe ε_t i η_t są względem siebie niezależne tj. $\varepsilon_t \perp \eta_s$ dla $t,\ s = 1,\ ...,\ T$.



Rysunek 1. Charakterystyki procesu STUR, w którym $\phi_0 = -0.01,\ \phi_1 = 0,4,\ \sigma^2 = 0,1$ i $\omega^2 = 0,01$. Lewy, górny panel przedstawia jego realizację, kolejny u góry pokazuje pierwsze przyrosty, dolny, lewy zawiera informacje na temat funkcji autokorelacji kwadratów pierwszych przyrostów, tj. $(\Delta y_t)^2$, natomiast prawy, dolny panel przedstawia histogram pierwszych przyrostów oraz funkcję gęstości rozkładu normalnego

Źródło: obliczenia własne.

Gdy $\delta_t = 0$ dla $t = 1, ..., T$, to model (1)-(3) redukuje się do dobrze znanego w literaturze przedmiotu modelu błądzenia losowego tj. $y_t = y_{t-1} + \varepsilon_t$, którego pierwsze przyrosty tworzą proces zintegrowany w stopniu zerowym.

Niektóre, nieznane dotąd własności procesu STUR przedstawił kilka lat temu Yoon [21]. Okazuje się, że proces STUR jest procesem ściśle stacjonarnym, natomiast nie jest kowariancyjnie stacjonarny. Dodatkowo nie ma skończonych momentów, w tym nawet średniej oraz ma bardzo grube ogony, co wskazuje na znaczne odstępstwo jego rozkładu od rozkładu normalnego. Również w odróżnieniu od standardowych procesów zawierających pierwiastek jednostkowy, proces STUR nie jest kowariancyjnie stacjonarny dla dowolnej różnicy tj. $\left\{\Delta^d y_t\right\}$, gdzie $d = 1, 2, ...$

Na rysunku 1 przedstawiono przykładową realizację procesu STUR, w którym przyjęto następujące założenia odnośnie parametrów modelu: $\phi_0 = -0,01$, $\phi_1 = 0,4$, $\sigma^2 = 0,1$, $\omega^2 = 0,01$ oraz $y_0 = 0$ i $\delta_0 = 0$. Oprócz jego realizacji, zamieszczono również pierwsze różnice, obliczone zgodnie ze wzorem $\Delta y_t = y_t - y_{t-1}$, wartości funkcji autokorelacji dla $(\Delta y_t)^2$ wraz z górną granicą 95% przedziału ufności, a także histogram pierwszych przyrostów i funkcję gęstości rozkładu normalnego.

Można łatwo zauważyć, że procesy STUR mogą generować szeregi czasowe, których własności zbliżone są do typowych własności szeregów finansowych. W tym przypadku, przebieg pierwszych przyrostów wskazuje wyraźnie na efekt skupiania zmienności (ang. *volatility clustering*), który polega na tym, że zarówno małe, jak i duże zmiany cen instrumentu bazowego następują seriami. Bardzo ciekawie przedstawia się również funkcja autokorelacji kwadratów pierwszych przyrostów, która wykazuje się istotną zależnością aż do 20-ego opóźnienia włącznie. Również histogram, zamieszczony w prawym, dolnym panelu potwierdza większą koncentrację i grubsze ogony niż w rozkładzie normalnym. Wszystkie wymienione cechy są szeroko znane i omawiane w literaturze przedmiotu jako typowe własności szeregów finansowych (zob.[2]).

Rozszerzmy model STUR w wersji Grangera i Swansona, tak aby składniki losowe η_t (dla $t = 1, ..., T$) tworzyły proces o strukturze SV. Oznacza to, że $\{\delta_t, t = 1, ..., T\}$ jest procesem autoregresyjnym rzędu pierwszego:

$$\delta_t = \phi_0 + \phi_1 \delta_{t-1} + \eta_t, \quad \eta_t \sim N\left(0, \omega_t^2\right), \tag{4}$$

w którym warunkową wariancję ω_t^2 można opisać jako:

$$\omega_t^2 = \omega^2 \exp\left(h_{sygnał,t}^{SV}\right),$$

gdzie:

$h_{sygnał,t}^{SV} = \rho_{sygnał} h_{sygnał,t-1}^{SV} + \zeta_{sygnał,t}$ oraz $\rho_{sygnał} \in (-1, 1)$, a także $\zeta_{sygnał,t} \overset{iid}{\sim} N\left(0, \gamma_{sygnał}^2\right)$, $t = 1, ..., T$.

Model o równaniach (1)-(2) i (4) jest zatem uogólnieniem podstawowego modelu STUR w wersji Grangera i Swansona [5], ponieważ dopuszcza możliwość zmiennej warunkowej wariancji w procesie pierwiastka jednostkowego. Pozostałe założenia

są identyczne jak w standardowym modelu STUR. Model ten oznaczono symbolem STUR-G-SV.

Niech $\theta = \left(\phi_0,\, \phi_1,\, \sigma^2,\, \omega^2,\, \rho_{sygnał},\, \gamma^2_{sygnał}\right)'$ będzie wektorem nieznanych parametrów w modelu STUR-G-SV (równania (1)-(2) i (4)). Dodatkowo przez $\delta = (\delta_1,\, ...,\, \delta_T)'$ i $h^{SV}_{sygnał} = \left(h^{SV}_{sygnał,\,1},\, ...,\, h^{SV}_{sygnał,\,T}\right)'$ oznaczmy wektory zmiennych ukrytych.

Przy założeniu normalności składników losowych, gęstość wektora obserwacji w modelu STUR-G-SV jest następująca:

$$p\left(y|\, \delta,\, \delta_0,\, h^{SV}_{sygnał},\, h^{SV}_{sygnał,\,0},\, \theta,\, y_0\right) = \prod_{t=1}^{T} f_N\left(y_t|\, e^{\delta_t} y_{t-1},\, \sigma^2\right), \tag{5}$$

gdzie:

$y = (y_1,\, ...,\, y_T)'$ jest wektorem obserwacji, δ_0, $h^{SV}_{sygnał,\,0}$ i y_0 są wartościami początkowymi, natomiast symbol $f_N(\cdot|\, c,\, d)$ oznacza gęstość rozkładu normalnego o średniej c i wariancji d.

Jako rozkłady *a priori* parametrów θ można przyjąć następujące rozkłady właściwe:

$$p(\theta) = p(\phi_0)\, p(\phi_1)\, p\left(\sigma^2\right) p\left(\omega^2\right) p\left(\rho_{sygnał}\right) p\left(\gamma^2_{sygnał}\right), \tag{6}$$

gdzie:

$$p(\phi_0) = f_N\left(\phi_0|\, 0,\, \sigma^2_{\phi_0}\right),\ p(\phi_1) \propto f_N\left(\phi_1|\, \mu_{\phi_1},\, \sigma^2_{\phi_1}\right) I_{(-1,\,1)}(\phi_1),\ p\left(\sigma^2\right) = f_{IG}\left(\sigma^2|\, c_\sigma,\, d_\sigma\right),$$

$$p\left(\omega^2\right) = f_{IG}\left(\omega^2|\, c_\omega,\, d_\omega\right), p\left(\rho_{sygnał}\right) \propto f_N\left(\rho_{sygnał}|\, \mu_\rho,\, \sigma^2_\rho\right) I_{(-1,\,1)}\left(\rho_{sygnał}\right),$$

$$p\left(\gamma^2_{sygnał}\right) = f_{IG}\left(\gamma^2_{sygnał}|\, c_\gamma,\, d_\gamma\right).$$

Symbol $I_{(-1,\,1)}(\cdot)$ oznacza funkcje wskaźnikową przedziału $(-1,\, 1)$, natomiast $f_{IG}(\cdot|c,\, d)$ jest gęstością odwróconego rozkładu gamma o średniej $d/(c-1)$ dla $c > 1$ i wariancji $d^2/\left[(c-1)^2 (c-2)\right]$ dla $c > 2$, w postaci:

$$f_{IG}(x|c,\, d) = \frac{d^c}{\Gamma(c)} x^{-(c+1)} \exp\left(-\frac{d}{x}\right), \quad x > 0, \quad c > 0, \quad d > 0. \tag{7}$$

Warunkowe gęstości rozkładów *a priori* zmiennych ukrytych są iloczynami gęstości rozkładów normalnych:

$$p\left(\delta|\, \delta_0,\, h^{SV}_{sygnał},\, h^{SV}_{sygnał,\,0},\, \theta\right) = \prod_{t=1}^{T} f_N\left(\delta_t|\, \phi_0 + \phi_1 \delta_{t-1},\, \omega^2_t\right), \tag{8}$$

$$p\left(h^{SV}_{sygnał}|\, h^{SV}_{sygnał,\,0},\, \theta\right) = \prod_{t=1}^{T} f_N\left(h^{SV}_{sygnał,\,t}|\, \rho_{sygnał} h^{SV}_{sygnał,\,t-1},\, \gamma^2_{sygnał}\right). \tag{9}$$

Poniżej przedstawiono układ warunkowych rozkładów *a posteriori* stosowanych w schemacie Gibbsa. Parametry σ^2, ω^2i $\gamma^2_{sygnał}$ mają następujące warunkowe rozkłady *a posteriori*:

$$p\left(\sigma^2|\, y,\, y_0,\, \delta\right) = f_{IG}\left(\sigma^2|\, \frac{T+2c_\sigma}{2},\, \frac{\sum_{t=1}^{T}\left(y_t - e^{\delta_t}y_{t-1}\right)^2 + 2d_\sigma}{2}\right), \tag{10}$$

$$p\left(\omega^2|\, y,\, \delta,\, \delta_0,\, h^{SV}_{sygnał},\, \phi_0,\, \phi_1\right) = f_{IG}\left(\omega^2|\, \frac{T+2c_\omega}{2},\, \sum_{t=1}^{T}\frac{(\delta_t - \phi_0 - \phi_1\delta_{t-1})^2}{2\exp\left(h^{SV}_{sygnał,\,t}\right)} + d_\omega\right), \tag{11}$$

$$p\left(\gamma^2_{sygnał}|y,\ h^{SV}_{sygnał},\, h^{SV}_{sygnał,\,0},\, \rho_{sygnał}\right) =$$

$$f_{IG}\left(\gamma^2_{sygnał}|\, \frac{T+2c_\gamma}{2},\, \frac{\sum_{t=1}^{T}\left(h^{SV}_{sygnał,\,t} - \rho_{sygnał}h^{SV}_{sygnał,\,t-1}\right)^2 + 2d_\gamma}{2}\right). \tag{12}$$

Wyraz wolny ϕ_0 ma warunkowy rozkład normalny, którego gęstość można przedstawić następująco:

$$p\left(\phi_0|\, y,\, \delta,\, \delta_0,\, h^{SV}_{sygnał},\, \phi_1,\, \omega^2\right) = f_N\left(g_{\phi_0},\, Var_{\phi_0}\right) \tag{13}$$

o średniej i wariancji równej odpowiednio:

$$g_{\phi_0} = \frac{\sigma^2_{\phi_0}\sum_{t=1}^{T}\frac{(\delta_t - \phi_1\delta_{t-1})}{\exp\left(h^{SV}_{sygnał,\,t}\right)}}{\sigma^2_{\phi_0}\sum_{t=1}^{T}\frac{1}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \omega^2} \text{ i } Var_{\phi_0} = \frac{\omega^2\sigma^2_{\phi_0}}{\sigma^2_{\phi_0}\sum_{t=1}^{T}\frac{1}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \omega^2}.$$

Również współczynnik autoregresji ma warunkowy rozkład *a posteriori*, którego gęstość jest proporcjonalna do gęstości rozkładu normalnego:

$$p\left(\phi_1|\, y,\, \delta,\, \delta_0,\, \phi_0,\, \omega^2,\, h^{SV}_{sygnał}\right) \propto f_N\left(\phi_1|\, g_{\phi_1},\, Var_{\phi_1}\right) \tag{14}$$

o średniej i wariancji równej odpowiednio:

$$g_{\phi_1} = \frac{\sigma^2_{\phi_1}\sum_{t=1}^{T}\frac{\delta_{t-1}(\delta_t - \phi_0)}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \mu_{\phi_1}\omega^2}{\sigma^2_{\phi_1}\sum_{t=1}^{T}\frac{\delta^2_{t-1}}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \omega^2} \text{ i } Var_{\phi_1} = \frac{\omega^2\sigma^2_{\phi_1}}{\sigma^2_{\phi_1}\sum_{t=1}^{T}\frac{\delta^2_{t-1}}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \omega^2}.$$

Gęstość warunkowego rozkładu *a posteriori* dla współczynnika autokorelacji $\rho_{sygnał}$ można wyrazić jako:

$$p\left(\rho_{sygnał}|\, y,\, h^{SV}_{sygnał},\, h^{SV}_{sygnał,\,0},\, \gamma^2_{sygnał}\right) \propto f_N\left(g_\rho,\, Var_\rho\right) \tag{15}$$

o średniej i wariancji równej odpowiednio:

$$g_\rho = \frac{\sigma_\rho^2 \sum_{t=1}^{T} h^{SV}_{sygnał,\,t-1} h^{SV}_{sygnał,\,t} + \mu_\rho \gamma^2_{sygnał}}{\sigma_\rho^2 \sum_{t=1}^{T} h^{2,\,SV}_{sygnał,\,t-1} + \gamma^2_{sygnał}} \text{ i } Var_\rho = \frac{\sigma_\rho^2 \gamma^2_{sygnał}}{\sigma_\rho^2 \sum_{t=1}^{T} h^{2,\,SV}_{sygnał,\,t-1} + \gamma^2_{sygnał}}.$$

Warunkowy rozkład *a posteriori* nieobserwowalnej zmiennej $\delta_t \quad (t = 1, ..., T)$ nie należy do znanej rodziny rozkładów, a jego gęstość wyraża się wzorem:

$$p\left(\delta_t | y, \delta_{t-1}, \delta_0, h^{SV}_{sygnał,\,t}, \theta\right) \propto \exp\left\{-\frac{y_{t-1}^2}{2\sigma^2}\left(e^{\delta_t} - \frac{y_t}{y_{t-1}}\right)^2 - \frac{\vartheta(\delta_t)}{2\omega_t^2}\left(\delta_t - \frac{\tau(\delta_t)}{\vartheta(\delta_t)}\right)^2\right\}. \quad (16)$$

Wartości $\vartheta(\delta_t)$ i $\tau(\delta_t)$ przedstawia tablica 1.

Tablica 1.

Wartości $\vartheta(\delta_t)$ i $\tau(\delta_t)$

t	$\vartheta(\delta_t)$	$\tau(\delta_t)$
$t = 1, ..., T-1$	$1 + \phi_1^2$	$\phi_1(\delta_{t-1} + \delta_{t+1}) + \phi_0(1 - \phi_1)$
$t = T$	1	$\phi_1 \delta_{t-1} + \phi_0$

Źródło: Jones i Marriott [8]

Z kolei gęstość warunkowego rozkładu *a posteriori* zmiennych ukrytych $h^{SV}_{sygnał,\,t}$ $(t = 1, ..., T)$ ma następującą postać:

$$p\left(h^{SV}_{sygnał,\,t} | y, \delta_0, \delta_t, \delta_{t-1}, h^{SV}_{sygnał,\,0}, h^{SV}_{sygnał,\,t-1}, h^{SV}_{sygnał,\,t+1}, \theta\right) \propto$$

$$\frac{1}{\exp\left(h^{SV}_{sygnał,\,t}/2\right)} \exp\left\{-\frac{(\delta_t - \phi_0 - \phi_1 \delta_{t-1})^2}{2\omega_t^2}\right\} \frac{1}{\exp\left(h^{SV}_{sygnał,\,t}\right)} \exp\left\{-\frac{1}{2Var_\gamma}\left(h^{SV}_{sygnał,\,t} - s_t\right)^2\right\}, \quad (17)$$

gdzie s_t i Var_γ wyrażają się wzorami:

$$s_t = \rho_{sygnał}\left(h^{SV}_{sygnał,\,t+1} + h^{SV}_{sygnał,\,t-1}\right)/\left(1 + \rho^2_{sygnał}\right), \quad Var_\gamma = \frac{\gamma^2_{sygnał}}{1 + \rho^2_{sygnał}}, \text{ dla } t = 1, ..., T-1 \quad (18)$$

oraz:

$$s_t = \rho_{sygnał} h^{SV}_{sygnał,\,t-1}, \quad Var_\gamma = \gamma^2_{sygnał}, \text{ dla } t = T, \text{ por. [18].} \quad (19)$$

3. MODEL STUR W WERSJI LEYBOURNE'A

Kolejną specyfikację STUR zaproponowali Leybourne, McCabe i Mills [10] (zob. również [11] i [19]). Omawiany tu model STUR jest szczególnym przypadkiem modelu autoregresyjnego z losowymi współczynnikami autoregresji (ang. *Random Coefficient*

Autoregressive; RCA), który przedstawiono i omówiono w [13], [20] oraz w [6]. W literaturze polskiej, w podejściu klasycznym (teorio-próbkowym), estymację modeli RCA przedstawia Górka [7].

Najprostszy model STUR w wersji Leybourne'a definiowany jest nastepujaco:

$$y_t = \delta_t y_{t-1} + \varepsilon_t, \quad \varepsilon_t \overset{iid}{\sim} N\left(0,\, \sigma^2\right), \tag{20}$$

$$\delta_t = 1 + \eta_t, \quad \eta_t \overset{iid}{\sim} N\left(0,\, \omega^2\right). \tag{21}$$

Również w tym przypadku składniki losowe ε_t i η_t są względem siebie niezależne tj. $\varepsilon_t \perp \eta_s$ dla $t, s = 1, ..., T$.

Dla $\omega^2 = 0$ proces (20)-(21) redukuje się do procesu błądzenia losowego. Natomiast jeżeli $\omega^2 > 0$ to mamy do czynienia z procesem STUR, którego średnia zawiera pierwiastek jednostkowy. Proces (20)-(21), podobnie jak proces STUR w wersji Grangera i Swansona, nie jest stacjonarny kowariancyjnie.

Rozważmy uogólnienie modelu STUR w wersji Leybourne'a tzn. takie, w którym składniki losowe są procesami zmienności stochastycznej:

$$\Delta y_t = \eta_t y_{t-1} + \varepsilon_t, \tag{22}$$

gdzie:

$\varepsilon_t \sim N\left(0,\, \sigma_t^2\right)$, $\sigma_t^2 = \sigma^2 \exp\left(h_{szum,t}^{SV}\right)$, $h_{szum,t}^{SV} = \rho_{szum} h_{szum,t-1}^{SV} + \zeta_{szum,t}$ oraz $\rho_{szum} \in (-1,\, 1)$, a także $\zeta_{szum,t} \sim N\left(0,\, \gamma_{szum}^2\right)$, $t = 1, ..., T$.

Zmienny parametr η_t można opisać jako:

$$\eta_t = \phi_1 \eta_{t-1} + \xi_t, \quad \eta_0 = 0, \quad \xi_t \sim N\left(0,\, \omega_t^2\right), \quad t = 1, ..., T, \tag{23}$$

gdzie:

$\phi_1 \in (-1,\, 1)$, $\omega_t^2 = \omega^2 \exp\left(h_{sygnał,t}^{SV}\right)$, $h_{sygnał,t}^{SV} = \rho_{sygnał} h_{sygnał,t-1}^{SV} + \zeta_{sygnał,t}$ oraz $\rho_{sygnał} \in (-1,\, 1)$, a także $\zeta_{sygnał,t} \sim N\left(0,\, \gamma_{sygnał}^2\right)$, $t = 1, 2, ..., T$ oraz $\varepsilon_t \perp \eta_s$ dla $t, s = 1, ..., T$.

Zmienny parametr η_t podlega tu stacjonarnemu procesowi autoregresyjnemu ze składnikiem losowym o strukturze SV. Model (22)-(23) oznaczono symbolem STUR-L-SV.

Podobnie jak we wcześniejszym przykładzie, jako gęstość rozkładu *a priori* wektora parametrów $\theta = \left(\sigma^2,\, \rho_{szum},\, \gamma_{szum}^2,\, \phi_1,\, \omega^2,\, \rho_{sygnał},\, \gamma_{sygnał}^2\right)'$ można przyjąć iloczyn gęstości jego składowych:

$$p(\theta) = p\left(\sigma^2\right) p\left(\rho_{szum}\right) p\left(\gamma_{szum}^2\right) p\left(\phi_1\right) p\left(\omega^2\right) p\left(\rho_{sygnał}\right) p\left(\gamma_{sygnał}^2\right), \tag{24}$$

gdzie:

$$p\left(\sigma^2\right) = f_{IG}\left(\sigma^2 | \, c_\sigma,\, d_\sigma\right), \quad p\left(\rho_{szum}\right) \propto f_N\left(\rho_{szum} | \, \mu_\rho,\, \sigma_\rho^2\right) I_{(-1,\,1)}\left(\rho_{szum}\right),$$

$$p\left(\gamma^2_{szum}\right) = f_{IG}\left(\gamma^2_{szum}|\, c_\gamma,\, d_\gamma\right), \quad p\left(\phi_1\right) \propto f_N\left(\phi_1|\, \mu_{\phi_1},\, \sigma^2_{\phi_1}\right) I_{(-1,\,1)}\left(\phi_1\right),$$

$$p\left(\omega^2\right) = f_{IG}\left(\omega^2|\, c_\omega,\, d_\omega\right), \quad p\left(\rho_{sygnał}\right) \propto f_N\left(\rho_{sygnał}|\, \mu_\rho,\, \sigma^2_\rho\right) I_{(-1,\,1)}\left(\rho_{sygnał}\right),$$

$$p\left(\gamma^2_{sygnał}\right) = f_{IG}\left(\gamma^2_{sygnał}|\, c_\gamma,\, d_\gamma\right).$$

Wariancje σ^2, ω^2 i γ^2_{szum} mają następujące warunkowe gęstości:

$$p\left(\sigma^2|\, y,\, y_0,\, \eta,\, h^{SV}_{szum}\right) = f_{IG}\left(\sigma^2|\, \frac{T + 2c_\sigma}{2},\, \sum_{t=1}^{T} \frac{(\Delta y_t - \eta_t y_{t-1})^2}{2\exp\left(h^{SV}_{szum,\,t}\right)} + d_\sigma\right), \tag{25}$$

$$p\left(\omega^2|\, y,\, \eta,\, \eta_0,\, \phi_1,\, h^{SV}_{sygnał}\right) = f_{IG}\left(\omega^2|\, \frac{T + 2c_\omega}{2},\, \sum_{t=1}^{T} \frac{(\eta_t - \phi_1\eta_{t-1})^2}{2\exp\left(h^{SV}_{sygnał,\,t}\right)} + d_\omega\right), \tag{26}$$

$$p\left(\gamma^2_{szum}|y,\ h^{SV}_{szum},\, h^{SV}_{szum,\,0},\, \rho_{szum}\right) = f_{IG}\left(\gamma^2_{szum}|\, \frac{T + 2c_\gamma}{2},\, \frac{\sum_{t=1}^{T}\left(h^{SV}_{szum,\,t} - \rho_{szum} h^{SV}_{szum,\,t-1}\right)^2 + 2d_\gamma}{2}\right), \tag{27}$$

gdzie $\eta = (\eta_1,\, \eta_2,\, ...,\, \eta_T)'$.

Warunkowy rozkład *a posteriori* dla współczynnika autoregresji w równaniu stanu ϕ_1 ma gęstość:

$$p\left(\phi_1|\, y,\, \eta,\, \eta_0,\, \omega^2,\, h^{SV}_{sygnał},\, h^{SV}_{sygnał,\,0}\right) \propto f_N\left(\phi_1|\, g_{\phi_1},\, Var_{\phi_1}\right), \tag{28}$$

o średniej i wariancji równej odpowiednio:

$$g_{\phi_1} = \frac{\sigma^2_{\phi_1} \sum_{t=1}^{T} \frac{\eta_{t-1}\eta_t}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \mu_{\phi_1}\omega^2}{\sigma^2_{\phi_1} \sum_{t=1}^{T} \frac{\eta^2_{t-1}}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \omega^2} \text{ i } Var_{\phi_1} = \frac{\omega^2\sigma^2_{\phi_1}}{\sigma^2_{\phi_1} \sum_{t=1}^{T} \frac{\eta^2_{t-1}}{\exp\left(h^{SV}_{sygnał,\,t}\right)} + \omega^2}.$$

Parametry γ^2_{szum}, ρ_{szum}, $\gamma^2_{sygnał}$, $\rho_{sygnał}$ mają warunkową gęstość *a posteriori* identyczną jak w równaniach (12) i (15), natomiast gęstości warunkowych rozkładów *a posteriori* zmiennych ukrytych $h^{SV}_{sygnał,\,t}$ i $h^{SV}_{szum,\,t}$ $(t = 1,\, ...,\, T)$ mają następującą postać:

$$p\left(h^{SV}_{sygnał,\,t}|\, y,\, \eta_t,\, \eta_{t-1},\, \eta_0,\, h^{SV}_{sygnał,\,t-1},\, h^{SV}_{sygnał,\,t+1},\, h^{SV}_{sygnał,\,0},\, \theta\right) \propto$$

$$\frac{1}{\exp\left(h^{SV}_{sygnał,\,t}/2\right)} \exp\left\{-\frac{(\eta_t - \phi_1\eta_{t-1})^2}{2\omega^2_t}\right\} \frac{1}{\exp\left(h^{SV}_{sygnał,\,t}\right)} \exp\left\{-\frac{1}{2Var_\gamma}\left(h^{SV}_{sygnał,\,t} - s_t\right)^2\right\}, \tag{29}$$

gdzie s_t i Var_γ mają identyczną postać jak w równaniach (18) i (19), natomiast warunkowy rozkład dla $h^{SV}_{szum,\,t}$ ma gęstość wyrażoną następującym wzorem:

$$p\left(h^{SV}_{szum,\,t}|\, y,\, y_0,\, \eta_t,\, h^{SV}_{szum,\,t-1},\, h^{SV}_{szum,\,t+1},\, h^{SV}_{szum,\,0}\,\theta\right) \propto$$

$$\frac{1}{\exp\left(h^{SV}_{szum,\,t}/2\right)} \exp\left\{-\frac{(\Delta y_t - \eta_t y_{t-1})^2}{2\sigma_t^2}\right\} \frac{1}{\exp\left(h^{SV}_{szum,\,t}\right)} \exp\left\{-\frac{1}{2Var_\gamma}\left(h^{SV}_{szum,\,t} - s_t\right)^2\right\}, \tag{30}$$

gdzie s_t i Var_γ można przedstawić jako:

$s_t = \rho_{szum}\left(h^{SV}_{szum,\,t+1} + h^{SV}_{szum,\,t-1}\right)/\left(1+\rho^2_{szum}\right)$, $Var_\gamma = \dfrac{\gamma^2_{szum}}{1+\rho^2_{szum}}$, dla $t = 1, ..., T-1$

oraz $s_T = \rho_{szum} h^{SV}_{szum,\,t-1}$, $Var_\gamma = \gamma^2_{szum}$, dla $t = T$.

Ze względu na możliwość przedstawienie modelu (22)-(23) za pomocą reprezentacji przestrzeni stanów, charakterystyki *a posteriori* nieobserwowalnej zmiennej η_t $(t = 1, ..., T)$ można obliczyć za pomocą filtru Kalmana lub losując z następującego warunkowego rozkładu *a posteriori*:

$$p\left(\eta_t|\Delta y_t,\, y_{t-1},\, \eta_{t-1},\, \eta_{t+1},\, h^{SV}_{sygnał},\, \theta\right) = f_N\left(\eta_t|\, g_\eta,\, Var_\eta\right), \text{ dla } t = 1, ..., T-1, \tag{31}$$

gdzie:

$$g_\eta = \frac{\omega_t^2 y_{t-1}\Delta y_t + \phi_1\sigma_t^2\left(\eta_{t-1}+\eta_{t+1}\right)}{\sigma_t^2\left(1+\phi_1^2\right)+\omega_t^2 y_{t-1}^2} \text{ i } Var_\eta = \frac{\omega_t^2\sigma_t^2}{\sigma_t^2\left(1+\phi_1^2\right)+\omega_t^2 y_{t-1}^2}.$$

Dla ostatniej obserwacji $t = T$ mamy:

$$p\left(\eta_t|\Delta y_t,\, y_{t-1}, \eta_{t-1},\, h^{SV}_{sygnał},\, \theta\right) = f_N\left(\eta_t|\, g_\eta,\, Var_\eta\right), \tag{32}$$

gdzie:

$$g_\eta = \frac{\omega_t^2 y_{t-1}\Delta y_t + \phi_1\sigma_t^2\eta_{t-1}}{\sigma_t^2+\omega_t^2 y_{t-1}^2} \text{ i } Var_\eta = \frac{\omega_t^2\sigma_t^2}{\sigma_t^2+\omega_t^2 y_{t-1}^2}.$$

Estymację warunkowej wariancji w modelach stochastycznej zmienności, w literaturze polskiej przedstawia Pajor [18]. Filtr Kalmana w podejściu bayesowskim omawia Koop [9].

4. MODELE DWULINIOWE ZAWIERAJĄCE STOCHASTYCZNY PIERWIASTEK JEDNOSTKOWY

Modele dwuliniowe po raz pierwszy zostały zaproponowane przez Grangera i Andersena [4]. W literaturze polskiej ich własności omówione są m.in. w [1] i [2]. W niniejszej części artykułu omówiono model dwuliniowy zawierający stochastyczny pierwiastek jednostkowy zaproponowany przez Francqa, Makarovą i Zakoïana [3]. Model ten oznaczony symbolem ECM(p)-BL(q) nawiązuje bezpośrednio do znanych wcześniej modeli dwuliniowych ma jednak nieco inną specyfikację i własności:

$$\Delta y_t = \psi_1 y_{t-1} + \varphi_1\Delta y_{t-1} + ... + \varphi_p\Delta y_{t-p} + v_t, \tag{33}$$

$$v_t = \left(1 + \lambda_1 v_{t-1} + ... + \lambda_q v_{t-q}\right)\eta_t, \quad \eta_t \overset{iid}{\sim} N\left(0, \sigma^2\right), \quad t = 1, ..., T. \tag{34}$$

Warunek konieczny i dostateczny, aby proces (33)-(34) był stacjonarny kowariancyjnie ma postać:

$$\sum_{i=1}^{q} \lambda_i^2 \sigma^2 < 1. \tag{35}$$

Wiadomo, że procesy dwuliniowe można zakwalifikować jako aproksymację liniowych modeli o losowych parametrach, co z kolei oznacza, że są one pewną formą pośrednią między procesami ARMA ze stałymi i losowymi parametrami (por. [1]). W tym przypadku model dwuliniowy (33)-(34) bezpośrednio nawiązuje do modeli, w których parametry strukturalne są losowe. Łatwo można wykazać, że dla $p = 0$ i $q = 1$ oraz po przeprowadzeniu elementarnych przekształceń, można go zapisać jako model zawierający stochastyczny pierwiastek jednostkowy:

$$\Delta y_t = \delta_t y_{t-1} + \varepsilon_t, \quad t = 1, ..., T, \tag{36}$$

gdzie losowy parametr jest wyrażony jako:

$$\delta_t = 1 + \psi + \lambda_1 \eta_t, \tag{37}$$

a ε_t jest składnikiem losowym nieskorelowanym z y_{t-i} dla $i > 0$, o postaci:

$$\varepsilon_t = (1 - \lambda_1 (1 + \psi) y_{t-2}) \eta_t. \tag{38}$$

Niech Ψ_{t-1} oznacza zbiór wszystkich informacji o procesie dostępnych do chwili $t - 1$ włącznie. Dla $p = 0$ i $q = 1$ warunkowa średnia $\mu_t = E(\Delta y_t | \Psi_{t-1})$ i warunkowa wariancja $h_t = Var(\Delta y_t | \Psi_{t-1})$ w modelu (36)-(38) mają postać: $\mu_t = \psi y_{t-1}$ i $h_t = (1 + \lambda_1 v_{t-1})^2 \sigma^2$.

Załóżmy, że składniki losowe ε_t w modelu BL(1) mają rozkład normalny. W tym przypadku gęstość wektora obserwacji jest następująca:

$$p(\Delta y | v, \theta, y_0) = \prod_{t=1}^{T} f_N\left(\Delta y_t | \psi y_{t-1}, (1 + \lambda_1 v_{t-1})^2 \sigma^2\right), \tag{39}$$

gdzie $\Delta y = (\Delta y_1, \Delta y_2, ..., \Delta y_T)'$, $v_= (v_0, v_1, ..., v_{T-1})'$ oraz $\theta = \left(\psi, \lambda_1, \sigma^2\right)'$, a także $\psi \in R$, $\lambda_1 \in R$ i $\sigma^2 \in R_+$. Dodatkowo przestrzeń parametrów jest ucięta przez warunek $\lambda_1^2 \sigma^2 < 1$.

Jako rozkłady *a priori* dla parametrów ψ i λ_1 przyjęto rozkłady normalne, natomiast dla wariancji σ^2 - odwrócony rozkład gamma.

Układ pełnych, warunkowych rozkładów *a posteriori* dla parametrów w modelu BL(1) ma postać:

$$p\left(\psi | y, \lambda_1, \sigma^2\right) \propto f_N\left(g_\psi, Var_\psi\right), \tag{40}$$

gdzie:

$$g_{\psi} = \frac{\sigma_{\psi}^2 \sum_{t=1}^{T} \frac{\Delta y_t y_{t-1}}{(1+\lambda_1 v_{t-1})^2} + \sigma^2 \mu_{\psi}}{\sigma_{\psi}^2 \sum_{t=1}^{T} \frac{y_{t-1}^2}{(1+\lambda_1 v_{t-1})^2} + \sigma^2} \text{ i } Var_{\psi} = \frac{\sigma^2 \sigma_{\psi}^2}{\sigma_{\varphi_1}^2 \sum_{t=1}^{T} \frac{y_{t-1}^2}{(1+\lambda_1 v_{t-1})^2} + \sigma^2},$$

$$p\left(\sigma^2 | y, \lambda_1, \sigma^2, v, y_0\right) = f_{IG}\left(\sigma^2 | \frac{T + 2c_{\sigma}}{2}, \sum_{t=1}^{T} \frac{(\Delta y_t - \psi y_{t-1})^2}{2(1 + \lambda_1 v_{t-1})^2} + d_{\sigma}\right), \tag{41}$$

$$p\left(\lambda_1 | y, \psi, \sigma^2\right) \propto \prod_{t=1}^{T} f_N\left(\Delta y_t | \psi y_{t-1}, (1 + \lambda_1 v_{t-1})^2 \sigma^2\right) \times f_N\left(\lambda_1 | \mu_{\lambda}, \sigma_{\lambda}^2\right). \tag{42}$$

Dla ψ i σ^2 warunkowe gęstości *a posteriori* należą do standardowej rodziny rozkładów. Z kolei losowania z warunkowego rozkładu *a posteriori* dla parametru λ_1 wymagają stosowania dodatkowej metody, wewnątrz procedury Gibbsa.

5. ESTYMACJA I TESTOWANIE MODELI STUR DLA WYBRANYCH KURSÓW WALUTOWYCH

W niniejszej części artykułu przedstawiono wyniki estymacji wybranych specyfikacji STUR, uzyskane na podstawie analizy dziennych kursów walutowych. Są to: model STUR w wersji Grangera i Swansona (równania (1)-(2) i (4)) – STUR-G-SV – M_1, model STUR w wersji Leybourne'a (równania (22)-(23)) – STUR-L-SV-M_2 oraz model dwuliniowy (równania (36)-(38))-BL(1)-M_3. Dodatkowo, w celach porównawczych obliczono charakterystyki *a posteriori* dwóch standardowych modeli zmienności. Są to: model zmienności stochastycznej (M_4):

$$\Delta y_t = \varepsilon_t, \quad \varepsilon_t \sim N\left(0, \sigma_t^2\right), \quad t = 1, ..., T, \tag{43}$$

gdzie:

$\varepsilon_t \sim N\left(0, \sigma_t^2\right)$, $\sigma_t^2 = \sigma^2 \exp\left(h_{szum,t}^{SV}\right)$, $h_{szum,t}^{SV} = \rho_{szum} h_{szum,t-1}^{SV} + \zeta_t$, $\rho_{szum} \in (-1, 1)$, a także $\zeta_t \overset{iid}{\sim} N\left(0, \gamma_{szum}^2\right)$, $\varepsilon_t \perp \zeta_s$ dla $t, s = 1, 2, ..., T$

oraz najprostszy model GARCH (symbol M_5):

$$\Delta y_t = \varepsilon_t, \quad \varepsilon_t \sim N(0, h_t), \quad t = 1, ..., T, \tag{44}$$

$$h_t = a_0 + a_1 \varepsilon_{t-1}^2 + b_1 h_{t-1}, \tag{45}$$

gdzie: $a_0 > 0$, $a_1 \geqslant 0$, $b_1 \geqslant 0$ oraz $a_1 + b_1 < 1$.

Badane szeregi składają się z 1013 obserwacji i obejmują okres od 2 stycznia 2006 do 31 grudnia 2009 roku. Analiza dotyczy następujących walut europejskich wyrażonych w złotych: franka szwajcarskiego (CHF), euro (EUR) i funta szterlinga

(GBP). Przed przystąpieniem do badań oryginalne szeregi czasowe poddano transformacji zgodnie z następującym wzorem $y_t = 100 \ln(P_t)$, gdzie P_t oznacza cenę zamknięcia.

Moc wyjaśniającą poszczególnych specyfikacji obliczono metodą Newtona i Raftery'ego [12].

Dla wszystkich wariancji jako rozkład *a priori* przyjęto odwrócony rozkład gamma z parametrami $c = 0,01$ i $d = 0,01$. Rozkład *a priori* współczynnika autoregresji jest uciętym rozkładem normalnym o średniej zero i odchyleniu standardowym równym jeden. Również stała ϕ_0 ma rozkład *a priori* będący rozkładem normalnym o średniej równej zero i wariancji równej jeden.

Dla początkowych stanów, tj. δ_0 i η_0 przyjęto wartość równą zero. Podobnie wartości początkowe w modelu SV $\left(h^{SV}_{szum,\,(sygnał),\,0}\right)$ są również traktowane jako ustalone i równe zero.

Tablica 2 przedstawia logarytmy dziesiętne czynników Bayesa obliczone pomiędzy wybranymi modelami STUR i GARCH a modelem SV. Uzyskane wyniki pokazują, że dla wszystkich szeregów czasowych model STUR w wersji Grangera i Swansona (STUR-GR-SV) uzyskał zdecydowanie największe szanse *a posteriori*. Przy założeniu jednakowych prawdopodobieństw *a priori*, model STUR-G-SV uzyskał prawdopodobieństwo *a posteriori* od 154 do 199 rzędów wielkości większe niż standardowy model SV i od 11 do 65 rzędów wielkości większe niż drugi w rankingu model STUR-L-SV. Różnice w logarytmach czynników Bayesa obliczone między modelami STUR (STUR-G-SV oraz STUR-L-SV) a modelem stochastycznej zmienności są zatem bardzo duże. Przeprowadzone badania empiryczne wyraźnie więc potwierdziły istnienie stochastycznego pierwiastka jednostkowego dla dziennych kursów walutowych.

Tablica 2.

Logarytmy dziesiętne czynników Bayesa obliczone między poszczególnymi modelami i modelem stochastycznej zmienności (SV, M_4)

	Model			
	M_1 STUR-GR-SV	M_2 STUR-L-SV	M_3 BL	M_5 GARCH
	CHF			
$\log_{10}\left(B_{M_i\mid M_4}\right)$	154,0050	97,8153	-175,7827	-16,1337
	EUR			
$\log_{10}\left(B_{M_i\mid M_4}\right)$	159,0340	147,3740	-191,2360	-14,1240
	GBP			
$\log_{10}\left(B_{M_i\mid M_4}\right)$	199,3030	134,2846	-143,0070	-11,4630

Źródło: obliczenia własne

Model dwuliniowy we wszystkich przypadkach ma najniższą moc wyjaśniającą, skutkiem czego zajmuje piąte, ostatnie miejsce w rankingach. Model stochastycznej

zmienności zajął trzecie miejsce w rankingach, za modelami STUR i przed modelem GARCH, który miał przedostatnie miejsce w rankingach. Większa moc wyjaśniająca modeli SV w stosunku do modeli GARCH jest typowa dla tych specyfikacji (zob. [14])

Tablica 3.

Informacje *a posteriori* o parametrach w modelu stochastycznej zmienności uzyskane na podstawie analizy kursu walutowego w okresie od 2.01.2006 do 31.12.2009

	Parametr		
Waluta	σ^2	ρ_{szum}	γ^2_{szum}
CHF	0,39380 (0,10890)	0,99220 (0,00426)	0,01810 (0,00616)
EUR	0,32950 (0,09440)	0,99350 (0,00349)	0,01574 (0,00521)
GBP	0,47670 (0,12920)	0,99350 (0,00367)	0,01314 (0,00496)

W pierwszej linii umieszczono wartość oczekiwaną rozkładu *a posteriori,* natomiast w drugiej linii w nawiasach okrągłych podano odchylenie standardowe Źródło: obliczenia własne

W tablicy 3 zamieszczono punktowe oceny *a posteriori* parametrów w standardowym modelu stochastycznej zmienności (równanie 43). Uzyskane wyniki są typowe dla dziennych szeregów finansowych. Wartości oczekiwane *a posteriori* współczynnika autokorelacji ρ_{szum} są bardzo bliskie wartości jeden, co przy względnie małym rozproszeniu wskazuje na dużą persystencję zmienności kursów walutowych. Podobnie oceny *a posteriori* parametrów w modelu GARCH (równania (44)-(45)) są typowe dla tego typu danych (zob. tablica 4), a suma parametrów a_1 i b_1 ma wartość oczekiwaną *a posteriori* bliską wartości jeden, co również wskazuje na dużą persystencję zmienności badanych szeregów.

Tablica 4.

Informacje *a posteriori* o parametrach w modelu GARCH uzyskane na podstawie analizy kursu walutowego w okresie od 2.01.2006 do 31.12.2009

	Parametr			
Waluta	a_0	a_1	b_1	$a_1 + b_1$
CHF	0,00851 (0,00284)	0,08956 (0,01904)	0,89990 (0,02001)	0,98941 (0,00552)
EUR	0,00452 (0,00123)	0,09202 (0,01581)	0,89743 (0,01564)	0,98945 (0,00546)
GBP	0,00532 (0,00164)	0,07667 (0,01454)	0,91570 (0,01443)	0,99237 (0,00388)

W pierwszej linii umieszczono wartość oczekiwaną rozkładu *a posteriori,* natomiast w drugiej linii w nawiasach okrągłych podano odchylenie standardowe Źródło: obliczenia własne

Kolejna tablica, z numerem 5, zawiera informacje *a posteriori* na temat parametrów w modelu dwuliniowym. Jak wiadomo z części czwartej artykułu, model ten można również interpretować jako model ze stochastycznym pierwiastkiem jednostkowym.

Tablica 5.
Informacje *a posteriori* o parametrach w modelu dwuliniowym uzyskane na podstawie analizy kursu walutowego w okresie od 2.01.2006 do 31.12.2009

	Parametr		
Waluta	ψ	σ^2	λ_1
CHF	0,00009 (0,00035)	1,01400 (0,04520)	0,01931 (0,01251)
EUR	0,00003 (0,00018)	0,63000 (0,02808)	-0,00336 (0,01433)
GBP	-0,00013 (0,00017)	0,81520 (0,03635)	0,00779 (0,01375)

W pierwszej linii umieszczono wartość oczekiwaną rozkładu *a posteriori,* natomiast w drugiej linii w nawiasach okrągłych podano odchylenie standardowe Źródło: obliczenia własne

Uzyskane wyniki nie potwierdzają jednak przydatności tej specyfikacji dla badanych szeregów czasowych. W żadnym z trzech przypadków parametr λ_1, odpowiedzialny za zmienność warunkowej wariancji, nie jest położony w rejonach odległych od zera. Wartość oczekiwana *a posteriori* jest zbliżona do wartości odchylenia standardowego *a posteriori* lub jest o jeden rząd wielkości mniejsza i w związku z tym rozkład *a posteriori* ma bardzo duże względne rozproszenie. Powyższe wnioski potwierdza uzyskany ranking modeli (por. tablica 2), w którym model BL(1) we wszystkich trzech przypadkach zajął ostatnie miejsce.

Tablica 6.
Informacje *a posteriori* o parametrach modelu STUR-G-SV uzyskane na podstawie analizy kursu walutowego w okresie od 2.01.2006 do 31.12.2009

	Waluta					
Parametr	ω^2	σ^2	ϕ_0	ϕ_1	$\rho_{sygnał}$	$\gamma^2_{sygnał}$
CHF	0,00231 (0,00101)	0,22800 (0,02279)	0,00012 (0,00037)	-0,06131 (0,03451)	0,99740 (0,00132)	0,14150 (0,04520)
EUR	0,00220 (0,00108)	0,12400 (0,01082)	0,00005 (0,00023)	-0,04054 (0,04071)	0,99780 (0,00097)	0,14080 (0,04369)
GBP	0,00175 (0,00062)	0,14710 (0,02217)	-0,00014 (0,00023)	-0,06760 (0,04134)	0,99760 (0,00109)	0,09777 (0,02955)

W pierwszej linii umieszczono wartość oczekiwaną rozkładu *a posteriori,* natomiast w drugiej linii w nawiasach okrągłych podano odchylenie standardowe Źródło: obliczenia własne

W tablicy 6 zamieszczono wyniki estymacji parametrów w modelu STUR w wersji Grangera i Swansona (model STUR-G-SV). W pierwszej linii znajdują się wartości oczekiwane rozkładów *a posteriori*, a w nawiasach okrągłych podano wartości odchyleń standardowych. Dla wszystkich szeregów czasowych, współczynnik autokorelacji ϕ_1 ma wartość oczekiwaną zlokalizowaną blisko wartości zero, co przy dużym względnym rozproszeniu wskazuje, że zmienny parametr nie wykazuje zależności autoregresyjnej. Zupełnie inaczej przestawia się rozkład *a posteriori* parametru $\rho_{sygnał}$, który odpowiedzialny jest za persystencję zmienności stochastycznego pierwiastka jednostkowego.

Jego wartość oczekiwana *a posteriori* jest bliska wartości jeden, co oznacza, że losowy parametr jest procesem zmienności stochastycznej.

Tablica 7.
Informacje *a posteriori* o parametrach w modelu STUR-L-SV uzyskane na podstawie analizy dziennych stóp zmian kursu walutowego w okresie od 2.01.2006 do 31.12.2009

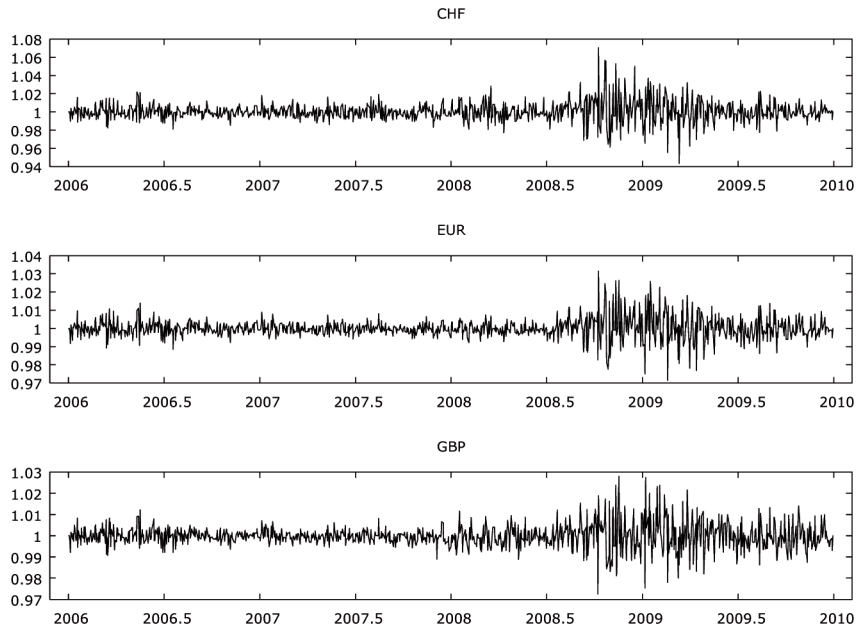
	Parametr						
Waluta	ω^2	σ^2	ρ_{szum}	γ^2_{szum}	ϕ_1	$\rho_{sygnał}$	$\gamma^2_{sygnał}$
CHF	0,0020 (0,0010)	0,2047 (0,0416)	0,2311 (0,5388)	0,0674 (0,0861)	-0,0892 (0,0441)	0,9972 (0,0014)	0,1256 (0,0470)
EUR	0,0016 (0,0007)	0,1418 (0,0337)	-0,0327 (0,5144)	0,1119 (0,1335)	-0,0506 (0,0564)	0,9977 (0,0010)	0,0972 (0,0341)
GBP	0,0016 (0,0007)	0,1418 (0,0337)	-0,0327 (0,5144)	0,1119 (0,1335)	-0,0425 (0,0524)	0,9977 (0,0010)	0,0972 (0,0341)

W pierwszej linii umieszczono wartość oczekiwaną rozkładu *a posteriori,* natomiast w drugiej linii w nawiasach okrągłych podano odchylenie standardowe Źródło: obliczenia własne

W tablicy 7 przedstawiono informacje *a posteriori* dotyczące ocen parametrów w modelu STUR-L-SV. Dla wszystkich szeregów czasowych wartość oczekiwana *a posteriori* współczynnika autoregresji $\rho_{sygnał}$ jest bliska wartości jeden, co potwierdza zasadność traktowania $\{\eta_t,\ t = 1,\ ...,\ T\}$ jako procesu zmienności stochastycznej. Możemy zaobserwować istotny współczynnik autokorelacji w strukturze SV dla zmiennego parametru, ale już nie dla składnika losowego w równaniu obserwacji. Wydaje się zatem, że tylko jeden z nich jest procesem o strukturze SV. Oznacza to, że tylko stochastyczny pierwiastek jednostkowy ma zmienną w czasie warunkową wariancję. Drugi składnik losowy ma wariancję, która jest homoskedastyczna. Powyższe wnioski potwierdza również czynnik Bayesa obliczony pomiędzy modelami STUR-L-SV i SV. Okazuje się bowiem, że bardziej prawdopodobny jest model, w którym losowa, warunkowa średnia traktowana jest jako proces SV niż model, w którym odchylenia od warunkowej, deterministycznej średniej opisane są jako proces zmienności stochastycznej. Zmienność stochastycznego pierwiastka jednostkowego łatwo zauważyć na rysunku 2, na którym przedstawiono mediany *a posteriori* parametru $a_t = \exp(\delta_t)$ $(t = 1, ..., 1013)$. Przedstawione realizacje obliczone zostały na podstawie najbardziej prawdopodobnego modelu, czyli modelu STUR w wersji Grangera i Swansona (STUR-G-SV).

Ponieważ rozpiętości przedziałów o najwyższej gęstości *a posteriori* (ang. *Highest Posterior Density intervals; HPD*), we wszystkich przypadkach, były bardzo niewielkie, co potwierdza statystyczną istotność pierwiastka stochastycznego, w celu większej przejrzystości wykresów zrezygnowano z ich prezentacji. Jak widać na rysunku 2 pierwiastek jednostkowy w modelu STUR-G-SV oscyluje wokół wartości jeden i charakteryzuje się wyraźnym, okresowym nasileniem zmienności.

Szczególnie wyraźnie uwidacznia się to w drugiej połowie 2008 i w pierwszym kwartale 2009, kiedy to miało miejsce załamanie się głównych światowych indeksów giełdowych oraz bardzo silne osłabienie złotego.

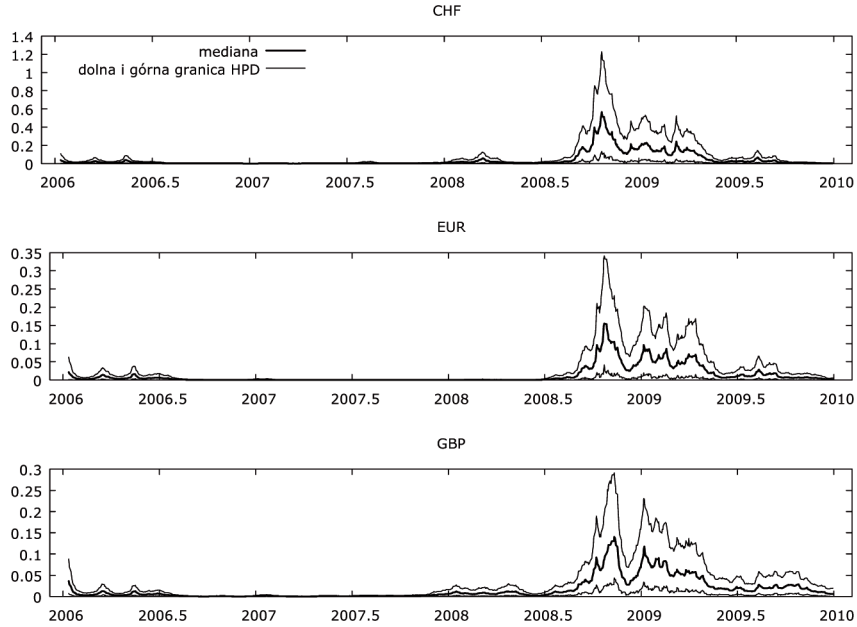


Rysunek 2. Mediany rozkładu *a posteriori* parametru $a_t = \exp(\delta_t)$ dla $t = 1, ..., 1013$ obliczone na podstawie realizacji kursu franka szwajcarskiego (CHF), euro (EUR) i funta szterlinga (GBP) względem złotego Źródło: obliczenia własne

Na rysunku 3 przedstawiono mediany *a posteriori* warunkowej wariancji ω_t^2 ($t = 1, ..., T$) wraz z 95% przedziałami o najwyższej gęstości *a posteriori*, obliczone na podstawie modelu STUR w wersji Grangera i Swansona.

Na podstawie wyników zamieszczonych na rysunku 3 łatwo można zauważyć, że od września 2008 nastąpił gwałtowny wzrost zmienności pierwiastka jednostkowego. Okresowe nasilenia zmienności losowego parametru mają swoje uzasadnienie w zaistniałych czynnikach rynkowych, ponieważ zgodnie z informacjami publikowanymi przez NBP, w okresie od września 2008 do maja 2009 roku nastąpiło znaczne osłabienie nominalnego kursu złotego względem innych walut ze względu na czynniki globalne i oczekiwania wskazujące na głębsze niż wcześniej zakładano spowolnienie wzrostu gospodarczego w 2009 roku. Kolejną przyczyną były informacje o pogarszaniu się salda na rachunku obrotów bieżących[2]. Znaczny wzrost niepewności można również zauważyć poprzez zwiększenie się rozpiętości przedziałów *a posteriori*.

[2] Raport o inflacji – luty 2009 r., www.nbp.pl (14.12.2009).



Rysunek 3. Warunkowa wariancja ω_t^2 dla $t = 1, ..., 1013$ w modelu STUR w wersji Grangera i Swansona, w którym losowy parametr jest procesem o strukturze SV (model STUR-G-SV), obliczona dla realizacji kursu franka szwajcarskiego (CHF), euro (EUR) i funta szterlinga (GBP) względem złotego. Pogrubiona linia przedstawia mediany *a posteriori* warunkowej wariancji, z kolei zwykłe linie oznaczają dolny i górny kraniec 95% przedziałów o najwyższej gęstości *a posteriori* (HPD)

Źródło: obliczenia własne

6. ZAKOŃCZENIE

Zasadniczym celem tego artykułu było przedstawienie i analiza empirycznych własności wybranych modeli zawierających stochastyczny pierwiastek jednostkowy. Wyprowadzono i omówiono również układy rozkładów *a posteriori*, które są niezbędne w podejściu bayesowskim. Badania empiryczne dotyczyły dziennych obserwacji najważniejszych walut europejskich wyrażonych w złotych w okresie od stycznia 2006 do grudnia 2009 roku.

W artykule dokonano przeglądu wybranych specyfikacji STUR. Są to niektóre już znane od dawna modele takie jak: model Grangera i Swansona [5], tzw. model Leybourne'a [10] oraz model dwuliniowy opracowany niedawno przez Francqa, Makarovą i Zakoïana [3]. Przeprowadzone badania empiryczne wyraźnie wykazały, że modele STUR, a zwłaszcza model Grangera i Swansona, uzyskały znacznie większe prawdopodobieństwo *a posteriori* niż standardowe modele zmienności tj. SV i GARCH.

W modelach STUR w wersji Grangera i Swansona oraz Leybourne'a zaproponowano, aby stochastyczny pierwiastek jednostkowy był procesem zmienności stochastycznej. Podejście to wydaje się zasadne, ponieważ badania empiryczne potwierdziły,

że stochastyczny pierwiastek jednostkowy nie ma stałej, lecz zmienną wariancję warunkową, a jego zmienność nasila się wraz z występowaniem czynników rynkowych.

Przeprowadzone badania nie potwierdziły przydatności modeli dwuliniowych dla dziennych kursów walutowych.

Katedra Ekonometrii i Statystyki
Wydział Nauk Ekonomicznych i Zarządzania
Uniwersytet Mikołaja Kopernika w Toruniu

LITERATURA

[1] Bruzda J. (2003), *Procesy dwuliniowe i procesy GARCH w modelowaniu finansowych szeregów czasowych*, Przegląd Statystyczny, 2, 73-95.
[2] Doman M., Doman R. (2004), *Ekonometryczne modelowanie dynamiki polskiego rynku finansowego*, Wyd. AE w Poznaniu, Poznań.
[3] Francq Ch., Makarova S., Zakoïan J.M. (2008), *A class of stochastic unit-root bilinear processes: Mixing properties and unit-root test*, Journal of Econometrics, 142, 312-326.
[4] Granger W.J.C., Andersen A. P. (1978), *An Introduction to Bilinear Time Series Models*, Vandenhoeck and Ruprecht, Göttingen.
[5] Granger C.W.J., Swanson N.R. (1997), *An Introduction to stochastic unit–root process*, Journal of Econometrics, 80, 35-62.
[6] Granger W.J.C., Teräsvirta T. (1993), *Modeling Nonlinear Economic Relationships*, Oxford University Press, Oxford.
[7] Górka J. (2007), *Modele autoregresyjne z losowymi parametrami*, Procesy STUR. Modelowanie i zastosowanie do finansowych szeregów czasowych, red. Osińska, M., Wyd. TNOiK, Toruń, 13-34.
[8] Jones C.R., Marriott J.M. (1999), *A Bayesian analysis of stochastic unit root models*, Bayesian Statistics, 6, 785-794.
[9] Koop G. (2003), *Bayesian Econometrics*, John Wiley & Sons.
[10] Leybourne S.J., McCabe B.P.M., Mills T.C. (1996), *Randomized unit root processes for modelling and forecasting financial time series: theory and applications*, Journal of Forecasting, 15, 253-270.
[11] Leybourne S.J., McCabe B.P.M., Tremayne A.R. (1996), *Can economic time series be differenced to stationarity?* Journal of Business and Economic Statistics, 14, 435-446.
[12] Newton M.A., Raftery A.E. (1994), *Approximate Bayesian inference by the weighted likelihood bootstrap (with discussion)*, Journal of the Royal Statistical Society, B, 56, 3-48.
[13] Nicholls D.F., Quinn B.G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Springer-Verlag, New York.
[14] Osiewalski J., Pajor A., Pipien M. (2004), *Bayesowskie modelowanie i prognozowanie indeksu WIG z wykorzystaniem procesów GARCH i SV*, XX Seminarium Ekonometryczne im. Profesora Zbigniewa Pawlowskiego, red. Zelias, A., Wyd. AE w Krakowie, Kraków, 17-39.
[15] Osińska M. (red) (2007), *Procesy STUR. Modelowanie i zastosowanie do finansowych szeregów czasowych*, Wyd. TNOiK, Toruń.
[16] Osińska M. (2004), *Stochastic unit roots process – properties and application*, 30-rd International Conference MACROMODELS'04, red. Welfe A., Welfe W., Wyd. UŁ w Łodzi, Łódź, 169-179.
[17] Osińska M., Górka J. (2005), *Identyfikacja nieliniowości w ekonomicznych szeregach czasowych*. Analiza symulacyjna, Dynamiczne modele ekonometryczne, Toruń, 35-44.
[18] Pajor A. (2003), *Procesy zmienności stochastycznej SV w bayesowskiej analizie finansowych szeregów czasowych*, Wyd. AE w Krakowie, Kraków.

[19] Sollis R., Leybourne S.J., Newbold P. (2000), *Stochastic unit roots modelling of stock price indices*, Applied Financial Economics, 10, 311-315.
[20] Tsay R.S. (2005), *Analysis of Financial Time Series*, Wiley, John and Sons, Inc., Hoboken, New Jersey.
[21] Yoon G. (2006), *A note on some properties of STUR processes*, Oxford Bulletin of Economics and Statistics, 68, 2, 253-260.

WYBRANE MODELE ZAWIERAJACE STOCHASTYCZNY PIERWIASTEK JEDNOSTKOWY W ANALIZIE KURSÓW WALUTOWYCH

S t r e s z c z e n i e

W artykule omówiono wybrane specyfikacje i własności modeli zawierających stochastyczny pierwiastek jednostkowy (STUR). Celem artykułu było dokonanie przeglądu różnych modeli STUR, opracowanie metod numerycznych wykorzystywanych w schemacie wnioskowania bayesowskiego w kontekście rozważanych modeli oraz porównanie ich mocy wyjaśniającej. Przykład empiryczny, zawarty w artykule, dotyczył dziennych notowań kursu walut obcych wyrażonych w złotych. Uzyskane wyniki wskazują, że najbardziej prawdopodobny okazał się model STUR w wersji Grangera i Swansona, który uzyskał znaczną przewagę nad standardowym modelem stochastycznej zmienności.

STOCHASTIC UNIT ROOT MODELS WITH APPLICATION TO DAILY EXCHANGE RATES

S u m m e r y

The paper presents the three different stochastic unit root models (STUR), proposed by Granger and Swanson (1997), by Leybourne, McCabe and Mills (1996) and finally by Francq, Makarova, Zakoïan (2008). The main purpose is to develop an MCMC algorithm for Bayesian estimation. We are also interested how the different specifications of the stochastic unit root affect the explanatory power of a set of competing models. We apply that these computational methods to daily exchange rates of foreign currencies in zlotys, namely Swiss franc, Pound sterling and Euro. The model selection and posterior estimates provide strong evidence in favor of the STUR models.

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

MICHAŁ KOLUPA
JOANNA PLEBANIAK

KONSEKWENCJE MAJORYZOWANIA MACIERZY KORELACJI PRZEZ MACIERZ UNIWERSALNĄ NA TLE NIERÓWNOŚCI HELLWIGA

W roku 1976 Profesor Zdzisław Hellwig sformułował hipotezę:
Jeżeli jest spełniona nierówność macierzowa:

$$\mathbf{0}(k) < \mathbf{R}(k) \leqslant \mathbf{G}(k), \tag{1}$$

gdzie $\mathbf{0}(k)$, $\mathbf{R}(k)$ oraz $\mathbf{G}(k)$ są macierzami stopnia k odpowiednio macierzą zerową, macierzą korelacji oraz macierzą uniwersalną, to model ekonometryczny określony przez regularną parę korelacyjną $(\mathbf{R}(k), \mathbf{R}_0(k))$ jest koincydentny.

Nierówność (1) nazywamy nierównością Hellwiga. Występujący w niej zapis:

$$\mathbf{R}(k) \leqslant \mathbf{G}(k) \tag{2}$$

oznacza, że macierz korelacji $\mathbf{R}(k)$ jest majoryzowana przez macierz uniwersalną $\mathbf{G}(k)$.

W niniejszej pracy przedstawimy konsekwencje, które wynikają z nierówności (2) występującej w nierówności Z. Hellwiga dotyczącej jakości mierzonej współczynnikiem $r^2_{\mathbf{R}}(k)$ modelu określonego przez regularna parę korelacyjną $(\mathbf{R}(k), \mathbf{R}_0(k))$.

Dla zwiększenia czytelności rozważań przypominamy definicje zarówno macierzy $\mathbf{R}(k)$ jak i $\mathbf{G}(k)$.

Macierz $\mathbf{R}(k)$ jest macierzą korelacji o elementach r_{ij}, które oznaczają współczynniki korelacji pomiędzy zmiennymi objaśniających Z_i oraz Z_j modelu, którego zmienną objaśnianą jest zmienna Y, czyli model:

$$Y = \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + e \tag{3}$$

określonego przez parę korelacyjną $(\mathbf{R}(k), \mathbf{R}_0(k))$, zaś $\mathbf{R}_0(k) = [r_j]$ jest wektorem korelacji, natomiast jego składowe r_j są równe współczynnikom korelacji pomiędzy parą zmiennych (Y, Z_j).

Para korelacyjna $(\mathbf{R}(k), \mathbf{R}_0(k))$ jest regularną parą korelacyjną to znaczy, że:

$$0 < r_1 \leqslant r_2 \leqslant \ldots \leqslant r_k < 1. \tag{4}$$

Z kolei macierz uniwersalna $\mathbf{G}(k)$ ma elementy g_{ij} zdefiniowane wzorem:

$$g_{ij} = \begin{cases} 1 & \text{dla} \quad i = j \\ r_i r_j & \text{dla} \quad i \neq j \end{cases} \tag{5}$$

natomiast współczynniki r_i oraz r_j są współczynnikami korelacji pomiędzy parami zmiennych (Y, Z_i) oraz (Y, Z_j).

Hipoteza Z. Hellwiga jest od 1993 roku twierdzeniem tegoż autora ponieważ w tym roku została udowodniona ([3]).

Dodajmy jeszcze, że koincydencja, regularna para korelacyjna, macierz korelacji i macierz uniwersalna zostały wprowadzone do ekonometrii przez Z. Hellwiga (patrz [1]).

Tak więc, jeżeli jest spełniona nierówność (1), to wektor $\mathbf{A}(k)$ spełniający układ równań:

$$\mathbf{R}(k)\mathbf{A}(k) = \mathbf{R}_0(k) \tag{6}$$

jest wektorem o dodatnich składowych.

Miarą jakości modelu określonego przez regularną parę korelacyjną $(\mathbf{R}(k), \mathbf{R}_0(k))$ jest współczynnik $r^2_{\mathbf{R}}(k)$ postaci:

$$r^2_{\mathbf{R}}(k) = \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}^{-1}(k)\mathbf{R}_0(k) = \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{A}(k). \tag{7}$$

Odnotujmy jeszcze, że współczynnik $r^2_{\mathbf{G}}(k)$ stanowi miarę modelu określonego przez regularną parę korelacyjną $(\mathbf{G}(k), \mathbf{R}_0(k))$, natomiast wektor $\mathbf{B}(k)$ spełnia układ równań:

$$\mathbf{G}(k)\mathbf{B}(k) = \mathbf{R}_0(k) \tag{8}$$

Dowodzi się ([2]), że składowa b_i, $i = 1, 2, \ldots, k$ wektora $\mathbf{B}(k)$ spełniającego układ (8) ma postać:

$$b_i = \frac{r_i}{1 - r_i^2}\varphi^2_{\mathbf{G}}(k), \tag{9}$$

gdzie:

$$\varphi^2_{\mathbf{G}}(k) = 1 - \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{G}^{-1}(k)\mathbf{R}_0(k). \tag{10}$$

Ponieważ para korelacyjna $(\mathbf{G}(k), \mathbf{R}_0(k))$ jest regularna, przeto wektor $\mathbf{B}(k)$ ma wszystkie składowe dodatnie.

Jak wiadomo ([2]) prawdziwa jest równość:

$$\mathbf{G}(k) = \mathbf{M}(k) + \mathbf{R}_0(k)\mathbf{R}_0^{\mathrm{T}}(k), \tag{11}$$

gdzie:

$$\mathbf{M}(k) = \operatorname{diag}\left\{ \begin{array}{cccc} 1 - r_1^2 & 1 - r_2^2 & \ldots & 1 - r_k^2 \end{array} \right\}. \tag{12}$$

Wówczas na podstawie nierówności mamy:

$$\mathbf{R}(k) \leqslant \mathbf{M}(k) + \mathbf{R}_0(k)\mathbf{R}_0^{\mathrm{T}}(k), \tag{13}$$

a ponieważ wektor $\mathbf{A}(k) > 0$, przeto (13) dostajemy:

$$\mathbf{R}(k)\mathbf{A}(k) \leqslant \mathbf{M}(k)\mathbf{A}(k) + \mathbf{R}_0(k)\mathbf{R}_0^{\mathrm{T}}(k)\mathbf{A}(k), \tag{14}$$

przeto na podstawie (5) i (6) otrzymujemy:

$$\mathbf{R}_0(k) \leqslant \mathbf{M}(k)\mathbf{A}(k) + \mathbf{R}_0(k)r_{\mathbf{R}}^2(k), \tag{15}$$

albo:

$$\mathbf{R}_0\left[1 - r_{\mathbf{R}}^2(k)\right] \leqslant \mathbf{M}(k)\mathbf{A}(k), \tag{16}$$

czyli:

$$\mathbf{R}_0(k)\varphi_{\mathbf{R}}^2(k) \leqslant \mathbf{M}(k)\mathbf{A}(k), \tag{17}$$

ale:

$$\mathbf{M}^{-1}(k) = \operatorname{diag}\left\{ \frac{1}{1-r_1^2} \quad \frac{1}{1-r_2^2} \quad \cdots \quad \frac{1}{1-r_k^2} \right\}, \tag{18}$$

przeto:

$$\mathbf{A}(k) \geqslant \mathbf{M}^{-1}(k)\mathbf{R}_0(k)\varphi_{\mathbf{R}}^2(k) \tag{19}$$

skąd na podstawie (9) dostajemy:

$$a_i \geqslant \frac{r_i}{1-r_i^2}\varphi_{\mathbf{R}}^2(k) = \frac{r_i}{1-r_i^2}\varphi_{\mathbf{G}}^2(k)\frac{\varphi_{\mathbf{R}}^2(k)}{\varphi_{\mathbf{G}}^2(k)} \qquad \mathrm{i} = 1, 2, \ldots, k \tag{20}$$

czyli:

$$a_i\varphi_{\mathbf{G}}^2(k) \geqslant b_i\varphi_{\mathbf{R}}^2(k) \qquad \mathrm{i} = 1, 2, \ldots, k \tag{21}$$

albo:

$$\mathbf{A}(k)\varphi_{\mathbf{G}}^2(k) \geqslant \mathbf{B}(k)\varphi_{\mathbf{R}}^2(k) \tag{22}$$

Ponieważ $r_i > 0$, $i = 1, 2, \ldots, k$, przeto z (21) mamy:

$$r_i a_i\varphi_{\mathbf{G}}^2(k) \geqslant r_i b_i\varphi_{\mathbf{R}}^2(k) \qquad \mathrm{i} = 1, 2, \ldots, k \tag{23}$$

ale:

$$\sum_{i=1}^{k} r_i a_i = r_{\mathbf{R}}^2(k), \tag{24}$$

$$\sum_{i=1}^{k} r_i b_i = r_{\mathbf{G}}^2(k), \tag{25}$$

przeto z (23) wynika, że:

$$r_{\mathbf{R}}^2(k)\varphi_{\mathbf{G}}^2(k) \geqslant r_{\mathbf{G}}^2(k)\varphi_{\mathbf{R}}^2(k) \tag{26}$$

a stąd po prostych nie wymagających objaśnień przekształceniach dostajemy:

$$r_{\mathbf{R}}^2(k) \geqslant r_{\mathbf{G}}^2(k) \tag{27}$$

Nierówność (26) oznacza, że w przypadku majoryzowania macierzy $\mathbf{R}(k)$ przez macierz $\mathbf{G}(k)$ jakość modelu określonego przez regularną parę korelacyjną ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) jest nie większa niż jakość modelu określonego przez regularną parę korelacyjną ($\mathbf{G}(k)$, $\mathbf{R}_0(k)$).

Na podstawie (9) i (24) mamy:

$$r^2_{\mathbf{G}}(k) = \sum_{i=1}^{k} r_i b_i = \varphi^2_{\mathbf{G}}(k) \sum_{i=1}^{k} \frac{r_i^2}{1 - r_i^2}, \tag{28}$$

czyli:

$$r^2_{\mathbf{G}}(k) = \left[1 - r^2_{\mathbf{G}}(k)\right] \sum_{i=1}^{k} \frac{r_i^2}{1 - r_i^2}, \tag{29}$$

$$r^2_{\mathbf{G}}(k) \left[1 + \sum_{i=1}^{k} \frac{r_i^2}{1 - r_i^2}\right] = \sum_{i=1}^{k} \frac{r_i^2}{1 - r_i^2}. \tag{30}$$

Ostatecznie:

$$r^2_{\mathbf{G}}(k) = \frac{Q}{1 + Q}, \tag{31}$$

gdzie:

$$Q = \sum_{i=1}^{k} \frac{r_i^2}{1 - r_i^2}. \tag{32}$$

Wobec tego na podstawie (7) oraz (27) mamy:

$$r^2_{\mathbf{R}}(k) \geqslant \frac{Q}{1 + Q}. \tag{33}$$

Tak więc jakość modelu określonego przez regularną parę korelacyjną ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) dla którego jest spełniona nierówność Z. Hellwiga jest co najmniej równa ułamkowi $\dfrac{Q}{1 + Q}$.

Na zakończenie odwołajmy się do przykładu ilustrującego przedstawione rozważania.

Przykład

Dana jest regularna para korelacyjna ($\mathbf{R}(3)$, $\mathbf{R}_0(3)$), gdzie:

$$\mathbf{R}(3) = \begin{bmatrix} 1 & 0,01 & 0,02 \\ 0,01 & 1 & 0,04 \\ 0,02 & 0,04 & 1 \end{bmatrix}, \qquad \mathbf{R}_0(3) = \begin{bmatrix} 0,1 \\ 0,2 \\ 0,3 \end{bmatrix}. \tag{34}$$

Model określony przez tę parę jest koincydentny, ponieważ jest spełniona nierówność Z. Hellwiga:

$$\mathbf{0}(3) < \mathbf{R}(3) \leqslant \mathbf{G}(3), \tag{35}$$

gdzie:

$$\mathbf{G}(3) = \begin{bmatrix} 1 & 0,02 & 0,03 \\ 0,02 & 1 & 0,06 \\ 0,03 & 0,06 & 1 \end{bmatrix}. \tag{36}$$

Następnie:

$$\sum_{i=1}^{k} \frac{r_i^2}{1-r_i^2} = \frac{0,0001}{1-0,0001} + \frac{0,0004}{1-0,0004} + \frac{0,0009}{1-0,0009} \approx 0,001401 \tag{37}$$

zaś:

$$1 + \sum_{i=1}^{k} \frac{r_i^2}{1-r_i^2} \approx 1,001401 \tag{38}$$

czyli:

$$r_{\mathbf{R}}^2(k) \geqslant \frac{0,001401}{1,001401} \approx 0,0014. \tag{39}$$

a to oznacza, że jakość modelu określonego przez regularną parę daną wzorem (34) jest co najmniej równa 0,0014. Jest to również istotna konsekwencja majoryzowania macierzy korelacji przez macierz uniwersalną na tle nierówności Z. Hellwiga.

prof. zw. dr hab. Michał Kolupa, Politechnika Radomska w Radomiu
dr hab. Joanna Plebaniak, prof. SGH, Szkoła Główna Handlowa w Warszawie

LITERATURA

[1] Hellwig Z., (1976), *Przechodniość relacji skorelowania zmiennych losowych i płynące stąd wnioski ekonometryczne,* „Przegląd Statystyczny" nr 3, Warszawa.
[2] Kolupa M., (1977), *Dowód pewnego twierdzenia Z. Hellwiga i pewne własności macierzy uniwersalnej,* „Przegląd Statystyczny" nr 7, Warszawa.
[3] Kolupa M., (1993), *Dowód hipotezy Z. Hellwiga*, „Przegląd Statystyczny" nr 2, Warszawa.

KONSEKWENCJE MAJORYZOWANIA MACIERZY KORELACJI PRZEZ MACIERZ UNIWERSALNĄ NA TLE NIERÓWNOŚCI HELLWIGA

S t r e s z c z e n i e

W pracy wykazano, że jeżeli jest spełniona nierówność Z. Hellwiga:

$\mathbf{0}(k) < \mathbf{R}(k) \leqslant \mathbf{G}(k)$,

gdzie $\mathbf{0}(k)$, $\mathbf{R}(k)$ oraz $\mathbf{G}(k)$ są macierzami stopnia k odpowiednio zerowej, korelacji oraz uniwersalnej, to jakość modelu określonego przez regularną parę korelacyjną ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) mierzona współczynnikiem $r^2_{\mathbf{R}}(k)$ jest co najmniej równa:

$$\frac{Q}{1+Q}$$

gdzie:

$$Q = \sum_{i=1}^{k} \frac{r_i^2}{1-r_i^2},$$

zaś $r_i = r(Y, Z_i)$, $i = 1, 2, \ldots, k$ jest współczynnikiem korelacji pomiędzy parą zmiennych Y oraz Z_i, $i = 1, 2, \ldots, k$.

W pracy udowodniono, że:

$$\mathbf{A}(k)\varphi^2_{\mathbf{G}}(k) \geqslant \mathbf{B}(k)\varphi^2_{\mathbf{R}}(k)$$

gdzie $\mathbf{A}(k)$ oraz $\mathbf{B}(k)$ są rozwiązaniami układów:

$\mathbf{R}(k)\mathbf{A}(k) = \mathbf{R}_0(k)$

$\mathbf{G}(k)\mathbf{B}(k) = \mathbf{R}_0(k)$

zaś $\varphi^2_{\mathbf{G}}(k)$ oraz $\varphi^2_{\mathbf{R}}(k)$ obliczamy według wzorów:

$$\varphi^2_{\mathbf{G}}(k) = 1 - \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{G}^{-1}(k)\mathbf{R}_0(k),$$

$$\varphi^2_{\mathbf{R}}(k) = 1 - \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}^{-1}(k)\mathbf{R}_0(k),$$

które mierzą jakość modeli określonych przez regularne pary korelacyjne odpowiednio ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) oraz ($\mathbf{G}(k)$, $\mathbf{R}_0(k)$). Główny wynik zamieszczony w niniejszej pracy orzeka, że jeżeli macierz korelacji jest majoryzowana przez macierz uniwersalną i jest spełniona nierówność Z. Hellwiga, to jakość modelu określonego przez regularną parę korelacyjną ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) jest co najmniej równa jakości modelu określonego przez parę korelacyjną ($\mathbf{G}(k)$, $\mathbf{R}_0(k)$).

Słowa kluczowe: macierz korelacji, macierz uniwersalna, nierówność Z. Hellwiga

THE CONSEQUENCES OF THE MAJORISATION OF THE CORRELATION MATRIX BY THE UNIVERSAL MATRIX BASED ON HELLWIG'S INEQUALITY

S u m m a r y

In the paper it was proved that if there is Hellwig's inequality:

$\mathbf{0}(k) < \mathbf{R}(k) \leqslant \mathbf{G}(k)$,

where $\mathbf{0}(k)$, $\mathbf{R}(k)$ and $\mathbf{G}(k)$ are accordingly, the zero matrix, the correlation matrix and the universal matrix, then the quality of the model defined by regular correlation pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) measured by the factor $r^2_{\mathbf{R}}(k)$ is at least equal:

$$\frac{Q}{1+Q}$$

where:

$$Q = \sum_{i=1}^{k} \frac{r_i^2}{1 - r_i^2},$$

and $r_i = r(Y, Z_i)$, $i = 1, 2, \ldots, k$ is a correlation coefficient between a pair of variables Y and Z_i, $i = 1, 2, \ldots, k$.

In the paper it was proved, that:

$$\mathbf{A}(k)\varphi_{\mathbf{G}}^2(k) \geqslant \mathbf{B}(k)\varphi_{\mathbf{R}}^2(k)$$

where vector $\mathbf{A}(k)$ and vector $\mathbf{B}(k)$ are solution to the equations:

$\mathbf{R}(k)\mathbf{A}(k) = \mathbf{R}_0(k)$

$\mathbf{G}(k)\mathbf{B}(k) = \mathbf{R}_0(k)$

where $\varphi_{\mathbf{G}}^2(k)$ and (10) $\varphi_{\mathbf{R}}^2(k)$ are calculated according to the formulas:

$$\varphi_{\mathbf{G}}^2(k) = 1 - \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{G}^{-1}(k)\mathbf{R}_0(k),$$

$$\varphi_{\mathbf{R}}^2(k) = 1 - \mathbf{R}_0^{\mathrm{T}}(k)\mathbf{R}^{-1}(k)\mathbf{R}_0(k),$$

which measure the quality of the models determined by the regular correlation pairs accordingly ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) and ($\mathbf{G}(k)$, $\mathbf{R}_0(k)$). The main solution included in this work states that if the correlation matrix is majorised by the universal matrix and Hellwig's inequality is satisfied then the quality of the model defined by the regular correlation pair ($\mathbf{R}(k)$, $\mathbf{R}_0(k)$) is at least equal to the quality of the model defined by the regular universal pair ($\mathbf{G}(k)$, $\mathbf{R}_0(k)$).

Key words: correlation matrix, universal matrix, Hellwig's inequality

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

KRZYSZTOF NAJMAN

GRUPOWANIE DYNAMICZNE Z WYKORZYSTANIEM SIECI GNG

1. WSTĘP

Od początku lat 90-tych XX wieku obserwuje się stały, dynamiczny wzrost liczby baz danych i zbieranych w nich informacji. Rozwój gospodarek wielu krajów, spadek cen komputerów, większa liczba sprawnych aplikacji, rozwój różnych typów sieci komputerowych Intranet, a także rozwój globalnej sieci Internet spowodowały z jednej strony wzrost zapotrzebowania na informacje, a z drugiej stworzyły możliwości ich zbierania i przechowywania. Rozwój ten przyczynia się do poszerzania wiedzy o otaczającym nas świecie. Umożliwia między innymi lepsze zrozumienie zachodzących procesów społecznych, ekonomicznych, sprawniejsze zarządzanie państwem czy przedsiębiorstwem. Sama jednak rosnąca liczba zbieranych danych nie wpływa bezpośrednio na zwiększanie się zasobów wiedzy. Dane muszą być przetworzone w informacje, a dopiero te w wiedzę. Paradoksalnie, proces pozyskiwania wiedzy nie jest obecnie łatwiejszy, mimo ogromnego wzrostu liczby dostępnych danych. Sama objętość i własności istniejących baz stają się poważnym problemem w analizie zgromadzonych danych.

Jednym z problemów związanych z analizą danych zawartych we współczesnych bazach jest duża zmienność ich zawartości. Rejestracja nowych jednostek może następować w sposób ciągły, dynamicznie zmieniając obraz obserwowanej populacji. Dynamicznie może się także zmieniać struktura grupowa rejestrowanych jednostek. Wraz z upływem czasu i napływem nowych danych znane i dobrze określone skupienia mogą tracić na znaczeniu lub rozmyć się w innych. Skupienia zawierające nieliczne jednostki, słabo określone (rozmyte) mogą z czasem stać się bardziej liczne, dobrze określone a nawet dominujące. Mogą pojawić się również całkowicie nowe skupienia. Możliwe jest także, aby dane posiadały przypisany im czas ważności, po którym tracą swoje znaczenie i wpływ na bieżącą strukturę populacji.

Powyższe własności bazy danych powodują, że sama struktura grupowa populacji może być dynamiczna i zmieniać się w sposób ciągły. Aby poprawnie obserwować proces zmian struktury grupowej badanej populacji należy dokonywać grupowania obiektów i korekt w opisie skupień także w sposób ciągły – wraz z napływaniem nowych danych. Konieczne jest zastosowanie metody grupowania zdolnej do reagowania na każdą nową informację i automatycznie dokonującej niezbędnych korekt w opisie istniejącej struktury grupowej. Opis ten musi być dostępny w dowolnym momencie istnienia bazy danych.

Celem prezentowanych badań jest weryfikacja możliwości zastosowania samouczącej się sieci neuronowej o zmiennej strukturze typu *Growing Neural Gas* w dynamicznym grupowaniu jednostek.

2. GRUPOWANIE STATYCZNE A GRUPOWANIE DYNAMICZNE

W literaturze dotyczącej metod grupowania danych pojęcie grupowania dynamicznego pojawia się głównie w kontekście analizy szeregów czasowych (Wang X., Smith K., Hyndman R. [1997], Zamir O., Grouper O.E. [1999], Guha S., Mishra N., Motwani R., O'Callaghan L. [2000], Babcock B., Babu S., Datar M., Motwani R., Widom J. [2002]). Staje się ono elementem szczególnej analizy korelacji (Fenn D.J., Porter M.A., Mucha P.J., McDonald M., Williams S., Johnson N.F., Jones N.S. [2010]). Obserwuje się jednocześnie znaczną liczbę szeregów czasowych i analizuje ich podobieństwo, klasyfikując poszczególne szeregi do niewielkiej liczby klas. Samo pojęcie grupowania dynamicznego jest tu jednak rozumiane bardzo wąsko. Dynamiczne są, bowiem jedynie badane szeregi czasowe. Sam proces grupowania jest statyczny. Stosuje się tu klasyczne metody grupowania hierarchicznego lub metodę k-średnich. Rozszerzenie powyższego podejścia do grupowania dynamicznego można znaleźć w pracy: Guedalia I.D., London M., Werman M. [1999] gdzie autorzy dzielą metody grupowania na metody wsadowe (batch) i ciągłe (online). O metodach wsadowych mówi się wtedy, gdy dokonuje się grupowania po wprowadzeniu do bazy danych większego zbioru jednostek[1]. W takim przypadku ignoruje się wszystkie pośrednie struktury grupowe, które mogły się pojawić między momentami dokonywania analizy. O grupowaniu ciągłym mówi się, gdy grupowanie jest dokonywane każdorazowo po wprowadzeniu do bazy nowej jednostki. Autorzy ograniczają jednak swoją definicję zakładając, że nowe dane pojawiają się w czasie pojedynczo – w jednej chwili jedna jednostka. Zakładają także, że nowe jednostki mogą się pojawić jedynie w stałych odstępach czasu (Guedalia I.D., London M., Werman M. [1999]). Łatwo zauważyć zbieżność podejść z analizą szeregów czasowych, w których w danej chwili może się pojawić jedynie jedna nowa obserwacja. Grupowanie online jest więc także wąskim rozumieniem grupowania dynamicznego.

Wskazane wyżej próby definicji grupowania dynamicznego wydają się bardzo ograniczone. Nie biorą pod uwagę najważniejszego elementu odpowiadającego za dynamizm procesu grupowania – zmianę jednostek i ich struktury grupowej w czasie grupowania. W powyższych definicjach dynamiczne są jedynie grupowane obiekty, jak w grupowaniu szeregów czasowych, lub uważa się za dynamiczne sekwencyjne powtarzanie procesu grupowania, jak w grupowaniu ciągłym.

Wydaje się, że aby pełniej opisać rozpatrywany problem należałoby wziąć pod uwagę kryterium stałości jednostek i ich struktury grupowej w czasie procesu grupowania. Korzystając z tego kryterium można wyróżnić dwa podejścia do grupowania

[1] Grupowanie powtarza się, gdy do bazy zostanie wprowadzona z góry ustalona liczba jednostek, nie wiadomo jednak, kiedy się to stanie, lub grupowanie powtarza się, co pewien ustalony czas o ile zostanie przekroczona zakładana minimalna liczba nowych jednostek.

i to niezależnie od typu grupowanych obiektów: statyczne i dynamiczne. O grupowaniu statycznym można mówić, gdy podczas procesu grupowania ani zbiór jednostek ani ich struktura grupowa nie zmieniają się. W tej definicji mieszczą się wszystkie powyższe podejścia do grupowania (także te nazywane dynamicznymi). O grupowaniu dynamicznym można mówić wtedy, gdy w czasie procesu grupowania zmieniają się zarówno grupowane jednostki jak i ich struktura grupowa.

Aby grupowanie dynamiczne możliwe było do praktycznego zrealizowania, stosowana metoda grupowania powinna spełniać przynajmniej cztery warunki. Metoda taka musi być:

1. szybka,
2. oszczędna,
3. autonomiczna,
4. zapewniająca wysoką jakość grupowania dla dowolnej konfiguracji przestrzennej jednostek.

Warunek szybkości jest kluczowy. Jeżeli w ciągu sekundy w bazie danych może zmienić się kilkadziesiąt lub kilkaset jednostek, także zmiany struktury grupowej muszą być uwzględnione w tym samym czasie. Metoda musi być także oszczędna, ponieważ grupowaniu podlegają duże bazy danych, coraz częściej liczące miliony przypadków. Metoda grupowania musi oszczędzać zasoby pamięci komputera. Macierz odległości dla 15000 jednostek zajmuje 1717 Mb pamięci RAM. Musi być także oszczędna obliczeniowo. Jeżeli algorytm grupowania wymaga dużej liczby obliczeń, to może wymagać użycia superkomputera lub będzie zbyt wolny, aby nadążyć za zmianami w bazie danych. Metoda powinna być wysoce autonomiczna gdyż sama szybkość zmian w bazie danych powoduje, że ewentualna ingerencja w algorytm lub jego parametry powinna być ograniczona do minimum. W szczególności algorytm taki powinien autonomicznie ustalać liczbę skupień, powinien być niewrażliwy na pojedyncze jednostki nietypowe a jednocześnie szybko tworzyć skupienie, gdy liczba obiektów nietypowych rośnie. Może to, bowiem oznaczać pojawienie się nowej prawidłowości. Metoda musi się także charakteryzować bardzo dobrymi własnościami uzyskanej struktury grupowej niezależnie od konfiguracji przestrzennej jednostek

3. SAMOUCZĄCA SIĘ SIEĆ NEURONOWA GNG

Konstrukcję sieci samouczącej się typu Growing Neural Gas zaproponował B. Fritzke w 1994. (Fritzke B. [1994]). Jest to rozwinięcie znanych wcześniej sieci samouczących się typu SOM (*Self Organizing Map*) (Kohonen T. [1997]). Zasadniczą wadą sieci SOM jest stała struktura założona a priori przez badacza. Jest to szczególnie uciążliwe w dużych badaniach empirycznych gdzie może istnieć wiele skupień o złożonej strukturze przestrzennej. Aby w takim przypadku poprawnie wyróżnić skupienia sieć SOM może wymagać dużej liczby neuronów – trudno jednak przewidzieć ile. Konsekwencją dużych rozmiarów sieci jest długi czas uczenia się sieci i spadająca

wraz ze wzrostem skali badanego problemu efektywność procesu grupowania (Migdał Najman K., Najman K. [2008], Najman K. [2009]).

Istotą budowy sieci GNG jest konstrukcja sieci maksymalnie oszczędnej, bez zbędnych neuronów, rozłożonych jedynie w tej części przestrzeni, w której znajdują się obserwowane jednostki. Sieć w procesie samouczenia się powinna sama korygować swoją strukturę zgodnie z potrzebami. Dla prostej struktury grupowej (np. kilka sferycznych, separowalnych skupień) sieć powinna mieć niewielkie rozmiary niezależnie od liczby obserwowanych obiektów. Rozmiar sieci powinien wzrastać jedynie wraz ze wzrostem liczby skupień i stopniem komplikacji ich struktury przestrzennej (np. skupienia niesferyczne, słabo separowalne).

Wydaje się, że cele powyższe udało się osiągnąć w konstrukcji sieci GNG i algorytmie samouczenia się sieci[2]. Na podstawie wyników badań teoretycznych i empirycznych można sądzić, że sieci GNG spełniają warunek oszczędności, autonomiczności i jakości grupowania (Qin A.K., Suganthan P.N. [2004], Najman K. [2009, 2010]). Do zweryfikowania pozostaje kluczowy element dobrego algorytmu grupowania dynamicznego – to jest szybkość jego działania przy założeniu zachowania przez sieć warunków oszczędności, autonomiczności i wysokiej jakości grupowania.

4. EKSPERYMENT BADAWCZY

W celu weryfikacji możliwości zastosowania sieci neuronowej typu GNG w grupowaniu dynamicznym przygotowano eksperyment.[3] Wygenerowano dynamiczną bazę danych[4], złożoną z 10 do 15000 różnych jednostek, z których w jednym momencie istnieje maksymalnie 7500 jednostek. Baza ma zmienną strukturę grupową od 1 do 10 skupień (maksymalnie 9 skupień istniejących jednocześnie). Jednostki w skupieniach mają ograniczony czas istnienia. Gdy w bazie jest już 7500 jednostek, wraz z pojawianiem się nowych jednostek te istniejące najdłużej zostają systematycznie usuwane. Wszystkie jednostki opisane są dwoma cechami[5] mierzonymi na skali ilorazowej. Struktura grupowa wszystkich jednostek została zaprezentowana na rysunku 1.

Do grupowania dynamicznego zastosowano sieć GNG o nieograniczonej maksymalnej liczbie neuronów, maksymalnym błędzie kwantyzacji równym zero i nieograniczonej liczbie iteracji uczących. Parametry te gwarantują, że proces samouczenia się sieci będzie ciągły. Zostanie on zatrzymany sztucznie 100 iteracji uczących po wprowadzeniu do bazy ostatniej jednostki. Nowy neuron wstawiany jest co 70 itera-

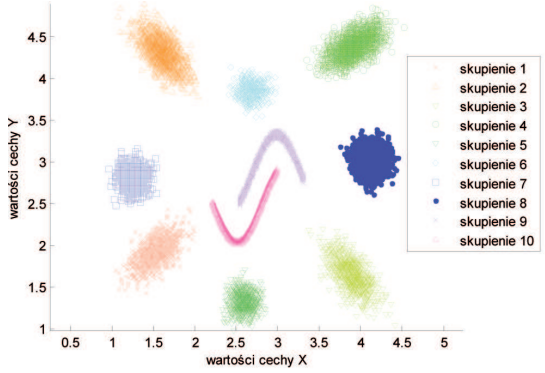
[2] Szczegóły algorytmu budowy sieci GNG, wraz ze schematem blokowym znajdują się w artykule Najman K. [2009].

[3] Niestety, nie jest publicznie dostępna empiryczna, dynamiczna baza danych.

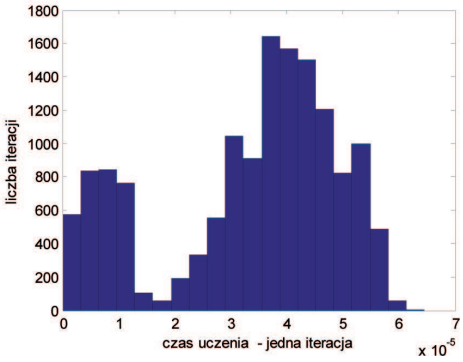
[4] Baza danych jest tu dynamiczna w takim sensie, w jakim zdefiniowano proces grupowania dynamicznego. Będzie się ona zmieniała w czasie procesu grupowania.

[5] Problem jest jedynie dwuwymiarowy z powodu niemożliwości wizualizacji sieci GNG dla wyższej liczby wymiarów. Sieć zachowuje ten sam wymiar neuronów, co wymiar przestrzeni cech badanych jednostek.

Rysunek 1 Struktura grupowa



Rysunek 2 Rozkład czasu uczenia jednej iteracji



Źródło: opracowanie własne

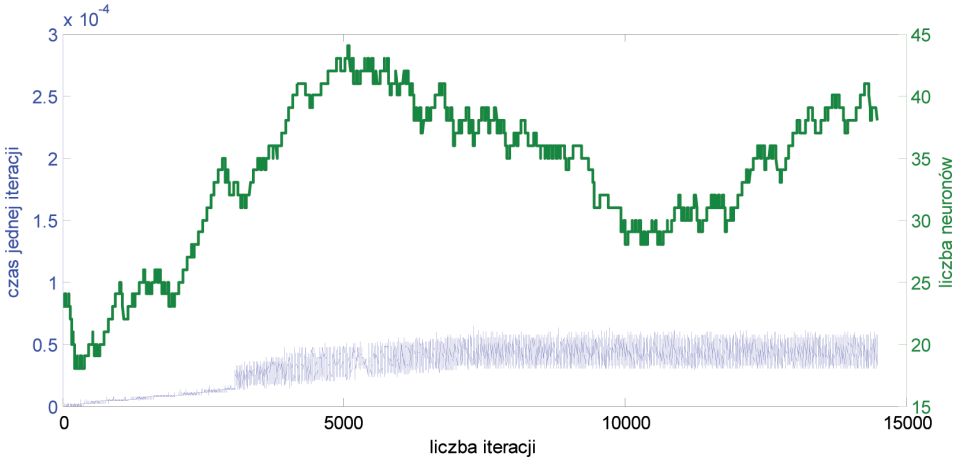
cji przy maksymalnym czasie życia neuronu równym 71 iteracji[6]. Zastosowano stały krok uczenia neuronu wygrywającego równy 0,01 i drugiego najlepszego równy 0,009. Pierwsza z wartości jest nieznacznie większa niż przeciętna odległość euklidesowa między obserwowanymi jednostkami, co zapewnia dużą szybkość przesuwania się neuronu zwycięskiego w przestrzeni. Druga wartość jest rzędu przeciętnej odległości między jednostkami, co pozwala na dokładne dopasowanie neuronu do obserwowanej jednostki. Po wprowadzeniu do bazy nowej jednostki po 142 iteracjach uczących (dwukrotnie wstawiony nowy neuron i usunięty najstarszy) dokonywany jest pomiar zgodności uzyskanej struktury grupowej ze wzorcem w oparciu o skorygowany współczynnik Rand'a.[7] Mierzony jest czas wykonania każdej iteracji uczącej.

Ponieważ jednostek w bazie jest 15000, a po wstawieniu każdej wykonywane są 142 iteracje uczące, to w eksperymencie sieć wykonała łącznie 142*15000=2130000 iteracji uczących. Łączny czas wszystkich iteracji uczących wyniósł niemal dokładnie 70 minut, co oznacza, że przeciętny czas jednej iteracji uczącej wyniósł 0,000033 minuty. Sieć jest w stanie wykonać ponad 30328 iteracji uczących na minutę i ponad 507 na sekundę. Oznacza to w praktyce ponad 3 procesy dodawania i usuwania neuronu na sekundę. Ponieważ jeden neuron może odpowiadać za poprawne grupowanie wielu jednostek, oznacza to możliwość uwzględnienia wielu nowych jednostek co sekundę. Sieć powinna zauważyć w tym czasie zmianę w strukturze grupowej wywołaną przez nowe jednostki. Rozkład czasów wykonania pojedynczej iteracji uczącej został zaprezentowany na rysunku 2. Jest on wyraźnie bimodalny. Pierwsze maksimum dotyczy jedynie początkowego okresu pracy sieci. Gdy jednostek jest bardzo niewiele

[6] Dłuższy czas życia neuronu niż częstotliwość wstawiania nowych neuronów powoduje zwykle, że sieć szybciej reaguje na pojawienie się nowych jednostek a jednocześnie mniej chętnie pozbywa się zbędnych neuronów.

[7] Każda sekwencja składająca się ze wstawienia nowej jednostki do bazy, 142 iteracji uczących sieci i prezentacji wyników jest w badaniu nazwana iteracją roboczą. Na wszystkich wykresach prezentowane są liczby iteracji roboczych.

Rysunek 3 Liczba neuronów i czas uczenia jednej iteracji



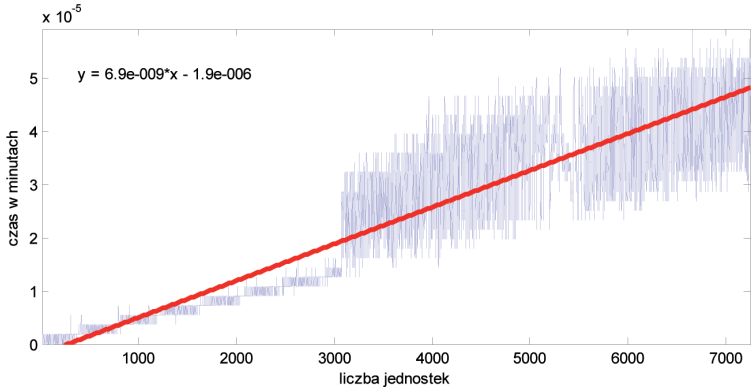
Źródło: opracowanie własne

i neuronów także, czasy uczenia są bardzo krótkie i wynoszą poniżej 0,00001 minuty. Przeciętne czasy wykonania pojedynczej iteracji uczącej dla wszystkich 15000 iteracji roboczych pokazano na rysunku 3. Ich zróżnicowanie rośnie w początkowej fazie budowy sieci. Wzrost ten nie jest jednak bezpośrednio związany z liczbą neuronów a raczej ze strukturą grupową jednostek. W początkowej fazie jest tylko jedno a później dwa skupienia, które są sferyczne i gęste. Neuronów jest niewiele stąd duża szybkość uczenia się i małe wahania czasu pojedynczej iteracji uczącej. Wyraźny skok następuje w momencie, gdy rozpoczynają się pojawiać jednostki należące do skupień o złożonej strukturze przestrzennej (dwa u-kształtne skupienia w centrum rysunku 1). Gdy sieć jest większa a struktura grupowa bardziej złożona sieć uczy się szybko w iteracjach, w których nie są wymieniane neurony a wolniej, gdy wstawiany lub usuwany jest neuron. Czas potrzebny na usunięcie neuronu jest krótszy niż czas potrzebny na jego wstawienie, co jest związane z liczbą koniecznych do wykonania operacji[8]. W ten sposób powstają 3 różne typy iteracji, różniące się także czasem wykonania, co jest widoczne na rysunku. Istotną obserwacją jest także niewielka zależność czasu iteracji od liczby neuronów i liczby jednostek. Wzrost czasu uczenia jednej iteracji jest w przybliżeniu liniową funkcją liczby grupowanych jednostek. Zależność ta zaprezentowana jest na rysunku 4.

W całym czasie pracy sieci maksymalna liczba neuronów wyniosła 44. Jest to bardzo niewielka liczba wskazująca, że jeden neuron odwzoruje nawet 170 jednostek w bazie danych. Przez większość czasu pracy sieci liczba neuronów była jeszcze

[8] Aby usunąć neuron należy znaleźć ten, który przekracza maksymalny czas życia, usunąć go i połączyć ze sobą neurony, z którymi połączony był ten usuwany. Aby wstawić nowy neuron należy znaleźć neuron o największym błędzie kwantyzacji i drugi najgorzej dopasowany, wstawić między nie nowy neuron, wyznaczyć dla niego przeciętny błąd kwantyzacji na podstawie błędów neuronów, między które jest wstawiany nowy, zaktualizować macierz czasu życia neuronów (powiększyć jej rozmiar).

Rysunek 4 Zależność czasu uczenia jednej iteracji od liczby grupowanych jednostek



Źródło: opracowanie własne

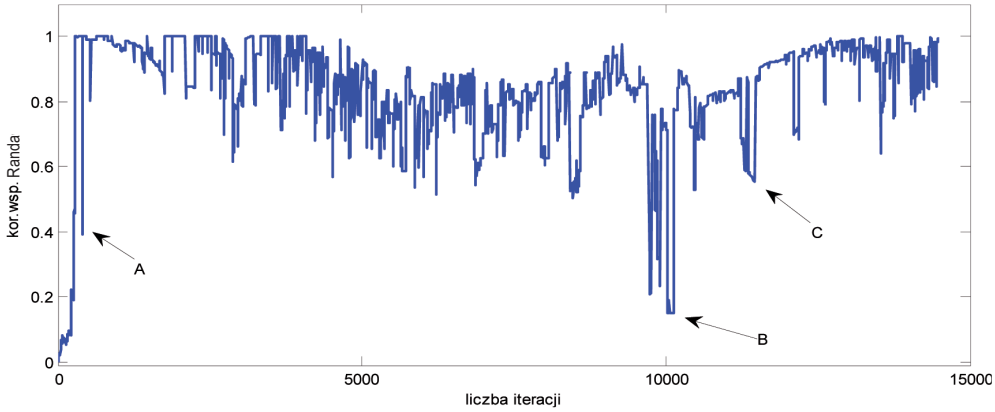
mniejsza i liczyła przeciętnie 33 neurony. Liczba ta w niewielkim tylko stopniu zależy od liczby jednostek w bazie. Obserwując liczbę neuronów w kolejnych iteracjach roboczych (por. rysunek 3) łatwo zaobserwować, że wzrost liczby neuronów zależy w największym stopniu od złożoności samej struktury grupowej obserwowanych jednostek. W początkowej fazie pracy sieci, gdy liczba jednostek rośnie i kształtuje się 5 pierwszych skupień liczba neuronów systematycznie rośnie. Powyżej pięciotysięcznej iteracji liczba jednostek już nie wzrasta i przez pewien czas nie komplikuje się także struktura grupowa. Można wtedy zaobserwować spadek liczby neuronów. Ponowny wzrost liczby neuronów obserwowany w końcowej fazie pracy sieci jest ponownie związany ze zwiększeniem się komplikacji w strukturze grupowej jednostek. Jest to spowodowane zanikaniem początkowych skupień. Sieć przeznacza część neuronów na odwzorowywanie jednostek w starych skupieniach, nawet gdy istnieje tam tylko kilka jednostek, a jednocześnie coraz lepiej odwzorowuje jednostki z nowych dużych skupień.

Interesujące jest kształtowanie się wartości skorygowanego współczynnika Randa w kolejnych iteracjach roboczych[9]. Wartość współczynnika Randa w początkowych iteracjach jest niska, co związanej jest z losowymi początkowymi współrzędnymi neuronów. Z tego powodu sieć wymaga pewnej liczby iteracji, aby zacząć rozpoznawać jednostki. Po około 220 iteracjach neurony przesunęły się w kierunku obserwowanych jednostek i następuje skokowy wzrost wartości współczynnika Randa do 1. Po niewielkiej liczbie iteracji następuje gwałtowny spadek do poziomu 0,4 (oznaczony na rysunku 5 symbolem A) co jest związane z pojawieniem się nowych jednostek w nowym skupieniu. Już po jednej iteracji roboczej sieć nauczyła się rozpoznawać nowe skupienie. Sytuacja taka powtarza się wielokrotnie w całym czasie pracy sieci. Skokowe zmiany wartości współczynnika Randa są zawsze wywołane dużą zmianą w strukturze grupowej jednostek. Szczególnie trudne momenty zostały zaznaczone na

[9] Wartości skorygowanego współczynnika Randa zawierają się w przedziale <0,1>, gdzie 0 oznacza całkowitą niezgodność badanego podziału ze wzorcem a 1 całkowitą zgodność.

rysunku 5 symbolami B i C. We wszystkich tych momentach następuje znaczna zmiana struktury grupowej jednostek. Struktura grupowa w iteracji roboczej oznaczonej literą B jest przedstawiona na rysunku 6.

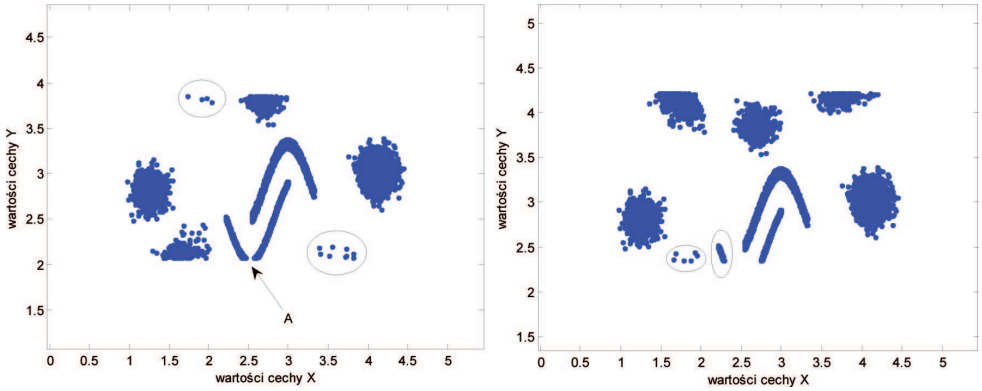
Rysunek 5 Wartości skorygowanego współczynnika Randa



Rysunek 1. Źródło: opracowanie własne

Rysunek 6 Struktura grupowa jednostek po 10000 iteracji

Rysunek 7 Struktura grupowa jednostek po 11000 iteracji



Rysunek 2. Źródło: opracowanie własne

Jest to bardzo trudna struktura grupowa. Jedno ze starych skupień zanika i składa się z zaledwie 10 jednostek (oznaczone elipsą w prawym dolnym rogu na rys.6). Jednocześnie pojawia się nowe skupienie (oznaczone elipsą w górnym lewym rogu na rys.6). W tym samym czasie u-kształtne skupienie w centrum wykresu zaczyna zanikać i rozdziela się na dwa niezależne skupienia (oznaczone symbolem A). Podobnie w momencie oznaczonym na rysunku 5 literą C następuje znacząca zmiana struktury grupowej pokazana na rysunku 7, gdy zanikają jednocześnie dwa skupienia (oznaczone

elipsami). Warto zauważyć, że gdy przez pewną liczbę iteracji struktura grupowa jest względnie stabilna (zmieniają się jednostki, ale nie zmieniają skupienia) sieć szybko poprawia swoją jakość i uzyskuje bardzo dobre wyniki grupowania. Przeciętna wartość współczynnika Randa wynosi w całym eksperymencie 0,85. Przez ponad 10% iteracji roboczych jego wartość jest równa 1. Przez 42% iteracji roboczych wartość współczynnika przekracza 0,9. Można uznać, że są to dobre wyniki grupowania jak na tak złożoną sytuację.

5. PODSUMOWANIE

W badaniach dotyczących zastosowania samouczącej się sieci neuronowej o zmiennej strukturze typu GNG skupiano się dotychczas, na jakości uzyskanej struktury grupowej. Wykazano eksperymentalnie, że o ile skupienia są separowalne sieć tego typu może bezbłędnie określić strukturę grupową jednostek niezależnie od konfiguracji przestrzennej skupień. W badaniach tych podkreślano także wysoką autonomiczność sieci GNG. Nie wymaga ona ustalenia liczby skupień, może pracować w sposób ciągły, ponieważ nie wymaga ustalenia żadnych parametrów zatrzymujących pracę sieci. Wymaga jednak zdefiniowania a priori wartości kilku parametrów. Można to zrobić jedynie eksperymentalnie gdyż nie ma ogólnych, formalnych przesłanek ustalania ich wartości. Sieć ta z samego założenia jest oszczędna w stosunku do sieci o stałej liczbie neuronów, gdyż inteligentnie i dynamicznie dopasowuje swoją strukturę do rzeczywistych potrzeb. Nie prowadzono jednak badań nad szybkością uzyskiwania wyników przez sieć GNG oceniając jedynie, że szybkość ta jest wystarczająca do jej praktycznego stosowania. Ponieważ w grupowaniu dynamicznym szybkość działania jest kluczowym elementem wymaga on znacznie głębszego zbadania. Niestety ze względu na dużą komplikację samego algorytmu samouczenia się sieci GNG wyznaczenie jego złożoności obliczeniowej[10] jest niezwykle trudne. Aby ocenić jej szybkość działania w praktycznych zastosowaniach posłużono się eksperymentem. Nie ma on siły dowodu formalnego, niemniej wskazuje, że w badaniach empirycznych zbliżonych do warunków eksperymentu jest to metoda bardzo szybka. Oceniono wpływ liczby neuronów, liczby i struktury przestrzennej skupień oraz liczby grupowanych jednostek na szybkość uczenia się sieci GNG. Na przeciętnym komputerze[11] sieć GNG może wykonać tysiące iteracji uczących na sekundę nawet w najmniej korzystnych warunkach branych pod uwagę w eksperymencie. Mimo zróżnicowanej liczby skupień istniejących w danym momencie, złożonej struktury grupowej części skupień, wzrost czasu wykonania jednej iteracji, w zależności od liczby jednostek i stopnia komplikacji grupowania jest

[10] Dla niektórych algorytmów istnieje możliwość wyliczenia łącznej liczby elementarnych działań matematycznych (najczęściej dodawania) niezbędnych do jego realizacji. Liczba ta jest nazywana złożonością obliczeniową algorytmu.

[11] Wszystkie obliczenia zostały wykonane na komputerze wyposażonym w procesor: Intel(R) Core(TM) i5 CPU M 450 @ 2.40GHz

w eksperymencie niemal liniowy. Pozwala to ocenić szybkość pracy sieci w innych podobnych badaniach, lecz o innej liczbie grupowanych jednostek.

Warto zauważyć, że w prezentowanych badaniach nie uwzględniono wpływu liczby cech opisujących jedną jednostkę (problem wymiarowości) na szybkość uczenia się sieci. Wydaje się także, że w grupowaniu dynamicznym istnieje możliwość wprowadzenia zmiennych wartości niektórych parametrów sieci GNG (np. czas życia neuronu, liczba iteracji między kolejnymi momentami wprowadzania nowego neuronu do sieci, krok uczenia neuronów), zależnych od bieżącej dynamiki zmian liczby jednostek w bazie i dynamiki zmian ich struktury grupowej. Problemy te będą elementem dalszych badań autora.

Przedstawiony powyżej eksperyment wskazuje, że samoucząca się sieć neuronowa typu GNG spełnia wszystkie wymagania stawiane metodzie grupowania dynamicznego i warto włączyć ją do zbioru procedur stosowanych przez badaczy w analizie skupień, szczególnie w grupowaniu dynamicznym.

LITERATURA

[1] Babcock B., Babu S., Datar M., Motwani R., Widom J. [2002], *Models and issues in data stream systems. In Proceedings of the Twentyfirst ACMSIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems*, June 3-5, Madison, Wisconsin, USA, 1-16, ACM.

[2] Fenn D.J., Porter M.A., Mucha P.J., McDonald M., Williams S., Johnson N.F., Jones N.S. [2010], *Dynamical Clustering of Exchange Rates*, arXiv:0905.4912v2.

[3] Fritzke B. [1994], *Growing cell structures - a self-organizing network for unsupervised and supervised learning*, Neural Networks, 7, 9, 1441-1460.

[4] Guedalia I.D., London M., Werman M. [1999], *An on-line agglomerative clustering method for non-stationary data*, Neural Computation, 11, 2, 521-540.

[5] Guha S., Mishra N., Motwani R., O'Callaghan L. [2000], *Clustering data streams*. In IEEE Symposium on Foundations of Computer Science (FOCS), 359-366.

[6] Kohonen T. [1997], *Self-Organizing Maps*, Springer Series in Information Sciences, Springer-Verlag, Berlin Heidelberg.

[7] Migdał Najman K., Najman K., Data Analysis [2008], *Machine Learning and Applications*, Applying the Kohonen Self-organizing Map Networks to Selecting Variables, Studies in Classification, Data Analysis and Knowledge Organization, Presisach C., Burkhardt H., Schmidt-Thieme L., Decker R., Springer Verlag Berlin Heidelberg, 45-54.

[8] Najman K. [2009], *Zastosowanie nienadzorowanych sieci neuronowych typu Growing Neural Gas w analizie skupień*, Prace Naukowe UE we Wrocławiu, nr 47, 196-205.

[9] Najman K. [2010], *Ocena wpływu parametrów sterujących procesem samouczenia się sieci GNG na ich zdolność do separowania skupień*, Klasyfikacja i analiza danych – teoria i zastosowania Taksonomia 17, Prace Naukowe UE we Wrocławiu nr 17, 296-2010.

[10] Qin A. K. i Suganthan P. N. [2004], *Robust growing neural gas algorithm with application in cluster analysis*, Neural Networks, 17, 8-9, 1135-1148.

[11] Wang X., Smith K. Hyndman R. [1997], *Characteristic-based clustering for time series data*, Data mining and knowledge discovery, 13, 3, 335-364.

[12] Zamir O., Etzioni O. [1999], Grouper: A dynamic clustering interface to web search results, *Proceeding of WWW8*, Toronto, Canada.

GRUPOWANIE DYNAMICZNE Z WYKORZYSTANIEM SIECI GNG

S t r e s z c z e n i e

Od początku lat 90-tych XX wieku obserwuje się stały, dynamiczny wzrost liczby baz danych i zbieranych w nich informacji. Obserwuje się też stały wzrost zapotrzebowania na informacje, a z drugiej strony stały wzrost możliwości ich zbierania i przechowywania. Jedną z własności niektórych baz danych jest ich dynamiczna, zmieniająca się w czasie struktura grupowa. W artykule przedstawiono przegląd podstawowych koncepcji grupowania dynamicznego i zaproponowano jego nową definicję. Wskazano także praktyczną metodę realizacji grupowania dynamicznego opartą na samouczącej się sieci neuronowej typu GNG. Przedstawiono wyniki badań symulacyjnych nad własnościami takiej sieci w grupowaniu dynamicznym.

słowa kluczowe: analiza skupień, grupowanie dynamiczne, sieci GNG

THE DYNAMIC GROUPING USING GNG NEURAL NETWORK

S U M M A R Y

Since early 90s of 20th century has seen a steady and dynamic growth databases and collected information. There has also been a steady increase in demand for information, on the other hand growth in collection and storage information. One of the properties of some databases is their dynamic, changing during the group structure. The article presents an overview of the basic concepts of dynamic grouping and its proposed new definition. It was also a practical method to implement dynamic grouping based on self-learning neural network type of GNG. The results of simulation studies are presented in a dynamic grouping.

keywords: cluster analysis, dynamical clustering, GNG networks

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

ROBERT KAPŁON

MODELE ANALIZY CZYNNIKOWEJ Z DWOMA ZMIENNYMI UKRYTYMI

1. WPROWADZENIE

Model analizy czynnikowej jest często wykorzystywaną techniką analizy danych wielowymiarowych. W zależności od postawionego celu można (por. [16]):wyodrębnić ukrytą, bezpośrednio nieobserwowalną strukturę zmiennych, zredukować wymiar przestrzeni zmiennych wraz z graficzną ich prezentacją.

W metodzie tej, jak i wielu innych, przyjmuje się *implicite* założenie o jednorodności obserwacji. Jeśli istnieją przesłanki poddające w wątpliwość takie założenie, wtedy należy poszukiwać pewnych modyfikacji modelu, aby jak najdokładniej odzwierciedlić strukturę danych. Jedno z podejść oparte jest na mieszaninie rozkładów. Jego istota sprowadza się do włączenia dodatkowej zmiennej ukrytej, która dzieli badaną zbiorowość na klasy. W konsekwencji możliwa jest jednoczesna estymacja parametrów w każdej klasie.

W pracy [7]zaproponowano taki model wraz z procedurą estymacji parametrów. Uwzględniono tam tylko przypadek ogólny, tzn. założono różne macierze ładunków czynnikowych oraz wariancji specyficznej. Jeśli nałoży się na nie pewne ograniczenia, to można otrzymać modele prostsze, które z jednej strony – mogą z prawie jednakową precyzją odwzorowywać strukturę obserwacji co modele bardziej złożone, z drugiej natomiast – mogą wykazywać większą stabilność estymacji. O tej stabilności wspomina się przy okazji szacowania parametrów mieszanki wielowymiarowych rozkładów normalnych (por. [13]).

Warto nadmienić, że w tzw. klasyfikacji opartej na modelach reprezentujących zadaną klasę (*model-basedclustering*), dokonuje się pewnych uproszczeń macierzy kowariancji, jeśli pochodzi ona z wielowymiarowego rozkładu normalnego. Propozycję taką przedstawili Banfield iRafteryw pracy [3], dokonując dekompozycji spektralnej macierzy. W ten sposób dokonali jej reparametryzacjiwyróżniając: orientację, kształt i rozmiar klasy. To pozwoliło na specyfikację 8 modeli. Bazując na tych ustaleniach Celeux i Govaert[5] rozszerzyli tę liczbę do 14.

Powyższe uwagi składają się na cel niniejszego opracowania jakim jest identyfikacja i budowa modeli analizy czynnikowej z dwoma zmiennymi ukrytymi, zależna od postaci macierzy ładunków czynnikowych i macierzy wariancji specyficznej. Ponadto zostanie zaproponowana procedura estymacji dla każdego modelu oraz kryteria, jakimi można się posłużyć, przy wyborze jednego z nich.

2. MODELE ANALIZY CZYNNIKOWEJ

Model analizy czynnikowej uwzględniający niejednorodność obserwacji można przedstawić wpostaci [7]:

$$\mathbf{y}_i = \boldsymbol{\mu}_c + \boldsymbol{\Lambda}_c \mathbf{b}_i + \mathbf{e}_i,$$

gdzie: $[\mathbf{y}_i]_{px1}$ – wektor obserwacji, $[\boldsymbol{\mu}_c]_{px1}$– wektor wartości przeciętnych,$[\boldsymbol{\Lambda}_c]_{pxq}$ – macierz ładunków czynnikowych, $[\mathbf{b}_i]_{qx1}$ – zmienna ukryta, $[\mathbf{e}_i]_{px1}$ – składnik losowy. Przy założeniu, że obserwacja i należy do klasy $K_c (c = 1, \ldots, C)$, rozkłady warunkowe mają postać:

$$\begin{aligned}
(\mathbf{b}_i | i \in K_c) &\sim \mathcal{N}_q(\mathbf{0}, \mathbf{I}), \\
(\mathbf{e}_i | i \in K_c) &\sim \mathcal{N}_p(\mathbf{0}, \boldsymbol{\Psi}_c), \\
(\mathbf{a}_i | \mathbf{b}_i; i \in K_c) &\sim \mathcal{N}_p(\boldsymbol{\mu}_c + \boldsymbol{\Lambda}_c \mathbf{b}_i, \boldsymbol{\Psi}_c),
\end{aligned}$$

$$gdzie : \boldsymbol{\Psi}_c = diag(\sigma_{1c}^2, \ldots, \sigma_{pc}^2).$$

Niech $p(\pi_c)$ będzie prawdopodobieństwem przynależności obserwacji i do klasy K_c. Wtedy rozkład wektora zaobserwowanych odpowiedzi można zapisać ogólnie:

$$f(\mathbf{y}_i) = \sum_{c=1}^{C} \int f(y_i | \mathbf{b}_i, \pi_c) f(\mathbf{b}_i | \pi_c) p(\pi_c) \mathrm{d}\mathbf{b}_i,$$

$$f(\mathbf{y}_i) = \sum_{C}^{\sum} c = 1, \pi_c) f(\mathbf{b}_i | \pi_c) p(\pi_c) \mathrm{d}\mathbf{b}_i$$

lub, przy uwzględnieniu warunkowych rozkładów normalnych, w postaci:

$$f(\mathbf{y}_i) \propto \sum_{c=1}^{C} p(\pi_c) |\boldsymbol{\Gamma}_c|^{-\frac{1}{2}} \exp\left[-\frac{1}{2} (\mathbf{y}_i - \boldsymbol{\mu}_c)^T \boldsymbol{\Gamma}_c^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_c) \right],$$

$$\boldsymbol{\Gamma}_c = \boldsymbol{\Lambda}_c \boldsymbol{\Lambda}_c^T + \boldsymbol{\Psi}_c.$$

W zależności od ograniczeń, jakie nakłada się na składowe macierzy kowariancji $\boldsymbol{\Gamma}_c$, wyróżnia się osiem modeli. W każdym z nich macierz ładunków czynnikowych może być taka sama ($\boldsymbol{\Lambda}_c = \boldsymbol{\Lambda}$) lub różnić się między klasami. W wypadku macierzy wariancji specyficznej, oprócz różnic miedzy klasami, uwzględnić można różnice wewnątrz klas. W konsekwencji pojawiają się 4 możliwości: brak jakichkolwiek ograniczeń – $\boldsymbol{\Psi}_c$, ograniczenia dotyczą wyłącznie różnicwewnątrz klas – $\boldsymbol{\Psi}_c = \boldsymbol{\Psi}$ lub różnic między klasami – $\boldsymbol{\Psi}_c = \sigma_c^2 \mathbf{I}$, wariancja we wszystkich klasach i wewnątrz nich jest identyczna $\boldsymbol{\Psi}_c = \sigma^2 \mathbf{I}$. W tabeli 1 przedstawiono w sposób syntetyczny, na jakie parametry nałożono ograniczenia.

Tabela 1.

Ograniczenia macierzy ładunków i wariancji

Ograniczenia na Λ_c	**Ograniczenia naΨ_c**	
	Między klasami	**Wewnątrz klas**
U	U	U
C	U	U
U	C	U
C	C	U
U	U	C
C	U	C
U	C	C
C	C	C

C – ograniczenie, U – brak ograniczenia.
Źródło: opracowanie własne.

3. ESTYMACJA PARAMETRÓW MODELI

W wypadku mieszaniny rozkładów, z jaką mamy tutaj do czynienia, do estymacji parametrów wykorzystuje się najczęściej algorytm **EM.** Jest to iteracyjna procedura, w której wyróżnia się dwa zasadnicze kroki [11]. W pierwszym oblicza się warunkową wartość oczekiwaną logarytmu funkcji wiarygodności dla kompletnego zbioru danych:

$$Q(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = \mathbb{E}\left[\log L_c\left(\boldsymbol{\Theta}|\mathbf{Y}, \mathbf{B}, \mathbf{Z}\right)\Big|\mathbf{Y}, \boldsymbol{\Theta}^{(t)}\right].$$

Zbiór ten nazywamy kompletnym, po wprowadzeniu dodatkowej zmiennej z_{ic}, która ma rozkład wielomianowy indeksowany parametrem $p(\boldsymbol{\pi}) = (p(\pi_c), \ldots, p(\pi_c))$. Przyjmuje ona wartość 1, gdy obserwacja i pochodzi z klasy K_c, lub 0 w przeciwnym wypadku.

Dla modelu analizy czynnikowej odpowiednie funkcje Q mają postać (por. [7]):

$$Q_{p(\boldsymbol{\pi})}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}p(\pi_c) \tag{1}$$

$$Q_{\boldsymbol{\mu}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\left[\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\mathbf{y}_i - \frac{1}{2}\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c - \boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)})\right] \tag{2}$$

$$Q_{\boldsymbol{\Lambda}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)} - \frac{1}{2}tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})\right] \tag{3}$$

$$Q_{\boldsymbol{\Psi}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\Big[-\log|\boldsymbol{\Psi}_c^{-1}| + (\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_c) - \\ -2(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)} + tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})\Big] \tag{4}$$

gdzie:

$$\tau_{ic}^{(t)} = \frac{p^{(t)}(\pi_c)|\boldsymbol{\Lambda}_c^{(t)}\boldsymbol{\Lambda}_c^{(t)T} + \boldsymbol{\Psi}_c^{(t)}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}M_{ic}^{(t)}\right]}{\sum_c p^{(t)}(\pi_c)|\boldsymbol{\Lambda}_c^{(t)}\boldsymbol{\Lambda}_c^{(t)T} + \boldsymbol{\Psi}_c^{(t)}|^{-\frac{1}{2}}\exp\left[-\frac{1}{2}M_{ic}^{(t)}\right]},$$
$$M_{ic}^{(t)} = \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t)}\right)^T(\boldsymbol{\Lambda}_c^{(t)}\boldsymbol{\Lambda}_c^{(t)T} + \boldsymbol{\Psi}_c^{(t)})^{-1}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t)}\right),$$

natomiast $[\mathbf{v}_{ic}^{(t)}]_{qx1} = \mathbb{E}(\mathbf{b}_i|\mathbf{y}_i, z_{ic})$ oraz $[\mathbf{W}_{ic}^{(t)}]_{qxq} = \mathbb{E}(\mathbf{b}_i\mathbf{b}_i^T|\mathbf{y}_i, z_{ic})$.

W drugim kroku maksymalizuje się funkcję Q ze względu na nieznane parametry $\boldsymbol{\Theta}$,

$$\boldsymbol{\Theta}^{(t+1)} = \arg\max_{\boldsymbol{\Theta}} Q\left(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)}\right), \tag{5}$$

przyjmując, że $\boldsymbol{\Theta}^{(t)}$ są znane, gdyż zostały oszacowane w iteracji (t).

Rozwiązanie (5) będzie różniło się miedzy modelami. Z tego też względu każdy z nich zostanie rozpatrzony osobno. Ponieważ nie wprowadza się żadnych ograniczeń ze względu na prawdopodobieństwa przynależności do klas oraz wektor wartości średnich, więc proces ich szacowania sprowadza się do jednego, ogólnego przypadku.

Wprowadzając mnożnik Lagrange'a do funkcji (1) nietrudno pokazać, że

$$p(\pi_c)^{(t+1)} = \frac{\sum_{i=1}^{n}\tau_{ic}^{(t)}}{n}. \tag{6}$$

W wypadku wartości przeciętnych należy zróżniczkować funkcję (2). Ponieważ

$$\frac{\partial(\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\mathbf{y}_i)}{\boldsymbol{\mu}_c} = \boldsymbol{\Psi}_c^{-1}\mathbf{y}_i, \quad \frac{\partial(\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c)}{\partial\boldsymbol{\mu}_c} = 2\boldsymbol{\Psi}_c^{-1}\boldsymbol{\mu}_c, \quad \frac{\partial(\boldsymbol{\mu}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)})}{\boldsymbol{\mu}_c} = \boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)},$$

i z założenia, że $\partial Q_{\boldsymbol{\mu}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)})/\partial\boldsymbol{\mu}_c = 0$, więc estymator wartości średnich w iteracji $(t+1)$ ma postać:

$$\boldsymbol{\mu}_c^{(t+1)} = \sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\Lambda}_c^{(t)}\mathbf{v}_{ic}^{(t)}\right)\left(np(\pi_c)^{(t+1)}\right)^{-1}. \tag{7}$$

Oczywiście, nałożone ograniczenia na macierze ładunków czynnikowych i wariancji specyficznej należy uwzględnić we wzorze (6) i (7).

Model UUU:

w modelu tym nie przyjmuje się żadnych ograniczeń na macierze ładunków czynnikowych i wariancji specyficznej. Odpowiednie składowe pochodnej funkcji (3) mają postać:

$$\frac{\partial(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)}}{\partial\boldsymbol{\Lambda}_c} = \boldsymbol{\Psi}_c^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_c)\mathbf{v}_{ic}^{(t)T}, \quad \frac{\partial tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})}{\partial\boldsymbol{\Lambda}_c} = 2\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)}. \quad (8)$$

Rozwiązując równanie $\partial Q_{\boldsymbol{\Lambda}}(\boldsymbol{\Theta}|\boldsymbol{\Theta}^{(t)})/\partial\boldsymbol{\Lambda}_c = 0$, otrzymuje się w iteracji $(t+1)$:

$$\boldsymbol{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\mathbf{v}_{ic}^{(t)T}\left(\sum_{i=1}^{n}\tau_{ic}^{(t)}\mathbf{W}_{ic}^{(t)}\right)^{-1}$$

Różniczkując z kolei składowe funkcji (4) ze względu na odwrotną macierz wariancji specyficznej otrzymano:

$$\frac{\partial\log|\boldsymbol{\Psi}_c^{-1}|}{\partial\boldsymbol{\Psi}_c^{-1}} = \boldsymbol{\Psi}_c, \quad \frac{\partial(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}(\mathbf{y}_i - \boldsymbol{\mu}_c)}{\partial\boldsymbol{\Psi}_c^{-1}} = (\mathbf{y}_i - \boldsymbol{\mu}_c)(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T),$$
$$\frac{\partial(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{v}_{ic}^{(t)}}{\partial\boldsymbol{\Psi}_c^{-1}} = (\mathbf{y}_i - \boldsymbol{\mu}_c)\mathbf{v}_{ic}^{(t)T}\boldsymbol{\Lambda}_c^T, \quad \frac{\partial tr(\boldsymbol{\Lambda}_c^T\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)})}{\partial\boldsymbol{\Psi}_c^{-1}} = \boldsymbol{\Lambda}_c\mathbf{W}_{ic}^{(t)T}\boldsymbol{\Lambda}_c^T.$$

Po rozwiązaniu odpowiedniego równania estymator największej wiarygodności w iteracji $(t+1)$ można zapisać w postaci:

$$\boldsymbol{\Psi}_c^{(t+1)} = \frac{1}{np(\pi_c)^{(t+1)}}diag\left[\sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\boldsymbol{\Lambda}_c^{(t+1)T}\right)\right]$$

Model CUU:

nakłada się graniczenie na macierz ładunków czynnikowych, czyli $\boldsymbol{\Lambda}_c = \boldsymbol{\Lambda}$. Wykorzystując funkcję (3) oraz obliczone wcześnie pochodne (8) otrzymuje się równanie, które jest podstawą do oszacowania ładunków czynnikowych:

$$\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\boldsymbol{\Psi}_c^{-1}\mathbf{P}_{ic}^{(t)} = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\boldsymbol{\Psi}_c^{-1}\boldsymbol{\Lambda}\mathrm{W}_{ic}^{(t)},$$
$$\mathbf{P}_{ic}^{(t)} = (\mathbf{y}_i - \boldsymbol{\mu}_c)\,\mathbf{v}_{ic}^{(t)T}.$$

Niech wiersz j macierzy $\mathbf{P}_{ic}^{(t)}$ zostanie oznaczony w postaci $\mathbf{P}_{ic}^{(t)}[j,]$, wtedy powyższe równanie, dla wiersza j, można zapisać:

$$\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2}\mathbf{P}_{ic}^{(t)}[j,] = \boldsymbol{\Lambda}[j,]\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2}\mathbf{W}_{ic}^{(t)},$$

i tym samym wyznaczyć w kolejnej iteracji wartości:

$$\mathbf{\Lambda}^{(t+1)}[j,] = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2(t)}\mathbf{P}_{ic}^{(t)}[j,]\left(\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_{jc}^{-2(t)}\mathbf{W}_{ic}^{(t)}\right)^{-1}.$$

W wypadku estymatora wariancji specyficznej można bezpośrednio, po niewielkiej korekcie ze względu na macierz ładunków, wykorzystać wynik otrzymany dla modelu *UUU*:

$$\mathbf{\Psi}_c^{(t+1)} = \frac{1}{np(\pi_c)^{(t+1)}}diag\left[\sum_{i=1}^{n}\tau_{ic}^{(t)}\left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\mathbf{\Lambda}^{(t+1)T}\right)\right]$$

Model CUC:

konsekwencją nałożonych ograniczeń jest $\mathbf{\Lambda}_c = \mathbf{\Lambda}$ oraz $\mathbf{\Psi}_c = \sigma_c^2\mathbf{I}$. Wykorzystując (3) i (8) otrzymano równanie:

$$\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2}(\mathbf{y}_i - \boldsymbol{\mu}_c)\mathbf{v}_{ic}^{(t)T} = \mathbf{\Lambda}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2}\mathbf{W}_{ic}^{(t)},$$

z którego wyznaczono:

$$\mathbf{\Lambda}^{(t+1)} = \sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2(t)}(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)})\mathbf{v}_{ic}^{(t)T}\left(\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(t)}\sigma_c^{-2(t)}\mathbf{W}_{ic}^{(t)}\right)^{-1}.$$

W wypadku wariancji przekształcono funkcję (4) do postaci:

$$\begin{aligned} Q_{\mathbf{\Psi}}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)}) = -\frac{1}{2}\sum_{i=1}^{n}\sum_{c=1}^{C}\tau_{ic}^{(k)}\Big[-p\log\sigma_c^{-2} + \sigma_c^{-2}(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)(\mathbf{y}_i - \boldsymbol{\mu}_c) - \\ -2\sigma_c^{-2}(\mathbf{y}_i^T - \boldsymbol{\mu}_c^T)\mathbf{\Lambda}_c\mathbf{v}_{ic}^{(t)} + \sigma_c^{-2}tr(\mathbf{\Lambda}_c^T\mathbf{\Lambda}_c\mathbf{W}_{ic}^{(t)})\Big]. \end{aligned} \tag{9}$$

Ponieważ $\partial Q_{\mathbf{\Psi}_c}(\mathbf{\Theta}|\mathbf{\Theta}^{(t)})/\partial\sigma_c^{-2} = 0$, więc w iteracji $(t+1)$ estymator wariancji ma postać:

$$\sigma_c^{2(t+1)} = \frac{1}{p\cdot n\cdot p(\pi_c)^{(t+1)}}\sum_{i=1}^{n}\tau_{ic}^{(t)}\left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T})\left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}^{(t+1)}\mathbf{v}_{ic}^{(t)}\right) + tr(\mathbf{\Lambda}^{(t+1)T}\mathbf{\Lambda}^{(t+1)}\mathbf{W}_{ic}^{(t)})\right].$$

Model CCU:

ograniczenia dotyczą obu macierzy, tzn. $\mathbf{\Lambda}_c = \mathbf{\Lambda}$ oraz $\mathbf{\Psi}_c = \mathbf{\Psi}$. Bazując na wynikach otrzymanych dla modelu bez ograniczeń (*UUU*), macierz ładunków czynnikowych ma postać:

$$\mathbf{\Lambda}^{(t+1)} = \sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)}\right)^{-1},$$

podczas gdy macierz wariancjiprzedstawia się następująco:

$$\mathbf{\Psi}^{(t+1)} = \frac{1}{n} diag\left[\sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\mathbf{\Lambda}^{(t+1)T}\right)\right].$$

Model UCU:

macierz wariancji specyficznej jest taka sama w każdej klasie, a więc $\mathbf{\Psi}_c = \mathbf{\Psi}$, natomiast na macierz ładunków czynnikowych nie nakłada się żadnych ograniczeń. Z tego względu postać estymatora ładunków czynnikowych będzie taka sama jak w modelu *UUU*:

$$\mathbf{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)}\right)^{-1}.$$

Z kolei estymator wariancji specyficznej:

$$\mathbf{\Psi}^{(t+1)} = \frac{1}{n} diag\left[\sum_{i=1}^{n}\sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}\right)\left(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T} - \mathbf{v}_{ic}^{(t)T}\mathbf{\Lambda}_c^{(t+1)T}\right)\right].$$

Model UUC:

ograniczenie w tym modelu dotyczy tylko macierzy wariancji specyficznej: $\mathbf{\Psi}_c = \sigma_c^2\mathbf{I}$. Ze względu na podobieństwo do szacowania parametrów modelu *CUC*, wystarczy drobna modyfikacja, by zaadoptować tamto rozwiązanie i otrzymać:

$$\mathbf{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n} \tau_{ic}^{(t)} (\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)}\right)^{-1},$$

$$\sigma_c^{2(t+1)} = \frac{1}{p \cdot np(\pi_c)^{(t+1)}} \sum_{i=1}^{n} \tau_{ic}^{(t)} \left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T})\left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}_c^{(t+1)}\mathbf{v}_{ic}^{(t)}\right) + tr(\mathbf{\Lambda}_c^{(t+1)T}\mathbf{\Lambda}_c^{(t+1)}\mathbf{W}_{ic}^{(t)})\right].$$

Model UCC:

jest to model podobny do wcześniejszego (*UUC*) z tą różnicą, że wariancja specyficzna jest taka sama wewnątrz jak i między klasami: $\mathbf{\Psi}_c = \sigma^2\mathbf{I}$. Wzór na oszacowanie ładunków czynnikowych będzie więc identyczny jak w modelu *UUU*:

$$\mathbf{\Lambda}_c^{(t+1)} = \sum_{i=1}^{n} \tau_{ic}^{(t)} (\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)} \right)^{-1},$$

natomiast estymator wariancji, po koniecznej modyfikacji ma postać:

$$\sigma^{2(t+1)} = \frac{1}{p \cdot n} \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T}) \left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}_c^{(t+1)} \mathbf{v}_{ic}^{(t)} \right) + tr(\mathbf{\Lambda}_c^{(t+1)T} \mathbf{\Lambda}_c^{(t+1)} \mathbf{W}_{ic}^{(t)}) \right].$$

Model CCC:

ostatni z rozważanych modeli jest najmniej rozbudowany, w sensie liczby parametrów, gdyż nałożone zostały wszystkie możliwe ograniczenia: $\mathbf{\Lambda}_c = \mathbf{\Lambda}, \mathbf{\Psi}_c = \sigma^2\mathbf{I}$. Wykorzystującwyniki otrzymane dla poprzednich modeli otrzymano:

$$\mathbf{\Lambda}^{(t+1)} = \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \left(\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)} \right) \mathbf{v}_{ic}^{(t)T} \left(\sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \mathbf{W}_{ic}^{(t)} \right)^{-1},$$

$$\sigma^{2(t+1)} = \frac{1}{p \cdot n} \sum_{i=1}^{n} \sum_{c=1}^{C} \tau_{ic}^{(t)} \left[(\mathbf{y}_i^T - \boldsymbol{\mu}_c^{(t+1)T}) \left((\mathbf{y}_i - \boldsymbol{\mu}_c^{(t+1)}) - 2\mathbf{\Lambda}^{(t+1)} \mathbf{v}_{ic}^{(t)} \right) + tr(\mathbf{\Lambda}^{(t+1)T} \mathbf{\Lambda}^{(t+1)} \mathbf{W}_{ic}^{(t)}) \right].$$

Iteracyjny charakter procedury estymacji powoduje, że w każdym kolejnym kroku t zwiększa się o 1. Ponieważ takiego postępowania nie można prowadzić w nieskończoność, dlatego należy zaproponować kryterium stopu, kiedy to dodatkowy krok estymacji nie przynosi istotnych korzyści. Wśród propozycji można wskazać podejścia bazujące na różnicy w wartościach estymowanych parametrów lub wartościach samej funkcji. Jeśli owe różnice będą mniejsze od założonego progu, wtedy procedurę należy przerwać (por. [10]).

4. WYBÓR MODELU

Podstawowe pytanie dotyczące kwestii wyboru modelu sprowadza się do rozstrzygnięcia, czy model w którym nałożono pewne ograniczenia na parametry nie będzie znacznie odbiegał – w sensie dopasowania do danych – od modelu, w którym tych ograniczeń nie ma, lub jest ich mniej. W przypadku tzw. modeli hierarchicznych, można wykorzystać test oparty na ilorazie funkcji wiarygodności. Gdy nie są one hierarchiczne, to wobec niespełnienia warunków regularności, takie podejście jest niewłaściwe, gdyż prowadzi do nieznajomości rozkładu statystyki testowej (por. [12]).

Można próbować aproksymować ten rozkład wykorzystując podejście bootstrapowe [14], lecz czasochłonność obliczeń sprawia, że poszukuje się innych rozwiązań.

Wśród nich znajdują się kryteria informacyjne, które można sklasyfikować w następujące grupy: kryteria ze szkoły Akaike, kryteria bayesowskie i kryteria klasyfikacyjne. W zależności od wykorzystywanej klasy weryfikowanych modeli, ich wiarygodność się zmienia. Przykładowo, w wielomianowym modelu logitowym opartym na mieszkankach rozkładów kryterium informacyjne Akaike *AIC* jest bardziej wiarygodne od kryterium bayesowskiego *BIC* [2], natomiast dla mieszkanek rozkładów normalnych jest odwrotnie [12].

W pracy [8] przeprowadzonotakie badania symulacyjne dotyczące analizy czynnikowej z dwoma zmiennymi ukrytymi. Okazało się, że największą trafnością co do wyboru właściwego modelu odznaczały się dwa kryteria: *ICL BIC* oraz *BIC*. Trochę gorzej wypadło kryterium zgodne *AIC* i kryterium *AIC*. Pozostałe z rozważanych kryteriów (*NEC, CLC*) wykazały się dość mała trafnością. Choć badania te dotyczyły modelu *UCU*, to jednak warto skorzystać z tych wyników i skupić się tylko na czterech kryteriach:

- kryterium Akaike [1]:

$$AIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + 2r,$$

- zgodne kryterium Akaike [4]:

$$CAIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + r\log n + r,$$

- kryterium bayesowskie [9]:

$$BIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + r\log n,$$

- łączne kryterium bayesowskie i klasyfikacyjne [12]:

$$ICLBIC = -2\log L\left(\hat{\boldsymbol{\Theta}}\right) + 2EN\left(\hat{\boldsymbol{\tau}}\right) + r\log n,$$

$$EN\left(\hat{\boldsymbol{\tau}}\right) = -\sum_{i=1}^{n}\sum_{c=1}^{C}\hat{\boldsymbol{\tau}}_{ic}\log\hat{\boldsymbol{\tau}}_{ic},$$

gdzie r jest liczbą szacowanych parametrów, natomiast $L\left(\hat{\boldsymbol{\Theta}}\right)$ funkcją wiarygodności. Spośród konkurencyjnych modeli wybiera się ten, dla którego obliczone kryterium informacyjne ma najmniejszą wartość. Należy odnotować, że opisywane modele mogą mieć różne warianty zależnie od liczby klas C oraz liczby czynników q. Nie zmienia to jednak procedury postępowania przy wyborze.

5. PRZYKŁAD EMPIRYCZNY – DANE O SATYSFAKCJI

Aby zilustrować użyteczność omawianych modeli, wykorzystano rzeczywiste dane – zaczerpnięte z pracy [6] – odnoszące się do opinii klientów wyrażonej na temat

1. Jakość produktów
2. Aktywność e-commerce
3. Wsparcie techniczne
4. Rozwiązywanie skarg
5. Reklama
6. Linia produktowa
7. Wizerunek
8. Konkurencyjność cen
9. Gwarancje
10. Nowe produkty
11. Zamówienia i faktury
12. Elastyczność cen
13. Szybkość dostawy

współpracy z pewną firmą działającą w branży przemysłowej. Ocenie poddano 13 zmiennych, wykorzystując jedenastostopniową skalę.

Do oceny porównawczej, rozważanych w części teoretycznej modeli, włączono dodatkowo wariant model *CCU*, który powstaje poprzez nałożenie ograniczeń na wartości średnie. W konsekwencji średnie nie będą różniły się między klasami. Tak sformułowany model można uznać za klasyczną postać modelu analizy czynnikowej. Jest to więc dobry punkt odniesienia, gdyż taki model jest zaimplementowany w każdym pakiecie statystycznym (np. SPSS, SAS, STATISTICA, R), co z kolei przekłada się na częste jego wykorzystanie. Wydaje się zasadne, z punktu widzenia praktycznych zastosowań, porównanie takiego modelu z modelami rozważanymi w pracy.

Procedura estymacji każdego modelu wymaga określenia *ex ante* liczby klas oraz liczby ładunków czynnikowych. Ponieważ liczba estymowanych parametrów jest ściśle powiązana z liczbą klas, a rozmiar próby jest niezbyt duży (200 obserwacji), dlatego przyjęto maksymalną liczbę klas na poziomie dwóch. W wypadku ładunków czynnikowych rozważano modele z 2, 3 i 4 czynnikami. Odpowiednie programy szacujące parametry modeli napisano w środowisku **R** [15].

W tabeli 2 podano wartości logarytmów funkcji wiarygodności oraz kryteriów informacyjnych. Dzięki temu można dokonać formalnego rozstrzygnięcia co do postaci modelu. Nie zamieszczono dwóch modeli: *CUC* i *UCC*. Wynika to z prostej obserwacji, że model bardziej rozbudowany*UUC* jest jednym z najgorszych w sensie wartości kryteriów informacyjnych.

Pierwsze spostrzeżenie jakie nasuwa się w odniesieniu do tabeli 2 to bardzo wysokie wartości kryteriów informacyjnych modelu klasycznego. Również i logarytm funkcji wiarygodności znacznie odstaje od pozostałych. To utwierdza w przekonaniu, że model ten jest nieodpowiedni, dlatego wyboru należy poszukiwać wśród modeli bardziej złożonych.

Przyjęcie ograniczenia na wartości wariancji specyficznej wewnątrz klas – skutkującetym, że każda zmienna ma taką samą wariancję (choć może się ona różnić między klasami) – powoduje gorsze dopasowanie takich modeli do danych w stosunku do pozostałych modeli. Szczególnie może zaskakiwać model *UUC*, w którym wartości średnie i ładunki czynnikowe szacowane są dla każdej z dwóch klas; po-

Tabela 2.

Wartości funkcji wiarygodności i kryteriów informacyjnych w zależności od modelu

Klasyczny	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-4690	-4605	-4598
AIC	9417	9271	9267
BIC	9581	9538	9577
CAIC	9619	9600	9649
ICL BIC	-	-	-

UCU	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-3156	-3045	-2891
AIC	6401	6202	5915
BIC	6788	6683	6482
CAIC	6878	6795	6614
ICL BIC	6807	6689	6485

CCC	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-3711	-3606	-3489
AIC	7475	7275	7051
BIC	7703	7551	7369
CAIC	7756	7615	7443
ICLBIC	7717	7559	7374

UUC	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-3670	-3540	-3418
AIC	7418	7182	6958
BIC	7758	7616	7478
CAIC	7837	7717	7599
ICLBIC	7765	7620	7482

CCU	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-3207	-3084	-2957
AIC	6478	6244	6000
BIC	6758	6571	6370
CAIC	6823	6647	6456
ICL BIC	6773	6585	6374

CUU	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-3204	-3069	-2955
AIC	6487	6226	6009
BIC	6822	6609	6435
CAIC	6900	6698	6534
ICL BIC	6883	6611	6438

UUU	Liczba ładunków czynnikowych		
	2	3	4
log$L$	-3126	-2974	-2865
AIC	6355	6073	5875
BIC	6798	6611	6498
CAIC	6901	6736	6643
ICL BIC	6814	6617	6499

Źródło: Opracowanie własne.

dobnie wariancja. Pomimo swojej elastyczności wartości kryteriów informacyjnych jak i logarytmy funkcji wiarygodności znacznie odstają od pozostałych modeli.
Najbardziej ogólny model *UUU*, na który nie nałożono żadnych ograniczeń, cechuje się najniższą wartością kryterium *AIC*. Biorąc pod uwagę pozostałe kryteria – pamiętając jednocześnie o tendencji kryterium *AIC* do wyboru modeli zbytnio rozbudowanych oraz wynikach badań symulacyjnych (wspomnianych w pkt. 4) – modelem najbardziej preferowanym byłby model *CCU* o czterech czynnikach wspólnych.
Wartości oszacowanych ładunków czynnikowych zamieszczono w tabeli 3. Kierując się rekomendacjami w zakresie ich praktycznej przydatności (por. [6]) przyjęto, że minimalna wartość ładunku dla danej zmiennej powinna być nie mniejsza niż 0,4. Dopiero wtedy taka zmienna może współtworzyć dany czynnik.

Tabela 3.

Wartości ładunków czynnikowych

	Czynnik			
Zmienna	**1**	**2**	**3**	**4**
Linia produktowa	0,96	0,17	-0,08	0,10
Szybkość dostawy	0,83	0,13	0,52	0,10
Rozwiązywanie skarg	0,74	0,12	0,46	0,10
Zamówienia i faktury	0,59	0,16	0,47	0,14
Elastyczność cen	0,32	-0,03	0,94	0,06
Reklama	0,24	0,58	0,14	-0,02
Wizerunek	0,20	0,97	0,06	0,10
Aktywność e-commerce	0,16	0,77	0,05	0,04
Gwarancje	0,09	0,10	-0,02	0,99
Nowe produkty	0,06	0,03	0,17	0,01
Wsparcie techniczne	0,05	0,03	0,00	0,84
Konkurencyjność cen	-0,01	-0,01	-0,03	0,01
Jakość produktów	-0,04	0,19	0,09	0,00

Źródło: Opracowanie własne.

Interpretacja czynnika 2 i 4 nie nastręcza większych problemów, gdyż można byłoby te wymiary opisać odpowiednio jako: marketing i serwis. Problematyczne staje się jednak nazwanie czynnika 1 i 3, gdyż trzy zmienne mają wysokie wartości ładunków na każdym z tych czynników. Zmienne które wyraźnie różnicują czynniki to linia produktowa i elastyczność cen; tutaj też ładunki są najwyższe. Być może

pewnym rozwiązaniem byłoby utożsamienie czynników z tymi zmiennymi. Jest to jeden ze sposobów interpretacji czynników (por. [6]).

6. PODSUMOWANIE

Przedstawione w pracy modele analizy czynnikowej wraz z procedurą estymacji parametrów pozwalają, przynajmniej w założeniu, rozszerzyć instrumentarium badawcze o dodatkowe modele. Modele te cechuje większa elastyczność, co może być szczególnie przydatne w analizie niejednorodnych zbiorów danych.
Jak można było zaobserwować na przykładzie danych o satysfakcji, klasyczny model analizy czynnikowej nie był odpowiedni, jeśli decyzję oprzeć na kryteriach informacyjnych. Zauważono również, znaczne obniżenie wartości informacyjnej modelu, jeśli przyjęto (w ramach rozważanych 8 modeli) równą wariancję specyficzną dla każdej zmiennej. Sytuacja nie uległa poprawie, jeśli wariancje różniły się między klasami.
Wydaje się, że egzemplifikacja proponowanych modeli z wykorzystaniem większych zbiorów danych, cechujących się większą niejednorodnością powinna stanowić podstawę dalszych badań.

Politechnika Wrocławska

LITERATURA

[1] Akaike H. (1973), *Information theory and anextension of the maximum likelihoodprinciple,* [w:] Petrov B.N., Csaki F. (Eds.), Second international symposium on information theory (pp.). Budapest: AcademiaiKiado, s. 267-281.
[2] Andrews L., Currim I.S. (2003), *A Comparison of segment retention criteria for finite mixture logit models*, Journal of Marketing Research, **40**(2), s. 235-243.
[3] Banfield, J.D.,Raftery, A.E. (1993). *Model-based Gaussian and non-Gaussian clustering,* Biometrics, **49**, s. 803-821.
[4] Bozdogan, H. (1987), *Model selection and Akaike's information criterion (AIC): The general theory and its analytical extensions*. Psychometrika, **52**, s. 345-370.
[5] Celeux G., Govaert G. (1995), *Gaussian Parsimonious Clustering Models*, Pattern Recognition, **28**(5), s. 781-793.
[6] Hair J.F., Black W.C., Babin B.J., Anderson R.E., (2010), *Multivariate Data Analysis,* Prentice Hall, New York.
[7] Kapłon R. (2004), *Estymacja parametrów modelu czynnikowego wykorzystującego klasy ukryte*, [w:] Jajuga K., Walesiak M. „Klasyfikacja i analiza danych – teoria i zastosowania". Taksonomia nr 11, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, s. 204-211.
[8] Kapłon R. (2007), *O liczbie klas w modelu analizy czynnikowej z dwoma zmiennymi ukrytymi,* [w:] Jajuga K., Walesiak M. „Klasyfikacja i analiza danych – teoria i zastosowania". Taksonomia nr 14, Prace Naukowe Akademii Ekonomicznej we Wrocławiu, s. 253-260.
[9] Kass R. E., Raftery A. E. (1995), *Bayes factors,*Journal of the American Statistical Association, **90**, s. 773-795.
[10] Luenberger D.G., Ye Y. (2008), *Linear and Nonlinear Programming*, Springer.
[11] McLachlan G.J., Krishnan T. (1997), *The EM Algorithm and Extensions*. New York: Wiley.

[12] McLachlan G.J., Peel. D. (2000), *Finite Mixture Models*, New York: Wiley.
[13] McLachlan, G.J, Basford K. (1988),*Mixturemodels*. New York: Marcel Dekker.
[14] McLachlan, G.J. (1987), *On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture,Journal of the Royal Statistical Society Series C (Applied Statistics),* **36**, s. 318-324.
[15] R Development Core Team (2011), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.
[16] Walesiak M. (1996), *Metody analizy danych marketingowych,* Wydawnictwo Naukowe PWN, Warszawa.

MODELE ANALIZY CZYNNIKOWEJ Z DWOMA ZMIENNYMI UKRYTYMI

S t r e s z c z e n i e

Celem analizy czynnikowej jest redukcja zmiennych poprzez ich zastąpienie mniejszą liczbą czynników, które traktowane są jako konstrukty lub zmienne ukryte. Niestety, jeśli mamy do czynienia z niejednorodnym zbiorem danych, estymacja jednego zbioru wartości średnich, ładunków czynnikowych czy wariancji specyficznych może prowadzić do błędnych wniosków. Jednym ze sposobów radzenia sobie z tą niejednorodnością jest wprowadzenie dodatkowej zmiennej ukrytej do modelu analizy czynnikowej. W konsekwencji zakłada się, że obserwacje pochodzą z dwóch lub więcej podpopulacji, których struktura jest nieznana.
W zależności od ograniczeń nałożonych na macierze ładunków czynnikowych i wariancji specyficznej, można otrzymać różne warianty modelu. W pracy przedstawiono 8 modeli analizy czynnikowej, wraz z propozycją procedury estymacji parametrów algorytmem **EM**. Zwrócono również uwagę na problem w wykorzystaniu testów statystycznych, opartych na ilorazie wiarygodności, wskazując na kryteria informacyjne jako alternatywne podejście. Proponowane podejście zilustrowano przykładem, wykorzystując w tym celu rzeczywiste dane dotyczące satysfakcji.

FACTOR ANALYSIS MODELS WITH TWO LATENT VARIABLES

A b s t r a k t

The goal of factor analysis is to reduce the redundancy among variables by using smaller number of factors that are treated as constructs or latent variables. Unfortunately, if we face with data heterogeneity, the estimates of a single set of means, factor loadings and specific variances may be misleading.One way of accounting for unobserved heterogeneity is to include another latent variable in a factor analysis model. As a consequence, the observations in a samples are assumed to arise from two or more subpopulations that are mixed in unknown proportions.
Since putting some restrictions on parameters such as factor loadings and specific variancesone can get more parsimonious models. Therefore, the purpose of this paper is to present the eight factor analysis models. Methods of optimization to derive the maximum likelihood estimates based on EM algorithm as well as model selection procedure are considered. Proposed approach is illustrated by using a set of data referring to preferences.

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

JADWIGA KOSTRZEWSKA

INTERPRETACJA W MODELACXCH TOBITOWYCH

1. WPROWADZENIE

Model tobitowy został wprowadzony w 1958 roku przez Jamesa Tobina (por. [24]) i służy do opisu zmiennej zależnej cenzurowanej. Model ten można traktować jako ciągłe uogólnienie modelu probitowego przez włączenie analizy regresji do opisu obserwowalnej części zmiennej zależnej. W kolejnych latach po pojawieniu się artykułu Jamesa Tobina w literaturze przedmiotu powstało wiele modeli wywodzących się od modelu tobitowego, w tym modele selekcji wprowadzone przez Jamesa Heckmana (por. [6], [7], [8]). W 1984 r. Takeshi Amemiya (por. [1]) dokonał klasyfikacji modeli występujących w literaturze przedmiotu wyodrębniając pięć głównych typów modeli opisujących zmienną zależną cenzurowaną[1]. Kryterium podziału była postać funkcji wiarygodności. T. Amemiya klasyfikowane modele ogólnie nazwał modelami tobitowymi. Model wprowadzony przez Jamesa Tobina określił mianem standardowego modelu tobitowego lub modelu tobitowego typu I. Model selekcji oryginalnie rozważany przez J. Heckmana to model tobitowy typu III.

Standardowy model tobitowy stosowany był przez Jamesa Tobina do opisu wydatków na dobra luksusowe (por. [24]). Modele zmiennej zależnej cenzurowanej, a w szczególności modele selekcji, rozwijały się m.in. w analizie zagadnień związanych z rynkiem pracy – co zawdzięczamy w dużej mierze Jamesowi Heckmanowi[2]. Znane są takie zastosowania modeli zmiennej cenzurowanej jak: badanie (rocznej) liczby godzin pracy zamężnych kobiet z uwzględnieniem samoselekcji do grupy pracujących, analiza wysokości wynagrodzeń z uwzględnieniem samoselekcji do próby pracujących, modele stosowane w ocenie: lokalnych i rządowych programów doskonalenia zawodowego bezrobotnych, programów pomocy w poszukiwaniu pracy i programów zasiłków dla bezrobotnych. Na uwagę zasługują także analizy wynagrodzeń w zależności od przynależności do związku zawodowego czy w zależności od liczby lat edukacji. Modele selekcji stosowane były także do oceny występowania dyskryminacji płacowej.

[1] Opis modeli zob. także [5] oraz [17]. W pracy [14] zaprezentowano model tobitowy typu I, w [15] – model tobitowy typu II. Zastosowanie wybranych modeli tobitowych (typu II lub V) zob. [16] oraz [18].

[2] W 2000 r. James Heckman za rozwój teorii i metod analizy zagadnień związanych z selekcją próby otrzymał nagrodę im. Alfreda Nobla w dziedzinie nauk ekonomicznych (wraz z Danielem McFaddenem, który został wyróżniony za rozwój teorii i metod analizy dyskretnego wyboru).

Obecnie modele tobitowe są często wykorzystywane, także w polskiej literaturze przedmiotu. Stosując je trzeba zwrócić uwagę na poprawną interpretację otrzymanych wyników, w szczególności na fakt, że interpretacji podlegają wpływy krańcowe, które w ogólnym przypadku nie są równe poszczególnym parametrom modelu. W literaturze przedmiotu zwracali na to uwagę m.in. W. Greene [3], J.F. McDonald i R.A. Moffitt [22], D.W. Roneck [23]. Ponadto wpływy krańcowe zmiennych binarnych mogą być inne niż zmiennych ciągłych, tzn. przy ich wyznaczaniu należy skorzystać z innych wzorów.

Celem artykułu jest przedstawienie matematycznych podstaw poprawnej interpretacji oraz wnioskowania w oparciu o modele z rodziny tobitowych. Część z zaprezentowanych w artykule wzorów na wpływy krańcowe zmiennych objaśniających w modelach tobitowych jest wyprowadzeniem własnym autorki, przy czym ogólne zasady ich wyznaczania są znane w literaturze od dawna.

2. MODELE TOBITOWE[3]

Modele tobitowe służą do opisu zmiennej zależnej podlegającej cenzurowaniu lub nieprzypadkowej selekcji z cenzurowaniem[4], będącej pewnego rodzaju uogólnieniem cenzurowania (por. np. [20]). Niektóre modele występują także w wersji dla zmiennej uciętej. Poniżej krótko omówiono wybrane modele tobitowe.

Niech Y^* będzie zmienną ukrytą (*latent variable*), tzn. taką zmienną, której wartości zazwyczaj nie są obserwowalne lub są obserwowalne w zależności od pewnych warunków (kryterium obserwowalności). Symbolem Y oznaczono zmienną uciętą, cenzurowaną lub podlegającą nieprzypadkowej selekcji, która przyjmuje wartości takie, jak zmienna Y^*, o ile są one obserwowane, tzn. o ile spełnione są określone warunki. W przeciwnym wypadku zmiennej Y zostaje nadana pewna umowna wartość progowa (a lub c) lub jej wartości nie są obserwowalne (albo umyślnie ignorowane). Przez n oznaczono liczbę elementów w próbie, przy czym dla wygody zapisu przyjęto, że pierwszych n_1 elementów ($i = 1, \ldots, n_1$) w próbie odnosi się do dokładnie zaobserwowanych wartości, natomiast następne $(n - n_1)$ elementów ($i = (n_1 + 1), \ldots, n$) – do przypisanych wartości umownych (o ile istnieją).

Niech ponadto Z^* oznacza zmienną, od której zależy obserwowalność zmiennej Y^* (Y). Zmienna Z^* nie musi być bezpośrednio obserwowalna. Przyjmuje się, że zmienna utworzona na bazie zmiennej ukrytej Z^* może być zmienną dychotomiczną, uciętą, cenzurowaną lub zmienną nie będącą zmienną ograniczoną (jej wartości są zwyczajnie dostępne). Przyjmuje się oznaczenie I dla zmiennej binarnej, w przeciwnym wypadku

[3] Przy opisie modeli tobitowych w tym oraz następnym punkcie artykułu korzystano z prac [14], [15], [16], [17], [18].

[4] Szerzej definicje zmiennej uciętej, cenzurowanej oraz podlegającej nieprzypadkowej selekcji omówiono np. w pracach [16], [17].

stosuje się oznaczenie Z:

$$I_i = \begin{cases} 1 & \text{gdy } z_i^* > a \\ 0 & \text{gdy } z_i^* \leqslant a \end{cases} \quad \text{lub } z_i = \begin{cases} z_i^* & \text{gdy } z_i^* > a \\ a \textit{ lub} \text{ brak} & \text{gdy } z_i^* \leqslant a \end{cases}, \tag{1}$$

gdzie a – wartość progowa dla zmiennej Z^*.

Model tobitowy typu I (krócej: model tobitowy[5]) służy do opisu zmiennej zależnej cenzurowanej. Po ewentualnym przenumerowaniu obserwacji można go zapisać w postaci (por. np. [1], [4], [21], a także [5], [17]):

$$y_i = \begin{cases} y_i^* & \text{gdy } y_i^* > c, \quad i = 1, \ldots, n_1 \\ c & \text{gdy } y_i^* \leqslant c, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{2}$$

gdzie:

$y_i^* = \mathbf{X}_i\boldsymbol{\beta} + u_i$ – równanie regresji,

$\mathbf{X}_i$ – $(k+1)$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $X = \{X_1, \ldots, X_k\}$, odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element wektora $\mathbf{X}_i$ jest równy jedności,

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ – wektor (pionowy) parametrów równania regresji,

u_i – składniki losowe równania, niezależne i pochodzące z tego samego rozkładu o wartości oczekiwanej równej 0 oraz stałej wariancji σ_u. Najczęściej zakłada się, że składniki losowe podlegają rozkładowi normalnemu i tak przyjęto w artykule.

Model tobitowy dla cenzurowania lewostronnego, z wartością progową $c = 0$ oraz przy założeniu rozkładu normalnego składników losowych nosi nazwę standardowego modelu tobitowego[6].

W standardowym modelu tobitowym zarówno prawdopodobieństwo osiągnięcia przez zmienną wartości powyżej progu, $P(y_i > 0|\mathbf{X}_i)$, jak i warunkowa wartość oczekiwana zmiennej uciętej, $E(y_i|y_i > 0, \mathbf{X}_i)$, zależą od tego samego zbioru zmiennych X. Naturalnym uogólnieniem jest dopuszczenie innego zbioru zmiennych (W) wpływającego na prawdopodobieństwo osiągnięcia przez zmienną zależną wartości powyżej progu oraz innego (X) – na warunkową wartość oczekiwaną $E(y_i|y_i > 0, \mathbf{X}_i)$. To pozwala na rozważenie modeli tobitowych opisujących zmienną zależną podlegającą nieprzypadkowej selekcji z cenzurowaniem lub ucięciem. Modele te należą do szerokiej rodziny tzw.

[5] W literaturze przedmiotu na ogół stosuje się nazwę „*model tobitowy*" na określenie modelu tobitowego typu I. Pojęcie „*modele tobitowe*" (w liczbie mnogiej) na ogół odnosi się do rodziny modeli tobitowych obejmujących modele tobitowe typu I–V zgodnie z klasyfikacją zaproponowaną przez T. Amemiyę [1] oraz inne modele wywodzące się od nich.

[6] W artykule rozważono tylko modele opisujące zmienną zależną cenzurowana lewostronnie. Nie wiąże się to ze startą ogólności: jeśli bowiem zmienna Y jest cenzurowana prawostronnie, to $-Y$ jest cenzurowana lewostronnie. Ponadto w dalszej części artykułu również bez straty ogólności przyjęto c=0 (wystarczy dokonać przesunięcia zmiennej tzn. Y–c). W dalszej części, także bez straty ogólności, przyjęto wartość progową a=0.

modeli selekcji[7]. Rodzina ta w szczególności zawiera modele nazwane przez T. Amemiyę modelami tobitowymi typu II i III (por. [1]) wraz z modyfikacjami oraz modele tobitowe typu IV i V – ich dwuwymiarowe odpowiedniki. Ogólnie można powiedzieć, że modele selekcji to modele opisujące zmienną zależną, której obserwowalność jest uzależniona od pewnych warunków opisanych przez inną zmienną (lub inne zmienne). Zagadnienie można rozważać nie tylko w kontekście modeli tobitowych, ale także w powiązaniu z modelami zmiennej zależnej jakościowej opisującej wiele kategorii (uporządkowanych lub nie) czy modelami zmiennej zależnej liczącej.

Modele tobitowe typu II–V znajdują zastosowanie, gdy pewna – zazwyczaj duża – część wartości zmiennej zależnej (zmiennych zależnych) gromadzi się (skupia) w zerze, lub innej wartości progowej c, nie ze względu na cenzurowanie, lecz w wyniku selekcji według pewnego nielosowego kryterium tzn. w wyniku cenzurowania zmiennej zależnej przez wartości innej, skorelowanej z nią zmiennej binarnej I (w modelach typu II i V) lub cenzurowanej Z (w modelach typu III i IV). Mechanizm selekcji, opisywany przez zmienną I lub Z, może zależeć od cech badanych jednostek lub otoczenia, a w szczególności może być związany z niekoniecznie świadomym podejmowaniem indywidualnych decyzji i dokonywaniem wyborów przez badane jednostki.

Model tobitowy typu II definiuje się w następujący sposób (por. np. [1], a także [17]):

$$y_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta} + u_i & \text{gdy} \quad I_i = 1, \quad i = 1, \ldots, n_1 \\ 0 \text{ lub brak} & \text{gdy} \quad I_i = 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{3}$$

gdzie:

$I_i = \begin{cases} 1 & \text{gdy } z_i^* > 0 \\ 0 & \text{gdy } z_i^* \leqslant 0 \end{cases}$ – zmienna binarna opisująca kryterium obserwowalności,

$z_i^* = \mathbf{W}_i\boldsymbol{\gamma} + v_i$ – równanie selekcji,

$\mathbf{W}_i$ – $(m{+}1)$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $W = \{W_1, \ldots, W_m\}$, odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element wektora $\mathbf{W}_i$ jest równy jedności,

$\gamma = (\gamma_0, \gamma_1, \ldots, \gamma_m)'$ – wektor (pionowy) parametrów równania selekcji,

v_i – składniki losowe równania selekcji,

$\mathbf{X}_i, \mathbf{W}_i, \boldsymbol{\beta}, u_i,$ – określone jak wcześniej, przy czym najczęściej zakłada się, że: $(u_i, v_i) \sim N\left(0,\ 0;\ \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$, gdzie ρ – współczynnik korelacji między u_i i v_i.

W modelu tobitowym typu II kryterium obserwowalności opisane jest przez zmienną binarną I. Z tego względu przyjmuje się $\sigma_v = 1$, podobnie jak w modelu probitowym.

[7] Opis zagadnienia nieprzypadkowej selekcji oraz odpowiednich modeli można znaleźć np. w pracach J. Heckmana ([6], [7], [8], [9], [10], [11]).

Model tobitowy z selekcją (*tobit model with sample selection*, *double-hurdle model*) jest połączeniem modelu tobitowego i zagadnienia selekcji. Model ten można nazwać zmodyfikowanym modelem tobitowym typu II (modelem tobitowym typu II o zmodyfikowanym kryterium obserwowalności). Modyfikacja polega na tym, że obecnie muszą być spełnione dwa warunki, aby wartości zmiennej zależnej były obserwowalne. Pierwszy warunek, tak jak w modelu tobitowym typu II, nałożony jest na zmienną binarną I, drugi – na zmienną zależną: jej wartości muszą być większe od zera (lub innej wartości progowej), aby można je było zaobserwować. Zmodyfikowany model tobitowy typu II (model tobitowy z selekcją) można zapisać w postaci (por. np. [2] s. 4, [12], [13]):

$$y_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta} + u_i & \text{gdy } \mathbf{X}_i\boldsymbol{\beta} + u_i > 0 \wedge I_i = 1 \\ 0 \text{ lub brak} & \text{gdy } \mathbf{X}_i\boldsymbol{\beta} + u_i \leqslant 0 \vee I_i = 0 \end{cases} \quad (i = 1, \ldots, n), \tag{4}$$

przy oznaczeniach jak przy wzorze (3).

Model tobitowy typu III definiuje się w podobny sposób jak model typu II (por. np. [1], a także [17]):

$$y_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta} + u_i & \text{gdy} \quad z_i > 0, \quad i = 1, \ldots, n_1 \\ 0 \text{ lub brak} & \text{gdy} \quad z_i = 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{5}$$

gdzie tym razem zmienna opisująca kryterium obserwowalności jest postaci:

$$z_i = \begin{cases} z_i^* & \text{gdy} \quad z_i^* > 0 \\ 0 & \text{gdy} \quad z_i^* \leqslant 0 \end{cases},$$

przy pozostałych oznaczeniach jak wcześniej.

Model tobitowy typu III jest jednym z pierwszych modeli selekcji. W literaturze przedmiotu nazywany jest także modelem selekcji, modelem doboru próby lub modelem heckitowym (*selectivity models, sample selection models, self-selection models, heckit model*).

W zależności od tego, czy i kiedy wartości zmiennej I (lub Z) oraz zmiennych ze zbiorów X i W są dostępne, można wyróżnić ucięty lub cenzurowany model tobitowy typu II (lub III). Jeśli dla wszystkich $i = 1, \ldots, n$ dostępne są wartości zmiennej I (lub Z) oraz zmiennych ze zbioru W, wówczas model jest cenzurowanym modelem tobitowym typu II (lub III) lub, ogólniej, cenzurowanym modelem selekcji. Wartości zmiennych ze zbioru X mogą być dostępne albo dla wszystkich $i = 1, \ldots, n$ albo tylko dla $i : I_i = 1$ (tzn. $i = 1, \ldots, n_1$). Jeżeli wartości zmiennej I (lub Z) oraz zmiennych ze zbioru W są dostępne tylko dla $i = 1, \ldots, n_1$, model (3) lub (5) jest uciętym modelem tobitowym typu II lub III odpowiednio.

Modele tobitowe typu II lub III dopuszczają ujemne wartości dla zmiennej Y, co jest charakterystyczne dla uciętych i cenzurowanych modeli selekcji. Ponadto w modelach tych zakłada się występowanie korelacji ρ między składnikami losowymi równania

regresji (u_i) i równania selekcji (v_i). Korelacja istotnie różna od zera oznacza występowanie związku między prawdopodobieństwem, że wartość zmiennej zależnej jest obserwowana, i warunkową wartością oczekiwaną zmiennej przy warunku, że jest ona obserwowana.

Modele tobitowe typu IV lub V powstają z modelu tobitowego typu III lub II, odpowiednio, dla zmiennej Y_1, przez uwzględnienie dodatkowej zmiennej cenzurowanej Y_2. Kryterium obserwowalności jest wspólne dla zmiennych Y_1 i Y_2, lecz wartości obu zmiennych nie mogą być obserwowalne równocześnie. W modelu tobitowym typu IV kryterium obserwowalności jest opisane przez zmienną cenzurowaną Z, natomiast w modelu tobitowym typu V – przez zmienną binarną I. Z konstrukcji wymagane jest, aby wartości zmiennej Z lub I oraz zmiennych objaśniających ze zbioru W były dostępne dla wszystkich $i = 1, \ldots, n$.

Model tobitowy typu IV jest postaci (por. [1], a także [5], [17]):

$$y_{1i} = \begin{cases} \mathbf{X}_i^{(1)}\boldsymbol{\beta}^{(1)} + u_{1i} & \text{gdy} \quad z_i > 0 \\ \text{brak} & \text{gdy} \quad z_i = 0 \end{cases}, \qquad y_{2i} = \begin{cases} \mathbf{X}_i^{(2)}\boldsymbol{\beta}^{(2)} + u_{2i} & \text{gdy} \quad z_i = 0 \\ \text{brak} & \text{gdy} \quad z_i > 0 \end{cases}, \tag{6}$$

gdzie:

$\mathbf{X}_i^{(1)}$, $\mathbf{X}_i^{(2)}$ – wektory (poziome) wartości zmiennych objaśniających ze zbioru $X^{(1)} = \{X_1^{(1)}, \ldots, X_{k_1}^{(1)}\}$ lub $X^{(2)} = \{X_1^{(2)}, \ldots, X_{k_2}^{(2)}\}$ odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element tych wektorów jest równy jedności,

$\boldsymbol{\beta}^{(1)} = (\beta_0^{(1)}, \beta_1^{(1)}, \ldots, \beta_{k_1}^{(1)})'$ lub $\boldsymbol{\beta}^{(2)} = (\beta_0^{(2)}, \beta_1^{(2)}, \ldots, \beta_{k_2}^{(2)})'$ – wektory (pionowe) parametrów równania regresji dla Y_1 lub równania regresji dla Y_2,

$z_i = \begin{cases} z_i^* & \text{gdy} \quad z_i^* > 0, \quad i = 1, \ldots, n_1 \\ 0 & \text{gdy} \quad z_i^* \leqslant 0, \quad i = (n_1 + 1), \ldots, n \end{cases}$ – kryterium obserwowalności wspólne dla zmiennych Y_1 i Y_2,

$z_i^*, \mathbf{W}_i, \gamma$ – jak wcześniej,

u_{1i}, u_{2i}, v_i – składniki losowe: równania regresji dla Y_1, dla Y_2 oraz równania selekcji; niezależne, pochodzące z trójwymiarowego rozkładu normalnego[8]:

$$(u_{1i}, u_{2i}, v_i) \sim N\left(0, 0, 0; \begin{bmatrix} \sigma_1^2 & 0 & \rho_1\sigma_1\sigma_v \\ 0 & \sigma_2^2 & \rho_2\sigma_2\sigma_v \\ \rho_1\sigma_1\sigma_v & \rho_2\sigma_2\sigma_v & \sigma_v^2 \end{bmatrix}\right), \text{ przy czym:}$$

σ_1 (σ_2) – odchylenie standardowe składnika losowego u_{1i} (odpowiednio u_{2i}),

ρ_1 – współczynnik korelacji między składnikami losowymi v_i oraz u_{1i},

ρ_2 – współczynnik korelacji między składnikami losowymi v_i oraz u_{2i}.

[8] Równania opisujące zmienne Y_1 oraz Y_2 są niezależne, gdyż są to równania dla dwóch wykluczających się warunków (nie ma możliwości, aby oba warunki były spełnione jednocześnie), stąd zera w macierzy wariancji-kowariancji.

W podobny sposób definiuje się model tobitowy typu V (por. [1], a także [5], [17]):

$$y_{1i} = \begin{cases} \mathbf{X}_i^{(1)}\boldsymbol{\beta}^{(1)} + u_{1i} & \text{gdy} \quad I_i = 1 \\ \text{brak} & \text{gdy} \quad I_i = 0 \end{cases}, \qquad y_{2i} = \begin{cases} \mathbf{X}_i^{(2)}\boldsymbol{\beta}^{(2)} + u_{2i} & \text{gdy} \quad I_i = 0 \\ \text{brak} & \text{gdy} \quad I_i = 1, \end{cases} \tag{7}$$

przy czym kryterium obserwowalności dla zmiennych Y_1 i Y_2 opisuje zmienna binarna:

$$I_i = \begin{cases} 1 & \text{gdy} \quad z_i^* > 0, \quad i = 1, \ldots, n_1 \\ 0 & \text{gdy} \quad z_i^* \leqslant 0, \quad i = (n_1 + 1), \ldots, n, \end{cases}$$

przy oznaczeniach jak wcześniej.

Podobnie jak w przypadku modelu tobitowego typu II można rozważyć modyfikację modelu typu V przez dołożenie dodatkowych warunków na zmienne zależne Y_1 oraz Y_2. Postać zmodyfikowanego modelu tobitowego typu V można znaleźć w pracy [17].

Dotychczas w artykule zaprezentowano modele tobitowe przy założeniu homoskedastyczności składnika losowego. W praktyce stosuje się także modele tobitowe przy założeniu heteroskedastyczności składnika losowego równania regresji. Można rozważyć heteroskedastyczność o multiplikatywnej postaci tzn.:

$$\sigma_i = \sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha}), \tag{8}$$

gdzie:

$\mathbf{H}_i - J$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $H = \{H_1, \ldots, H_J\}$ w równaniu heteroskedastyczności,

$\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_J)'$ – J-elementowy wektor (pionowy) parametrów równania heteroskedastyczności.

Definicja heteroskedastyczności sprawia, że przy $\boldsymbol{\alpha} = \mathbf{0}$, dla wszystkich i zachodzi $\sigma_i = \sigma_u$, czyli występuje homoskedastyczność.

Prezentacja metod estymacji w modelach tobitowych wykracza poza ramy niniejszego artykułu. W tym miejscu warto wspomnieć jedynie, że na ogół stosuje się metodę największej wiarygodności (por. np. [1], [4]), która zdaje egzamin także w wypadku heteroskedastyczności składnika losowego (po odpowiedniej modyfikacji funkcji wiarygodności). Metoda największej wiarygodności wymaga założenia o normalności rozkładu składników losowych. Przy skomplikowanych modelach o funkcji wiarygodności posiadającej więcej niż jedno ekstremum stosuje się dwustopniową metodę Heckmana (por. np. [1]). W literaturze przedmiotu można odnaleźć także metody semiparametryczne i nieparametryczne. Metody te jednak nie są pozbawione pewnych wad.

3. WARUNKOWE WARTOŚCI OCZEKIWANE

W modelach tobitowych można rozważyć różne wartości oczekiwane zmiennej zależnej: $E(Y^*)$, $E(Y)$, a także wartość oczekiwaną zmiennej Y pod warunkiem, że jej

wartości są obserwowalne (np. $E(Y|Y > 0)$ w modelu tobitowym typu I). We wszystkich modelach tobitowych wartość oczekiwana zmiennej wejściowej Y^* jest postaci[9]:

$$E(y_i^*|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta} \;\; (\text{lub} \;\; E(y_i^*|\mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta}) \tag{9}$$

Z modelem tobitowym (typu I) związane są:

— wartość oczekiwana „niezerowych" wartości zmiennej zależnej:

$$E(y_i\,|y_i > 0, \mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta} + \sigma_u\lambda(-\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}), \tag{10}$$

gdzie, w wypadku ograniczenia lewostronnego:

$$\lambda(\alpha) = \frac{\varphi(\alpha)}{1 - \Phi(\alpha)} \tag{11}$$

nazywane jest lambdą Heckmana[10];

— wartość oczekiwana wszystkich wartości zmiennej zależnej:

$$E(y_i|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta} \cdot \Phi(\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}) + \sigma_u\varphi(\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}). \tag{12}$$

Między wartościami oczekiwanymi (10) i (12) zachodzi następujący związek:

$$E(y_i|\mathbf{X}_i) = E(y_i|y_i > 0, \mathbf{X}_i) \cdot \Phi(\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}). \tag{13}$$

Wartości oczekiwane (10) oraz (12) różnią się od wartości oczekiwanej zmiennej wejściowej $E(y_i^*|\mathbf{X}_i) = \mathbf{X}_i\boldsymbol{\beta}$. Różnice te są tym większe, im większy jest procent obserwacji ocenzurowanych (tzn. takich, że $y_i = 0$).

W modelach tobitowych w zależności od celu prowadzanego badania można wyznaczyć różne (warunkowe) wartości oczekiwane zmiennej zależnej. W modelu tobitowym typu II danym wzorem (3) wartość oczekiwana zmiennej zależnej pod warunkiem, że jej wartości są obserwowalne jest postaci:

$$E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = E(y_i|z_i^* > 0, \mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta} + \rho\sigma_u\lambda(-\mathbf{W}_i\boldsymbol{\gamma}). \tag{14}$$

Można ponadto wyznaczyć wartość oczekiwaną zmiennej zależnej zgodnie z formułą:

$$E(y_i|\mathbf{X}_i, \mathbf{W}_i) = (\mathbf{X}_i\boldsymbol{\beta} + \rho\sigma_u\lambda(-\mathbf{W}_i\boldsymbol{\gamma})) \cdot \Phi(\mathbf{W}_i\boldsymbol{\gamma}). \tag{15}$$

[9] Z tego względu, że modele zmiennej zależnej cenzurowanej, służą do opisu mikrodanych (danych indywidualnych, jednostkowych), wyrażenie $E(y_i)$ rozumiane jest jako wartość oczekiwana zmiennej Y dla i-tej badanej jednostki (osoby, przedsiębiorstwa itp.). W artykule przyjęto taką notację wzorując się na literaturze przedmiotu (por. np. [21], a także [5]).

[10] Dla ograniczenia prawostronnego lambdę Heckmana definiuje się jako: $\lambda(\alpha) = -\dfrac{\varphi(\alpha)}{\Phi(\alpha)}$.

Stosownie do celu badania do wyznaczenia wartości teoretycznych zmiennej zależnej na podstawie otrzymanego modelu wykorzystuje się wartość oczekiwaną (14) lub (15). Jeżeli przedmiotem zainteresowania jest opis badanego zjawiska w wybranej na podstawie kryterium selekcji próbie, wówczas należy skorzystać ze wzoru (14).

W modelu tobitowym typu II o zmodyfikowanym kryterium obserwowalności (por. (4)) warunkowe wartości oczekiwane odnoszące się do zmiennej zależnej Y są postaci:

$$E(y_i|y_i > 0, I_i = 1, \mathbf{X}_i, \mathbf{W}_i) =$$

$$= \mathbf{X}_i\boldsymbol{\beta} + \frac{\sigma_u \cdot \left(\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}) \Phi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma} - \rho \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}}{\sqrt{1-\rho^2}} \right) + \rho\varphi(\mathbf{W}_i\boldsymbol{\gamma}) \Phi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u} - \rho\, \mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}} \right) \right)}{\Phi_2(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}, \mathbf{W}_i\boldsymbol{\gamma}; \rho)} \tag{16}$$

oraz

$$E(y_i|\mathbf{X}_i, \mathbf{W}_i) = E(y_i|y_i > 0, I_i = 1, \mathbf{X}_i, \mathbf{W}_i) \cdot \Phi_2(\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}, \mathbf{W}_i\boldsymbol{\gamma}; \rho), \tag{17}$$

gdzie $\Phi_2(\,\cdot\,, \cdot\,; \rho)$ jest dystrybuantą dwuwymiarowego rozkładu normalnego o średnich równych zeru, odchyleniach standardowych równych jedności i współczynniku korelacji ρ.

Można również rozważyć następującą wartość oczekiwaną:

$$E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = E(y_i|\mathbf{X}_i, \mathbf{W}_i)/\Phi(\mathbf{W}_i\boldsymbol{\gamma}). \tag{18}$$

W modelu tobitowym typu III (por. (5)) można wyznaczyć następujące wartości oczekiwane:

$$E(y_i|z_i > 0, \mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta} + \rho\sigma_u\lambda(-\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}), \tag{19}$$

$$E(y_i|\mathbf{X}_i, \mathbf{W}_i) = \left(\mathbf{X}_i\boldsymbol{\beta} + \rho\sigma_u\lambda(-\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v})\right) \cdot \Phi(\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}). \tag{20}$$

Podstawiając $\sigma_v = 1$ w powyższych formułach, otrzymuje się wzory na wartości oczekiwane w modelu tobitowym typu II (por. wzory (14) i (15)).

Wzory na warunkowe wartości oczekiwane zmiennych zależnych Y_1 oraz Y_2 w modelu tobitowym typu IV (por. (6)) wynikają ze wzorów na odpowiednie wartości oczekiwane w modelu tobitowym typu III i są postaci[11]:

— dla zmiennej Y_1:

$$E(y_{1i}|z_i > 0, \mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i^{(1)}\boldsymbol{\beta}^{(1)} + \rho_1\sigma_1\lambda(-\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}), \tag{21}$$

$$E(y_{1i}|\mathbf{X}_i, \mathbf{W}_i) = \left(\mathbf{X}_i^{(1)}\boldsymbol{\beta}^{(1)} + \rho_1\sigma_1\lambda(-\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v})\right) \cdot \Phi(\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}), \tag{22}$$

[11] Dla wygody we wzorach poniżej przyjęto oznaczenie $\mathbf{X}_i = (\mathbf{X}_i^{(1)}, \mathbf{X}_i^{(2)})$.

— dla zmiennej Y_2:

$$E(y_{2i}|z_i = 0, \mathbf{X}_i, \mathbf{W}_i) = E(y_{2i}|z_i^* \leqslant 0, \mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i^{(2)}\boldsymbol{\beta}^{(2)} - \rho_2\sigma_2\lambda(\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}), \tag{23}$$

$$E(y_{2i}|\mathbf{X}_i, \mathbf{W}_i) = E(y_{2i}|z_i = 0, \mathbf{X}_i, \mathbf{W}_i) \cdot P(z_i^* \leqslant 0|\mathbf{X}_i, \mathbf{W}_i) =$$

$$= \left(\mathbf{X}_i^{(2)}\boldsymbol{\beta}^{(2)} - \rho_2\sigma_2\lambda(\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v})\right) \cdot \Phi(-\tfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}). \tag{24}$$

Wartości oczekiwane zmiennych zależnych w modelu tobitowym typu V danym wzorem (7) wyznacza się zgodnie ze wzorami (21)–(24) dla modelu typu IV, z podstawieniem $\sigma_v = 1$.

4. WPŁYWY KRAŃCOWE ZMIENNYCH OBJAŚNIAJĄCYCH W MODELACH TOBITOWYCH

Stosując modele tobitowe należy pamiętać, że interpretacja parametrów modelu jest inna niż w zwykłym modelu regresji (por. np. [3], [22], [23]). Wynika to z faktu, że każda z (warunkowych) wartości oczekiwanych (10), (12), (14)–(24) jest różna od wartości oczekiwanej (9). W rezultacie pochodna cząstkowa (warunkowej) wartości oczekiwanej zmiennej zależnej w ogólnym przypadku nie jest równa parametrowi stojącemu przy odpowiedniej zmiennej objaśniającej. W artykule wpływy krańcowe dla kolejnych modeli tobitowych wyznaczono obliczając pochodne cząstkowe z odpowiedniej wartości oczekiwanej zmiennej zależnej. W literaturze przedmiotu można odnaleźć postać niektórych spośród nich, lecz na ogół dla bardziej skomplikowanych modeli nie są one podawane.

Ponieważ wartości oczekiwane (10) i (12) w standardowym modelu tobitowym różnią się od wartości oczekiwanej zmiennej wejściowej Y^* danej wzorem (9), w celu wyznaczenia wpływu zmiennej objaśniającej na zamienną zależną Y należy obliczyć wpływy krańcowe (pochodne cząstkowe). Prawe strony wyrażeń (10) i (12) w sposób nieliniowy zależą od zmiennej, której wpływ na zmienną zależną jest badany, zatem również pochodna cząstkowa zależy od tej zmiennej. Przydatne jest oznaczenie symbolem δ pochodnej z funkcji λ:

$$\delta(\alpha) := \frac{\partial\lambda(\alpha)}{\partial\alpha} = \lambda^2(\alpha) - \alpha\lambda(\alpha). \tag{25}$$

Odpowiednie pochodne cząstkowe są postaci:

$$\frac{\partial E(y_i\,|y_i > 0, \mathbf{X}_i)}{\partial x_{ji}} = \beta_j\left(1 - \tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u} \cdot \lambda\left(-\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right) - \lambda^2\left(-\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right)\right) = \beta_j\left(1 - \delta\left(-\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right)\right), \tag{26}$$

$$\frac{\partial E(y_i|\mathbf{X}_i)}{\partial x_{ji}} = \beta_j\Phi\left(\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right). \tag{27}$$

Gdy zmienna objaśniająca X_j jest zmienną binarną, jej wpływ wyznacza się jako różnicę:

$$\frac{\Delta E(y_i|\mathbf{X}_i)}{\Delta x_{ji}} = E(y_i|\mathbf{X}_i, x_{ji} = 1) - E(y_i|\mathbf{X}_i, x_{ji} = 0). \tag{28}$$

Otrzymane wyniki są cenne ze względu na możliwość ich rzeczywistej interpretacji. Równanie (26) informuje, jaki wpływ wywierają zmiany wartości zmiennej objaśniającej X_j na zmiany w niezerowych wartościach zmiennej Y. Wpływ ten jest iloczynem parametru β_j oraz wyrażenia w nawiasie, które jest zawsze dodatnie (por. [7]), a jego wartość wzrasta wraz ze wzrostem wartości $\mathbf{X}_i\boldsymbol{\beta}/\sigma_u$. Zatem dla wyższych wartości $\mathbf{X}_i\boldsymbol{\beta}/\sigma_u$ zmiany w wartości oczekiwanej $E(y_i\,|y_i > 0, \mathbf{X}_i)$ są większe.

Równanie (27) wyznacza wpływ zmiennej X_j na zmienną Y. Im mniej w próbie jest wartości ocenzurowanych (tzn. im większe jest prawdopodobieństwo $P(y_i^* > 0) = \Phi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u})$), tym zmiany w wartościach zmiennej zależnej generowane przez zmiany zmiennej objaśniającej X_j są bliższe β_j. Jeżeli prawdopodobieństwo $\Phi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u})$ jest bliskie jedności, tzn. w badanej próbie jest bardzo mały odsetek danych ocenzurowanych, wówczas wpływ zmiennej X_j na zmienną Y może być nieistotnie różny od β_j. Zatem wyznaczanie wpływów krańcowych w modelu tobitowym jest ważne szczególnie wtedy, gdy w próbie jest duży odsetek obserwacji „zerowych" (ocenzurowanych).

Ponadto na bazie wzoru (27) można wnioskować, że stosunek wpływów krańcowych $j-$tej i $l-$tej zmiennej objaśniającej na zmienną Y jest ilorazem parametrów stojących przy tych zmiennych, tzn. wynosi $\beta_j \big/ \beta_l$.

Przydatne jest wyznaczenie pochodnej cząstkowej z prawdopodobieństwa $\Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right)$:

$$\frac{\partial \Phi\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right)}{\partial x_{ji}} = \frac{\beta_j}{\sigma_u}\varphi\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}\right). \tag{29}$$

Równanie (29) określa, jak zmiany zmiennej objaśniającej X_j wpływają na zmiany prawdopodobieństwa $P(y_i > 0|\mathbf{X}_i)$, tzn. jak wpływają na prawdopodobieństwo „przejścia" wartości zmiennej zależnej z przedziału poniżej poziomu zera, do przedziału powyżej zera (z nieobserwowalnego do obserwowalnego).

Znak wpływów krańcowych (26)–(29) jest taki sam, jak znak parametru przy zmiennej X_j. Wielkości wpływów krańcowych zależą nie tylko od wielkości zmian zmiennej X_j, ale także od wartości wszystkich zmiennych objaśniających.

W praktyce interpretację wpływów krańcowych podaje się np. wyliczając odpowiednie wartości dla średniej arytmetycznej, mediany, modalnej lub innej wartości zmiennych $X_1, \ldots, X_k$. Można również dla wyznaczonego estymatora wektora parametrów $\boldsymbol{\beta}$ dla każdego i wyznaczyć wartości pochodnych cząstkowych zadanych wzorami podanymi powyżej, a następnie wyliczyć z nich średnią i dopiero te wartości zinterpretować jako wpływ zmiennej objaśniającej X_j na wartość oczekiwaną zmiennej zależnej cenzurowanej lub uciętej, albo na wspomniane wyżej prawdopodobieństwo. Wartość

parametru β_j interpretuje się jako wpływ zmiennej objaśniającej X_j na ukrytą zmienną Y^* (np. skłonność lub użyteczność[12], której nie można zaobserwować).

W modelu tobitowym typu II punktem wyjścia do interpretacji są wartości oczekiwane (14) oraz (15) zmiennej zależnej Y podlegającej nieprzypadkowej selekcji. W modelach tobitowych typu II–V zmienną objaśniającą, której wpływ na zmienną zależną jest badany, dla wygody oznaczono przez ξ_j, gdyż będzie to zmienna objaśniająca ze zbioru X lub ze zbioru W, albo zmienna należąca do części wspólnej tych zbiorów. Aby określić wpływ j–tej zmiennej objaśniającej ξ_j na zmienną zależną Y trzeba wziąć pod uwagę, do którego ze zbiorów ta zmienna należy. Wyróżnia się następujące wpływy krańcowe:

- *wpływ pośredni* zmiennej ξ_j występującej tylko wśród zmiennych objaśniających równania selekcji ($\xi_j = W\backslash X$). Wpływ ten następuje pośrednio przez $\varphi(.)$ i/lub $\Phi(.)$, co oznacza, że na ogół wielkość wpływu pośredniego zależy od wartości zmiennej ξ_j. Wpływ pośredni wynika z faktu, że zmienna zależna jest zmienną podlegającą nieprzypadkowej selekcji;
- *wpływ bezpośredni* zmiennej ξ_j występującej tylko wśród zmiennych objaśniających równania regresji ($\xi_j \in X\backslash W$). Wpływ bezpośredni w większości przypadków nie zależy od wartości zmiennej ξ_j;
- *wpływ łączny* zmiennej ξ_j występującej zarówno w równaniu regresji i równaniu selekcji ($\xi_j \in X \cap W$). Dla wygody zapisu przyjęto ten sam numer (j) zmiennej, gdy występuje ona w zbiorze X i ten sam, gdy występuje w zbiorze W. Wpływ łączny zmiennej $\xi_j \in X \cap W$ na zmienną zależną na ogół jest sumą wpływu pośredniego i bezpośredniego tej zmiennej. W takim wypadku we wzorach zamieszczonych poniżej wpisano słowo „suma" (suma wpływów cząstkowych $\xi_j \in X\backslash W$ oraz $\xi_j = W\backslash X$).

Wzory opisujące wielkość wpływu krańcowego pośredniego, bezpośredniego oraz łącznego zmiennej ξ_j na zmienną zależną w modelu tobitowym typu II są postaci:

$$\frac{\partial E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i)}{\partial \xi_{ji}} = \begin{cases} -\gamma_j \rho \sigma_u \delta(-\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j & \text{gdy } \xi_j \in X\backslash W \\ \text{suma} & \text{gdy } \xi_j \in X \cap W, \end{cases} \tag{30}$$

$$\frac{\partial E(y_i|\mathbf{X}_i, \mathbf{W}_i)}{\partial \xi_{ji}} = \begin{cases} \gamma_j \varphi(\mathbf{W}_i\boldsymbol{\gamma})(\mathbf{X}_i\boldsymbol{\beta} - \rho\sigma_u \mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j \Phi(\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in X\backslash W \\ \text{suma} & \text{gdy } \xi_j \in X \cap W. \end{cases} \tag{31}$$

Wartość $\sigma_u \delta(-\mathbf{W}_i\boldsymbol{\gamma})$ jest zawsze dodatnia. Stąd, jeżeli $\rho > 0$, znak parametru stojącego przy zmiennej objaśniającej w równaniu selekcji jest przeciwny niż znak pośredniego

[12] Por. także [5], s. 16-17, o interpretacji zmiennych jakościowych w świetle użyteczności lub skłonności. Interpretacje te w analogiczny sposób przekładają się na zmienne cenzurowane.

wpływu krańcowego tej zmiennej na zmienną zależną Y wyznaczonego zgodnie z (30). Jeżeli $\rho < 0$, to znaki te są takie same.

Gdy $j-$ta zmienna objaśniająca jest zmienną binarną, jej wpływ wyznacza się jako różnicę:

$$\frac{\Delta E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i)}{\Delta \xi_{ji}} = E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i, \xi_{ji} = 1) - E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i, \xi_{ji} = 0) \quad (32)$$

lub, odpowiednio,

$$\frac{\Delta E(y_i|\mathbf{X}_i, \mathbf{W}_i)}{\Delta \xi_{ji}} = E(y_i|\mathbf{X}_i, \mathbf{W}_i, \xi_{ji} = 1) - E(y_i|\mathbf{X}_i, \mathbf{W}_i, \xi_{ji} = 0), \quad (33)$$

gdzie ξ_j może być zmienną ze zbioru X lub W. Przy tym zachodzi:

$$\frac{\Delta E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i)}{\Delta \xi_{ji}} = \begin{cases} \rho\sigma_u \left(\lambda(-\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \lambda(-\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j & \text{gdy } \xi_j \in X\backslash W \\ \text{suma} & \text{gdy } \xi_j \in X \cap W, \end{cases} \quad (34)$$

gdzie $\mathbf{W}_i^{(1)}$ i $\mathbf{W}_i^{(0)}$ – wektory wartości zmiennych ze zbioru W z podstawieniem w miejsce $j-$tej zmiennej wartości 1 lub 0, odpowiednio. Podobnie w wypadku wpływu zmiennej binarnej wyznaczonego wzorem (33):

$$\frac{\Delta E(y_i|\mathbf{X}_i, \mathbf{W}_i)}{\Delta \xi_{ji}} =$$
$$= \begin{cases} \mathbf{X}_i\boldsymbol{\beta}\left(\Phi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \Phi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) + \rho\sigma_u \left(\varphi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \varphi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j\Phi(\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in X\backslash W \\ \mathbf{X}_i^{(1)}\boldsymbol{\beta} \cdot \Phi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \mathbf{X}_i^{(0)}\beta \cdot \Phi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma}) + \rho\sigma_u \left(\varphi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \varphi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) & \text{gdy } \xi_j \in X \cap W, \end{cases}$$

gdzie $\mathbf{X}_i^{(1)}$ i $\mathbf{X}_i^{(0)}$ – wektory wartości zmiennych ze zbioru X z podstawieniem w miejsce $j-$tej zmiennej wartości 1 lub 0, odpowiednio.

W szczególności interesujący jest wynik, że wpływy krańcowe zmiennej $\xi_j \in X\backslash W$ są takie same, gdy zmienna ta jest ciągła lub gdy jest zmienną binarną.

W modelu tobitowym typu II przy założeniu heteroskedastyczności składnika losowego o multiplikatywnej postaci, zgodnie ze wzorem (8), odpowiednie wpływy krańcowe mają postać[13]:

[13] Wartości oczekiwane $E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i)$ oraz $E(y_i|\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i)$w modelu tobitowym typu II przy założeniu multiplikatywnej heteroskedastyczności składnika losowego są postaci (14) oraz (15), odpowiednio, z podstawieniem $\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})$ w miejsce σ_u.

$$\frac{\partial E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i)}{\partial \xi_{ji}} =$$
$$= \begin{cases} -\gamma_j \rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\delta(-\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in W\backslash(X \cup H) \\ \beta_j & \text{gdy } \xi_j \in X\backslash(W \cup H) \\ \alpha_j \rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\lambda(-\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in H\backslash(X \cup W), \end{cases} \tag{35}$$

a gdy $\xi_j \in (X \cap H)\backslash W$, $\xi_j \in (W \cap H)\backslash X$, $\xi_j \in (X \cap W)\backslash H$ lub $\xi_j \in X \cap W \cap H$, wpływy krańcowe są równe odpowiedniej sumie,

oraz:

$$\frac{\partial E(y_i|\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i)}{\partial \xi_{ji}} =$$
$$= \begin{cases} \gamma_j \varphi(\mathbf{W}_i\boldsymbol{\gamma})(\mathbf{X}_i\boldsymbol{\beta} - \rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in W\backslash(X \cup H) \\ \beta_j \Phi(\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in X\backslash(W \cup H) \\ \alpha_j \rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\varphi(\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in H\backslash(X \cup W), \end{cases} \tag{36}$$

a gdy $\xi_j \in (X \cap H)\backslash W$, $\xi_j \in (W \cap H)\backslash X$, $\xi_j \in (X \cap W)\backslash H$ lub $\xi_j \in X \cap W \cap H$, wpływy krańcowe są równe odpowiedniej sumie.

Podobnie jak w modelu tobitowym typu II przy założeniu homoskedastyczności składnika losowego także tutaj, gdy $j-$ta zmienna objaśniająca jest zmienną binarną, jej wpływ wyznacza się jako odpowiednią różnicę, analogicznie jak we wzorach (32) i (33), lecz tym razem ξ_j może być zmienną ze zbioru X, W lub H. Zachodzi:

$$\frac{\Delta E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i)}{\Delta \xi_{ji}} =$$
$$= \begin{cases} \rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})[\lambda(-\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \lambda(-\mathbf{W}_i^{(0)}\boldsymbol{\gamma})] & \text{gdy } \xi_j \in W\backslash(X \cup H) \\ \beta_j & \text{gdy } \xi_j \in X\backslash(W \cup H) \\ \rho\sigma_u \lambda(-\mathbf{W}_i\boldsymbol{\gamma})[\exp(\mathbf{H}_i^{(1)}\boldsymbol{\alpha}) - \exp(\mathbf{H}_i^{(0)}\boldsymbol{\alpha})] & \text{gdy } \xi_j \in H\backslash(X \cup W) \\ \rho\sigma_u[\exp(\mathbf{H}_i^{(1)}\boldsymbol{\alpha})\lambda(-\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \exp(\mathbf{H}_i^{(0)}\boldsymbol{\alpha})\lambda(-\mathbf{W}_i^{(0)}\boldsymbol{\gamma})] & \text{gdy } \xi_j \in (W \cap H)\backslash X \\ \beta_j + \rho\sigma_u[\exp(\mathbf{H}_i^{(1)}\boldsymbol{\alpha})\lambda(-\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \exp(\mathbf{H}_i^{(0)}\boldsymbol{\alpha})\lambda(-\mathbf{W}_i^{(0)}\boldsymbol{\gamma})] & \text{gdy } \xi_j \in X \cap W \cap H \end{cases} \tag{37}$$

a gdy $\xi_j \in (X \cap H)\backslash W$ lub $\xi_j \in (X \cap W)\backslash H$, wpływy krańcowe są równe odpowiedniej sumie. Przyjęto oznaczenia: $\mathbf{W}_i^{(1)}$ i $\mathbf{W}_i^{(0)}$ ($\mathbf{H}_i^{(1)}$ i $\mathbf{H}_i^{(0)}$) – wektor wartości zmiennych ze zbioru W (H) z podstawieniem w miejsce $j-$tej zmiennej wartości 1 lub 0, odpowiednio.

Wpływy krańcowe zmiennej binarnej w modelu tobitowym typu II przy założeniu heteroskedastyczności składnika losowego wyznaczone w oparciu o wartość oczekiwaną $E(y_i|\mathbf{X}_i, \mathbf{W}_i, \mathbf{H}_i)$ są postaci:

$$\frac{\Delta E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\Delta\xi_{ji}} =$$

$$= \begin{cases} \mathbf{X}_i\boldsymbol{\beta}\left(\Phi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \Phi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) + & \\ \quad +\rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\left(\varphi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \varphi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) & \text{gdy } \xi_j \in W\backslash(X\cup H) \\ \beta_j\Phi(\mathbf{W}_i\boldsymbol{\gamma}) & \text{gdy } \xi_j \in X\backslash(W\cup H) \\ \rho\sigma_u\lambda(-\mathbf{W}_i\boldsymbol{\gamma})(\exp(\mathbf{H}_i^{(1)}\boldsymbol{\alpha}) - \exp(\mathbf{H}_i^{(0)}\boldsymbol{\alpha})) & \text{gdy } \xi_j \in H\backslash(X\cup W) \\ \mathbf{X}_i^{(1)}\boldsymbol{\beta}\cdot\Phi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \mathbf{X}_i^{(0)}\boldsymbol{\beta}\cdot\Phi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma}) + & \\ \quad +\rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\left(\varphi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \varphi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) & \text{gdy } \xi_j \in (X\cap W)\backslash H \\ \mathbf{X}_i\boldsymbol{\beta}\left(\Phi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \Phi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) + & \\ +\rho\sigma_u\left(\exp(\mathbf{H}_i^{(1)}\boldsymbol{\alpha})\varphi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) - \exp(\mathbf{H}_i^{(0)}\boldsymbol{\alpha})\varphi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right) & \text{gdy } \xi_j \in (W\cap H)\backslash X \\ \left(\mathbf{X}_i^{(1)}\boldsymbol{\beta} + \rho\sigma_u\exp(\mathbf{H}_i^{(1)}\boldsymbol{\alpha})\lambda(\mathbf{W}_i^{(1)}\boldsymbol{\gamma})\right)\cdot\Phi(\mathbf{W}_i^{(1)}\boldsymbol{\gamma}) + & \\ \quad -\left(\mathbf{X}_i^{(0)}\boldsymbol{\beta} + \rho\sigma_u\exp(\mathbf{H}_i^{(0)}\boldsymbol{\alpha})\lambda(\mathbf{W}_i^{(0)}\boldsymbol{\gamma})\right)\cdot\Phi(\mathbf{W}_i^{(0)}\boldsymbol{\gamma}) & \text{gdy } \xi_j \in X\cap W\cap H \end{cases} \tag{38}$$

przy oznaczeniach analogicznych jak wcześniej, przy czym wpływ krańcowy zmiennej $\xi_j \in (X\cap H)\backslash W$ jest sumą wpływów krańcowych $\xi_j \in X\backslash(W\cup H)$ i $\xi_j \in H\backslash(X\cup W)$.

W modelu tobitowym typu II o zmodyfikowanym kryterium obserwowalności (tobitowym z selekcją) wpływy krańcowe wyznaczone jako odpowiednie pochodne cząstkowe z wartości oczekiwanej zadanej wzorem (16) lub (17) są następującej postaci:

$$\frac{\partial E(y_i|y_i>0, I_i=1, \mathbf{X}_i,\mathbf{W}_i)}{\partial\xi_{ji}} = \\ = \frac{1}{\Phi_2^2}\left(\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i)}{\partial\xi_{ji}}\cdot\Phi_2 - E(y_i|\mathbf{X}_i,\mathbf{W}_i)\cdot\frac{\partial\Phi_2}{\partial\xi_{ji}}\right), \tag{39}$$

gdzie[14]:

$$\Phi_2 = \Phi_2\left(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}, \mathbf{W}_i\boldsymbol{\gamma}; \rho\right),$$

$$\frac{\partial\Phi_2}{\partial\xi_{ji}} = \begin{cases} \gamma_j\varphi(\mathbf{W}_i\boldsymbol{\gamma})\Phi\left(\dfrac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u} - \rho\,\mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}}\right) & \text{gdy } \xi_j \in W\backslash X \\ \frac{1}{\sigma_u}\beta_j\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u})\Phi\left(\dfrac{\mathbf{W}_i\boldsymbol{\gamma} - \rho\,\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}}{\sqrt{1-\rho^2}}\right) & \text{gdy } \xi_j \in X\backslash W \\ \text{suma} & \text{gdy } \xi_j \in X\cap W \end{cases}$$

[14] Pochodne cząstkowe z dystrybuanty dwuwymiarowego rozkładu normalnego o średnich równych zeru, odchyleniach standardowych równych jedności oraz współczynniku korelacji ρ wynoszą: $\frac{d}{dx}\Phi_2(x,y,\rho) = \varphi(x)\Phi\left(\frac{y-\rho x}{\sqrt{1-\rho^2}}\right)$ oraz $\frac{d}{dy}\Phi_2(x,y,\rho) = \varphi(y)\Phi\left(\frac{x-\rho y}{\sqrt{1-\rho^2}}\right)$.

oraz

$$\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i)}{\partial \xi_{ji}} =$$

$$= \begin{cases} \gamma_j \left(\varphi(\mathbf{W}_i\boldsymbol{\gamma})\Phi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}-\rho\,\mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}}\right)(\mathbf{X}_i\boldsymbol{\beta}-\rho\sigma_u\mathbf{W}_i\boldsymbol{\gamma}) + \right. \\ \left. + \frac{\sigma_u}{\sqrt{1-\rho^2}}\left(\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u})\varphi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma}-\rho\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}}{\sqrt{1-\rho^2}}\right) - \rho^2\varphi(\mathbf{W}_i\boldsymbol{\gamma})\varphi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}-\rho\,\mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}}\right)\right)\right) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j\left(\Phi_2 + \frac{\rho}{\sqrt{1-\rho^2}}\left(\varphi(\mathbf{W}_i\boldsymbol{\gamma})\varphi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}-\rho\,\mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}}\right) - \varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u})\varphi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma}-\rho\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u}}{\sqrt{1-\rho^2}}\right)\right)\right) & \text{gdy } \xi_j \in X\backslash W \end{cases}$$

a gdy $\xi_j \in X \cap W$, wpływy krańcowe są odpowiednią sumą.

W modelu tobitowym typu II o zmodyfikowanym kryterium przy założeniu heteroskedastyczności składnika losowego o multiplikatywnej postaci odpowiednie wpływy krańcowe mają postać[15]:

$$\begin{aligned} &\frac{\partial E(y_i|y_i>0, I_i=1,\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\partial \xi_{ji}} = \\ &= \frac{1}{\Phi_2^2}\left(\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\partial \xi_{ji}}\cdot\Phi_2 - E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)\cdot\frac{\partial\Phi_2}{\partial\xi_{ji}}\right), \end{aligned} \tag{40}$$

gdzie:

$$\Phi_2 = \Phi_2(\tfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})}, \mathbf{W}_i\boldsymbol{\gamma}; \rho),$$

$$\frac{\partial\Phi_2}{\partial\xi_{ji}} = \begin{cases} \gamma_j\varphi(\mathbf{W}_i\boldsymbol{\gamma})\Phi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})}-\rho\,\mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}}\right) & \text{gdy } \xi_j \in W\backslash(X\cup H) \\ \frac{\beta_j}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})}\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})})\Phi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma}-\rho\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})}}{\sqrt{1-\rho^2}}\right) & \text{gdy } \xi_j \in X\backslash(W\cup H) \\ \frac{-\alpha_j\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})}\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})})\Phi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma}-\rho\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})}}{\sqrt{1-\rho^2}}\right) & \text{gdy } \xi_j \in H\backslash(X\cup W) \end{cases}$$

oraz $\dfrac{\partial\Phi_2}{\partial\xi_{ji}}$ jest równe odpowiedniej sumie, gdy $\xi_j \in (X\cap H)\backslash W$, $\xi_j \in (W\cap H)\backslash X$, $\xi_j \in (X\cap W)\backslash H$ lub $\xi_j \in X\cap W\cap H$. A ponadto:

- dla $\xi_j \in W\backslash(X\cup H)$:

[15] Warunkowa wartość oczekiwana $E(y_i|y_i>0, I_i=1,\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)$ w zmodyfikowanym modelu tobitowym typu II przy założeniu heteroskedastyczności składnika losowego jest postaci (16) z podstawieniem $\sigma_u\exp(\mathbf{H}_i\boldsymbol{\alpha})$ w miejsce σ_u.

$$\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\partial \xi_{ji}} = \gamma_j \left(\varphi(\mathbf{W}_i\boldsymbol{\gamma})\Phi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})} - \rho\, \mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}} \right) (\mathbf{X}_i\boldsymbol{\beta} - \rho\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})\mathbf{W}_i\boldsymbol{\gamma}) + \right.$$
$$\left. + \frac{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})}{\sqrt{1-\rho^2}} \left(\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})})\varphi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma} - \rho\, \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})}}{\sqrt{1-\rho^2}} \right) - \rho^2 \varphi(\mathbf{W}_i\boldsymbol{\gamma})\varphi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})} - \rho\, \mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}} \right) \right) \right)$$

- dla $\xi_j \in X \backslash (W \cup H)$:

$$\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\partial \xi_{ji}} =$$
$$= \beta_j \left(\Phi_2 + \frac{\rho}{\sqrt{1-\rho^2}} \left(\varphi(\mathbf{W}_i\boldsymbol{\gamma})\varphi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})} - \rho\, \mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}} \right) - \varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})})\varphi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma} - \rho\, \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})}}{\sqrt{1-\rho^2}} \right) \right) \right)$$

- dla $\xi_j \in H \backslash (X \cup W)$:

$$\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\partial \xi_{ji}} =$$
$$= \alpha_j \sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha}) \left(\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})})\Phi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma} - \rho\, \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})}}{\sqrt{1-\rho^2}} \right) + \rho\varphi(\mathbf{W}_i\boldsymbol{\gamma})\Phi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})} - \rho\, \mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}} \right) \right) +$$
$$+ \alpha_j \cdot \mathbf{X}_i\boldsymbol{\beta} \cdot \frac{\rho}{\sqrt{1-\rho^2}} \left(\varphi(\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})})\varphi\left(\frac{\mathbf{W}_i\boldsymbol{\gamma} - \rho\, \frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})}}{\sqrt{1-\rho^2}} \right) - \varphi(\mathbf{W}_i\boldsymbol{\gamma})\varphi\left(\frac{\frac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u \exp(\mathbf{H}_i\boldsymbol{\alpha})} - \rho\, \mathbf{W}_i\boldsymbol{\gamma}}{\sqrt{1-\rho^2}} \right) \right),$$

a gdy $\xi_j \in (X \cap H)\backslash W$, $\xi_j \in (W \cap H)\backslash X$, $\xi_j \in (X \cap W)\backslash H$ lub $\xi_j \in (X \cap W \cap H)$, wyrażenie $\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i,\mathbf{H}_i)}{\partial \xi_{ji}}$ jest równe odpowiedniej sumie.

W modelu tobitowym typu II o zmodyfikowanym kryterium obserwowalności przy założeniu homo- lub heteroskedastyczności dla binarnej $j-$tej zmiennej objaśniającej wpływy krańcowe wyznacza się analogicznie jak w modelu tobitowym typu II, tzn. jako różnicę odpowiednich wartości oczekiwanych (por. wzory (32) i (33)), przy czym przy założeniu heteroskedastyczności ξ_j może być zmienną ze zbioru X, W lub H. Wzory te redukują się tylko do nieznacznie prostszej postaci, stąd nie zostały tutaj przytoczone.

W modelu tobitowym typu III wzory służące do oceny wpływu krańcowego $j-$tej zmiennej niezależnej na zmienną Y są analogiczne, jak w modelu tobitowym typu II, z tym, że obecnie σ_v jest dowolne, niekoniecznie równe 1. Tak jak poprzednio należy rozważyć trzy przypadki: $\xi_j \in W\backslash X$ – aby wyznaczyć wpływ pośredni, $\xi_j \in X\backslash W$ –

aby wyznaczyć wpływ bezpośredni, oraz $\xi_j \in X \cap W$ – aby wyznaczyć łączny wpływ $j-$tej zmiennej objaśniającej. Wpływy krańcowe oblicza się zgodnie z formułami:

$$\frac{\partial E(y_i|z_i>0,\mathbf{X}_i,\mathbf{W}_i)}{\partial \xi_{ji}} = \begin{cases} -\gamma_j \rho \dfrac{\sigma_u}{\sigma_v}\delta(-\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j & \text{gdy } \xi_j \in X\backslash W \\ \text{suma} & \text{gdy } \xi_j \in X \cap W, \end{cases} \tag{41}$$

$$\frac{\partial E(y_i|\mathbf{X}_i,\mathbf{W}_i)}{\partial \xi_{ji}} = \begin{cases} \gamma_j \dfrac{\sigma_u}{\sigma_v}\varphi(\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v})\left(\dfrac{\mathbf{X}_i\boldsymbol{\beta}}{\sigma_u} - \rho\dfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}\right) & \text{gdy } \xi_j \in W\backslash X \\ \beta_j \Phi(\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) & \text{gdy } \xi_j \in X\backslash W \\ \text{suma} & \text{gdy } \xi_j \in X \cap W. \end{cases} \tag{42}$$

Jeżeli $j-$ta zmienna objaśniająca jest zmienną binarną jej wpływ wyznacza się, analogicznie jak w modelu tobitowym typu II, jako odpowiednią różnicę (por. wzory (32) i (33).

Wzory służące do oceny wpływu $j-$tej zmiennej objaśniającej na zmienne Y_1 i Y_2 w modelu tobitowym typu IV są podobne jak w poprzednich modelach. Odpowiednie wartości oczekiwane oraz ich pochodne cząstkowe względem $j-$tej zmiennej zapisano poniżej. Analogicznie jak poprzednio $\xi_j^{(1)}$ to zmienna, której wpływ na zmienną Y_1 jest badany; $\xi_j^{(2)}$ to zmienna, której wpływ na zmienną Y_2 jest badany. Rozważono wpływy krańcowe pośrednie, bezpośrednie i łączne.

- Wpływy krańcowe zmiennej $\xi_j^{(1)}$ na zmienną Y_1 wynoszą:

$$\frac{\partial E(y_{1i}|z_i>0,\mathbf{X}_i,\mathbf{W}_i)}{\partial \xi_{ji}^{(1)}} = \begin{cases} -\gamma_j \rho_1 \dfrac{\sigma_1}{\sigma_v}\delta(-\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) & \text{gdy } \xi_j^{(1)} \in W\backslash X^{(1)} \\ \beta_j^{(1)} & \text{gdy } \xi_j^{(1)} \in X^{(1)}\backslash W \\ \text{suma} & \text{gdy } \xi_j^{(1)} \in X^{(1)} \cap W, \end{cases} \tag{43}$$

$$\frac{\partial E(y_{1i}|\mathbf{X}_i,\mathbf{W}_i)}{\partial \xi_{ji}^{(1)}} = \begin{cases} \gamma_j \dfrac{\sigma_1}{\sigma_v}\varphi(\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v})\left(\dfrac{\mathbf{X}_i^{(1)}\boldsymbol{\beta}^{(1)}}{\sigma_1} - \rho_1\dfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}\right) & \text{gdy } \xi_j^{(1)} \in W\backslash X^{(1)} \\ \beta_j^{(1)} \Phi(\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) & \text{gdy } \xi_j^{(1)} \in X^{(1)}\backslash W \\ \text{suma} & \text{gdy } \xi_j^{(1)} \in W \cap X^{(1)}. \end{cases} \tag{44}$$

- Wpływy krańcowe zmiennej $\xi_j^{(2)}$ na zmienną Y_2 wynoszą:

$$\frac{\partial E(y_{2i}|z_i=0,\mathbf{X}_i,\mathbf{W}_i)}{\partial \xi_{ji}^{(2)}} = \begin{cases} -\gamma_j \rho_2 \dfrac{\sigma_2}{\sigma_v}\delta(\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) & \text{gdy } \xi_j^{(2)} \in W\backslash X^{(2)} \\ \beta_j^{(2)} & \text{gdy } \xi_j^{(2)} \in X^{(2)}\backslash W \\ \text{suma} & \text{gdy } \xi_j^{(2)} \in X^{(2)} \cap W, \end{cases} \tag{45}$$

$$\frac{\partial E(y_{2i}|\mathbf{X}_i, \mathbf{W}_i)}{\partial \xi_{ji}^{(2)}} =$$

$$= \begin{cases} -\gamma_j \dfrac{\sigma_2}{\sigma_v} \varphi(\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) \left(\dfrac{\mathbf{X}_i^{(2)}\boldsymbol{\beta}^{(2)}}{\sigma_2} - \rho_2 \dfrac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v} \right) & \text{gdy } \xi_j^{(2)} \in W \backslash X^{(2)} \\ \beta_j^{(2)} \Phi(-\frac{\mathbf{W}_i\boldsymbol{\gamma}}{\sigma_v}) & \text{gdy } \xi_j^{(2)} \in X^{(2)} \backslash W \\ \text{suma} & \text{gdy } \xi_j^{(2)} \in X^{(2)} \cap W. \end{cases} \tag{46}$$

Jeśli j–ta zmienna objaśniająca jest zmienną binarną, wpływy krańcowe wyznacza się jako odpowiednie różnice, analogicznie jak przy zaprezentowanych wcześniejszych modelach.

W modelu tobitowym typu V oraz zmodyfikowanym modelu tobitowym typu V wzory służące do oceny wpływu zmiennych objaśniających $\xi_j^{(1)}$ lub $\xi_j^{(2)}$ na zmienne Y_1 lub Y_2, odpowiednio, są takie same, jak w modelu tobitowym typu IV (por. wzory (43)–(46)), z podstawieniem $\sigma_v = 1$.

Odpowiednie wzory dla modelu tobitowego typu V o zmodyfikowanym kryterium obserwowalności można uzyskać na podstawie wzoru (39) dla zmodyfikowanego modelu tobitowego typu II, uwzględniając cenzurowanie prawostronne przy wyznaczaniu wpływów krańcowych na zmienną Y_2, tzn. zmieniając znak przy parametrach $\boldsymbol{\gamma}$ oraz współczynniku korelacji ρ_2 na przeciwny.

Podobnie w modelu tobitowym typu V (oraz jego zmodyfikowanej wersji) przy założeniu heteroskedastyczności[16] składników losowych w równaniach regresji dla zmiennych Y_1 oraz zmiennej Y_2 należy postępować, jak w modelu tobitowym typu II (oraz jego modyfikacji) przy założeniu heteroskedastyczności. W modelu tobitowym typu V z heteroskedastycznym składnikiem losowym w celu wyznaczania wpływów krańcowych zmiennych objaśniających na zmienną Y_1 należy zastosować bezpośrednio wzory (35)–(38) (w zmodyfikowanym modelu – wzór (40)), zaś na zmienną Y_2 – przed zastosowaniem wzorów należy zmienić znak parametrów $\boldsymbol{\gamma}$ oraz współczynnika korelacji ρ_2 na przeciwny.

5. PRZYKŁAD EMPIRYCZNY

W celu ilustracji dotychczasowych rozważań, poniżej zaprezentowano przykład zastosowania jednego z omawianych modeli, przedstawiono porównanie oszacowanych parametrów i wyliczonych wpływów krańcowych zmiennych objaśniających.

Analizę zwrócono w kierunku poszukiwania modelu poprawnie opisującego tygodniową liczbę godzin pracy zamężnych kobiet (zmienna zależna) z uwzględnieniem czynników wpływających na decyzję o podjęciu zatrudnienia. Podstawą badania były dane indywidualne pochodzące z BAEL prowadzonego co kwartał przez GUS. Analizowana próba obejmowała 5183 zamężnych kobiet w Polsce w I kwartale 2009 r.

[16] Postać heteroskedastycznego modelu tobitowego typu V można znaleźć np. w pracy [18].

Tygodniowy czas pracy zamężnych kobiet, zgodnie z definicją BAEL, jest dodatni tylko dla kobiet pracujących, natomiast dla kobiet bezrobotnych lub biernych zawodowo wynosi zero godzin. W próbie zamężnych kobiet dla 50,138% obserwacji w I kwartale 2009 roku wartości omawianej zmiennej były dodatnie. Dla pozostałych obserwacji odnotowano wartość zero. Tak duża reprezentacja wartości zerowych nie pozwala na ich odrzucenie i zastosowanie np. modelu regresji wielorakiej bez konsekwencji w postaci błędnie oszacowanych parametrów modelu. Ponadto to czy zamężne kobiety posiadały pracę czy jej nie posiadały nie jest losowe – było zależne od pewnych cech (zmiennych) opisujących te kobiety. Zmienna odzwierciedlająca tygodniową liczbę godzin pracy zamężnych kobiet pracujących jest zatem zmienną podlegającą selekcji (tutaj cenzurowanej). Przy jej opisie trzeba uwzględnić czynniki determinujące wybór przynależności do grupy pracujących lub do grupy nie posiadających pracy. Do opisu takiej zmiennej zależnej sięgnięto po model tobitowy typu II dany wzorem (3) przy założeniu heteroskedastyczności składnika losowego w multiplikatywnej postaci opisanej wzorem (8). Parametry modelu oszacowano metodą największej wiarygodności. Obliczenia wykonano w środowisku NLogit 9.0, przy czym korzystano z samodzielnie napisanych procedur wyznaczających wpływy krańcowe.

Poniżej zamieszczono wybrane fragmenty analizy tygodniowego czasu pracy zamężnych kobiet w Polsce przeprowadzonego przez autorkę artykułu w rozprawie doktorskiej (por. [19]). Omówiono wpływ na tygodniowy czas pracy zamężnych kobiet w badanym okresie następujących (wybranych) zmiennych objaśniających odzwierciedlających:

– wiek respondentki ze względu na ekonomiczne grupy wieku (produkcyjny mobilny lub młodszy: do 44 lat, produkcyjny niemobilny: 45–59 lat, poprodukcyjny: 60+ lat);

– poziom wykształcenia respondentki;

– sektor ekonomiczny instytucji będącej głównym miejscem pracy respondentki (rolniczy, przemysłowy, usługowy);

– wymiar czasu pracy (niepełny/pełny);

– wykonywanie pracy dodatkowej poza głównym miejscem pracy;

– rodzaj wykonywanej pracy (pracujący najemnie, na własny rachunek, jako pomagający nieodpłatnie członek rodziny).

Jak wspomniano w analizie uwzględniono także inne zmienne, jednak ze względu na ograniczone miejsce nie omówiono ich w niniejszym artykule (por. [19]).

W tabeli 1 zestawiono oszacowane parametry modelu tobitowego typu II przy założeniu heteroskedastyczności składnika losowego oraz wyznaczone wpływy krańcowe wybranych zmiennych objaśniających korzystając ze wzorów (35) oraz (37). Zastosowany model tobitowy typu II przy założeniu heteroskedastyczności składnika losowego składa się z trzech równań: równania selekcji (uczestnictwa), równania regresji oraz równania wariancji. W każdym z równań oszacowano parametry oraz wpływy krańcowe zmiennych objaśniających. W poszczególnych równaniach nie muszą występować te same zmienne i niekoniecznie te same zmienne będą statystycznie istotne.

Tabela 1.

Oszacowane parametry (w nawiasach podano błędy szacunku) oraz wpływy krańcowe pośrednie, bezpośrednie i łączne w modelu tobitowym typu II z heteroskedastycznym składnikiem losowym dla I kwartału 2009 r. – wybrane zmienne

Zmienna:	Równanie selekcji		Eównanie regresji		Równanie wariancji		Wpływ łączny
	Oszac. parametr	Wpływ pośredni	Oszac. parametr	Wpływ bezpośredni	Oszac. parametr	Wpływ pośredni	
sektor przemysłowy	×	×	7,400 (0,747)	7,400	×	×	7,400
sektor usługowy	×	×	7,663 (1,775)	7,663	×	×	7,663
pełny wymiar czasu pracy	×	×	17,211 (1,840)	17,211	×	×	17,211
rodzaj wyk. pracy: własny rachunek	×	×	4,404 (0,732)	4,404	×	×	4,404
rodzaj wyk. pracy: pomagający członek rodziny	×	×	4,377 (0,774)	4,377	×	×	4,377
wiek 45–59 lat	-0,117 (0,041)	-0,773	×	×	×	×	-0,773
wiek 60+ lat	-0,397 (0,062)	-2,689	2,798 (1,008)	2,798	×	×	0,109
wykształcenie wyższe[17]	0,344 (0,054)	2,178	-5,644 (0,869)	-5,644	0,152 (0,022)	-0,284	-5,008
praca dodatkowa	×	×	8,195 (0,747)	8,195	0,198 (0,027)	-0,386	7,809

Uwaga: znakiem × oznaczono zmienne statystycznie nieistotne lub nieujęte w danym równaniu modelu.

Źródło: obliczenia własne na podstawie danych BAEL (GUS) dla I kwartału 2009 roku.

Warto przyjrzeć się otrzymanym wynikom. Po pierwsze niekiedy oszacowania wpływów krańcowych są takie same jak parametrów (zob. pięć pierwszych zmiennych w tab. 1). Uwaga ta dotyczy tych zmiennych objaśniających, które występują tylko w równaniu regresji zastosowanego modelu i jest zgodna ze wzorami (35) oraz (37). Zatem wpływ zmiennych odzwierciedlających rodzaj sektora ekonomicznego, wymiar czasu pracy oraz reprezentujących rodzaj wykonywanej pracy na tygodniową liczbę godzin pracy zamężnych kobiet wyznaczony jest za pomocą oszacowanego parametru modelu. W I kwartale 2009 r. zmiana głównego miejsca pracy z sektora rolniczego (kategoria przyjęta jako bazowa) na sektor przemysłowy lub usługowy przeciętnie wiązała się ze zwiększeniem tygodniowego czasu pracy zamężnych kobiet o około 7,5 godz. (7,400 godz. oraz 7,663 godz., odpowiednio). Przejście z niepełnego na pełen wymiar

[17] Rozważono wykształcenie w następujących kategoriach: wykształcenie wyższe (licencjat, inżynier, magister i wyższe); policealne lub średnie zawodowe; średnie ogólnokształcące; zasadnicze zawodowe; gimnazjum, podstawowe, niepełne podstawowe i bez wykształcenia. Wykształcenie na poziomie gimnazjum lub niższym przyjęto jako kategorię bazową. W tab. 1 ujęto tylko wykształcenie wyższe. Pozostałe warianty zmiennej były statystycznie nieistotne, stąd nie przytoczono ich w tabeli.

czasu pracy skutkowało przeciętnie wydłużeniem czasu pracy o około 17 godz. (17,211 godz.). Wykonywanie pracy na własny rachunek lub w charakterze nieodpłatnie pomagającego członka rodziny, w porównaniu do pracujących najemnie (kategoria przyjęta jako bazowa), powodowało przeciętnie zwiększenie liczby godzin pracy o około 4,5 godz. (4,404 godz. oraz 4,337 godz., odpowiednio).

Jednakże, jeśli zmienna objaśniająca występuje także w równaniu selekcji lub równaniu wariancji, wówczas oszacowany łączny wpływ krańcowy tej zmiennej na ogół nie jest równy oszacowanemu parametrowi. Przykładowo zmiana wykształcenia z podstawowego (przyjętego jako kategoria bazowa) na wykształcenie na poziomie wyższym powodowała przeciętnie zmniejszenie tygodniowej liczby godzin pracy zamężnych kobiet o około 5 godz. (-5,008 godz.). Przy czym trzeba zauważyć, że za pośrednictwem równania selekcji zmienna ta wpływała na czas pracy dodatnio (2,178 godz.). Fakt ten można interpretować w następujący sposób: podniesienie wykształcenia do wykształcenia na poziomie wyższym wpływało na zwiększenie długości tygodniowego czasu pracy kobiet poprzez zwiększenie prawdopodobieństwa podjęcia pracy (przynależności do grupy pracujących), o czym mówi dodatni znak oszacowanego parametru w równaniu selekcji. W równaniu regresji oszacowany parametr i bezpośredni wpływ krańcowy tej zmiennej był taki sam (-5,644 godz.), lecz łączny wpływ zmiany na wykształcenie na poziomie wyższym na długość tygodniowego czasu pracy został skorygowany przez uwzględnienie równania selekcji oraz równania wariancji. Należy także zwrócić uwagę na znaczne różnice w oszacowaniach parametrów w tych równaniach oraz wpływów krańcowych za ich pośrednictwem (0,334 i 2,178 w równaniu selekcji, 0,152 i -0,284 w równaniu wariancji, por. tab. 1). Różnice te wskazują na skalę błędów w wypadku interpretacji bezpośrednio parametrów modelu, zamiast wpływów krańcowych.

Pozostałe zmienne wpływały na tygodniowy czas pracy kobiet w następujący sposób. Zmienna reprezentująca kobiety w wieku produkcyjnym niemobilnym (45–59 lat) była statystycznie istotna tylko w równaniu selekcji i jej wpływ na tygodniowy czas pracy był ujemny, niewielki. Ujemny znak oszacowanego parametru wskazuje na ujemny wpływ tej zmiennej na prawdopodobieństwo przynależności do grupy pracujących w porównaniu do kobiet w wieku produkcyjnym mobilnym (do 44 lat, kategoria przyjęta za bazową).

Podobnie sytuacja przedstawiała się w grupie kobiet w wieku poprodukcyjnym (60+ lat), przy czym ujemny wpływ tej zmiennej na tygodniowy czas pracy zamężnych kobiet za pośrednictwem równania selekcji był większy (-2,689 godz.). Wpływ ten był równoważony dodatnim wpływem tej zmiennej wynikającym z równania regresji (2,798 godz.) i w rezultacie łączny jej wpływ na czas pracy był niewielki (0,109 godz.).

Podjęcie dodatkowej pracy poza głównym miejscem pracy przyczyniało się przeciętnie do zwiększenia tygodniowego czasu pracy zamężnych kobiet o około 8 godz. (7,809 godz.). Zmienna ta występowała w równaniach regresji i wariancji.

W modelach zmiennej zależnej cenzurowanej oprócz parametrów modelu estymuje się także odchylenie standardowe składnika resztowego modelu regresji oraz współczynnik korelacji między składnikami losowymi równania selekcji i równania regre-

sji. W zastosowanym modelu tobitowym typu II przy założonej heteroskedastyczności składnika losowego estymator stałej multiplikatywnej w równaniu wariancji wynosił 11,175, natomiast estymator współczynnika korelacji: -0,778. Ponieważ współczynnik korelacji był ujemny, znak oszacowanych wpływów krańcowych i parametrów w równaniu selekcji był taki sam, zaś w równaniu wariancji – przeciwny (por. wzory (30) i (35)). Statystyczna istotność współczynnika korelacji wskazała na konieczność stosowania modelu tobitowego typu II (statystyczna nieistotność współczynnika korelacji oznaczałaby, że wystarczy zastosować model regresji wielorakiej do opisu zagadnienia).

6. PODSUMOWANIE

W artykule przedstawiono wzory umożliwiające poprawną interpretację wyników otrzymanych za pomocą wybranych modeli tobitowych. Niektóre z modeli rozważono w wersji o zmodyfikowanym kryterium obserwowalności lub przy dopuszczeniu heteroskedastyczności składnika losowego. W omawianych modelach wpływ krańcowy na ogół nie jest równy parametrowi odpowiadającemu zmiennej objaśniającej, której wpływ na badane zjawisko (zmienną zależną) jest analizowany. Ponadto należy uwzględnić czy zmienna objaśniająca, której wpływ na analizowane zjawisko jest badany, jest zmienną binarną (ogólnie dyskretną) czy ciągłą.

Przedstawiony przykład zastosowania modelu tobitowego typu II jako jednego z rodziny modeli zmiennej zależnej cenzurowanej ilustruje, w jaki sposób interpretuje się oszacowane wpływy krańcowe zmiennych objaśniających w omawianych modelach. Wskazuje także na różnice w wartościach oszacowanych parametrów i łącznych wpływów krańcowych poszczególnych zmiennych objaśniających.

Stosując modele z rodziny tobitowych należy posłużyć się odpowiednim oprogramowaniem. Trzeba zwrócić uwagę, czy wykorzystywany program wyznacza tylko parametry modelu czy także potrzebne do interpretacji wpływy krańcowe. Dobrym rozwiązaniem jest sięgnięcie po programy, które pozwalają na samodzielne zapisanie procedur zgodnie z zaprezentowanymi w artykule wzorami na wpływy krańcowe w poszczególnych modelach z rodziny tobitowych (np. NLogit, Gauss, R). Wówczas można mieć pewność, że otrzymane wyniki pozwolą na poprawną interpretację analizowanego zagadnienia.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

[1] Amemiya T. (1984), *Tobit Models: A survey*, Journal of Econometrics, vol. 24, s. 3-61.

[2] Flood L., Gråsjö U. (1998), *Regression Analysis and Time Use Data: A Comparison of Microeconometric Approaches with Data from the Swedish Time Use Survey (HUS)*, Working Papers in Economics, nr 5, School of Economics and Commercial Law, Göteborg University.

[3] Greene W. (1999), *Marginal Effects in the Censored Regression Model*, Econometrics Letters, nr 1(64).

[4] Greene W. (2003), *Econometric Analysis*, 5th ed., Prentice-Hall, Inc, New Jersey.
[5] Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, wyd. 2, Monografie i Opracowania 490, Oficyna Wydawnicza SGH, Warszawa.
[6] Heckman J. (1974), *Shadow Prices, Market Wages, and Labor Supply*, Econometrica, vol. 42, nr 4, s. 679–694.
[7] Heckman J. (1976), *The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models*, Annals of Economic and Social Measurement, nr 5/4, 1976, s. 475–492.
[8] Heckman J. (1979), *Sample Selection Bias as a Specification Error*, Econometrica, vol. 47, nr 1, s. 153-161.
[9] Heckman J. (1987), *Selection Bias and Self-Selection* [w:] *The New Palgrave: A Dictionary of Economics*, vol. 1-4, pod red. J. Eatwell, M. Milgate, P. Newman, Macmillan Press, London, s. 287-297.
[10] Heckman J. (1990), *Varieties of Selection Bias*, American Economic Review, 80, s. 313-318.
[11] Heckman J. (2003), *Microdata, Heterogeneity and the Evaluation of Public Policy* [w:] *Nobel Lectures in Economic Sciences 1996 – 2000*, T. Persson (ed.), World Scientific Publishing Co., Singapore, s. 255-322.
[12] Jones A.M. (1989), *A Double-Hurdle Model of Cigarette Consumption*, Journal of Applied Econometrics, vol. 4, s. 23-39.
[13] Jones A.M. (1992), *A Note on Computation of the Double-Hurdle Model with Dependence with an Application to Tobaco Expenditure*, Bulletin of Economic Research, vol. 44, nr 1, s. 67-74.
[14] Kostrzewska J. (2003), *Model tobitowy jako szczególny przypadek cenzurowanego modelu regresji* [w:] *Przestrzenno–czasowe modelowanie i prognozowanie zjawisk gospodarczych*, pod red. A. Zeliasia, Wyd. AE w Krakowie, Kraków, s. 397-404.
[15] Kostrzewska J. (2004), *Model regresji dla danych dobieranych do próby według nielosowego kryterium*, Zeszyty Naukowe AE w Krakowie, nr 666, Prace z zakresu prognozowania, Kraków, s. 111-117.
[16] Kostrzewska J. (2008), *Zastosowanie wybranych modeli tobitowych do opisu tygodniowej liczby godzin pracy* [w:] *Taksonomia 15. Klasyfikacja i analiza danych – teoria i zastosowania*, pod red. K. Jajugi i M. Walesiaka, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, Wrocław.
[17] Kostrzewska J. (2009), *Truncated and Censored Dependent Variables and Their Models* [w:] *Statistical Analysis of the Economic and Social Consequences of Transition Processes in Central-East European Countries*, ed. by J. Pociecha, Studia i Prace UEK, nr 6, UEK, Kraków.
[18] Kostrzewska J. (2011), *Analiza porównawcza tygodniowego czasu pracy zamężnych kobiet pracujących na własny rachunek lub najemnie*, [w:] *Osiągnięcia i perspektywy modelowania i prognozowania zjawisk społeczno-gospodarczych*, pod red. B. Pawełek, Wyd. UEK, Kraków.
[19] Kostrzewska J. (2011), *Modele regresji zmiennych cenzurowanych w analizie rynku pracy w Polsce*, manuskrypt, UEK, Kraków.
[20] Maddala G.S. (1987), *Censored Data Models* [w:] *The New Palgrave: A Dictionary of Economics*, ed. by J. Eatwell, M. Milgate, P. Newman, Macmillan Press, London, vol. 1, s. 384-385.
[21] Maddala G.S. (1994), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge (pierwsze wyd. w 1983 roku).
[22] McDonald J.F., Moffitt R.A. (1980), *The Uses of Tobit Analysis*, Review of Economics and Statistics, vol. LXII, s. 318-321.
[23] Roneck D.W. (1992), *Learning More from Tobit Coefficients: Extending a Comparative Analysis of Political Protests*, American Sociological Review, vol. 57, nr 4, s. 503-507.
[24] Tobin J. (1958), *Estimation of Relationships for Limited Dependent Variables*, Econometrica, vol. 26, s. 24-36.

INTERPRETACJA W MODELACH TOBITOWYCH

Streszczenie

W artykule krótko scharakteryzowano wybrane modele tobitowe zgodnie z klasyfikacja zaproponowaną przez T. Amemiyę ([1]), następnie zaprezentowano postaci wartości oczekiwanych dla poszczególnych modeli oraz wyznaczono wpływy krańcowe. Ze względu na specyfikę omawianych modeli wpływy krańcowe w ogólnym przypadku nie są równe odpowiednim parametrom. Zatem w celu uzyskania poprawnej interpretacji należy wyznaczyć pochodne cząstkowe z odpowiednich (warunkowych) wartości oczekiwanych zmiennej zależnej. Ponadto w artykule zwrócono uwagę na fakt, że wpływy krańcowe binarnych zmiennych objaśniających są inne niż ciągłych.

AN INTERPRETATION IN THE TOBIT MODELS

Summary

At first the tobit models according to T. Amemiya's classification ([1]) are presented in the paper. Next, expected values and marginal effects in these models are obtained. In the tobit models marginal effects are not equal to coefficients of the model in general. To get a proper interpretation in the considered models, it is necessary to obtain marginal effects as a partial deviation of the expected values. Moreover, marginal effects are different for binary and continuous variables.

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

KAMILA MIGDAŁ-NAJMAN

OCENA JAKOŚCI WYNIKÓW GRUPOWANIA – PRZEGLĄD BIBLIOGRAFII

1. WSTĘP

Jedną z ważnych zdolności człowieka jest umiejętność rozróżniania i rozpoznawania obiektów, zdarzeń czy faktów, a także ich łączenia w grupy. Wydaje się, że czynności te dla człowieka są czymś naturalnym, łatwym i prostym. Można powiedzieć, że wyodrębnianie jednostek podobnych w celu ich pogrupowania jest niemal zdolnością przyrodzoną człowieka. J.A. Hartigan powiedziałby, że jest to sposób myślenia o rzeczach a nie badanie rzeczy samych w sobie, poprzez czerpanie wiedzy z doświadczeń i umiejętności uzyskanych z wielu obszarów życia człowieka (Arabie P., Hubert L.J., Soete G.De [1996]). Ten sposób myślenia stał się podstawą rozwoju wielu dziedzin wiedzy, które zaproponowały w tym zakresie własne bogate terminologie i definicje.

Już 1737 roku Carolus Linnaeus w swoim dziele „*Genera Plantarum*" wskazuje, że cała wiedza, jaką rzeczywiście posiadamy o interesujących nas jednostkach zależy od stosowanych metod, które pozwalają na wyróżnienie podobnych do siebie grup jednostek. Im większe zróżnicowanie badanych jednostek obserwujemy, tym łatwiej nam przy pomocy tych metod pojąć złożoną strukturę badanego zjawiska. Im badane zbiory są bardziej liczne, tym ważniejsze i bardziej konieczne staje się posługiwanie się takimi metodami (Everitt B.S., Landau S., Leese M., Stahl D. [2011]). Ważnym impulsem do rozwoju metod grupowania była publikacja z 1963 roku dwóch biologów P.H. Sneatha i R.R. Sokala zatytułowana „*Principles of numerical taxonomy*". Monografia ta jest często uważana za przełomową i wskazującą nowe kierunki badań nad metodami grupowania.

Grupowanie jednostek jest zadaniem złożonym. Jest wiele czynników wpływających na uzyskane rozwiązanie. Do ważniejszych można zaliczyć: liczbę grupowanych jednostek (stosuje się inne metody grupowania zbiorów o kilkudziesięciu jednostkach a inne o setkach tysięcy), liczbę cech zmiennych opisujących daną jednostkę (tzw. problem wymiarowości), zastosowane skale pomiarowe wszystkich cech (skale mogą być identyczne dla wszystkich cech lub nie), strukturę przestrzenną jednostek (skupienia separowalne lub nie, separowalne liniowo lub nie, posiadające skupienie gęstości obiektów lub rozłożone równomiernie, i inne), istnienie braków danych czy istnienie wartości skrajnych (*outliers*). Każdy z tych czynników powoduje konieczność innego podejścia do problemu grupowania. Różnorodność ta jest także przyczyną istnienia dużej liczby algorytmów grupowania, często opartych na bardzo różnych pomysłach. Ponieważ zbiór jednostek można zwykle pogrupować na bardzo wiele sposobów jedne

z nich wydają się „lepsze" a inne „gorsze". Ponieważ pojęcia te są niejednoznaczne, konieczne stało się wypracowanie obiektywnych kryteriów oceny jakości wyróżnionych skupień.

W artykule podjęto próbę usystematyzowania wiedzy w zakresie istniejących metod oceny jakości struktury grupowej. Przedstawiono propozycje klasyfikacji metod oceny jakości grupowania i wskaźników ustalania liczby skupień.

2. KRYTERIA OCENY JAKOŚCI WYNIKÓW GRUPOWANIA

Wszelkiego rodzaju procedury oceny uzyskanych wyników grupowania znane są w literaturze światowej pod nazwą *cluster validity* (prawdziwość, wiarygodność, jakość, grupowania). Pojęcie to definiuje się jako proces oceny wiarygodności algorytmów grupowania przy zastosowaniu różnych warunków początkowych a także jako kwantyfikowalną ocenę uzyskanych rezultatów grupowania. Ponieważ istnieje bardzo wiele samych metod grupowania, w konsekwencji jest także wiele metod oceny uzyskanej struktury grupowej. Metody oceny struktury grupowej można podzielić na dwie zasadnicze grupy: metody wzorcowe i bezwzorcowe. Metody wzorcowe to te, w których uzyskany podział jednostek porównuje się z innym znanym podziałem. Istnieje więc wzorzec idealnego lub zakładanego podziału, z którym porównuje się podział bieżący. Wzorzec może pochodzić od eksperta, może wynikać z rozważań teoretycznych a może pochodzić z innych badań tej samej zbiorowości, tą samą metodą ale z innymi parametrami lub inną metodą. (Brun M., Sima C., Hua J., Lowey J., Carroll B., Suh E., Dougherty E.R. [2007]). Metody bezwzorcowe to te, które do oceny jakości klasyfikacji wykorzystują jedynie informacje pochodzące bezpośrednio z danych. Bierze się tu pod uwagę takie własności uzyskanych skupień jak ich spójność, separowalność, sferyczność czy gęstość.

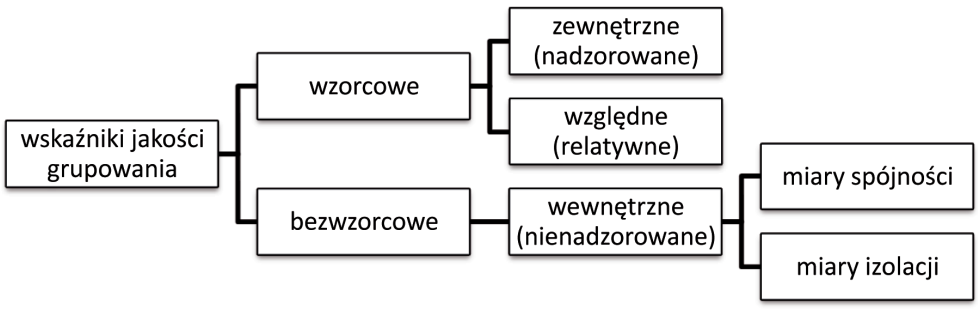
Ta ogólna idea wydaje się być akceptowana przez środowiska naukowe reprezentujące różne dyscypliny. Różnice sprowadzają się głównie do stosowanej nomenklatury lub stopnia szczegółowości opisu. Najbardziej znane propozycje pojęć pochodzą jeszcze z 1967 roku z prac MacQueena. Biorąc pod uwagę źródło pochodzenia informacji o strukturze grupowej, wyróżniono trzy klasy wskaźników (współczynników) jakości grupowania opartych o kryterium: zewnętrzne (*external criteria, external validation*), wewnętrzne (*internal criteria, internal validation*) i względne (*relative criteria, relative validation*) (Theodoridis S., Koutroumbas K. [1999, 2003], Halkidi M., Batistakis Y., Vazirgiannis M. [2001], Maimon O., Rokach L. [2005]).

Ze wskaźnikami biorącymi pod uwagę kryterium zewnętrzne mamy do czynienia, gdy uzyskana struktura grupowa i przynależność jednostek do skupień porównywana jest ze z góry założoną strukturą grupową badanego zbioru jednostek. Zakładana struktura jest odzwierciedleniem naszej wiedzy, wiedzy lub intuicji eksperta o analizowanym zbiorze. Wzorzec grupowania jest, więc zewnętrzny w stosunku do grupowanych jednostek. Ze wskaźnikami biorącymi pod uwagę kryterium wewnętrzne mamy do czynienia, gdy ocena rezultatów uzyskanej struktury grupowej dokonana jest w oparciu o infor-

macje pochodzące z samego analizowanego zbioru jednostek. Informacja o strukturze grupowej pochodzi wyłącznie z danych. Jest więc wewnętrzna w stosunku do nich. Ze wskaźnikami biorącymi pod uwagę kryterium względne mamy do czynienia, gdy ocena rezultatów uzyskanej struktury grupowej porównywana jest w stosunku do innej struktury grupowej uzyskanej w wyniku zastosowania tego samego algorytmu grupowania, ale z założonymi innymi parametrami (np. liczbą skupień). Ocena grupowania jest dokonywana względem innych uzyskanych podziałów populacji (Friedman H.P., Rubin J. [1967], Halkidi M., Batistakis Y., Vazirgiannis M. [2001], Duda R.O., Hart P.E., Stork D.G. [2002], Theodoridis S., Koutroumbas K. [2003], Brun M., Sima C., Hua J., Lowey J., Carroll B., Suh E., Dougherty E.R. [2007], Halkidi M., Gunopulos D., Vazirgiannis M., Kumar N., Domeniconi C. [2008]).

Przy powyższych definicjach wskaźniki zewnętrzne i względne należą do szerszej grupy wskaźników wzorcowych, a wskaźniki wewnętrzne do bezwzorcowych. Opierając się na tej samej idei, powyższą klasyfikację definiuje się czasem stosując nomenklaturę bardziej techniczną. Metody oparte o kryterium zewnętrzne są nazywane metodami nadzorowanymi (*supervised*), oparte o kryterium wewnętrzne – metodami nienadzorowanymi (*unsupervised*), a oparte o kryterium względne – metodami relatywnymi (*relative*). W ramach metod nienadzorowanych wyróżnia się dodatkowo dwie grupy: miary spójności skupień (*measures of cluster cohesion*), które określają jak blisko powiązane są ze sobą jednostki w skupieniu oraz miary rozdzielenia, izolacji (*measures of cluster separation*), które określają jak oddalone jest dane skupienie od innych (por. schemat 1).

Schemat 1 Wskaźniki jakości grupowania



Źródło: opracowanie własne

Informacja o jakości uzyskanych skupień jest w procesie grupowania danych potrzebna na dwóch etapach badań. Na etapie końcowym, gdy dokonano już grupowania jednostek konieczna jest ocena uzyskanego podziału. Informacja taka jest jednak konieczna także w samym procesie wyróżniania skupień. Wiele metod grupowania opiera się na pomyśle, aby w procesie iteracyjnym optymalizować wybrane kryterium podziału jednostek. Poszukuje się takiego podziału, który jest najlepszy z wybranego punktu widzenia. Ponieważ najlepszy podział to taki, który pozwala uzyskać skupienia o najwyższej jakości, konieczne jest zdefiniowanie kryterium oceny jakości skupień,

które będzie w procesie grupowania optymalizowane. Krytycznym etapem optymalizacji jest wybór liczby skupień, na którą należy zbiór jednostek rozdzielić. Nie można uzyskać dobrej jakości skupień dla błędnie ustalonej ich liczby. Jednocześnie liczba skupień stanowi parametr aprioryczny dla wielu często stosowanych w praktyce metod grupowania danych takich jak metoda k-średnich, k-medoid, rozmyta metoda c-średnich czy DBSCAN. Konieczne jest zastosowanie obiektywnego, ilościowego kryterium wyznaczania liczby skupień. Wskaźniki takie istnieją i nazywane są wskaźnikami jakości grupowania (*cluster validity indices CVIs, cluster separation index, validity measure, cluster validity measures, validity indices, cluster validity methods*). Z powyższych powodów problem ustalania liczby skupień należy także do klasy problemów oceny jakości grupowania. W dalszej części artykułu przyjęto podział wskaźników jakości grupowania oparty o kryterium: zewnętrzne, wewnętrzne i względne.

3. KRYTERIUM ZEWNĘTRZNE

Wskaźniki oceny jakości grupowania oparte o kryterium zewnętrzne mają szczególne zastosowanie w sytuacjach gdy: 1) badacz jest zainteresowany informacją jak bardzo podobne do siebie (zgodne) są podziały uzyskane w drodze zastosowania różnych algorytmów grupowania, 2) należy sprawdzić zgodność grupowań dokonanych tą samą metodą ale z różnymi parametrami, 3) konieczne jest porównanie skupień uzyskanych dla tych samych obiektów obserwowanych w różnych momentach czasu, 4) należy ocenić zgodność uzyskanego grupowania ze znaną, wzorcową strukturą grup.

W literaturze tematu proponuje się wiele współczynników zgodności podziałów, służących do oceny podobieństwa wyników różnych klasyfikacji. Do bardziej znanych i częściej stosowanych należą: współczynnik Jaccarda (*Jaccard Coefficient, Jaccard's Coefficient of Community)* (Jaccard P. [1908]), współczynnik Randa (*Rand Index, Rand Statistic)* (Rand W.M. [1971]), współczynnik Fowlkesa i Mallowsa (*Fowlkes and Mallows Index)* (Fowlkes E.B., Mallows C.L. [1983]), skorygowany współczynnik Randa (*Rand Adjusted Statistic, Modified Rand)* (Hubert L.J., Arabie P. [1985]). Propozycje współczynników zgodności grupowania przedstawiali również: Kulczynski S. [1927], Anderberg M.R. [1973], Arabie P., Boorman S.A. [1973], Baker F.B. [1974], Rohlf F.J. [1974, 1982], Hartigan J.A. [1975], Hubert L.J., Levin J.R. [1976], Szmigiel C. [1976], Goodman L.A., Kruskal W.H. [1979], Wallace D.L. [1983], Nowak E. [1985], Gordon A.D. [1987], Mirkin B. [1996], Ben-Hur A., Elisseeff A., Guyon I. [2002] i inni. Interesujące zastosowanie wskaźników Jaccarda, Randa, skorygowanego Randa oraz Fowlkesa i Mallowsa do oceny podobieństwa wyników grupowania w badaniu preferencji i zachowań komunikacyjnych mieszkańców Gdyni zaprezentowała K.Migdał Najman [2011].

Poza tymi najbardziej znanymi wskaźnikami można również spotkać inne propozycje współczynników zgodności podziałów, jak np. miarę F (*F-Measure, F-Score measure*), wskaźnik *purity*, metrykę Mirina (*Mirkin metric*), entropię (*entropy*), wskaźnik wzajemnej informacji (*mutual information*), miarę NMI (*normalized mutual informa-*

tion), współczynnik zmienności informacji (*variation of information*), współczynnik podziału (*partition coefficient*), miarę V (*V-measure*), współczynnik Minkowskiego (*Minkowski score*), błąd klasyfikacji (*classification error, classification metric, CE*), kryterium van Dongena (*van Dongen criterion*), współczynnik lokalnej precyzji (*micro-average precision*) i inne (por. Celeux G., Soromenho G. [1996], Holliday J.D., Hu C-Y., Willett P. [2002], Zhong S., Ghosh J. [2003], Mirkin B. [1996], Handl J., Knowles J. [2004], Zhao Y., Karypis G. [2004], Rubinov A.M., Soukhorokova N.V., Ugon J. [2006], Meilă M. [2007], Aliguliyev R.M. [2009], Seung-Seok C., Sung-Hyuk C., Tappert C.C. [2010], Albatineh A.N., Niewiadomska-Bugaj M. [2011], Labatut V., Cherifi H. [2011], Pitchandi P., Raju N. [2011], Rendón E., Abundez I.M., Gutierrez C., Diaz Zagal S., Arizmendi A., Quiroz E.M., Arzate E.H. [2011]).

Większość z proponowanych współczynników zgodności przyjmuje wartości z unormowanego przedziału, np. [0,1]. Wartość równą 0 osiągają, gdy dwa porównywane podziały są zupełnie niepodobne a wartość równą 1, gdy dwa porównywane podziały są identyczne.

Szerokie zastosowania wskaźników zgodności podziałów w swoich badaniach przedstawiali: Szultz J.V., Hubert L.J. [1979], Pal N.R., Biswas J. [1997], Bezdek J.C., Pal N.R. [1998], Yeung K.Y., Ruzzo W.L. [2001], Mali K., Mitra S. [2003], Bryan J. [2004], Gatnar E., Walesiak M. [2004], Hoffmann A., Motoda H., Scheffer T. (Eds.) [2005], Maimon O., Rokach L. [2005], Ayala G., Epifanio I., Simó A., Zapater V. [2006], Reulke R., Eckardt U., Flach B., Knauer U., Polthier K. (Eds.) [2006], Brun M., Sima C., Hua J., Lowey J., Carroll B., Suh E., Dougherty E.R. [2007], Falasconi M., Pardo M., Vezzoli M., Sberveglieri G. [2007], Migdał Najman K. [2007], Ding C., Li T., Peng W. [2008], Ming-Tso Chiang M., Mirkin B. [2010], Gurrutxaga I., Muguerza J., Arbelaitz O., Pérez J.M., Martín J.I. [2011], Albatineh A.N., Niewiadomska-Bugaj M. [2011] i inni.

4. KRYTERIUM WEWNĘTRZNE

Celem stosowania wskaźników jakości grupowania opartych na kryterium wewnętrznym jest poszukiwanie odpowiedzi na pytanie: na ile uzyskana struktura grupowa otrzymana w wyniku zastosowana danej techniki grupowania stanowi dobre podsumowanie informacji zawartej w danych? Wiele z metod grupowania np. hierarchiczne metody aglomeracyjne (Lance G.N., Williams W.T. [1966a, 1966b], Johnson S.C. [1967], Lance G.N., Williams W.T. [1967a, 1967b], Anderberg M.R. [1973], Gordon A.D. [1987]) oparte są na pomiarze stopnia podobieństwa (*similarity*) lub zróżnicowania (*dissimilarity*) jednostek w przestrzeni cech. To między innymi od właściwego zdefiniowania miary oceniającej stopień podobieństwa (*similarity measures*) lub zróżnicowania (odległości, *dissimilarity measures*) jednostek będzie zależał uzyskany wynik grupowania. (Sneath P.H.A., Sokal R.R. [1973], Anderberg M.R. [1973], Everitt B.S. [1993], Mahalanobis P.C. [1936], Spath H. [1980], Tanimoto T. [1958]). Zastosowanie jednej z metod aglomeracyjnego grupowania hierarchicznego do zdefiniowanej macie-

rzy podobieństwa pozwala na prezentację wyników klasyfikacji w formie graficznej w postaci tzw. dendrogramu lub drzewa połączeń (*dendrogram, tree diagram, hierarchical tree diagram*). Wykres taki prezentuje hierarchię łączenia jednostek w grupy oraz poziomy łączenia (*splitting, fusion levels*), na których jednostki te połączyły się po raz pierwszy (Hartigan J.A. [1967], Gordon A.D. [1987, 1999]). Dendrogram jest efektem zastosowania konkretnej strategii grupowania hierarchicznego i miary podobieństwa lub odległości (Florek K., Łukasiewicz J., Perkal J., Steinhaus H., Zubrzycki S. [1951], Sneath P.H.A. [1957], Sokal R.R., Michener C.D. [1958], McQuitty L.L. [1960, 1966, 1967], Sokal R.R., Sneath P.H.A. [1963], Gower J.C. [1967], Hartigan J.A. [1967], Podani J. [1989], Ward J.H. [1963], Wishart D. [1969]). W połowie lat 60-tych G.N. Lance i W.T. Williams [1966a, 1966b, 1967a, 1967b] opracowali ogólny schemat obliczania odległości między skupieniami biorący pod uwagę wszystkie znane metody aglomeracyjnego grupowania hierarchicznego[1]. W 1978 roku M. Jambu zaproponował poszerzenie tego schematu i umożliwił włączenie do niego jeszcze innych strategii (Jambu M. [1978]). Analizując własności metod aglomeracyjnego grupowania hierarchicznego i możliwość zastosowania w nim różnych miar odległości lub podobieństwa wynika, że możemy uzyskać wiele różnych hierarchii. W literaturze przedmiotu prezentowane są różne metody pozwalające na zmierzenie dopasowania dendrogramu wyznaczonego dla zadanej metody aglomeracyjnego grupowania hierarchicznego do macierzy odległości lub podobieństwa. Ocenę jakości grupowania można oprzeć na porównaniu uzyskanych wyników, prezentowanych np. w formie wyjściowej macierzy odległości do macierzy odległości uzyskanej dla danej strategii grupowania, która prezentuje poziomy łączenia, na których pary jednostek pojawiły się po raz pierwszy w tym samym skupieniu. Najbardziej znanym współczynnikiem pozwalającym na ocenę stopnia dopasowania między macierzą odległości $\mathbf{D}$ a macierzą dendrogramu $\mathbf{C}_{dendr}$ jest współczynnik korelacji kofenetycznej (*cophenetic correlation coefficient CPCC*), zaproponowany w 1962 przez R.R Sokala i F.J. Rohlfa. (Sokal R.R., Rohlf F.J. [1962]) Jego zastosowania można znaleźć w pracach: Rohlf F.J., Fisher D.R. [1968], Farris J.S. [1969], Kruskal J.B., Carroll J.D. [1969], Rohlf F.J. [1970], Holgersson M. [1978], Trakhtenbrot A., Kadmon R. [2006], Kuramae E.E., Robert V., Echavarri-Erasun C., Boekhout T. [2007], Mérigot B., Durbec J.P., Gaertner J.C. [2010]. Innym współczynnikiem pozwalającym na ocenę stopnia dopasowania dendrogramu do macierzy odległości (podobieństwa) jest współczynnik Goodmana-Kruskala (*Goodman-Kruskal gamma coefficient, gamma index*). Został on zaproponowany w 1954 roku (Goodman L.A., Kruskal W.H. [1954]) do oceny zgodności uporządkowań cech wyrażonych na skali porządkowej. W bogatej literaturze tematu proponuje się także inne współczynniki zgodności dla cech wyrażonych na skali porządkowej. Ich propozycje i szersze dyskusje przedstawiali: Kendall M.G. [1945], Somers R.H. [1962a,b], Costner H.L. [1965], Davis J.A. [1967], Hawkes R.K. [1971], Kim J. [1971], Cunningham K.M.,

[1] Metody objęte schematem przez Lancea i Williamsa: metoda najbliższego sąsiada, metoda najdalszego sąsiada, metoda średniej grupowej, metoda centroidalna, metoda mediany i metoda Warda.

Ogilvie J.C. [1972], Baker F.B. [1974], Hubert L.J. [1974], Rohlf F.J. [1974], Baker F.B., Hubert L.J. [1975], Günter S., Bunke H. [2003], Rousson [2007].

Wielu autorów proponuje także wskaźniki oparte na różnicach odległości (podobieństwa) w dwóch porównywanych macierzach: Kruskal J.B. [1964], Gower J.C. [1966, 1967, 1970], Guttman L. [1968], Hartigan J.A. [1967], Jardine C.J., Jardine N., Sibson R. [1967], Jardine N., Sibson R. [1968], Kruskal J.B., Carroll J.D. [1969], Sammon J.W. [1969], Anderson A.J.B. [1971], Sneath P.H.A., Sokal R.R. [1973], Everitt B. [1978], Balicki A. [2009], Kalinowski S.T. [2009].

Do oceny rezultatów uzyskanej klasyfikacji w oparciu o informacje pochodzące z samego analizowanego zbioru może również posłużyć tzw. wskaźnik sylwetkowy (*silhouette index - SI, silhouette coefficient, SIL index*) zaproponowany przez P.J. Rousseeuw w 1987 roku (Rousseeuw P.J. [1987]). Wartość wskaźnika SI można zinterpretować jako wskaźnik jakości otrzymanej struktury grupowej. Interesujące wykorzystanie wskaźnika SI można znaleźć w artykule D. Waneka [2003] do oceny rezultatów uzyskanej klasyfikacji klientów fińskiego przedsiębiorstwa produkującego domy z prefabrykatów.

Zastosowania indeksu SI w swoich badaniach przedstawiali: Kaufman L., Rousseeuw P.J. [1990], Raymond T.N., Jiawei H. [1994], Tibshirani R., Walther G., Hastie T. [2001], Bolshakova N., Azuaje F. [2003], Mali K., Mitra S. [2003], Sugar C.A., James B.M. [2003], Brun M., Sima C., Hua J., Lowey J., Carroll B., Suh E., Dougherty E.R. [2007], Saitta S., Raphael B., Smith I.F.C. [2007], Hennig C. [2008], Schepers J., Ceulemans E., Van Mechelen I. [2008], Lago-Fernández L.F., Corbacho F. [2010], Ming-Tso Chiang M., Mirkin B. [2010].

5. KRYTERIUM WZGLĘDNE

Biorąc pod uwagę względne kryterium oceny jakości grupowania ocena uzyskanej struktury grupowej dokonywana jest poprzez porównanie otrzymanego podziału z innymi wynikami grupowania. Podejście to wykorzystuje się także w samym procesie grupowania do ustalania liczby skupień. Przyjmuje się a priori pewną liczbę skupień (zwykle dwa), dokonuje grupowania i ocenia jakość uzyskanej struktury grupowej. Następnie przyjmuje się większą o jeden liczbę skupień i powtarza procedurę tak długo, aż uzyska się optymalną wartość miary jakości grupowania. Ten podział, który optymalizuje wartość przyjętego do oceny grupowania wskaźnika będzie uznany za najlepszy.

W 1953 roku R.L. Thorndike'a w czasopiśmie *Psychometrika* przedstawił subiektywną propozycję ustalania liczby klas[2]. Podejście Thorndike'a polegało na porównywaniu średnich wewnątrzklasowych przy ustalonej, różnej liczbie skupień. Au-

[2] Jest on twórcą pomysłu na iteracyjny algorytm optymalizacyjny podziału zbioru jednostek, nazwanego później przez McQueena metodą k-średnich. Niesłusznie uważa się, że to McQueen jest jego twórcą, gdy w rzeczywistości prowadził badania nad ulepszeniem już znanego pomysłu (ulepszanego do dziś) i nadał mu nazwę.

tor sugerował, że wraz ze wzrostem liczby klas następuje spadek wartości średniej wewnątrzklasowej. Proponował prezentowanie tej zależności w formie graficznej. Na wykresie, na osi odciętej prezentowane były kolejne podziały (zaczynając od dwóch klas, *number of clusters*) a na osi rzędnej średnia wewnątrzklasowa dla proponowanych różnych podziałów (*average within-cluster distance for different numbers of cluters*). Odczytywanie sugerowanej liczby klas następowało poprzez poszukiwanie na liniowym wykresie nagłego „spłaszczenia" wykresu (bądź posłużenie się kryterium „łokcia"), które następowało wraz ze wzrostem liczby klas (Thorndike R.L. [1953]). Podobne podejście proponowane było również przez innych autorów, między innymi J.C. Gowera w 1975 roku. E.M.L. Beale w artykule z 1969 roku pt. „*Euclidean cluster analysis*" prezentowanym w *Bulletin of the International Statistical Institute*, sugerował, że formalną metodą pozwalającą na ustalanie optymalnej liczby skupień może być test F (Fishera-Snedecora, równości wariancji). Rozważał, czy test F może zostać wykorzystany do sprawdzenia, że podział na większą liczbą klas jest „lepszy" niż podział na mniejszą liczbę klas. Prowadzone badania z zastosowaniem testu F wykazały skuteczność proponowanego podejścia, ale tylko w przypadku skupień dość dobrze separowalnych i hipersferycznych (Everitt B.S. [1979]). W 1969 roku L. Engelman i J.A. Hartigan testowali podejście oparte na maksymalizacji ilorazu międzygrupowej sumy kwadratów w stosunku do wewnątrzgrupowej sumy kwadratów dla zbioru jednostek, który został podzielony na dwa skupienia (Engelman L., Hartigan J.A. [1969]). W 1971 roku F.H.C. Marriott przeprowadził interesującą dyskusję nad wykorzystaniem jako kryterium ustalania liczby klas wyznacznika z macierzy **W** (macierz rozrzutu, dyspersji wewnątrzklasowej). Konfiguracja skupień, która minimalizowałaby iloczyn liczby klas podniesiony do potęgi drugiej i wartość wyznacznika macierzy **W** ($k^2 \cdot |\mathbf{W}|$) zostałaby uznana za optymalną liczbę klas (Marriott F.H.C. [1971]).

Od czasu badań Thorndike'a zaproponowano przynajmniej kilkadziesiąt wskaźników jakości grupowania stosowanych do ustalania liczby skupień. W 1985 roku ukazał się artykuł G.W. Milligana i M.C. Cooper, (Milligan G.W., Cooper M.C. [1985]) który do dzisiaj jest cytowany, jako podstawowe źródło wiedzy w tej dziedzinie. Ta wielość utrudnia zapoznanie się z nimi i wybór właściwego z punktu widzenia bieżących celów badawczych. Aby zadanie to ułatwić proponuje się ich klasyfikację uwzględniającą źródła i rodzaj informacji brany pod uwagę przy ich konstrukcji. Można wyróżnić tu 5 kryteriów (por. schemat 2):

1. Wskaźniki uwzględniające jedynie odległości między obiektami w skupieniu i między skupieniami, (między obiektami z różnych skupień, między centrami skupień). Należą do nich między innymi: wskaźnik Dunna (*Dunn's index*) (Dunn J.C. [1974]), wskaźnik sylwetkowy (*Global silhouette index, silhouette index-SI, silhouette coefficient, SIL index*) (Rousseeuw P.J. [1987]), wskaźnik C (*C-index*) (Hubert L.J., Levin J.R. [1976]). Interesujące zastosowanie wskaźnika SI zaprezentowano w artykule K.Migdał Najman, K.Najman [2006a] w klasyfikacji 329 miast Stanów Zjednoczonych ze względu na cechy sprzyjające inwestycjom mieszkaniowym.

2. Wskaźniki uwzględniające rozproszenie obiektów wewnątrz skupień i odległości między skupieniami, (między centrami skupień), np. wskaźnik geometryczny (*Geometrical index*) (Lam B.S.Y., Yan H. [2005, 2007]), wskaźnik Daviesa-Bouldina (*Davies-Bouldin index*) (Davies D.L., Bouldin D.W. [1979]). Interesujące zastosowanie wskaźnika Daviesa-Bouldina zaprezentowano w artykule K.Migdał Najman [2010] w klasyfikacji zwyczajów i prawidłowości zakupowych klientów.

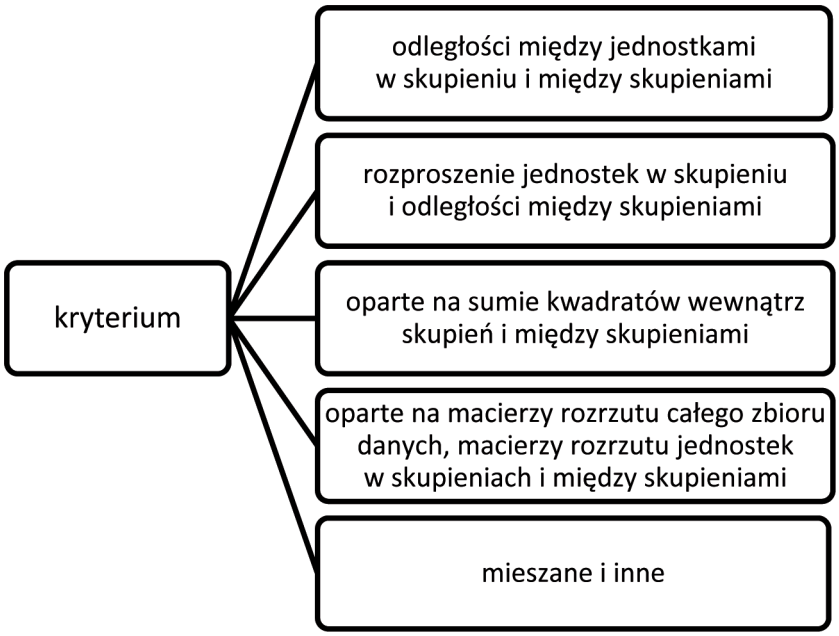
3. Wskaźniki oparte na sumie kwadratów wewnątrz skupień (*within-group sum of squares*) i między skupieniami (*between-group sum of squares*). Należą do nich: wskaźnik Balla-Halla (*Ball-Hall index*) (Ball G.H., Hall D.J. [1965]), wskaźnik Calińskiego-Harabasza (*Calinski-Harabasz index*) (Caliński T., Harabasz J.S. [1974]), wskaźnik Hartigana (*Hartigan index*) (Hartigan J.A. [1975]), wskaźnik Ratkowsky'ego-Lance'a (*Ratkowsky-Lance index*) (Ratkowsky D.A., Lance G.N. [1978]), wskaźnik Krzanowskiego-Lai (*Krzanowski-Lai index*) (Krzanowski W., Lai Y. [1988]), wskaźnik Xu (*Xu index*) (Xu L. [1997]). Interesujące zastosowanie wskaźnika Ratkowsky'ego-Lance'a zaprezentowano w artykule S. Dolnicara i F. Leischa [2004] w segmentacji turystów odwiedzających Austrię. Wyróżnione segmenty nazwano: turystyka aktywna, turystyka zdrowotna i turystyka relaksacyjna. Wskaźnik Krzanowskiego-Lai wykorzystano w artykule J.S. Larsona, E.T. Bradlowa i P.S. Fadera [2005] w klasyfikacji ścieżek zakupowych (*shopper travel path*) klientów pewnego supermarketu w zachodniej części Stanów Zjednoczonych.

4. Wskaźniki oparte na macierzy rozrzutu (*scatter matrix*) całego zbioru danych (**T),** macierzy rozrzutu jednostek w skupieniach (wewnątrzklasowej) (**W**) i między skupieniami (międzyklasowej) (**B**). Należą do nich: wskaźniki Friedmana-Rubina (*Trace*$\mathbf{W}^{(-1)}\mathbf{B}$**,** $|\mathbf{T}|$ / $|\mathbf{W}|$) (Friedman H.P., Rubin J. [1967]), wskaźnik Scotta-Symonsa (*Scott-Symons index*) (Scott A.J., Symons M.J. [1971]), wskaźnik Marriotta (*Marriott index*) (Marriott F.H.C. [1971]), wskaźnik TraceCov**W** (Milligan G.W., Cooper M.C. [1985]).

5. Wskaźniki mieszane i inne (np. oparte na rozmytej funkcji przynależności do skupień). Należą do nich: wskaźnik PC (*Bezdek's partition coefficient PC*) (Bezdek J.C. [1974a]), wskaźnik PE (*Partition entropy PE)* (Bezdek J.C. [1974b]), wskaźnik CE (*Bezdek's classification entropy*) (Bezdek J.C. [1981]), wskaźnik Xie-Beni (*Xie-Beni's separation measure*) (Xie X.L., Beni G. [1991]) i inne.

Problem ustalania optymalnej liczby klas prezentowali w pracach: Thorndike R.L. [1953], Ling R.F. [1972], Marriott F.H.C. [1971], Dunn J.C. [1973, 1974], Sneath P.H.A., Sokal R.R. [1973], Bezdek J.C. [1974a,b], Caliński T., Harabasz J.S. [1974], Baker F.B., Hubert L.J. [1975], Davies D.L., Bouldin D.W. [1979], Jain A.K., Dubes R.C [1979, 1988], Hubert L., Arabie P. [1985], Milligan G.W., Cooper M.C. [1985, 1987], Rousseeuw P.J. [1987], Krzanowski W.J., Lai Y.T. [1988], Gath I., Geva A.B. [1989], Kaufman L., Rousseeuw P.J. [1990], Wilson R., Spann M. [1990], Xie X.L., Beni G. [1991], Bensaid A.M., Hall L.O., Bezdek J.C., Clarke L.P., Silbiger M.L., Arrington J.A., Murtagh R.F. [1996], Celeux G., Soromenho G. [1996], Hardy A. [1996], Bezdek J.C., Pal N.R. [1998], Bandyopadhyay S., Maulik U. [2001], Halkidi

Schemat 2 Podstawowe kryteria podziału wskaźników jakości grupowania



Źródło: opracowanie własne

M., Batistakis Y., Vazirgiannis M. [2001], Tibshirani R., Walther G., Hastie T. [2001], Dimitriadou E., Dolničar S., Weingessel A. [2002], Dudoit S., Fridlyand J. [2002], Bolshakova N., Azuaje F. [2003], Mali K., Mitra S. [2003], Sugar C.A., James G.M. [2003], Chou C.H., Su M.C., Lai E. [2004], Sun H., Wang S., Jiang Q. [2004], Lam B.S.Y., Yan H. [2005, 2007], Migdał Najman K., Najman K. [2005, 2006, 2008], Steinley D. [2006], Walesiak M., Dudek A. [2006], Brun M., Sima C., Hua J., Lowey J., Carroll B., Suh E., Dougherty E.R. [2007], Cios K.J., Pedrycz W., Świniarski R.W., Kurgan L.A. [2007], Migdał Najman [2007], Halkidi M., Vazirgiannis M. [2008], Saitta S., Raphael B., Smith I.F.C. [2008], Wang J.S., Chiang J.C. [2008], Aliguliyev R.M. [2009], Muhlenbach F., Lallich S. [2009], Chicco G., Akilimali J.S. [2010], Lago-Fernández L.F., Corbacho F. [2010], Yue S., Wang J.S., Wu T., Wang H. [2010], Ming-Tso Chiang M., Mirkin B. [2010], Mirkin B. [2011] i inni.

B.S. Everitt, S. Landau, M. Leese w książce z 2001 roku „*Cluster analysis*" twierdzą, że nie ma optymalnego kryterium ustalania liczby klas w badanym zbiorze. Na podstawie badań symulacyjnych podobne wnioski zaprezentowali K. Migdał Najman, K. Najman [2005, 2006b, 2008], K. Migdał Najman [2007]. H.H. Bock w 1985 roku zastosował cztery testy do weryfikacji hipotezy zerowej stwierdzającej homogeniczność, jednorodność populacji wobec hipotezy alternatywnej zakładającej heterogeniczność, niejednorodność populacji. We wnioskach użył określenia, że żaden z czterech badanych rodzajów testów nie okazał się „idealny" (*no one of these tests is the „ideal"*)

w poszukiwaniu struktury grupowej. K. Jajuga potwierdził, że „Jest to problem nie rozwiązany dotychczas w sposób zadowalający" (Jajuga K. [1990]).

6. PODSUMOWANIE

Podobnie jak sam problem grupowania danych tak problem oceny jakości uzyskanej struktury grupowej jest bardzo złożony. Składa się na to duża liczba możliwych do uzyskania podziałów, wielość metod grupowania, z których każda posiada niemały czasem zbiór parametrów, wielość kryteriów i punktów widzenia na cel grupowania. Z tego powodu istnieje duża liczba wskaźników oceny jakości uzyskanej struktury grupowej. Tak jak różne są cele stawiane przed grupowaniem jednostek tak różne muszą być metody oceny ich realizacji. Nie istnieje jeden uniwersalny wskaźnik, który można stosować zawsze, niezależnie od rozwiązywanego problemu i zastosowanej metody. Każdy z nich w swojej konstrukcji uwzględnia jedynie część informacji o strukturze grupowej. Aby wybrać właściwy wskaźnik dla danego zastosowania konieczna jest znajomość szerokiego spektrum istniejących wskaźników i podstawowych ich własności.

W artykule dokonano syntetycznego przeglądu literatury tematu począwszy od prac P. Jaccarda z roku 1908 a skończywszy na pracach B. Mirkina z 2011 roku. Dokonano próby klasyfikacji znanych wskaźników jakości grupowania, uwzględniając kryteria pochodzące z różnych dyscyplin naukowych. W szczególności dokonano klasyfikacji wskaźników optymalnej liczby skupień jako podklasy wskaźników jakości grupowania. Wyniki prezentowanych badań powinny być użyteczne dla wszystkich zajmujących się problemami grupowania i klasyfikacji.

LITERATURA

[1] Albatineh A.N., Niewiadomska-Bugaj M. [2011], *MCS: A method for finding the number of clusters*, Journal of classification, 28, 2, 184-209.

[2] Aliguliyev R.M. [2009], *Performance evaluation of density-based clustering methods*, Information Sciences, 179, 20, 3583-3602.

[3] Anderberg M.R. [1973], *Cluster analysis for applications*, Academic Press, New York, San Francisco, London.

[4] Anderson A.J.B. [1971], *Numeric examination of multivariate soil samples*, Mathematical geology, 3, 1, 1-14.

[5] Arabie P., Boorman S.A. [1973], *Multidimensional scaling of measures of distance between partitions*, Journal of Mathematical Psychology, 10, 2, 148-203.

[6] Arabie P., Hubert L.J., Soete G.De (editors) [1996], *Clustering and classification*, World Scientific Publishing Co. Pte. Ltd., Singapore.

[7] Ayala G., Epifanio I., Simó A., Zapater V. [2006], *Clustering of spatial point patterns*, Computational Statistics & Data Analysis 50.

[8] Baker F.B. [1974], *Stability of two hierarchical grouping techniques, case 1: sensitivity to data errors*, Journal of the American Statistical Association, 69, 346, 440-445.

[9] Baker F.B., Hubert L.J. [1975], *Measurig the power of hierarchical cluster analysis*, Journal of the American Statistical Association, 70, 349, 31-38.

[10] Balicki A. [2009], *Statystyczna analiza wielowymiarowa i jej zastosowania społeczno-ekonomiczne*, Wydawnictwo Uniwersytetu Gdańskiego, Gdańsk.
[11] Ball G.H., Hall D.J. [1965], ISODATA, *A novel method of data analysis and pattern classification*, Technical report, Stanford Research Institute, Menlo Park, CA, (NTIS No. AD 699616).
[12] Bandyopadhyay S., Maulik U. [2001], *Nonparametric genetic clustering: comparison of validity indices*, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, 31, 1, 120-125.
[13] Ben-Hur A., Elisseeff A., Guyon I. [2002], *A stability based method for discovering structure in clustered data*, Pacific Symposium on Biocomputing 7, 6-17.
[14] Bensaid A.M., Hall L.O., Bezdek J.C., Clarke L.P., Silbiger M.L., Arrington J.A., Murtagh R.F. [1996], *Validity-guided (re)clustering with applications to image segmentation*, IEEE Transactions on, Fuzzy Systems, 4, 2, 112-123.
[15] Bezdek J.C. [1974a], *Numerical taxonomy with fuzzy sets*, Journal of Mathematical Biology, 1, 1, 57-71.
[16] Bezdek J.C. [1974b], *Cluster validity with fuzzy sets*, Journal of Cybernetics, 3, 58-72.
[17] Bezdek J.C. [1981], *Pattern recognition with fuzzy objective function algorithms*, Plenum Press, NY.
[18] Bezdek J.C., Pal N.R. [1998], *Some new indexes of cluster validity, Systems, Man, and Cybernetics, Part B: Cybernetics*, IEEE Transactions, 28, 3, 301-315.
[19] Bock H.H. [1985], *On some significance tests in cluster analysis*, Journal of classification, 2, 1, 77-108.
[20] Bolshakova N., Azuaje F. [2003], *Cluster validation techniques for genome expression data*, Signal Processing, 83, 4, 825-833.
[21] Brun M., Sima C., Hua J., Lowey J., Carroll B., Suh E., Dougherty E.R. [2007], *Model-based evaluation of clustering validation measures*, Pattern Recognition, 40, 3, 807-824.
[22] Bryan J. [2004], *Problems in gene clustering based on gene expression data*, Journal of multivariate analysis, 90, 1, 44-66.
[23] Caliński T., Harabasz J.S. [1974], *A dendrite method for cluster analysis*, Communications in Statistics – Theory and Methods, 3, 1, 1-27.
[24] Celeux G., Soromenho G. [1996], *An entropy criterion for assessing the number of clusters in a mixture model*, Journal of classification, 13, 2, 195-212.
[25] Chicco G., Akilimali J.S. [2010], *Renyi entropy-based classification of daily electrical load patterns*, IET Generation, Transmission & Distribution, 4, 6, 736-745.
[26] Chou C.H., Su M.C., Lai E. [2004], *A new cluster validity measure and its application to image compression*, Pattern Analysis & Applications, 7, 2, 205-220.
[27] Cios K.J., Pedrycz W., Świniarski R.W., Kurgan L.A. [2007], *Data mining, a knowledge discovery approach*, Springer Science+Business Media, LLC.
[28] Costner H.L. [1965], *Criteria for measures of association*, American Sociological Review, 30, 3, 341-353.
[29] Cunningham K.M., Ogilvie J.C. [1972], *Evaluation of hierarchical grouping techniques: a preliminary study*, The Computer Journal, 15, 3, 209-213.
[30] Davis J.A. [1967], *A partial coefficient for Goodman and Kruskal's gamma*, Journal of the American Statistical Association, 62, 317, 189-193.
[31] Davies D.L., Bouldin D.W. [1979], *A cluster separation measure*, IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-1, 2, 224-227.
[32] Dimitriadou E., Dolničar S., Weingessel A. [2002], *An examination of indexes for determining the number of clusters in binary data sets*, Psychometrika, 67, 3, 137-160.
[33] Ding C., Li T., Peng W. [2008], *On the equivalence between non-negative matrix factorization and probabilistic latent semantic indexing*, Computational statistics and data analysis, 52, 8, 3913-3927.

[34] Dolnicar S., Leisch F. [2004], *Segmenting markets by bagged clustering*, Australasian marketing journal, 12, 1, 51-65.
[35] Dubes R.C., Jain A.K. [1979], *Validity studies in clustering methodologies*, Pattern Recognition, 11, 4, 235-254.
[36] Duda R.O., Hart P.E., Stork D.G. [2002], *Pattern classification*, Wiley, New York.
[37] Dudoit S., Fridlyand J. [2002], *A prediction-based resampling method for estimating the number of clusters in a dataset*, Genome Biology, 3, 7, research 0036-research 0036.21.
[38] Dunn J.C. [1973], *A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters*, Journal of Cybernetics, 3, 3, 32-57.
[39] Dunn J.C. [1974], *Well separated clusters and optimal fuzzy partitions*, Journal of Cybernetics, 4, 1, 95-104.
[40] Engelman L., Hartigan J.A. [1969], *Percentage points of a test for clusters*, Journal of the American Statistical Association, 64, 328, 1647-1648.
[41] Everitt B. [1978], *Graphical techniques for multivariate data*, Heinemann Educational Books Ltd. London.
[42] Everitt B.S. [1979], *Unresolved problems in cluster analysis*, Biometrics, 35, 1, Perspectives in Biometry, 169-181.
[43] Everitt B.S. [1993], *Cluster Analysis*, Edward Arnold, London.
[44] Everitt B.S., Landau S., Leese M., Stahl D. [2011], *Cluster analysis*, 5th edition, John Wiley & Sons, Ltd., Chichester.
[45] Falasconi M., Pardo M., Vezzoli M., Sberveglieri G. [2007], *Cluster validation for electronic nose data*, Sensors and Actuators B: Chemical, 125, 2, 596-606.
[46] Farris J.S. [1969], *On the cophenetic correlation coefficient*, Systematic Biology, 18, 3, 279-285.
[47] Florek K., Łukasiewicz J., Perkal J., Steinhaus H., Zubrzycki S. [1951], *Taksonomia wrocławska*, Przegląd Antropologiczny, 17, 193-210.
[48] Fowlkes E.B., Mallows C.L. [1983], *A Method for Comparing two hierarchical clusterings*, Journal of the American Statistical Association, 78, 383, 553-569.
[49] Friedman H.P., Rubin J. [1967], *On some invariant criteria for grouping data*, Journal of the American Statistical Association, 62, 320, 1159-1178.
[50] Gath I., Geva A.B. [1989], *Unsupervised optimal fuzzy clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 11, 7, 773-781.
[51] Gatnar E., Walesiak M. [2004], *Metody statystycznej analizy wielowymiarowej w badaniach marketingowych*, AE Wrocław, 333-336.
[52] Goodman L.A., Kruskal W.H. [1954], *Measures of association for cross classifications*, Journal of the American Statistical Association, 49, 268, 732-764.
[53] Goodman L.A., Kruskal W.H. [1979], *Measures of association for cross classifications*, Sprinter-Verlag, New York, Heidelberg.
[54] Gordon A.D. [1987], *A review of hierarchical classification*, Journal of the Royal Statistical Society, Series A (General), 150, 2, 119-137.
[55] Gordon A.D. [1999], *Classification*, Chapman&Hall/CRC, Boca Raton.
[56] Gower J.C. [1966], *Some distance properties of latent root and vector methods used in multivariate analysis*, Biometrika, 53, 3/4, 325-338.
[57] Gower J.C. [1967], *A comparison of some methods of cluster analysis*, Biometrics, 23, 4, 623-638.
[58] Gower J.C. [1970], *Classification and geology*, Review of the International Statistical Institute, 38, 1, 35-41.
[59] Gurrutxaga I., Muguerza J., Arbelaitz O., Pérez J.M., Martín J.I. [2011], *Towards a standard methodology to evaluate internal cluster validity indices*, Pattern Recognition Letters, 32, 3, 505-515.
[60] Guttman L. [1968], *A general nonmetric technique for finding the smallest coordinate space for a configuration of points*, Psychometrika, 33, 2, 469-506.

[61] Günter S., Bunke H. [2003], *Validation indices for graph clustering*, Pattern Recognition Letters, 24, 8, 1107-1113.
[62] Halkidi M., Batistakis Y., Vazirgiannis M. [2001], *On clustering validation techniques*, Journal of Intelligent Information Systems, 17, 2-3, 107-145.
[63] Halkidi M., Gunopulos D., Vazirgiannis M., Kumar N., Domeniconi C. [2008], *A clustering framework based on subjective and objective validity criteria*, ACM Transactions on Knowledge Discovery from Data, 1, 4, 18, 18:1-18:25.
[64] Halkidi M., Vazirgiannis M. [2008], *NPCIu: an approach for clustering spatially extended objects*, Intelligent Data Analysis, 12, 6, 587-606.
[65] Handl J., Knowles J. [2004], *Multiobjective clustering with automatic determination of the number of clusters*, Technical Report TR- COMPSYSBIO-2004-02. UMIST, Manchester, UK.
[66] Hardy A. [1996], *On the number of clusters*, Computational Statistics & Data Analysis, 23, 1, 15, 83-96.
[67] Hartigan J.A. [1967], *Representation of similarity matrices by tree*, Journal of the American Statistical Association, 62, 320, 1140-1158.
[68] Hartigan J.A. [1975], *Clustering Algorithms*, New York: John Wiley.
[69] Hawkes R.K. [1971], *The multivariate analysis of ordinal measures*, The American Journal of Sociology, 76, 5, 908-926.
[70] Hennig C. [2008], *Dissolution point and isolation robustness: Robustness criteria for general cluster analysis methods*, Journal of multivariate analysis, 99, 6, 1154-1176.
[71] Hoffmann A., Motoda H., Scheffer T. (Eds.) [2005], *Discovery Science*, 8th International Conference, DS 2005, Singapore, October, Proceedings, Springer-Verlag Berlin Heidelberg, 302.
[72] Holgersson M. [1978], *The limited value of cophenetic correlation as a clustering criterion*, Pattern Recognition, 10, 4, 287-295.
[73] Holliday J.D., Hu C-Y., Willett P. [2002], *Grouping of coefficients for the calculation of intermolecular similarity and dissimilarity using 2D fragment bit-strings*, Combinatorial Chemistry & High Throughput Screening, 5, 2, 155-166.
[74] Hubert L.J. [1974], *Approximate evaluation techniques for the single-link and complete-link hierarchical clustering procedures*, Journal of the American Statistical Association, 69, 347, 698-704.
[75] Hubert L.J., Arabie P. [1985], *Comparing partitions*, Journal of Classification, 2, 1, 193-218.
[76] Hubert L.J., Levin J.R. [1976], *Evaluating object set partitions: Free-sort analysis and some generalizations*, Journal of Verbal Learning and Verbal Behavior, 15, 4, 459-470.
[77] Jaccard, P. [1908] *Nouvelles recherchés, sur la distribution florale.* Bulletin de la Société vaudoise des Sciences Naturelles, 44, 223-270.
[78] Jain A.K., Dubes R.C. [1988], *Algorithms for clustering data*, Englewood Cliffs NJ: Prentice Hall, Chapter 4.
[79] Jajuga K. [1990], *Statystyczna teoria rozpoznawania obrazów*, PWN, Warszawa.
[80] Jambu M. [1978], *Classification automatiqe pour l'analyse des donnees*, tom I, Paris Dunod.
[81] Jardine C.J., Jardine N., Sibson R. [1967], *The structure and construction of taxonomic hierarchies*, Mathematical Biosciences, 1, 2, 173-179.
[82] Jardine N., Sibson R. [1968], *The construction of hierarchic and non-hierarchic classification*, The computer journal, 11, 2, 177-184.
[83] Johnson S.C. [1967], *Hierarchical clustering schemes*, Psychometrika, 32, 3, 241-254.
[84] Kalinowski S.T. [2009], *How well to evolutionary trees describe genetic relationships among populations?*, Heredity, 102, 5, 506-513.
[85] Kaufman L., Rousseeuw P.J. [1990], *Finding groups in data: a introduction to cluster analysis*, Wiley, New York.
[86] Kendall M.G. [1945], *The treatment of ties in ranking problems*, Biometrika, 33, 3, 239-251.
[87] Kim J. [1971], *Predictive measures of ordinal association*, The American Journal of Sociology, 76, 5, 891-907.

[88] Kruskal J.B. [1964], *Nonmetric multidimensional scaling: a numerical method*, Psychometrika **29**, 2, 115-129.
[89] Kruskal J.B., Carroll J.D. [1969], *Geometrical models and badness-of-fit functions*. In Multivariate analysis (P.R.Krishnaiah, ed.), vol. II, (Proceedings of the 2. International Symposium on Multivariate Analysis held at Wright State Universaity, Dayton, Ohio, June 17-22, 1968), New York, Academic Press, 639-671.
[90] Krzanowski W.J., Lai Y.T. [1988], *A criterion for determining the number of groups in a data set using sum-of-squares clustering*, Biometrics, 44, 1, 23-34.
[91] Kulczynski S. [1927], *Die Pflanzenassociationen der Pienenen*, Bulletin International de L'Académie Polonaise des Sciences et des lettres, Classe des sciences mathématiques et naturelles, Série B, Supplément II, 2, 57-203.
[92] Kuramae E.E., Robert V., Echavarri-Erasun C., Boekhout T. [2007], *Cophenetic correlation analysis as a strategy to select phylogenetically informative proteins: an example from the fungal kingdom*, BMC Evolutionary Biology, 7, 134.
[93] Labatut V., Cherifi H. [2011], *Accuracy Measures for the comparision of classifiers*, ICIT, The 5^{th} International Conference on Information Technology (artykuł wygłoszony na konferencji ICIT).
[94] Lago-Fernández L.F., Corbacho F. [2010], *Normality-based validation for crisp clustering*, Pattern Recognition, 43, 3, 782-795.
[95] Lam B.S.Y., Yan H. [2005], *A new cluster validity index for data with merged clusters and different densities*, IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC), 1, 798-803.
[96] Lam B.S.Y., Yan H. [2007], *Assessment of microarray data clustering results based on a new geometrical index for cluster validity*, Soft Computing, 11, 4, 341-348.
[97] Lance G.N., Williams W.T. [1966a], *Computer programs for hierarchical polythetic classification ("Similarity analysis")*, The Computer Journal, 9, 1, 60-64.
[98] Lance G.N., Williams W.T. [1966b], *A generalized sorting strategy for computer classifications*, Nature 212, 218 (8 October), Letters to Nature.
[99] Lance G.N., Williams W.T. [1967a], *A general theory of classificatory sorting strategies*, I. Hierarchical systems, The Computer Journal, 9, 4, 373-380.
[100] Lance G.N., Williams W.T. [1967b], *A general theory of classificatory sorting strategies: II. Clustering systems*, The Computer Journal, 10, 3, 271-277.
[101] Larson J.S., Bradlow E.T., Fader P.S. [2005], *An exploratory look at supermarket shopping paths*, International Journal of Research in Marketing, 22, 4, 395-414.
[102] Ling R.F. [1972], *On the theory and consucion of k-clusters*, The computer journal, 15, 4, 326-332.
[103] Mahalanobis P.C. [1936], *On the generalised distance in statistics*, Proceedings of the National Institute of Sciences of India, 2, 1, 49-55.
[104] Maimon O., Rokach L. (editors) [2005], *The data mining and knowledge discovery handbook*, Halkidi M., Vazirgiannis M.: Quality assessment approaches in data mining, Springer.
[105] Mali K., Mitra S. [2003], *Clustering and its validation in a symbolic framework*, Pattern Recognition Letters, 24, 14, 2367-2376.
[106] Marriott F.H.C. [1971], *Practical problems in a method of cluster analysis*, Biometrics, 27, 3, 501-514.
[107] McQuitty L.L. [1960], *Hierarchical linkage analysis for the isolation of types*, Educational and Psychological Measurement, 20, 1, 55-67.
[108] McQuitty L.L. [1966], *Similarity analysis by reciprocal pairs for discrete and continuous data*, Educational and Psychological Measurement, 26, 4, 825-831.
[109] McQuitty L.L. [1967], *Expansion of similarity analysis by reciprocal pairs for discrete and continuous data*, Educational and Psychological Measurement, 27, 2, 253-255.
[110] Meilă M. [2007], *Comparing clusterings – an information based distance*, Journal of multivariate analysis, 98, 5, 873-895.

[111] Mérigot B., Durbec J.P., Gaertner J.C. [2010], Ecology, 91, 6, 1850-1859.
[112] Migdał Najman K. [2007], *Analiza podobieństwa wyników grupowania uzyskanych w oparciu o metodę k-średnich dla wybranych metod ustalania optymalnej liczby skupień*, Prace i Materiały WZ UG, 5, 601-610.
[113] Migdał Najman K. [2010], *Zastosowanie samouczącej się sieci neuronowej typu SOM w analizie koszykowej*, Taksonomia 17, Prace Naukowe UE we Wrocławiu, 305-315.
[114] Migdał Najman K. [2011], *Propozycja hybrydowej metody grupowania opartej na sieciach samouczących*, Referat wygłoszony na konferencji: SKAD 2011, Wągrowiec.
[115] Migdał Najman K., Najman K. [2005], *Analityczne metody ustalania liczby skupień*, Taksonomia 12, Prace Naukowe AE we Wrocławiu, 1076, 265-273.
[116] Migdał Najman K., Najman K. [2006a], *Wykorzystanie indeksu silhouette do ustalania optymalnej liczby skupień*, Wiadomości Statystyczne, 6, 1-10.
[117] Migdał Najman K., Najman K. [2006b], *Analityczne metody ustalania skupień w rozmytych zbiorach danych*, Taksonomia 13, Prace Naukowe AE we Wrocławiu, 1126, 159-167.
[118] Migdał Najman K., Najman K. [2007], *Charakterystyka mierników oceny podobieństwa wyników podziałów*, Prace i Materiały WZ UG, 3, 191-201.
[119] Migdał Najman K., Najman K. [2008], *Wykorzystanie wskaźnika Dunn'a do ustalania optymalnej liczby skupień*, Wiadomości Statystyczne, 11, 26-34.
[120] Milligan G.W., Cooper M.C. [1985], *An examination of procedures for determining the number of clusters in a data set*, Psychometrika, 50, 2, 159-179.
[121] Milligan G.W., Cooper M.C. [1987], *Methodology review: clustering methods*, Applied psychological measurement, 11, 4, 329-354.
[122] Ming-Tso Chiang M., Mirkin B. [2010], *Intelligent choice of the number of clusters in k-means clustering: an experimental study with different cluster spreads*, Journal of classification, 27, 1, 3-40.
[123] Mirkin B. [1996], *Mathematical classification and clustering*, Nonconvex optimization and its applications, Kluwer Academic Publishers, Dordrecht.
[124] Mirkin B. [2011], *Choosing the number of clusters*, WIREs Data Mining and Knowledge Discovery, vol. 1, May/June, John Wiley&Sons, Inc., 252-259.
[125] Muhlenbach F., Lallich S. [2009], *A new clustering algorithm based on regions of influence with self-detection of the best number of clusters*, Data Mining, 2009, ICDM'09, Ninth IEEE International Conference on Data Mining, Miami, Florida, 884-889.
[126] Nowak E. [1985], *Wskaźnik podobieństwa wyników podziału*, Przegląd Statystyczny, 1, 41-48.
[127] Pal N.R., Biswas J. [1997], *Cluster validation using graph theoretic concepts*, Pattern Recognition, 30, 6, 848-849.
[128] Pitchandi P., Raju N. [2011], *Improving the performance of multivariate Bernoulli model based documents clustering algorithms using transformation techniques*, Journal of Computer Science, 7, 5, 762-769.
[129] Podani J. [1989], *New combinatorial clustering methods*, Vegetatio (Plant ecology), 81, 61-77.
[130] Rand W.M. [1971], *Objective criteria for the evaluation of clustering methods*, Journal of the American Statistical Association, 66, 336, 846-850.
[131] Ratkowsky D.A., Lance G.N. [1978], *A criterion for determining the number of groups in a classification*, Australian Computer Journal, 10, 3, 115-117.
[132] Raymond T.N., Jiawei H. [1994], *Efficient and effective clustering methods for spatial data mining*, Tech. Rep. TR-93-13, University of British Columbia, Vancouver, B.C., Canada.
[133] Rendón E., Abundez I.M., Gutierrez C., Diaz Zagal S., Arizmendi A., Quiroz E.M., Arzate E.H. [2011], *A comparison of internal and external cluster validation indexes*, in Applications of mathematics & computer engineering, Zemliak A., Mastorakis N. (editors), American Conference on Applied Mathematics (AMERICAN-MATH'11), 5^{TH} WSEAS (World Scientific and Engine-

ering Academy and Society) International Conference on Computer Engineering and Applications (CEA'11), Puerto Morelos, Mexico, Published by WSEAS Press, 158-163.
[134] Reulke R., Eckardt U., Flach B., Knauer U., Polthier K. (Eds.) [2006], *Combinatorial Image Analysis*, 11^{th} International Workshop, IWCIA, Berlin, Germany, June, Springer-Verlag, 108-110.
[135] Rohlf F.J. [1970], *Adaptive hierarchical clustering schemes*, Systematic Biology, 19, 1, 58-82.
[136] Rohlf F.J. [1974], *Methods of comparing classifications*, Annual Review of Ecology and Systematics, 5, 1, 101-113.
[137] Rohlf F.J. [1982], *Consensus Indices for Comparing Classifications*, Mathematical Biosciences, 59.
[138] Rohlf F.J., Fisher D.R. [1968], *Tests for hierarchical structure in random data sets*, Systematic Biology, 17, 4, 407-412.
[139] Rousseeuw P.J. [1987], *Silhouettes: a graphical aid to the interpretation and validation of cluster analysis*, Journal of of computational and applied mathematics, 20, 1, 53-65.
[140] Rousson V. [2007], *The gamma coefficient revisited*, Statistics & Probability Letters, 77, 17, 1696-1704.
[141] Rubinov A.M., Soukhorokova N.V., Ugon J. [2006], *Classes and clusters in data analysis*, European Journal of Operational Research, 173, 3, 849-865.
[142] Saitta S., Raphael B., Smith I.F.C. [2007], *A bounded index for cluster validity*, Machine Learning and Data Mining in Pattern Recognition, Heidelberg, Germany, Springer, 174-187.
[143] Saitta S., Raphael B., Smith I.F.C. [2008], *A comprehensive validity index for clustering*, Intelligent Data Analysis, 12, 6, 529-548.
[144] Sammon J.W. [1969], *A nonlinear mapping for data structure analysis*, IEEE Transactions on computers, C-18, 5, 401-409.
[145] Schepers J., Ceulemans E., Van Mechelen I. [2008], *Selecting among multi-mode partitioning models of different complexities: a comparison of four model selection criteria*, Journal of Classification, 25, 1, 67-85.
[146] Schultz J.V., Hubert L.J. [1979], *A nonparametric test for the correspondence between two proximity matrices*, Journal of Educational Statistics, 1, 1, 59-67.
[147] Scott A.J., Symons M.J. [1971], *Clustering methods based on likelihood ratio criteria*, Biometrics, 27, 2, 387-397.
[148] Seung-Seok C., Sung-Hyuk C., Tappert C.C. [2010], *A survey of binary similarity and distance measures*, Journal of Systemics, Cybernetics & Informatics, 8, 1, 43-48.
[149] Sneath P.H.A. [1957], *The application of computers to taxonomy*, Journal of General Microbiology, 17, 201-226.
[150] Sneath P.H.A., Sokal R.R. [1973], *Numerical Taxonomy*. The principles and practice of numerical classification, W.H. Freeman & Company, San Francisco, CA.
[151] Sokal R.R., Michener C.D. [1958], *A statistical method for evaluating systematic relationships*, The University of Kansas Scientific Bulletin 38, 1409-1438.
[152] Sokal R.R., Rohlf F.J. [1962], *The comparison of dendrograms by objective methods*, Taxon, 11, 2, 33-40.
[153] Sokal R.R., Sneath P.H.A. [1963], *Principles of numerical taxonomy*, W.H. Freeman & Company, San Francisco, CA.
[154] Somers R.H. [1962a], *A new asymmetric measure of association for ordinal variables*, American Sociological Review, 27, 6, 799-811.
[155] Somers R.H. [1962b], *A similarity between Goodman and Kruskal's tau and Kendall's tau, with a partial interpretation of the latter*, Journal of the American Statistical Association, 57, 300, 804-812.
[156] Spath H. [1980], *Cluster Analysis Algorithms*, Ellis Horwood.
[157] Steinley D. [2006], *K-means clustering: a half-century synthesis*, British Journal of Mathematical and Statistical Psychology, 59, 1, 1-34.

[158] Sugar C.A., James G.M. [2003], *Finding the number of clusters in a dataset: an information theoretic approach*, Journal of the American Statistical Association, 98, 463, 750-763.
[159] Sun H., Wang S., Jiang Q. [2004], *FCM-based model selection algorithms for determining the number of clusters*, Pattern Recognition, 37, 10, 2027-2037.
[160] Szmigiel C. [1976], *Wskaźnik zgodności kryteriów podziału*, Przegląd Statystyczny, 4.
[161] Tanimoto T. [1958], *An elementary mathematical theory of classification and prediction*, Internal Report, IBM Corp.
[162] Theodoridis S., Koutroumbas K. [1999], *Pattern recognition*, Elsevier, Academic Press.
[163] Theodoridis S., Koutroumbas K. [2003], *Pattern recognition*, second edition, Elsevier, Academic Press.
[164] Thorndike R.L. [1953], *Who belongs in the family?*, Psychometrika, 18, 4, 267-276.
[165] Tibshirani R., Walther G., Hastie T. [2001], *Estimating the number of clusters in a data set via the gap statistic*, Journal of the Royal Statistical Society; Series B (Statistical Methodology), 63, 2, 411-423.
[166] Trakhtenbrot A., Kadmon R. [2006], *Effectiveness of environmental cluster analysis in representing regional species diversity*, Conservation Biology, 20, 4, 1087-1098.
[167] Walesiak M., Dudek A. [2006], *Symulacyjna optymalizacja wyboru procedury klasyfikacyjnej dla danego typu danych – charakterystyka problemu*, Zeszyty Naukowe Uniwersytetu Szczecińskiego nr 450, Prace Katedry Ekonometrii i Statystyki nr 17, 635-646.
[168] Wallace D.L. [1983], *A method for comparing two hierarchical clustering: comment*, Journal of the American Statistical Association, 78, 383, 569-576.
[169] Wanek D. [2003], *Fuzzy spatial analysis techniques in a business GIS environment*, ERSA 2003 Congress, University of Jyväskylä, Finland.
[170] Ward J.H. [1963], *Hierarchical grouping to optimize an objective function*, Journal of the American Statistical Association, 58, 301, 236-244.
[171] Wilson R., Spann M. [1990], *A new approach to clustering*, Pattern Recognition, 23, 12, 1413-1425.
[172] Wishart D. [1969], *An algorithm for hierarchical classifications*, Biometrics, 25, 1, 165-170.
[173] Xie X.L., Beni G. [1991], *A validity measure for fuzzy clustering*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 13, 8, 841-847.
[174] Xu L. [1997], *Bayesian Ying-Yang machine, clustering and number of clusters*, Pattern Recognition Letters, 18, 11-13, 1167-1178.
[175] Yeung K.Y., Ruzzo W.L. [2001], *Details of the adjusted Rand index and clustering algorithms*, suplement do artykułu: An empirical study on principal component analysis for clustering gene expression data, Bioinformatics, 17, 9, 763-774.
[176] Yue S., Wang J.S., Wu T., Wang H. [2010], *A new separation measure for improving the effectiveness of validity indices*, Information Sciences, 180, 5, 748-764.
[177] Zhao Y., Karypis G. [2004], *Empirical and theoretical comparisons of selected criterion functions for document clustering*, Machine Learning, 55, 3, 311-331.
[178] Zhong S., Ghosh J. [2003], *A comparative study of generative models for document clustering*, The University of Texas at Austin, http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.8.4583&rep=rep1&type=pdf.

OCENA JAKOŚCI WYNIKÓW GRUPOWANIA – PRZEGLĄD BIBLIOGRAFII

S t r e s z c z e n i e

W artykule dokonano syntetycznego przeglądu literatury tematu począwszy od prac P. Jaccarda z roku 1908 a skończywszy na pracach B. Mirkina z 2011 roku. Dokonano próby klasyfikacji znanych wskaźników jakości grupowania, uwzględniając kryteria pochodzące z różnych dyscyplin naukowych. W szczególności dokonano klasyfikacji wskaźników optymalnej liczby skupień jako podklasy wskaźników jakości grupowania. Wyniki prezentowanych badań powinny być użyteczne dla wszystkich zajmujących się problemami grupowania i klasyfikacji.

Słowa kluczowe: analiza skupień, wskaźniki jakości grupowania

CLUSTER VALIDITY MEASUREMENT – A BIBLIOGRAPHY REVIEW

S u m m a r y

In the article are presented the synthetic review of the literature from P. Jaccard in 1908 to B. Mirkin, 2011. In this paper, the concept and classification of cluster validity indices are proposed. There are presented classification of validity indices to find the optimal number of clusters. The results of this study should be useful for all concerned with the problems of classification.

Keywords: cluster analysis, cluster validity indices

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

BEATA BAL-DOMAŃSKA, JUSTYNA WILK

GOSPODARCZE ASPEKTY ZRÓWNOWAŻONEGO ROZWOJU WOJEWÓDZTW – WIELOWYMIAROWA ANALIZA PORÓWNAWCZA

1. WPROWADZENIE

Pojęcie rozwoju jest kluczowe dla gospodarki i nauk ekonomicznych. W definicjach tego pojęcia podkreśla się znaczenie zmian ilościowych i jakościowych zachodzących w danym obszarze (np. społecznym, gospodarczym). Zrównoważony rozwój jest procesem integrującym zjawiska społeczne, gospodarcze oraz środowiskowe w sposób zapewniający rozwój w długim okresie, dla obecnych i przyszłych pokoleń. Zawarta w definicji zrównoważonego rozwoju z 1987 roku w Raporcie Światowej Komisji Środowiska i Rozwoju ONZ wizja uwzględnia zarówno populację ludzką, jak i świat zwierząt i roślin, ekosystemy, zasoby naturalne Ziemi, surowce energetyczne, a także w sposób zintegrowany traktuje najważniejsze wyzwania stojące przed światem, takie jak walka z ubóstwem, równość płci, prawa człowieka i jego bezpieczeństwo, edukacja dla wszystkich, zdrowie, dialog międzykulturowy [3]. W kontekście zrównoważonego rozwoju zwraca się uwagę na trzy jego cechy: samopodtrzymywanie, trwałość i zrównoważenie [1].

Zrównoważony rozwój jest procesem mającym na celu poprawę jakości życia i dobrobytu pokoleń w długim okresie. Oceniając sytuację w tym zakresie należy zwrócić uwagę na dokonywany przez regiony postęp w kierunku wyznaczonych celów lub jego brak. Istotną rolę w procesie dążenia do zrównoważonego rozwoju odgrywa monitoring efektów podejmowanych działań. W artykule przedstawiono propozycję oceny i porównania sytuacji województw w kontekście gospodarczych aspektów zrównoważonego rozwoju z wykorzystaniem miar syntetycznych wykorzystujących wspólny wzorzec rozwoju. Narzędzie to wydaje się szczególnie przydatne do realizacji celu niniejszej pracy, a mianowicie oceny zróżnicowania poziomu i dynamiki zmian.

Analizę przeprowadzono w dwóch ujęciach – przestrzennym i czasowym. Pierwsze ujęcie dotyczy oceny zróżnicowania województw pod względem wartości wskaźników charakteryzujących kluczowe obszary ładu gospodarczego w 2005 i 2009 roku. Drugie ujęcie odnosi się do oceny postępów województw w kierunku zrównoważonego rozwoju w 2009 roku w odniesieniu do 2005 roku.

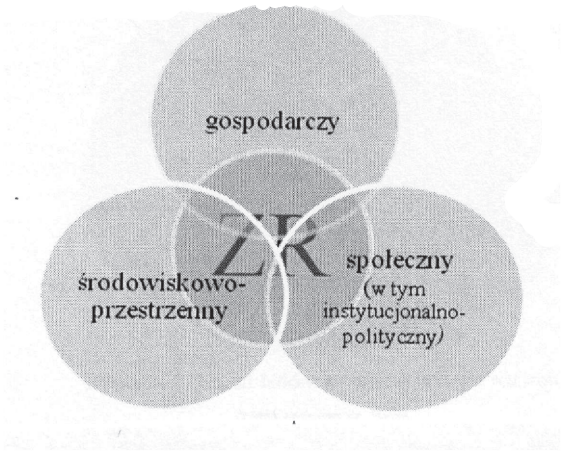
Struktura artykułu jest następująca: w pierwszej części określony został gospodarczy wymiar zrównoważonego rozwoju oraz przyjęta procedura badawcza. W drugiej części przedstawiono wyniki analizy sytuacji województw w obszarach opisujących ład

gospodarczy z wykorzystaniem podstawowych miar statystycznych oraz syntetycznego wzorca rozwoju.

2. GOSPODARCZY WYMIAR ZRÓWNOWAŻONEGO ROZWOJU

Zrównoważony rozwój jest zjawiskiem długotrwałym i złożonym. W założeniach Unii Europejskiej obejmuje on takie kwestie, jak rozwój społeczno-gospodarczy, zrównoważona konsumpcja i produkcja, włączenie społeczne, zmiany demograficzne, zdrowie publiczne, zmiany klimatu i gospodarowanie energią, zrównoważony transport, zasoby naturalne, dobre rządzenie i globalne partnerstwo [2].

Nadrzędnym celem procesu dążenia do zrównoważonego rozwoju, z uwzględnieniem tych obszarów, jest trwała poprawa jakości życia poprzez kształtowanie właściwych proporcji w wymiarze gospodarczym, środowiskowo-przestrzennym i społecznym, w tym instytucjonalno-politycznym [1]. Zjawiska te są w stosunku do siebie komplementarne i współzależne (rys. 1). W tej koncepcji rozwój gospodarczy rozpatrywany jest z uwzględnieniem potrzeb obywateli i rozwoju kapitału społecznego (np. równego dostępu do rynku pracy, podniesienia poziomu wykształcenia ludności), a także poszanowaniem środowiska naturalnego i korzystania z jego zasobów w sposób pozwalający na zachowanie funkcji ekosystemów w perspektywie długookresowej.



Rysunek 1. Układ ładów zrównoważonego rozwoju

Źródło: opracowanie własne

3. PROCEDURA BADAWCZA

Ocena zrównoważonego rozwoju nie jest zadaniem łatwym. Pod uwagę należy wciąć spektrum wzajemnie powiązanych zjawisk, których oddziaływanie często przekracza jeden aspekt. Dlatego zadaniem szczególnie trudnym jest zdefiniowanie wskaźników jednoznacznie obrazujących problem, mierzalnych, dla których istnieje pokrycie

informacyjne. Kolejnym problemem jest wskazanie optymalnych wartości wskaźnika. Dla większości wskaźników zrównoważonego rozwoju nie ma możliwości wyznaczenia ilościowych celów. Dzieje się tak m.in. z powodu braku w teorii jednoznacznych wskazówek, co do pożądanego poziomu zjawisk oraz trudności w określeniu dla każdego ze wskaźników wspólnego dla wszystkich jednostek (województw) poziomu docelowego.

W niniejszej pracy skupiono się na wybranych aspektach rozwoju gospodarczego, dla których jako przybliżenie progów docelowych wykorzystano najwyższe (dla stymulant) lub najniższe (dla destymulant) wartości. Nie oznacza to jednak, że wartości te są optymalne. Wyznaczają one najbardziej korzystne, realnie osiągnięte wartości. Może się zdarzyć, że najkorzystniejsze wartości, wciąż są duże niższe niż wartości pożądane (docelowe). W takich przypadkach wartości miar syntetycznych wskazują na bliższą pozycję względem wzorca niż wynikałoby to z sytuacji optymalnej.

Badaniem objęto zmiany wartości 15 cech diagnostycznych (por. tab. 1) charakteryzujących gospodarcze aspekty zrównoważonego rozwoju województw zgrupowane w pięciu obszarach:

1. Gospodarka,
2. Innowacyjność,
3. Gospodarstwa domowe,
4. Rynek pracy,
5. Oddziaływanie na środowisko.

Dla uzyskania porównywalności skali zjawisk wszystkie cechy diagnostyczne zostały wyrażone jako wartości relatywne w postaci wskaźników (struktury lub natężenia). Przy doborze cech diagnostycznych kierowano się ich merytoryczną przydatnością do oceny zrównoważonego rozwoju w przestrzeni wszystkich województw oraz dostępnością porównywalnych danych dla wytypowanych lat.

Ocena sytuacji zrównoważonego rozwoju województw w kontekście ładu gospodarczego objęła:

– analizę wskaźnikową każdego obszaru na podstawie wybranych statystyk opisowych,
– konstrukcję rankingów z wykorzystaniem syntetycznej miary rozwoju (SMR_{it}^{m}) w każdym z obszarów,
– ocenę postępu poszczególnych województw w kierunku zrównoważonego rozwoju w 2009 roku w porównaniu z sytuacją w 2005 roku.

Analiza miała charakter statyczny (w zakresie poziomu oraz międzywojewódzkiego zróżnicowania wartości wskaźników) oraz dynamiczny, obejmujący porównanie sytuacji województw w latach 2005 i 2009. W celu porównania zmian zachodzących w dwóch rozpatrywanych latach wykorzystano procedurę uwzględniającą wspólny wzorzec rozwoju[1].

[1] O konstrukcji miar ze wspólnym wzorcem rozwoju zob. szerzej w: M. Walesiak, *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wyd. AE, Wrocław 2006.

Podstawą konstrukcji rankingów w ramach obszarów i ustalenia pozycji poszczególnych województw było wyznaczenie wartości syntetycznej miary rozwoju (SMR_{it}^{m}). Procedura obliczeniowa objęła następujące kroki:

1. Normalizacja cech diagnostycznych z wykorzystaniem formuły unitaryzacji zerowanej:

$$z_{itj} = \frac{x_{itj} - \min\limits_{i} x_{itj}}{\max\limits_{i} x_{itj} - \min\limits_{i} x_{itj}}, \tag{1}$$

gdzie: z_{itj} – wartość j–tej cechy diagnostycznej (wskaźnika) $j = 1, 2,\ldots, k$ w i–tym obiekcie $i = 1, 2,\ldots, N$(województwie) w t-tym roku (t= 2005, 2009) po unitaryzacji; x_{itj} – realizacja j-tej cechy diagnostycznej w i-tym obiekcie; $\min x_{itj}$ ($\max x_{itj}$) – najmniejsza (największa) wartość j-tej cechy diagnostycznej x_{ij}.

W wyniku unitaryzacji wartości cech zawarte są w przedziale [0; 1]. Przy czym wartość 0 przyjmuje województwo, dla którego wskaźnik osiągnął w latach 2005 i 2009 wartość najniższą (minimalną), a 1 – wartość najwyższą. Po zastosowaniu unitaryzacji zerowanej zmienna mierzona jest na skali przedziałowej z umownym zerem na poziomie wartości minimalnej.

2. Ze względu na trudności z określeniem wartości optymalnej dla nominat bądź realizację wartości wskaźników znacznie poniżej potencjalnej wartości nominalnej, wszystkie wskaźniki potraktowano jako stymulanty bądź destymulanty. W związku z koniecznością ujednolicenia charakteru (preferencji) cech dokonano zamiany destymulant z^D na stymulanty z^S z wykorzystaniem formuły:

$$z_{itj}^{S} = 1 - z_{itj}^{D}. \tag{2}$$

3. Ustalenie współrzędnych obiektu-wzorca. Obiekt-wzorzec stanowią najkorzystniejsze wartości wskaźników uzyskane łącznie w latach 2005 i 2009. Jako wzorcowy przyjęto górny wzorzec rozwoju, tzn. za najkorzystniejsze wartości cech diagnostycznych w przypadku stymulant uznano wartości maksymalne, a destymulant – wartości minimalne. Wektor wzorca z_0 przyjmuje formę:

$$z_0 = \left[z_{01}\, z_{02}\, \ldots . z_{0j}\right]$$

$$\text{taki, że:} \quad \begin{array}{l} z_{0j} = \max z_{itj} \quad dla \quad stymulanty \\ z_{0j} = \min z_{itj} \quad dla \quad destymulant, \end{array} \tag{3}$$

natomiast antywzorca $z{-}_0$:

$$z_{-0} = \left[z_{-01}\ z_{-02}\ \ldots z_{-0j}\right]$$

$$\text{taki, że:} \quad \begin{array}{l} z_{-0j} = \min z_{itj} \quad dla \quad stymulanty \\ z_{-0j} = \max z_{itj} \quad dla \quad destymulanty, \end{array} \tag{4}$$

gdzie: z_{0j}– znormalizowana wartość (w naszym przypadku po unitaryzacji) j-tego wskaźnika we wzorcowym obiekcie, z_{itj} – znormalizowana wartość j-tego wskaźnika w i-tym obiekcie w okresie t.

4. Określenie dystansu i-tego województwa do obiektu-wzorca z wykorzystaniem odległości euklidesowej d_{i0} w m-tym ($m = 1, 2, \ldots, M$) obszarze, na podstawie formuły:

$$d_{it0}^{m} = \sqrt{\sum_{j=1}^{k} \left(z_{itj}^{m} - z_{0j}^{m}\right)^2}, \tag{5}$$

gdzie: d_{it0}^{m} – odległość euklidesowa między i-tym województwem w t-tym okresie i obiektem-wzorcem w m-tym obszarze.

5. Wyznaczenie wartości syntetycznej miary rozwoju (SMR_{it}^{m}) w m-tym obszarze w t-tym okresie dla i-tego województwa według formuły:

$$SMR_{it}^{m} = 1 - \frac{d_{it0}^{m}}{d_{0}^{m}}, \tag{6}$$

gdzie: d_{0}^{m} – odległość między wzorcem i antywzorcem w m-tym obszarze.

Wartości SMR_{it}^{m} są unormowane w przedziale [0, 1]. Najlepszą sytuację osiągnie województwo, dla którego wartość SMR_{it}^{m} wyniesie 1. Oznacza to, że województwo osiągnęło najkorzystniejsze (wzorcowe) wartości dla wszystkich wskaźników zrównoważonego rozwoju w danym obszarze. Poszczególne przedziały wartości SMR_{it}^{m} interpretowano jako sytuację:

– [0,0; 0,2] bardzo niekorzystną,
– (0,2; 0,4] niekorzystną,
– (0,4; 0,6] umiarkowaną,
– (0,6; 0,8] korzystną,
– (0,8; 1,0] bardzo korzystną.

Na podstawie SMR_{i2009}^{m} określono także przewagi i ograniczenia województw. Za przewagę uznano obszar, w którym województwo osiągnęło wartość na poziomie wyższym bądź równym 0,6. Natomiast jako zagrożenie określono obszar, dla którego wartość ta nie przekroczyła 0,4.

6. Do oceny postępów województw w m-tym obszarze w kierunku zrównoważonego rozwoju wykorzystano wartości przyrostów SMR_{it}^{m}. Jeżeli $SMR_{i2009}^{m} - SMR_{i2005}^{m} \geqslant 0,2$[2] uznano, iż w województwie wystąpił wyraźny postęp w kierunku zrównoważonego rozwoju. Natomiast, sytuację $SMR_{i2009}^{m} - SMR_{i2005}^{m} \leqslant -0,1$ uznano za zagrożenie dla rozwoju województwa (pogorszenie się sytuacji).
7. Na podstawie SMR_{it}^{m} obliczonych dla poszczególnych obszarów wyznaczono średnią wartość SMR_{it} w województwie:

[2] Za wystarczający można byłoby uznać 10% wzrost SMR_{it}^{m}, jednakże ze względu na relatywnie dużą liczbę województw osiągających taki wynik prezentację ograniczono do regionów najszybciej podążających ścieżką zrównoważonego rozwoju.

$$SMR_{it}^{=}\frac{1}{M}\sum_{m=1}^{M}SMR_{it}^{m}. \tag{7}$$

8. Zgodność otrzymanych w 2005 i 2009 roku uporządkowań województw porównano z wykorzystaniem współczynnika korelacji liniowej Pearsona.

4. WSTĘPNA ANALIZA WSKAŹNIKÓW ZRÓWNOWAŻONEGO ROZWOJU WOJEWÓDZTW

W tab. 1 przedstawiono wartości podstawowych statystyk opisowych oraz klasycznego współczynnika zmienności opartego na średniej arytmetycznej i odchyleniu standardowym dla wskaźników ZR według obszarów[3]. Jak wynika z analizy, większość wskaźników charakteryzuje się umiarkowanym stopniem zróżnicowania między województwami. Najmniejsze zróżnicowanie odnotowano w obszarze "Gospodarstwa domowe" i "Rynek pracy". W tych dwóch obszarach można uznać, że sytuacja między województwami jest wyrównana i nie występują znaczne dysproporcje rozwojowe.

Istotne różnice między województwami odnotowano dla wybranych wskaźników w obszarach:

1. „Gospodarka" – dla wskaźnika „Udział gospodarstw rolnych wielkopowierzchniowych w liczbie gospodarstw ogółem",
2. „Oddziaływanie na środowisko" – dla 2 wskaźników: „Emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych dla środowiska" oraz „Zużycie wody na potrzeby gospodarki narodowej i ludności na 1 mieszkańca".

W przypadku udziału gospodarstw wielkopowierzchniowych (często monokulturowych) w całkowitej liczbie gospodarstw rolnych, wysoki ich udział jest przejawem uproszczenia struktury krajobrazu i źródłem zanieczyszczenia wód oraz degradacji gleb, co zwykle oddziałuje negatywnie na różnorodność biologiczną i krajobraz rolniczy, a także „ekologiczności" produktów rolnych. Najwięcej gospodarstw wielko-powierzchniowych (powyżej 1% wszystkich gospodarstw rolnych przy przeciętnej wielkości dla Polski wynoszącej 0,35%) było w 2005 roku w województwach: zachodniopomorskim i warmińsko-mazurskim, w 2009 roku do tej grupy dołączyły: dolnośląskie, pomorskie i lubuskie.

Województwa charakteryzowały się dużym zróżnicowaniem także pod względem emisji zanieczyszczeń gazowych z zakładów szczególnie uciążliwych. W województwie śląskim (3074 t/km^2), o najwyższej emisji zanieczyszczeń gazowych, wielkość ta była około 53 krotnie wyższa niż w warmińsko-mazurskim – o najniższym poziomie emisji. Wysoką emisję odnotowano także w opolskim i łódzkim (powyżej 1400 t/km^2). Niska emisja charakteryzowała, oprócz warmińsko-mazurskiego, województwo podlaskie (poniżej 90 t/km^2).

[3] Źródłem danych w konstrukcji wskaźników był Bank Danych Lokalnych Głównego Urzędu Statystycznego.

Tablica 1. Wartości statystyk opisowych dla wskaźników charakteryzujących poszczególne obszary ładu gospodarczego

Obszar	Wskaźnik (charakter)	Rok	Statystyki opisowe			
			minimum	maksimum	mediana	współczynnik zmienności (%)
Gospodarka	PKB na 1 mieszkańca (stymulanta)	2005	17591,0 lubelskie	40817,0 mazowieckie	22857,5	23,1
		2008	21269,0 podkarpackie	49155,0 mazowieckie	26793,4	23,0
	Udział wartości dodanej brutto w sektorze usług w wartości dodanej brutto ogółem (stymulanta)	2005	56,8 opolskie	73,7 mazowieckie	62,8	6,9
		2008	57,7 opolskie	75,7 mazowieckie	63,4	7,1
	Udział gospodarstw rolnych wielko powierzchniowych w liczbie gospodarstw ogółem (%) (destymulanta)	2005	0,03 małopolskie, świętokrzyskie	2,12 zachodnio-pomorskie	0,36	107,8
		2009	0,03 małopolskie	1,96 zachodnio-pomorskie	0,44	94,4
Innowacyjność	Udział ludności aktywnej zawodowo pracującej z wykształceniem wyższym w ludności aktywnej zawodowo pracującej ogółem (%) (stymulanta)	2005	16,6 podkarpackie	28,9 mazowieckie	19,9	13,4
		2009	20,5 opolskie	33,5 mazowieckie	24,5	12,2
	Nakłady na działalność badawczo-rozwojową w relacji do PKB (%) (stymulanta)	2005	0,1 świętokrzyskie	1,1 mazowieckie	0,3	70,1
		2008	0,1 lubuskie	1,2 mazowieckie	0,4	66,0
	Udział produkcji sprzedanej wyrobów nowych/istotnie ulepszonych w przedsię-biorstwach przemysłowych w produkcji sprzedanej ogółem (%) (stymulanta)	2006	7,2 kujawsko-pomorskie	34,2 mazowieckie	12,8	50,6
		2008	6,2 podlaskie	37,4 pomorskie	10,3	58,1
Gospodarstwa domowe	Dochody do dyspozycji brutto w sektorze gospodarstw domowych na 1 mieszkańca (zł) – ceny stałe z 2005 roku (stymulanta)	2005	13079,0 podkarpackie	21982,0 mazowieckie	15963	13,5
		2008	16209,0 podkarpackie	26882,0 mazowieckie	19375	13,2
	Przeciętna miesięczna emerytura i renta brutto w relacji do przeciętnego miesięcznego wynagrodzenia brutto (%) (stymulanta)	2005	37,9 mazowieckie	58,6 śląskie	49,6	7,8
		2009	39,2 mazowieckie	58,5 śląskie	50,6	7,4

cd. Tablicy 1.

1	2	3	4	5	6	7
	Wskaźnik zagrożenia ubóstwem skrajnym – procent osób w gospodarstwach domowych znajdujących się poniżej minimum egzystencji (%) (destymulanta)	2005	8,4 opolskie	18,8 warmińsko-mazurskie	13,2	21,7
		2009	2,9 opolskie	10,6 świętokrzyskie	5,8	34,1
Rynek pracy	Wskaźnik zatrudnienia (%) (stymulanta)	2005	48,5 zachodnio-pomorskie	57,7 mazowieckie	52,1	5,3
		2009	54,9 zachodnio-pomorskie	64,8 mazowieckie	58,1	4,0
	Wskaźnik bezrobocia osób zarejestrowanych w wieku 55 lat i więcej (%) (destymulanta)	2005	2,0 małopolskie	5,0 warmińsko-mazurskie	3,8	25,1
		2009	2,2 wielkopolskie	5,0 warmińsko-mazurskie	3,9	21,6
	Stopa bezrobocia rejestrowanego wśród osób poszukujących pracy dłużej niż rok (%) (destymulanta)	2005	5,5 wielkopolskie	11,1 warmińsko-mazurskie	7,6	20,0
		2009	1,8 wielkopolskie	5,3 warmińsko-mazurskie	3,2	32,2
Oddziaływanie na środowisko	Emisja zanieczyszczeń gazowych z zakładów szczególnie uciążliwych (t/km^2) (destymulanta)	2005	62,0 warmińsko-mazurskie	3311,3 śląskie	511,5	105,7
		2009	59,6 warmińsko-mazurskie	3073,8 śląskie	483,2	104,2
	Zużycie wody na potrzeby gospodarki narodowej i ludności na 1 mieszkańca (dam^3) (destymulanta)	2005	0,1 podlaskie	0,9 zachodnio-pomorskie	0,1	96,3
		2009	0,1 podlaskie	1,0 świętokrzyskie	0,1	108,3
	Zużycie energii elektrycznej ogółem na 1 mieszkańca (MWh) (destymulanta)	2006	0,5 podkarpackie	0,8 mazowieckie	0,7	11,9
		2009	0,6 podkarpackie	0,9 mazowieckie	0,7	10,9

Źródło: opracowanie własne na podstawie danych BDL.

Duże dysproporcje wykazywał także wskaźnik zużycia wody na potrzeby gospodarki narodowej i ludności w przeliczeniu na 1 mieszkańca. W województwach o najwyższym zużyciu w 2009 roku wynosił od 0,53 dam ^{3} w mazowieckim do 0,97 dam ^{3} w świętokrzyskim. W województwach o najniższym zużyciu, tj. lubuskim, podlaskim, śląskim i warmińsko-mazurskim, wartość ta nie przekroczyła 0,9 dam ^{3}.

Jako pozytywne zjawisko należy uznać zmniejszanie się dysproporcji międzywojewódzkich. Dotyczyło to 10 spośród 15 rozpatrywanych wskaźników. Jedynie w przy-

padku 5 wskaźników odnotowano nieznaczny wzrost dysproporcji międzywojewódzkich w 2009 roku w relacji do 2005 roku, dotyczyło to w obszarze:
– „Gospodarka": udziału wartości dodanej brutto w sektorze usług w wartości dodanej brutto ogółem,
– „Innowacyjność": udziału produkcji sprzedanej wyrobów nowych/istotnie ulepszonych w przedsiębiorstwach przemysłowych w produkcji sprzedanej ogółem,
– „Gospodarstwa domowe": wskaźnika zagrożenia ubóstwem skrajnym,
– „Rynek pracy": stopy bezrobocia rejestrowanego wśród osób poszukujących pracy dłużej niż rok,
– „Oddziaływanie na środowisko": zużycia wody na potrzeby gospodarki narodowej i ludności na 1 mieszkańca.

5. OCENA OBSZARÓW ZRÓWNOWAŻONEGO ROZWOJU WOJEWÓDZTW W KONTEKŚCIE ŁADU GOSPODARCZEGO

Analizę postępów województw w realizacji koncepcji zrównoważonego rozwoju w kontekście ładu gospodarczego prowadzono wieloaspektowo. Objęła ona ocenę sytuacji województw, poziomu spójności terytorialnej oraz kierunku i tempa zachodzących zmian.

Wartości syntetycznej miary rozwoju dla województw w poszczególnych obszarach w latach 2005 i 2009 zestawiono w tab. 2. W relacji do 2005 roku sytuacja województw poprawiła się w obszarach „Gospodarstwa domowe" i „Rynek pracy", ponieważ każde województwo przybliżyło się do wzorca. W obszarze „Gospodarstwa domowe" największy postęp odnotowano dla województwa warmińsko-mazurskiego, natomiast w obszarze „Rynek pracy" dla województwa zachodniopomorskiego. W zakresie „Gospodarki" i „Innowacyjności" rozwój okazał się nierównomierny. Regres w obu obszarach odnotowano w województwach dolnośląskim i lubuskim, w pozostałych województwach sytuacja uległa poprawie. Niekorzystnie należy ocenić obszar „Oddziaływanie na środowisko", z tego względu że sytuacja wszystkich województw pogorszyła się, przy czym najbardziej w przypadku województwa warmińsko-mazurskiego.

Relatywnie małe zmiany usytuowania województw w rankingu zaobserwowano w obszarze „Oddziaływanie na środowisko", natomiast relatywnie duże nastąpiły w obszarze „Innowacyjność" (tab. 3). Województwa należące do czołówki w poszczególnych obszarach zwykle zajmowały podobną pozycję w 2009 roku jak w 2005 roku, podobna tendencja dotyczyła województw zamykających ranking.

Swoją pozycję w większości obszarów poprawiły województwa kujawsko-pomorskie i świętokrzyskie. Spadek lokat we wszystkich obszarach odnotowano w województwie lubuskim. Podobne lokaty jak w 2005 otrzymywały województwa łódzkie i małopolskie.

Tablica 2. Wartości SMR_{it}^{m} województw według obszarów w 2005 i 2009 roku

Województwo	Gospodarka			Innowacyjność			Gospodarstwa domowe			Rynek pracy			Oddziaływanie na środowisko		
	2005	2009	zmiana	2005	2009	zmiana	2005	2009	zmiana	2005	2009	zmiana	2005	2009	zmiana
Dolnośląskie	0,394	0,370	-0,024	0,303	0,275	-0,028	0,281	0,470	0,189	0,197	0,543	0,346	0,708	0,689	-0,019
Kujawsko-pomorskie	0,342	0,418	0,076	0,071	0,120	0,050	0,257	0,439	0,183	0,275	0,545	0,270	0,726	0,648	-0,078
Lubelskie	0,363	0,426	0,063	0,196	0,336	0,141	0,161	0,340	0,178	0,492	0,677	0,185	0,838	0,777	-0,062
Lubuskie	0,320	0,312	-0,008	0,143	0,110	-0,033	0,307	0,439	0,132	0,285	0,492	0,207	0,719	0,670	-0,049
Łódzkie	0,367	0,432	0,065	0,282	0,296	0,013	0,342	0,492	0,150	0,362	0,642	0,280	0,503	0,461	-0,042
Małopolskie	0,441	0,501	0,061	0,400	0,421	0,021	0,266	0,427	0,160	0,583	0,817	0,234	0,572	0,522	-0,050
Mazowieckie	0,834	0,972	0,138	0,821	0,643	-0,178	0,338	0,433	0,095	0,489	0,752	0,263	0,455	0,339	-0,116
Opolskie	0,214	0,272	0,058	0,074	0,089	0,015	0,284	0,427	0,143	0,340	0,580	0,240	0,522	0,460	-0,062
Podkarpackie	0,306	0,365	0,059	0,190	0,382	0,193	0,129	0,337	0,208	0,382	0,590	0,209	0,943	0,920	-0,023
Podlaskie	0,351	0,416	0,065	0,131	0,221	0,089	0,199	0,354	0,156	0,527	0,570	0,042	0,753	0,654	-0,099
Pomorskie	0,440	0,458	0,018	0,442	0,582	0,139	0,211	0,427	0,216	0,339	0,694	0,355	0,584	0,503	-0,080
Śląskie	0,310	0,414	0,104	0,332	0,349	0,017	0,440	0,587	0,148	0,382	0,720	0,338	0,335	0,314	-0,021
Świętokrzyskie	0,310	0,332	0,022	0,133	0,252	0,119	0,159	0,352	0,193	0,265	0,550	0,285	0,569	0,388	-0,181
Warmińsko-mazurskie	0,282	0,284	0,002	0,246	0,301	0,055	0,102	0,380	0,278	0,045	0,360	0,315	0,933	0,670	-0,263
Wielkopolskie	0,335	0,399	0,063	0,358	0,242	-0,116	0,299	0,481	0,182	0,545	0,834	0,289	0,556	0,494	-0,062
Zachodnio-pomorskie	0,243	0,321	0,078	0,158	0,216	0,058	0,290	0,483	0,193	0,041	0,420	0,378	0,441	0,399	-0,042

Źródło: opracowanie własne na podstawie danych BDL GUS

Tablica 3. Lokaty województw według obszarów w 2005 i 2009 roku

Województwo	Gospodarka			Innowacyjność			Gospodarstwa domowe			Rynek pracy			Oddziaływanie na środowisko		
	2005	2009	zmiana	2005	2009	zmiana	2005	2009	zmiana	2005	2009	zmiana	2005	2009	zmiana
Dolnośląskie	4	10	-6	6	9	-3	8	5	3	14	13	1	7	3	4
Kujawsko-pomorskie	8	6	2	16	14	2	10	6	4	12	12	–	5	7	-2
Lubelskie	6	5	1	9	6	3	13	15	-2	4	6	-2	3	2	1
Lubuskie	10	14	-4	12	15	-3	4	7	-3	11	14	-3	6	5	1
Łódzkie	5	4	1	7	8	-1	2	2	–	8	7	1	13	11	2
Małopolskie	2	2	–	3	3	–	9	11	-2	1	2	-1	9	8	1
Mazowieckie	1	1	–	1	1	–	3	8	-5	5	3	2	14	15	-1
Opolskie	16	16	–	15	16	-1	7	10	-3	9	9	–	12	12	–
Podkarpackie	13	11	2	10	4	6	15	16	-1	6	8	-2	1	1	–
Podlaskie	7	7	–	14	12	2	12	13	-1	3	10	-7	4	6	-2
Pomorskie	3	3	–	2	2	–	11	9	2	10	5	5	8	9	-1
Śląskie	11	8	3	5	5	–	1	1	–	7	4	3	16	16	–
Świętokrzyskie	12	12	–	13	10	3	14	14	–	13	11	2	10	14	-4
Warmińsko-mazurskie	14	15	-1	8	7	1	16	12	4	15	16	-1	2	4	-2
Wielkopolskie	9	9	–	4	11	-7	5	4	1	2	1	1	11	10	1
Zachodnio-pomorskie	15	13	2	11	13	-2	6	3	3	16	15	1	15	13	2

Objaśnienia: Lokata 1 – pozycja najkorzystniejsza w stosunku do pozostałych województw, 16 – najmniej korzystna.
Źródło: opracowanie własne na podstawie tab. 2.

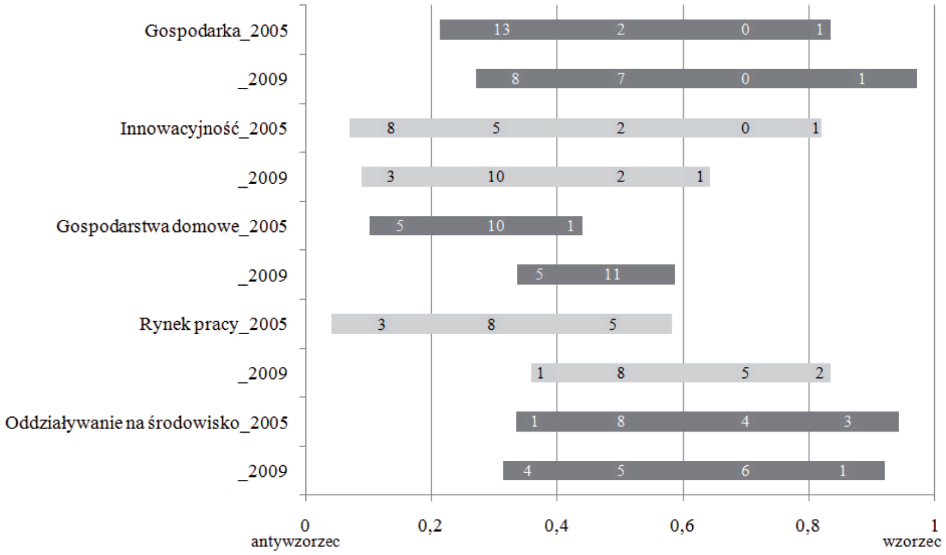
Tablica 4. Podstawowe miary statystyczne wyznaczone dla SMR_{it}^{m} województw według obszarów w 2005 i 2009 roku

Wyszczególnienie	Gospodarka			Innowacyjność			Gospodarstwa domowe			Rynek pracy			Oddziaływanie na środowisko		
	2005	2009	zmiana SMR_{it}^{m}	2005	2009	zmiana SMR_{it}^{m}	2005	2009	zmiana SMR_{it}^{m}	2005	2009	zmiana SMR_{it}^{m}	2005	2009	zmiana SMR_{it}^{m}
Minimum	0,214	0,272	-0,024	0,071	0,089	-0,178	0,102	0,337	0,095	0,041	0,360	0,042	0,335	0,314	-0,263
Maksimum	0,834	0,972	0,138	0,821	0,643	0,193	0,440	0,587	0,278	0,583	0,834	0,378	0,943	0,920	-0,019
Mediana	0,339	0,406	0,062	0,221	0,285	0,035	0,274	0,430	0,180	0,351	0,585	0,275	0,578	0,513	-0,062
Współczynnik zmienności (%)	36,813	37,282	x	67,408	49,255	x	33,875	15,066	x	45,300	21,082	x	26,912	29,358	x
Współczynnik korelacji liniowej Pearsona [-1, 1]	0,97		x	0,86		x	0,90		x	0,86		x	0,93		x

Objaśnienia: „x" – wyznaczenie wartości jest niemożliwe lub niecelowe.
Źródło: opracowanie własne na podstawie tab. 2.

W 2009 roku znacznie poprawił się poziom spójności regionalnej w obszarach „Gospodarstwa domowe" i „Rynek pracy" – współczynnik zmienności osiągnął o połowę niższe wartości w porównaniu z 2005 rokiem (tab. 4). Zmniejszyły się również dysproporcje międzywojewódzkie w zakresie „Innowacyjności", choć nadal zróżnicowanie terytorialne jest duże. W pozostałych obszarach („Gospodarka" oraz „Oddziaływanie na środowisko") nieco zmniejszył się poziom spójności regionalnej, pozostając na relatywnie niskim poziomie.

Rozrzut wartości syntetycznej miary rozwoju dla obszarów, a także liczebność województw w poszczególnych klasach (przedziałach wartości SMR_{it}^{m}) w latach 2005 i 2009 zaprezentowano na rys. 2. W 2009 roku uwidoczniło się dosyć duże tempo zmian i związane z nim przesunięcia większości województw do klasy o korzystniejszej sytuacji, w obszarach „Rynek pracy" i „Gospodarstwa domowe". Z kolei zróżnicowane tempo i kierunek zmian dotyczył obszarów „Gospodarka" i „Innowacyjność. W 2009 roku pogorszeniu uległa dotąd relatywnie dobra sytuacja większości województw w obszarze „Oddziaływanie na środowisko".



Rysunek 2. Rozrzut wartości SMR_{it}^{m} województw i liczebność klas według obszarów w latach 2005 i 2009

Źródło: opracowanie własne na podstawie tab. 2.

6. SYNTETYCZNA OCENA ZRÓWNOWAŻONEGO ROZWOJU WOJEWÓDZTW W ZAKRESIE ŁADU GOSPODARCZEGO

Syntetyczną ocenę postępów w dążeniu do zrównoważonego rozwoju przeprowadzono wyznaczając dla każdego województwa średnią wartość SMR_{it} (tab. 5). W 2009 roku sytuacja wszystkich województw uległa poprawie – każde z nich przesunęło się

w kierunku wzorca. Tempo rozwoju było jednak zróżnicowane. Największe przyrosty wartości średniego SMR_{it} odnotowano w województwie zachodniopomorskim, pomorskim i podkarpackim, najmniejsze zaś w mazowieckim, lubuskim i podlaskim. Kolejność województw w rankingu nie zmieniła się zasadniczo w stosunku do 2005 roku (wartość współczynnika korelacji liniowej Pearsona wyniosła 0,92). W obu okresach pierwszą lokatę utrzymało województwo mazowieckie, a drugą małopolskie. Na końcu rankingu znalazły się podobnie jak w 2005 roku województwa opolskie, zachodniopomorskie, świętokrzyskie i warmińsko-mazurskie.

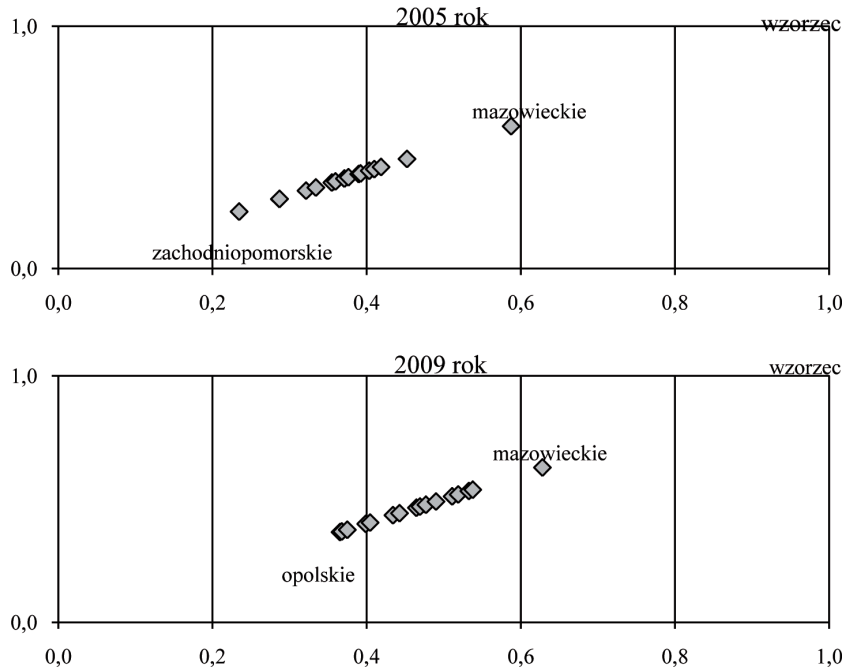
Tablica 5. Średnie wartości SMR_{it} oraz lokaty województw w 2005 i 2009 roku

Województwo	2005		2009		Zmiana SMR_{it}	Zmiana lokaty
	średni SMR_{it}	lokata	średni SMR_{it}	lokata		
Dolnośląskie	0,377	8	0,469	8	0,092	–
Kujawsko-pomorskie	0,334	12	0,434	11	0,100	1
Lubelskie	0,410	4	0,511	5	0,101	-1
Lubuskie	0,355	11	0,405	12	0,050	-1
Łódzkie	0,371	9	0,465	9	0,094	–
Małopolskie	0,452	2	0,538	2	0,086	–
Mazowieckie	0,587	1	0,628	1	0,041	–
Opolskie	0,287	15	0,366	16	0,079	-1
Podkarpackie	0,390	7	0,519	4	0,129	3
Podlaskie	0,392	6	0,443	10	0,051	-4
Pomorskie	0,403	5	0,533	3	0,130	2
Śląskie	0,360	10	0,477	7	0,117	3
Świętokrzyskie	0,287	14	0,375	14	0,088	–
Warmińsko-mazurskie	0,322	13	0,399	13	0,077	–
Wielkopolskie	0,419	3	0,490	6	0,071	-3
Zachodnio-pomorskie	0,235	16	0,368	15	0,133	1

Źródło: opracowanie własne na podstawie tab. 2 i tab. 3.

W 2009 roku nieco zwiększył się poziom spójności regionalnej – wartość współczynnika zmienności zmniejszyła się z 20,71% w 2005 do 15,17% w 2009 roku. Województwa przesunęły się w kierunku wzorca, na co wskazuje wzrost mediany (z 0,37 w 2005 na 0,47 w 2009 roku), a także wartości minimalnej (z 0,24 w 2005 do 0,37 w 2009 roku) i maksymalnej średniego SMR_{it} (z 0,59 w 2005 do 0,63 w 2009 roku). Sytuacja województw nieco poprawiła się – w 2005 roku większość wo-

jewództw charakteryzowała się sytuacją niekorzystną, a w 2009 roku sytuacją umiarkowaną (rys. 3).



Rysunek 3. Wykresy rozrzutu średniego SMR_{it} województw względem wzorca

Źródło: opracowanie własne na podstawie tab. 5.

W 2005 roku sytuacją umiarkowanie korzystną charakteryzowały się województwa mazowieckie, małopolskie, wielkopolskie, lubelskie i pomorskie (rys. 4). Na ich pozycję w zasadniczy sposób wpłynęły dobre warunki w obszarze „Gospodarka" i „Innowacyjność, a także „Rynek pracy" i „Oddziaływanie na środowisko".

Pozostałe województwa charakteryzowały się sytuacją niekorzystną, no co największy wpływ miały bardzo złe warunki w zakresie „Innowacyjności" oraz słaba sytuacja w obszarach „Gospodarstwa domowe" i „Rynek pracy".

W 2009 roku mimo tego, że ogólna sytuacja województw poprawiła się, tylko pozycję województwa mazowieckiego można określić jako korzystną, przede wszystkim z uwagi na dobre warunki w obszarach „Gospodarka" i „Innowacyjność" oraz „Rynek pracy" (rys. 5).

Grupę województw charakteryzującą się sytuacją umiarkowanie korzystną reprezentują województwa małopolskie, pomorskie, podkarpackie, lubelskie, wielkopolskie, śląskie, dolnośląskie, łódzkie, podlaskie, kujawsko-pomorskie i lubuskie. Wykazały się one relatywnie dobrą sytuacją w obszarach „Rynek pracy", „Oddziaływanie na środowisko" i „Gospodarstwa domowe".

Rysunek 4. Sytuacja województw w zakresie zrównoważonego rozwoju w kontekście ładu gospodarczego w 2005 roku

Źródło: opracowanie własne na podstawie tab. 5.

Rysunek 5. Sytuacja województw w zakresie zrównoważonego rozwoju w kontekście ładu gospodarczego w 2009 roku

Źródło: opracowanie własne na podstawie tab. 5.

Sytuację pozostałych województw, tj. warmińsko-mazurskiego, świętokrzyskiego, zachodniopomorskiego i opolskiego można określić jako niekorzystną w większości obszarów.

Biorąc pod uwagę wartości SMR_{it}^{m} i lokaty w poszczególnych obszarach określono korzyści (przewagi) i ograniczenia województw (tab. 6). Na podstawie przyrostów wartości SMR_{it}^{m} określono także szanse i zagrożenia województw na drodze do zrównoważonego rozwoju dla każdego obszaru.

Tablica 6. Korzyści i ograniczenia oraz szanse i zagrożenia województw w zakresie ładu gospodarczego

Lokata w 2009 r.	Województwo	Korzyści	Ograniczenia	Szanse	Zagrożenia
1	mazowieckie	*Gospodarka*, *Innowacyjność*, *Rynek pracy*	*Oddziaływanie na środowisko*	**Rynek pracy**, *Gospodarka*	*Innowacyjność*, *Oddziaływanie na środowisko*
2	małopolskie	*Rynek pracy*, *Gospodarka*, *Innowacyjność*	–	**Rynek pracy**	–
3	pomorskie	**Rynek pracy**, *Gospodarka*, *Innowacyjność*	–	*Gospodarstwa domowe*, *Rynek pracy*, *Innowacyjność*	–
4	podkarpackie	*Oddziaływanie na środowisko*	*Gospodarstwa domowe* , **Innowacyjność**, **Gospodarka**	*Gospodarstwa domowe*, **Rynek pracy**, *Innowacyjność*	–
5	lubelskie	*Oddziaływanie na środowisko*, **Rynek pracy**	*Gospodarstwa domowe*, **Innowacyjność**	*Innowacyjność*	–
6	wielkopolskie	*Rynek pracy*	**Innowacyjność**, **Gospodarka**	**Rynek pracy**	*Innowacyjność*
7	śląskie	**Rynek pracy**, *Gospodarstwa domowe*	*Oddziaływanie na środowisko*, **Innowacyjność**	**Rynek pracy**, *Gospodarka*	–
8	dolnośląskie	*Oddziaływanie na środowisko*	**Innowacyjność**, **Gospodarka**	*Rynek pracy*	*Gospodarka*, Innowacyjność
9	łódzkie	**Rynek pracy**, *Gospodarstwa domowe*	**Innowacyjność**	**Rynek pracy**	–
10	podlaskie	**Oddziaływanie na środowisko**	**Innowacyjność**, **Gospodarstwa domowe**	–	–
11	kujawsko-pomorskie	**Oddziaływanie na środowisko**	*Innowacyjność*	**Rynek pracy**	–
12	lubuskie	**Oddziaływanie na środowisko**	*Innowacyjność*, *Gospodarka*, *Rynek pracy*	**Rynek pracy**	*Gospodarka*, *Innowacyjność*
13	warmińsko-mazurskie	**Oddziaływanie na środowisko**	*Rynek pracy*, *Gospodarka*, **Innowacyjność**, **Gospodarstwa domowe**	**Rynek pracy**, *Gospodarstwa domowe*	*Oddziaływanie na środowisko*

cd. Tablicy 6.

14	świętokrzyskie	–	*Gospodarstwa domowe*, *Oddziaływanie na środowisko*, **Gospodarka**, **Innowacyjność**	**Rynek pracy**	*Oddziaływanie na środowisko*
15	zachodnio-pomorskie	*Gospodarstwa domowe*	**Gospodarka**, **Innowacyjność**, **Oddziaływanie na środowisko**, *Rynek pracy*	*Rynek pracy*, *Gospodarka*	–
16	opolskie	–	*Gospodarka*, *Innowacyjność*	**Rynek pracy**	–
Legenda	Podkreślenie	maksymalna wartość SMR_{it}^m	minimalna wartość SMR_{it}^m	maksymalny dodatni przyrost SMR_{it}^m	maksymalny ujemny przyrost SMR_{it}^m
	Wytłuszczenie	wartość $SMR_{it}^m \geqslant 0,6$	wartość $SMR_{it}^m \leqslant 0,4$	przyrost $SMR_{it}^m \geqslant 0,2$	przyrost $SMR_{it}^m \leqslant -0,1$
	Kursywa	1, 2, 3 lokata w rankingu	14, 15, 16 lokata w rankingu	3 największe przyrosty dodatnie	3 największe przyrosty ujemne

Źródło: opracowanie własne na podstawie tab. 2, tab. 3, tab. 5.

Wśród województw nie było lidera w zakresie wszystkich rozpatrywanych obszarów. Często województwa mocne w zakresie „Gospodarki" lub „Innowacyjności" były słabsze w obszarze „Oddziaływanie na środowisko", i odwrotnie. Z kolei wysoka pozycja w rankingu niekoniecznie oznaczała jednoczesne występowanie pozytywnych zmian wartościSMR_{it}^m; niektóre województwa mimo pozostawania na niskich lokatach dokonały znacznych postępów w danym obszarze.

Najlepszą ogólną sytuacją, oceniając na podstawie średniego SMR_{it}, zarówno w 2005 jak i 2009 roku, charakteryzowało się województwo **mazowieckie**, chociaż spośród wszystkich jednostek w 2009 roku w porównaniu do 2005 roku dokonało najmniejszego postępu (oceniając łącznie wszystkie obszary). W obu okresach było liderem w zakresie „Gospodarki" i „Innowacyjności". Charakteryzowało się również korzystną sytuacją w zakresie „Rynku pracy" (3. lokata). W stosunku do 2005 roku dokonało największego spośród wszystkich województw progresu w obszarze „Gospodarka", a także dużego postępu w obszarze „Rynek pracy". W województwie nastąpił największy, na tle pozostałych województw, regres w „Innowacyjności". Pomimo tego utrzymało pierwszą pozycję w tym zakresie. Istotny spadek lokat (5 lokat w dół) nastąpił w obszarze „Gospodarstwa domowe". Województwo w obu latach charakteryzowało się niekorzystną sytuacją w zakresie „Oddziaływania na środowisko" (15. lokata w 2009 roku) i wykazuje ono ryzyko pogorszenia się tej sytuacji w przyszłości.

Województwo **małopolskie** utrzymało w 2009 roku drugą lokatę. Wśród atutów województwa wskazać można korzystną sytuację i jednocześnie pozytywne tendencje w zakresie „Rynku pracy". Województwo uplasowało się na wysokich pozycjach w rankingu w obszarach „Gospodarka" i „Innowacyjność". Mimo dokonania znacz-

nych postępów w zakresie „Gospodarstw domowych" na tle pozostałych województw wypadło dosyć słabo i zajęło 11. lokatę.

Istotny postęp w dążeniu do zrównoważonego rozwoju charakteryzował województwo **pomorskie**. Sytuację województwa w większości badanych obszarów można określić jako umiarkowanie korzystną. Wzmocniło ono jednak znacznie swoją pozycję w obszarach „Gospodarstwa domowe", „Rynek pracy", a także „Innowacyjność". W zakresie „Rynku pracy" dokonało znacznego postępu i jednocześnie przeskoku aż o 5 lokat w górę. Poza tym zajmuje wysokie pozycje w aspektach „Gospodarka" i „Innowacyjność".

Bardzo pozytywne zmiany zaobserwowano w województwie **podkarpackim**, przede wszystkim w zakresie „Gospodarstw domowych" i „Rynku pracy", ale także „Innowacyjności". W obu porównywanych latach województwo było liderem w obszarze „Oddziaływanie na środowisko". Jednocześnie charakteryzuje się najmniej korzystnymi warunkami w aspekcie „Gospodarstwa domowe" (16. lokata). Uzyskało nienajlepsze wyniki oraz niską pozycję w zakresie „Gospodarki" (11. lokata), a także „Innowacyjności".

Usytuowanie województwa **lubelskiego** na tle pozostałych jest zadowalające. Charakteryzuje się ono pozytywnymi uwarunkowaniami w zakresie „Oddziaływania na środowisko" (2. lokata), a także „Rynku pracy". Jednak sytuację „Gospodarstw domowych" należy określić jako niekorzystną, mimo że województwo dokonało przesunięcia w kierunku wzorca. W ciągu kilku lat nastąpiła widoczna poprawa w obszarach „Rynek pracy", „Innowacyjność", ale w tym ostatnim sytuacja nadal jest niezadowalająca.

Znaczny spadek w rankingu (3 lokaty) charakteryzował województwo **wielkopolskie**. Odznacza się ono dużym zróżnicowaniem w poszczególnych obszarach. Dokonało znacznego postępu w zakresie „Rynku pracy" i jego sytuacja jest obecnie najkorzystniejsza na tle pozostałych województw. Wyraźne zmiany na korzyść nastąpiły również w przypadku „Gospodarstw domowych". Złą sytuację i bardzo niekorzystne tendencje województwo wykazało w obszarze „Innowacyjności"; odnotowało spadek w rankingu aż o 7 lokat. Poprawy wymaga również obszar „Gospodarka".

Swoją pozycję na tle pozostałych województw o 3 lokaty poprawiło **śląskie**. Zajmuje najwyższą pozycję w zakresie „Gospodarstw domowych" i dokonało dosyć dużego postępu w tym zakresie w stosunku do 2005 roku, choć daleko mu jeszcze do wzorca. Charakteryzuje się dobrą sytuacją w zakresie „Rynku pracy". Zrobiło znaczny postęp w tym obszarze w stosunku do 2005 roku i dzięki temu wspięło się o 3 lokaty w górę. Dosyć korzystne zmiany zaobserwowano w obszarze „Gospodarka"; w którym województwo odnotowało drugi (za mazowieckim) największy przyrost wartości SMR_{it}^{m} i przeskoczyło o 3 lokaty w górę. Najmniej korzystnie ze wszystkich województw wypadło w obu okresach w obszarze „Oddziaływanie na środowisko". Niekorzystna sytuacja dotyczy również „Innowacyjności".

Ogólną sytuację województwa **dolnośląskiego** można ocenić jako przeciętną (8. lokata). Bardzo istotne zmiany na korzyść dotyczyły obszaru „Rynek pracy. Choć zajmuje ono przeciętne pozycje w zakresie „Gospodarki" i „Innowacyjności", to charakteryzuje

się bardzo niekorzystnymi tendencjami w obu tych obszarach i odnotowało spadek lokat (odpowiednio o 6 i 3 pozycje). W obszarze „Oddziaływania na środowisko" awansowało o 4 lokaty (3. lokata) i wykazało najmniejszą ujemną zmianę SMR_{it}^{m} wśród województw (relatywnie najkorzystniejszą sytuację), zaś w obszarze „Gospodarstwa domowe" zyskało 3 lokaty w górę dokonując znacznego postępu w tym zakresie.

Ogólna sytuacja województwa **łódzkiego** uległa poprawie, ale jego pozycja w rankingu nie zmieniła się w stosunku do 2005 roku. Dosyć mocną stroną województwa jest obszar „Gospodarstwa domowe", gdzie zajmuje 2. lokatę i dokonało dosyć dużego postępu w stosunku do 2005 roku. Zajmuje również wysoką lokatę w zakresie „Gospodarki" (4. pozycja). Korzystne tendencje nastąpiły w tym województwie w obszarze „Rynek pracy"**.** Natomiast obszarem wymagającym poprawy jest „Innowacyjność".

Niewielki postęp i jednocześnie niekorzystne zmiany usytuowania w rankingu (4 lokaty w dół) charakteryzowały województwo **podlaskie**. Zdecydowanej poprawy wymaga w tym przypadku obszar „Innowacyjność" (12. lokata), a także „Gospodarstwa domowe" (13. lokata), choć odnotowano znaczny postęp w tym zakresie. Niewielki postęp i, co za tym idzie, największy spadek (o 7 lokat) zaobserwowano w obszarze „Rynek pracy".

Województwo **kujawsko-pomorskie** należy do grupy regionów o umiarkowanie korzystnej sytuacji w zakresie „Gospodarki" (6. lokata) i „Gospodarstw domowych" (6. lokata, 4 pozycje wyżej i znaczne zmiany na korzyść w stosunku do 2005 roku). Umiarkowanie dobrą sytuację i korzystne tendencje zaobserwowano w zakresie „Rynku pracy", choć województwo nadal nisko plasuje się w rankingu (12. lokata). Do słabych stron województwa należy obszar „Innowacyjność", natomiast do mocnych „Oddziaływanie na środowisko".

Województwo **lubuskie** znalazło się w piątce województw zamykających ranking. Poczyniło nieznaczny postęp w kierunku wzorca. Do słabych stron województwa należą „Innowacyjność" (15. lokata) i „Gospodarka" (14. lokata). Oba obszary charakteryzują się niekorzystnymi tendencjami i województwo odnotowało spadek lokat (odpowiednio o 3 i 4 lokaty w dół). Mimo nienajgorszej pozycji w rankingu (obecnie 7. pozycja, co oznacza spadek o 3 lokaty w stosunku do 2005 roku) oraz dokonania postępu, jako niekorzystną sytuację można również wskazać obszar „Gospodarstwa domowe". Pozytywne zmiany, mimo spadku o 3 lokaty, dotyczą „Rynku pracy".

Województwo **warmińsko-mazurskie** charakteryzuje się zróżnicowaną sytuacją w poszczególnych aspektach. Zajęło ostatnie miejsce w zakresie „Rynku pracy", ale jednocześnie zauważyć można korzystne tendencje w tym obszarze. Do jego słabych stron należy również „Gospodarka", a także „Innowacyjność". W obszarze „Oddziaływanie na środowisko" wykazało największy regres. Wykazuje najbardziej pozytywne zmiany w obszarze „Gospodarstwa domowe" (4. lokaty w górę), choć nadal sytuacja w tym obszarze nie jest najlepsza.

Sytuacja województwa **świętokrzyskiego** jest wyrównana w większości rozpatrywanych obszarów, tj. „Innowacyjność", „Gospodarstwa domowe" i „Oddziaływanie na środowisko". W przypadku „Innowacyjności" mimo znacznego postępu w kierunku

wzorca, województwo odnotowało spadek pozycji w rankingu o 3 lokaty. Województwo wykazało niekorzystne tendencje i spadek o 4 lokaty w obszarze „Oddziaływanie na środowisko". Pozytywne zmiany dokonane zostały na „Rynku pracy".

Dużym potencjałem gospodarczym wykazało się województwo **zachodniopomorskie**. Na tle pozostałych województw wypadło dosyć dobrze w obszarze „Gospodarstw domowych", w którym zajmuje 3. lokatę (o 3 pozycje wyżej niż w 2005 roku) i dokonało znacznego postępu w tym zakresie. Choć sytuacja w obszarze „Rynek pracy" jest nadal umiarkowanie korzystna, to województwo dokonało bardzo pozytywnych przemian w tym zakresie. Do jego słabych stron należy „Innowacyjność", „Gospodarka" (gdzie wykazało korzystne tendencje) i „Oddziaływanie na środowisko" (13. lokata).

Województwo **opolskie** uplasowało się na końcu rankingu. W obszarze „Rynek pracy" wykazało ono znaczny postęp w porównaniu do 2005 roku i korzystne zmiany w zakresie „Gospodarstw domowych", mimo tego w tym ostatnim obszarze odnotowało spadek o 3 lokaty. Charakteryzowało się najmniej korzystną sytuacją w obszarach „Gospodarka" i „Innowacyjność" (16. lokaty).

7. PODSUMOWANIE

Na podstawie przeprowadzonych badań można mówić o pewnym postępie województw w kierunku zrównoważonego rozwoju w kontekście ładu gospodarczego, choć nie była to radykalna zmiana. W 2005 roku sytuację większości województw można było określić jako niekorzystną, natomiast w 2009 roku jako umiarkowanie korzystną. Wynikało to zarówno z osiągniętego poziomu wskaźników, jak i poniekąd było wynikiem zastosowanej metody badawczej. Przy konstrukcji miar syntetycznych opartych na kilku zmiennych, rzeczą często spotykaną jest uśrednienie ich realizacji.

Najkorzystniejsze zmiany – ze względu na poprawę sytuacji województw oraz zwiększenie spójności terytorialnej – zaszły w obszarach „Gospodarstwa domowe" i „Rynek pracy".

W relacji do 2005 roku nastąpiły niewielkie przesunięcia województw w rankingu. W 2009 roku nadal liderem było województwo mazowieckie, choć jego dystans w stosunku do pozostałych województw uległ zmniejszeniu. Natomiast w grupie województw o najmniej korzystnej sytuacji, podobnie jak w 2005 roku, znalazły się zachodniopomorskie, opolskie, świętokrzyskie i warmińsko-mazurskie.

Największy postęp w kierunku wzorca charakteryzował województwo zachodniopomorskie. Relatywnie dużą zmianę na korzyść wykazały również województwa pomorskie, podkarpackie, śląskie, lubelskie i kujawsko-pomorskie. Jednocześnie większość z tych województw (oprócz lubelskiego) obniżyło swoją lokatę w rankingu. Oznacza to, że nawet tak duży postęp nie wystarczył, aby dorównać województwom o lepszej sytuacji.

Swoje usytuowanie w rankingu znacznie poprawiły województwa podkarpackie i śląskie. Relatywnie duży spadek lokat odnotowano w przypadku województwa pod-

laskiego oraz wielkopolskiego. Województwa te charakteryzowała niewielka poprawa sytuacji w większości obszarów.

Interpretując wyniki przeprowadzonych badań należy mieć na względzie specyfikę województw (m.in. usytuowanie geograficzne, profil gospodarczy) oraz ich wewnętrzne zróżnicowanie (spójność terytorialną), a także relatywnie krótki okres badania oraz przyjęte umownie wzorcowe wartości wskaźników, które nie zawsze są wartościami optymalnymi w danej dziedzinie.

Katedra Gospodarki Regionalnej,
Uniwersytet Ekonomiczny we Wrocławiu
Katedra Ekonometrii i Informatyki,
Uniwersytet Ekonomiczny we Wrocławiu

LITERATURA

[1] Borys T. (red.), *Wskaźniki zrównoważonego rozwoju*, Wyd. Ekonomia i Środowisko, Warszawa-Białystok 2005.
[2] Eurostat, http://epp.eurostat.ec.europa.eu/portal/page/portal/sdi/indicators (data dostępu: 23.02.2011).
[3] *Nasza wspólna przyszłość*, raport Światowej Komisji Środowiska i Rozwoju ONZ, Oxford: Oxford University Press, Oxford 1987.
[4] Walesiak M., *Uogólniona miara odległości w statystycznej analizie wielowymiarowej*, Wyd. AE, Wrocław 2006.

GOSPODARCZE ASPEKTY ZRÓWNOWAŻONEGO ROZWOJU WOJEWÓDZTW – WIELOWYMIAROWA ANALIZA PORÓWNAWCZA

S t r e s z c z e n i e

Zrównoważony rozwój jest nadrzędnym celem Unii Europejskiej. Zakłada podniesienie poziomu i jakości życia poprzez osiągnięcie kompromisu w zakresie rozwoju gospodarczego, społecznego i środowiskowego. W artykule dokonano oceny zrównoważonego rozwoju województw Polski w kontekście ładu gospodarczego w 2005 i 2009 roku. Województwa opisano zestawem wskaźników, które pogrupowano w pięć obszarów, tj. gospodarka, innowacyjność, gospodarstwa domowe, rynek pracy i oddziaływanie na środowisko. Analizowano sytuację województw, poziom spójności terytorialnej oraz kierunek zachodzących zmian. W tym celu zastosowano wybrane statystyki opisowe oraz porządkowanie liniowe z formułą wzorcową.

Słowa kluczowe: zrównoważony rozwój, ład gospodarczy, wielowymiarowa analiza statystyczna

ECONOMIC ASPECTS OF SUSTAINABLE DEVELOPMENT IN POLISH VOIVODESHIPS – MULTIVARIATE DATA ANALYSIS

A b s t r a c t

Sustainable development is basic and superior aim of European Union. It assumes growing standard and quality of life by achieving compromise in respect to economic, social and environmental development. In this paper an evaluation of sustainable development in Polish voivodeships with reference to economic order in 2005 and 2009 was performed. The voivodeships were described by a set of indicators grouped into five areas such as economy, innovativeness, households, labour market and impact upon the environment. The voivodeships situation, territorial cohesion and direction of changes were investigated. Selected description statistics and linear ordering with pattern formula were applied.

Keywords: sustainable development, economic order, multivariate data analysis

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

MICHAŁ KOLUPA, JOANNA PLEBANIAK

KILKA UWAG O WARTOŚCIACH WŁASNYCH MACIERZY KORELACJI

W niniejszej pracy, w nawiązaniu do pracy Kolupa, 2006, podajemy konstrukcję przedziału $[\alpha^{\mathbf{R}}, \beta^{\mathbf{R}}]$, w którym leżą wszystkie wartości własne macierzy korelacji $\mathbf{R}(k) = [r_{ij}]$ stopnia k, gdzie r_{ij} oznacza współczynnik korelacji pomiędzy standaryzowanymi zmiennymi Z_i oraz Z_j będącymi zmiennymi objaśniającymi jednorównaniowego liniowego modelu ekonometrycznego postaci:

$$Y = \alpha_1 Z_1 + \ldots + \alpha_k Z_k + e. \tag{1}$$

Wspomnianą konstrukcję przedziału $[\alpha^{\mathbf{R}}, \beta^{\mathbf{R}}]$ wykorzystujemy prezentując kryterium (dwie wersje), na podstawie którego rozstrzygamy o dodatniości wszystkich wartości własnych macierzy korelacji $\mathbf{R}(k)$.

Dodajmy jeszcze, że przedstawione niżej kryterium jest inne niż dotychczas znane i opisane w literaturze (Mostowski, Stark, 1968).

Jeżeli $\mathbf{A} = [a_{ij}]$ jest rzeczywistą macierzą symetryczną stopnia k, to wszystkie jej wartości własne leżą w przedziale $[\ \alpha, \beta\ ]$, gdzie:

$$\alpha = \frac{1}{k}\mathrm{tr}\mathbf{A} - \frac{1}{k}\sqrt{(k-1)[k\mathrm{tr}\mathbf{A}^2 - (\mathrm{tr}\mathbf{A})^2]}, \tag{2}$$

$$\beta = \frac{1}{k}\mathrm{tr}\mathbf{A} + \frac{1}{k}\sqrt{(k-1)[k\mathrm{tr}\mathbf{A}^2 - (\mathrm{tr}\mathbf{A})^2]}. \tag{3}$$

Oznaczmy symbolem d następujące wyrażenie:

$$d = \frac{1}{k}\sqrt{(k-1)[k\mathrm{tr}\mathbf{A}^2 - (\mathrm{tr}\mathbf{A})^2]} \tag{4}$$

(por. np. Kolupa, 2006).

Wówczas jak to wynika ze wzorów (2) i (3) mamy:

$$\alpha + \beta = \frac{2\mathrm{tr}\mathbf{A}}{k} \tag{5}$$

$$\alpha - \beta = -2d$$

Ponieważ macierz korelacji $\mathbf{R}(k) = [r_{ij}]_{kxk}$ jest rzeczywistą macierzą symetryczną przeto:

$$\frac{1}{k}\mathrm{tr}\mathbf{R}(k) = 1 \tag{6}$$

$$\mathrm{tr}\mathbf{R}^2(k) = k + \sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 \tag{7}$$

$$[\mathrm{tr}\mathbf{R}(k)]^2 = k^2 \tag{8}$$

Wobec tego:

$$k\mathrm{tr}\mathbf{R}^2(k) - [\mathrm{tr}\mathbf{R}(k)]^2 = k^2 + k\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 - k^2 = k\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 \tag{9}$$

co pociąga za sobą, że:

$$(k-1)\left[k\mathrm{tr}\mathbf{R}^2(k) - [\mathrm{tr}\mathbf{R}(k)]^2\right] = (k-1)k\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 \tag{10}$$

Wówczas na podstawie (2) i (3) uwzględniając rezultaty podane we wzorach (6), (7), (8), (9) i (10) mamy przedział $[\ \alpha^{\mathbf{R}}, \beta^{\mathbf{R}}]$, gdzie:

$$\alpha^{\mathbf{R}} = 1 - d^{\mathbf{R}}, \tag{11}$$

$$\beta^{\mathbf{R}} = 1 + d^{\mathbf{R}}, \tag{12}$$

zaś:

$$d^{\mathbf{R}} = \frac{1}{k}\sqrt{(k-1)k\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2}. \tag{13}$$

Oznacza to, że wszystkie wartości własne $\lambda_i^{\mathbf{R}}$ macierzy korelacji $\mathbf{R}(k)$ leżą w przedziale:

$$1 - d^{\mathbf{R}} \leqslant \lambda_i^{\mathbf{R}} \leqslant 1 + d^{\mathbf{R}}, i = 1, 2, \ldots, k. \tag{14}$$

Żądamy, aby $\lambda_i^{\mathbf{R}} > 0$, czyli:

$$1 - d^{\mathbf{R}} \geqslant 0 \tag{15}$$

albo:

$$k^2 \geqslant k^2 \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2 - k \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2. \tag{16}$$

Zapisujemy (16) w równoważnej postaci:

$$\left(1 - \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2\right) k^2 \geqslant -k \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2. \tag{17}$$

Ostatecznie:

$$\left(1 - \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2\right) k \geqslant - \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2. \tag{18}$$

Z nierówności (18) wyznaczamy k. Mogą zaistnieć dwa przypadki, które oznaczamy symbolami A oraz B.

Przypadek A:

$$sign\left(1 - \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2\right) = +1. \tag{19}$$

Wówczas:

$$k \geqslant u, \tag{20}$$

zaś:

$$u = \frac{-\sum\limits_{\substack{i,j \\ i\neq j}} r_{ij}^2}{1 - \sum\limits_{\substack{i,j \\ i\neq j}} r_{ij}^2}. \tag{21}$$

Nierówność (20) jest, to jak tu widać ze wzoru (21), zawsze spełniona.

Przypadek *B*:

$$sign\left(1 - \sum_{\substack{i,j \\ i\neq j}} r_{ij}^2\right) = -1. \tag{22}$$

Wówczas:

$$k \leqslant u, \tag{23}$$

gdzie u dane jest wzorem (21).

Tak więc w obydwu przypadkach (przypadek *A* oraz przypadek *B*) jest spełniona nierówność (15).

Oznacza to, że wówczas wszystkie wartości własne macierzy $\mathbf{R}(k)$ są dodatnie.

Dodajmy jeszcze, że warunek (19) ma miejsce wówczas, kiedy:

$$\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 < 1, \tag{24}$$

zaś warunek (22) zachodzi wówczas, kiedy:

$$\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 > 1. \tag{25}$$

Tak więc otrzymujemy dwie wersje kryterium na podstawie którego można rozstrzygnąć, czy wszystkie wartości własne macierzy korelacji $\mathbf{R}(k)$ są dodatnie.

Wersja 1

Jeżeli jest spełniony warunek (24), to wszystkie wartości własne macierzy korelacji $\mathbf{R}(k)$ są dodatnie (nierówność (19) jest zawsze spełniona).

Wersja 2

Jeżeli są spełnione warunek (22) i nierówność (23), to wszystkie wartości własne macierzy korelacji $\mathbf{R}(k)$ są dodatnie.

Odwołamy się do kilku przykładów ilustrujących opisane rezultaty.

Przykład 1

Dana jest macierz $\mathbf{R}(2)$ postaci:

$$\mathbf{B}(2) = \begin{bmatrix} 1 & 0,2 \\ 0,2 & 1 \end{bmatrix}. \tag{26}$$

Bezpośrednim rachunkiem sprawdzamy, że:

$$\det\left[\mathbf{R}(2) - \lambda\mathbf{I}(2)\right] = \det\begin{bmatrix} 1-\lambda & 0,2 \\ 0,2 & 1-\lambda \end{bmatrix} = 0, \tag{27}$$

czyli:

$$(1-\lambda)^2 - 0,04 = 0 \tag{28}$$

albo:

$$1 - 2\lambda + \lambda^2 - 0,04 = 0 \tag{29}$$

$$\lambda^2 - 2\lambda + 0,96 = 0 \tag{30}$$

$$\Delta = 4 - 4 \cdot 0,96 = 4 \cdot 0,04 \tag{31}$$

$$\sqrt{\Delta} = 2 \cdot 0,2 \tag{32}$$

$$\lambda_1 = \frac{2 - 2 \cdot 0,2}{2} = 1 - 0,2 = \frac{4}{5}, \tag{33}$$

$$\lambda_2 = \frac{2 + 2 \cdot 0,2}{2} = 1 + 0,2 = \frac{6}{5}. \tag{34}$$

Ze wzorów (11) i (12) wynika, że w omawianym przypadku dla macierzy (26) mamy:

$$\alpha^{\mathbf{R}} = 1 - \frac{1}{2}\sqrt{1 \cdot 2 \cdot 0,08} = 1 - \frac{1}{2}\sqrt{0,16} = \frac{4}{5}, \tag{35}$$

$$\beta^{\mathbf{R}} = 1 + \frac{1}{2}\sqrt{1 \cdot 2 \cdot 0,08} = 1 + \frac{1}{2}\sqrt{0,16} = \frac{6}{5}. \tag{36}$$

Wszystkie wartości własne macierzy $\mathbf{R}(2)$ danej wzorem (26) należą do przedziału $\left[\frac{4}{5}, \frac{6}{5}\right]$. A zatem są dodatnie.

Przykład 2 (ilustracja wersji 1)
Rozpatrzymy ponownie macierz $\mathbf{R}(2)$ daną wzorem (26). Mamy zatem:

$$\sum_{\substack{i,j \\ i \neq j}} r_{ij}^2 = (0,2)^2 + (0,2)^2 = 2 \cdot 0,04 = \frac{2}{25}, \tag{37}$$

czyli:

$$1 - \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2 = 1 - \frac{2}{25} = \frac{23}{25}. \tag{38}$$

Jest zatem spełniony warunek (19). Tym samym nierówność (20) jest zawsze spełniona. Oznacza to, że wszystkie wartości własne macierzy danej wzorem (26) są dodatnie, co sprawdziliśmy też bezpośrednio (patrz przykład 1).

Przykład 3 (ilustracja wersji 2)
Dana jest macierz:

$$\mathbf{R}(3) = \begin{bmatrix} 1 & 0,8 & 0 \\ 0,8 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{39}$$

Bezpośrednim rachunkiem sprawdzamy, że:

$$0 = \det\left[\mathbf{R}(3) - \lambda \mathbf{I}(3)\right] = \det \begin{bmatrix} 1-\lambda & 0,8 & 0 \\ 0,8 & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{bmatrix}, \tag{40}$$

czyli:

$$(1-\lambda)^3 - 0,64(1-\lambda) = 0, \tag{41}$$

stąd:

$$(1-\lambda)[\ (1-\lambda)^2 - 0{,}64 + 0{,}64\lambda] = 0,$$
$$(1-\lambda)[\ \lambda^2 - 2\lambda + 1 - 0{,}64 + 0{,}64\lambda] = 0,$$
$$(1-\lambda)[\ \lambda^2 - 1{,}36\lambda + 0{,}36] = 0,$$

stąd:

$$\lambda_1 = 1 \tag{42}$$

oraz:

$$\lambda^2 - 1{,}36\lambda + 0{,}36 = 0,$$

przeto:

$$\Delta = 1,8496 - 1,4400 = 0,4096$$

$$\sqrt{\Delta} = 0,64$$

$$\lambda_2 = \frac{1,36 - 0,64}{2} = 0,36 \tag{43}$$

$$\lambda_3 = \frac{1,36 + 0,64}{2} = 1. \tag{44}$$

A zatem wszystkie wartości własne macierzy $\mathbf{R}(3)$ danej wzorem (39) są dodatnie. Tym razem skorzystamy z wersji 2 kryterium, na podstawie którego można rozstrzygnąć czy wszystkie wartości własne macierzy $\mathbf{R}(3)$ danej wzorem (39) są dodatnie.

Zauważmy, że:

$$\sum_{\substack{i,j \\ i \neq j}} r_{ij}^2 = 0,64 + 0,64 = 1,28 \tag{45}$$

i wobec tego jest spełniony warunek (25), czyli:

$$1 - \sum_{\substack{i,j \\ i \neq j}} r_{ij}^2 = 1 - 1,28 = -0,28 \tag{46}$$

oraz nierówność (23), bo u dana jest wzorem (21). Zatem:

$$u = \frac{-1,28}{-0,28} \approx 4,57 > 0. \tag{47}$$

Wobec tego dla $k = 3$ (taki jest stopień macierzy $\mathbf{R}(3)$ danej wzorem (39)) mamy:

$$3 < 4,57. \tag{48}$$

A to oznacza, że wszystkie wartości własne macierzy $\mathbf{R}(3)$ danej wzorem (39) są dodatnie.

Sprawdziliśmy to bezpośrednio (patrz wzory (41), (42) i (43)).

Podamy teraz przedział $[\ \alpha^{\mathbf{R}}, \beta^{\mathbf{R}}]$, w którym leżą wszystkie wartości własne rozpatrywanej macierzy.

Dla macierzy $\mathbf{R}(3)$ danej wzorem (39) na podstawie wzorów (11), (12) i (13) mamy kolejno:

$$d^{\mathbf{R}} = \frac{1}{3}\sqrt{2 \cdot 3 \cdot 1,28} = \frac{\sqrt{3}}{3}\sqrt{2,56} = 0,5737 \cdot 1,6 \approx 0,91792,$$

przeto:

$$\alpha^{\mathbf{R}} = 1 - d^{\mathbf{R}} = 1 - 0{,}91792 = 0{,}08208,$$
$$\beta^{\mathbf{R}} = 1 + d^{\mathbf{R}} = 1 + 0{,}91792 = 1{,}91792,$$

a zatem wszystkie wartości własne $\lambda_i^{\mathbf{R}}$ $i = 1, 2, 3$ macierzy korelacji $\mathbf{R}(3)$ danej wzorem (29) należą do przedziału $[\alpha^{\mathbf{R}}, \beta^{\mathbf{R}}]$, czyli:

$$0{,}08208 < \lambda_i^{\mathbf{R}} \leqslant 1{,}91792, \; i = 1, 2, 3,$$

co stwierdziliśmy bezpośrednio (patrz wzory (41), (42) i (43)).

Zauważmy następnie, że podane rezultaty przenoszą się na przypadek macierzy uniwersalnej $\mathbf{G}(k) = [g_{ij}]$ stopnia k, dla której:

$$g_{ij} = \begin{cases} 1 & \text{dla} \quad i = j \\ r_i r_j & \text{dla} \quad i \neq j \end{cases} \tag{49}$$

gdzie r_p oznacza współczynnik korelacji pomiędzy parą zmiennych (Z_i, Z_j), $p = i, j$ oraz również macierzy uniwersalnej $\mathbf{Q}(k) = [q_{ij}]$ stopnia k, dla której:

$$q_{ij} = q_{ji} = \begin{cases} 1 & \text{dla} \quad i = j \\ \dfrac{r_i}{r_j} & \text{dla} \quad i \neq j \end{cases} \tag{50}$$

gdzie r_i oraz r_j mają te same znaczenie o którym informowaliśmy omawiając macierz uniwersalną.

Dodajmy jeszcze, że macierz uniwersalna i neutralna zostały wprowadzone do ekonometrii przez Z. Hellwiga (patrz np. Hellwig, 1976).

prof. zw. dr hab. Michał Kolupa,
Politechnika Radomska w Radomiu
dr hab. Joanna Plebaniak, prof. SGH.
Szkoła Główna Handlowa w Warszawie

LITERATURA

[1] Hellwig Z., (1976), *Przechodniość relacji skorelowania zmiennych losowych i płynące stąd wnioski ekonometryczne,* „Przegląd Statystyczny" nr 3, Warszawa.
[2] Kolupa M., (2006), *O konstrukcji przedziału dającego oszacowanie wartości własnych regularnej macierzy symetrycznej – Matematyka język uniwersalny* – AE Kraków.
[3] Mostowski A., Stark M , (1968), *Algebra liniowa*, PWN, Warszawa.

KILKA UWAG O WARTOŚCIACH WŁASNYCH MACIERZY KORELACJI

S t r e s z c z e n i e

W pracy wykazano, że wszystkie wartości własne macierzy korelacji $\mathbf{R}(k) = [r_{ij}]$ stopnia k leżą w przedziale $[\alpha^{\mathbf{R}}, \beta^{\mathbf{R}}]$, gdzie:

$$\alpha^{\mathbf{R}} = 1 - d^{\mathbf{R}}$$

$$\beta^{\mathbf{R}} = 1 + d^{\mathbf{R}}$$

zaś:

$$d^{\mathbf{R}} = \frac{1}{k}\sqrt{(k-1)k\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2}$$

Wartości własne macierzy korelacji są dodatnie, jeżeli tylko jest spełniony jeden z dwóch następujących warunków:
albo:

$$\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 < 1 \text{i} k \geqslant u$$

albo:

$$\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2 > 1$$

oraz :

$$k \leqslant u$$

gdzie:

$$u = \frac{-\sum\limits_{\substack{i,j \\ i\neq j}} r_{ij}^2}{1 - \sum\limits_{\substack{i,j \\ i\neq j}} r_{ij}^2}$$

Słowa kluczowe: macierz korelacji, wartości własne macierzy

SAME REMARKS ABOUT CHARACTERISTIC ROOTS OF MATRIX OF CORRELATION $\mathbf{R}(k) = [r_{ij}]_{kxk}$

S u m m a r y

The paper is devoted presentation construction the of the interval [$\alpha^{\mathbf{R}}$, $\beta^{\mathbf{R}}$] for all characteristic roots of matrix of correlation $\mathbf{R}(k) = [r_{ij}]_{kxk}$:

$$\alpha^{\mathbf{R}} = 1 - d^{\mathbf{R}}$$
$$\beta^{\mathbf{R}} = 1 + d^{\mathbf{R}}$$

where:

$$d^{\mathbf{R}} = \frac{1}{k}\sqrt{(k-1)k\sum_{\substack{i,j \\ i\neq j}} r_{ij}^2}$$

The all characteristic root $\lambda_i^{\mathbf{R}}$ are positive, if:

A)
$$\operatorname{sign}\left(1 - \sum_{\substack{i,j \\ i\neq j}} r_{ij}^2\right) = +1 \text{ and } k \geqslant u$$

or:

B)
$$\operatorname{sign}\left(1 - \sum_{\substack{i,j \\ i\neq j}} r_{ij}^2\right) = -1$$

and:

$$k \leqslant u$$

where:

$$u = \frac{-\sum\limits_{\substack{i,j \\ i\neq j}} r_{ij}^2}{1 - \sum\limits_{\substack{i,j \\ i\neq j}} r_{ij}^2}$$

Keywords: matrix of correlation, characteristic roof of matrix

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

JADWIGA KOSTRZEWSKA

PRZEGLĄD SEMIPARAMETRYCZNYCH METOD ESTYMACJI MODELI TOBITOWYCH TYPU I–III

1. WPROWADZENIE

Modele tobitowe rozwijają się od 1958 roku, kiedy to po raz pierwszy James Tobin w pracy [45] rozważył model regresji dla zmiennej zależnej cenzurowanej. Model ten nazwano modelem tobitowym od nazwiska jego twórcy oraz ze względu na pewne pokrewieństwo z modelem probitowym. W kolejnych latach następował rozwój modeli zmiennej zależnej ograniczonej (*limited dependent variable models*), tzn. modeli zmiennej zależnej uciętej (*truncated models*), cenzurowanej (*censored models*) lub podlegającej nieprzypadkowej selekcji[1] (*sample-selection models*, *selection models*). Takeshi Amemiya w pracy [1] dokonał przeglądu istniejących modeli i usystematyzował je ze względu na postać funkcji wiarygodności. W rezultacie wyodrębnił pięć głównych typów modeli tobitowych[2].

W celu estymacji parametrów w modelach tobitowych typu I–III (lub wyższego typu) na ogół stosuje się metodę największej wiarygodności. Metoda ta jednak wymaga założenia o normalności rozkładu składnika losowego. Metodę można stosować także w modelach tobitowych przy założeniu heteroskedastyczności składników losowych. W modelach z rodziny tobitowych o funkcji wiarygodności posiadającej więcej niż jedno ekstremum można stosować także dwustopniową metodę zaproponowaną przez J. Heckmana w pracy [11], która wymaga słabszych założeń. Metoda ta w pierwszym kroku wykorzystuje analizę probitową, zaś w drugim – korzysta z klasycznej metody najmniejszych kwadratów. Obecnie stosuje się ją zazwyczaj do wyliczenia wartości początkowych estymatorów zgodnych w procedurze iteracyjnej wyznaczania estymatorów największej wiarygodności.

[1] Zmienne ucięte (ang. *truncated variables*) to zmienne, których wartości są obserwowalne tylko w pewnym przedziale, a nie są brane pod uwagę wartości spoza niego. Zmienne cenzurowane (ang. *censored variable*s) to takie zmienne, których dokładne wartości są znane, gdy występują w pewnym przedziale, a nie są znane te wartości, które są spoza niego i wówczas zmiennej zostaje przypisana umowna wartość z brzegu przedziału. Jeżeli to, czy wartości badanej zmiennej są obserwowane czy nie, zależy od wartości innej skorelowanej z nią zmiennej, wówczas występuje tzw. nieprzypadkowa selekcja (selekcja, samoselekcja, ang. *self-selection, sample selection*), będąca uogólnieniem ucięcia lub cenzurowania. Pojęcie zmiennej uciętej, cenzurowanej i podlegającej nieprzypadkowej selekcji omówiono np. w pracach [21], [22].

[2] Przegląd modeli tobitowych można odnaleźć we wspomnianej pracy [1] oraz [9], [22], [24].

Gdy nie ma żadnych przesłanek odnośnie rozkładu składnika losowego modelu zmiennej zależnej cenzurowanej lub, gdy w wyniku przeprowadzenia odpowiedniego testu założenie o rozkładzie normalnym zostanie odrzucone, odpowiednim narzędziem do oszacowania modelu mogą okazać się metody semiparametryczne. Metody te wymagają słabszych założeń o rozkładzie, przy których można uzyskać estymatory o dobrych własnościach, tj. zgodne, asymptotycznie normalne. Niektóre z metod nie wymagają założenia o homoskedastyczności składnika losowego.

W literaturze anglojęzycznej metody semiparametryczne dla modeli zmiennej zależnej cenzurowanej (modeli tobitowych) omawiane są od szeregu lat. W artykule zaprezentowano niektóre metody estymacji semiparametrycznej oraz wskazano na ich przydatność przy weryfikacji założeń narzuconych na rozkład składnika losowego. Skupiono uwagę na metodach odnoszących się do modeli tobitowych typu I, II oraz III zgodnie z klasyfikacją T. Amemiyi (por. [1]).

2. MODELE TOBITOWE TYPU I–III

W celu prezentacji poszczególnych metod wprowadzono następujące oznaczenia[3]. Niech Y^* oraz Z^* będą zmiennymi ukrytymi (*latent variable*), tzn. zmiennymi, których wartości zazwyczaj nie są obserwowalne lub są obserwowalne w zależności od pewnych warunków (kryterium obserwowalności). Przez n oznaczono liczbę elementów w próbie, przy czym dla wygody zapisu przyjęto, że pierwszych n_1 elementów ($i = 1, \ldots, n_1$) w próbie odnosi się do dokładnie zaobserwowanych wartości zmiennej Y^*, natomiast następne $(n - n_1)$ elementów ($i = (n_1 + 1), \ldots, n$) – do przypisanych wartości umownych (o ile istnieją).

Model tobitowy typu I jest postaci[4] (por. [24]):

$$y_i = \begin{cases} y_i^* & \text{gdy } y_i^* > 0, \quad i = 1, \ldots, n_1 \\ c & \text{gdy } y_i^* \leqslant 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{1}$$

gdzie:

$y_i^* = \mathbf{X}_i \boldsymbol{\beta} + u_i$ – równanie regresji,

$\mathbf{X}_i$ – $(k+1)$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $X = \{X_1, \ldots, X_k\}$, odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element wektora $\mathbf{X}_i$ jest równy jedności,

$\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta_k)'$ – wektor (pionowy) parametrów równania regresji,

u_i – składniki losowe równania, niezależne i pochodzące z tego samego rozkładu o wartości oczekiwanej równej 0 oraz stałej wariancji σ_u. Najczęściej (w celu estymacji metodą największej wiarygodności) zakłada się, że składniki losowe podlegają rozkładowi normalnemu.

[3] Zob. pracę [24].

[4] W artykule rozważono modele z wartością progową równą zero. Nie wiąże się to ze stratą ogólności.

Model tobitowy typu II jest postaci (por. [24]):

$$y_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta} + u_i & \text{gdy} \quad I_i = 1, \quad i = 1, \ldots, n_1 \\ 0 \text{ lub brak} & \text{gdy} \quad I_i = 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{2}$$

gdzie:

$I_i = \begin{cases} 1 & \text{gdy } z_i^* > 0 \\ 0 & \text{gdy } z_i^* \leqslant 0 \end{cases}$ – zmienna binarna opisująca kryterium obserwowalności,

$z_i^* = \mathbf{W}_i\boldsymbol{\gamma} + v_i$ – równanie selekcji,

$\mathbf{W}_i$ – $(m+1)$-elementowy wektor (poziomy) wartości zmiennych objaśniających ze zbioru $W = \{W_1, \ldots, W_m\}$, odnotowanych dla i–tej obserwacji (jednostki), przy czym pierwszy element wektora $\mathbf{W}_i$ jest równy jedności,

$\boldsymbol{\gamma} = (\gamma_0, \gamma_1, \ldots, \gamma_m)'$ – wektor (pionowy) parametrów równania selekcji,

v_i – składniki losowe równania selekcji,

$\mathbf{X}_i, \mathbf{W}_i, \beta, u_i,$ – określone jak wcześniej, przy czym najczęściej (w celu estymacji metodą największej wiarygodności) zakłada się, że:

$(u_i, v_i) \sim N\left(0,\ 0;\ \begin{bmatrix} \sigma_u^2 & \rho\sigma_u\sigma_v \\ \rho\sigma_u\sigma_v & \sigma_v^2 \end{bmatrix}\right)$, gdzie ρ – współczynnik korelacji między u_i i v_i.

W modelu tobitowym typu II kryterium obserwowalności opisane jest przez zmienną binarną I. Z tego względu przyjmuje się $\sigma_v = 1$, podobnie jak w modelu probitowym.

Model tobitowy typu III jest postaci (por. [24]):

$$y_i = \begin{cases} \mathbf{X}_i\boldsymbol{\beta} + u_i & \text{gdy} \quad z_i > 0, \quad i = 1, \ldots, n_1 \\ 0 \text{ lub brak} & \text{gdy} \quad z_i = 0, \quad i = (n_1 + 1), \ldots, n, \end{cases} \tag{3}$$

gdzie tym razem zmienna opisująca kryterium obserwowalności jest postaci:

$$z_i = \begin{cases} z_i^* & \text{gdy} \quad z_i^* > 0 \\ 0 & \text{gdy} \quad z_i^* \leqslant 0 \end{cases},$$

przy pozostałych oznaczeniach jak wcześniej.

Przegląd modeli tobitowych (modeli zmiennej zależnej cenzurowanej) można znaleźć m.in. w pracach: [1], [9], [27], [28], a także [22], [24].

3. ESTYMACJA SEMIPARAMETRYCZNA W MODELU TOBITOWYM TYPU I

Wśród wielu metod semiparametrycznych stosowanych w wypadku modelu tobitowego typu I najbardziej popularne są metody wprowadzone przez J.L. Powella (zob. [38], [39], [40], [41]) oraz B.E. Honoré'a i J.L. Powella ([16]). Są to m.in.

metody: CLAD, STLS/SCLS, ICLAD oraz ICLS. W metodach tych określa się postać parametryczną równania regresji, ale nie narzuca się postaci parametrycznej rozkładu składnika losowego. Zaprezentowane metody semiparametryczne dają estymatory o dobrych własnościach dopuszczając dowolny rozkład składnika losowego i/lub jego heteroskedastyczność, przy pewnych słabszych założeniach. Poniżej metody te omówiono dla modelu tobitowego typu I z cenzurowaniem w zerze.

METODA NAJMNIEJSZYCH ODCHYLEŃ BEZWZGLĘDNYCH Z CENZUROWANIEM (*censored least absolute deviations*, *censored* LAD, CLAD). Metoda CLAD, wprowadzona przez J.L. Powella w pracy [38], jest metodą estymacji semiparametrycznej stosowaną w modelu tobitowym typu I. Przy dość ogólnych założeniach estymatory otrzymane metodą CLAD są zgodne i asymptotycznie normalne. Spośród wymaganych założeń należy wspomnieć następujące[5]:

- składniki losowe u_i są niezależne, o tym samym rozkładzie, przy czym ich rozkład może być dowolnym rozkładem, w którym mediana równa jest zero (warunkowo od $\mathbf{X}_i$),
- procent obserwacji „zerowych" (ocenzurowanych) nie jest zbyt wysoki,
- dla obserwacji „zerowych" odpowiadające im zmienne $\mathbf{X}_i$ nie są współliniowe.

Jeżeli w próbie jest znaczący procent obserwacji „zerowych", można zastosować estymator oparty na cenzurowanej regresji kwantylowej (ang. *censored quantile regression*) bazując na kwantylu wyższego rzędu niż mediana[6] (por. [39]). Główną zaletą metody jest fakt, że dopuszcza dowolny, w tym również asymetryczny (por. [5], s. 34), rozkład składnika losowego oraz heteroskedastyczność[7]. Metoda CLAD charakteryzuje się dużą złożonością numeryczną.

Celem metody CLAD jest uzyskanie estymatora wektora parametrów $\boldsymbol{\beta}$. Estymator ten otrzymuje się wyznaczając wektor $\hat{\boldsymbol{\beta}}_{CLAD}$ minimalizujący wyrażenie:

$$\frac{1}{n}\sum_{i=1}^{n}|y_i - \max\{0; \mathbf{X}_i\boldsymbol{\beta}\}|, \tag{4}$$

przy czym mediana zmiennej zależnej wynosi: $Me(y_i) = \max\{0; \mathbf{X}_i\boldsymbol{\beta}\}$.

Sposób wyznaczenia zgodnego estymatora asymptotycznej macierzy kowariancji można znaleźć w pracach J.L. Powella (zob. [38], [41]). W wypadku estymatorów otrzymanych metodą CLAD w modelu tobitowym typu I można stosować test Walda[8] odnośnie ewentualnych restrykcji nałożonych na parametry.

Metoda CLAD oraz inne metody semiparametryczne oparte na kwantylach nie są odpowiednie do estymacji modeli regresji uciętej.

---

[5] Szczegóły można odnaleźć w pracy J.L. Powella [38], s. 307.

[6] Estymator CLAD jest szczególnym przypadkiem estymatora opartego na cenzurowanej regresji kwantylowej.

[7] Przy heteroskedastycznym składniku losowym estymatory otrzymane metodą CLAD nadal pozostaną zgodne i asymptotycznie normalne (por. [38], s. 308).

[8] Postać testu Walda można odnaleźć np. w [28].

METODA NAJMNIEJSZYCH KWADRATÓW Z SYMETRYCZNYM PRZYCINANIEM lub CENZUROWANIEM (ang. *symmetrically trimmed least squares*, STLS; *symmetrically censored least squares*, SCLS). Metoda ta, również zaproponowana przez J.L. Powella w pracy [40], znajduje zastosowanie zarówno w modelu regresji uciętej (metoda STLS) jak i cenzurowanej (metoda SCLS). W niniejszym artykule omówiono tylko wersję metody dla cenzurowania. Metoda SCLS jest prostsza w zastosowaniach niż metoda CLAD, ale daje estymatory asymptotycznie mniej efektywne (por. [17], s. S65).

J.L. Powell w pracy [40] zauważył, że cenzurowanie lub ucięcie zmiennej zależnej sprawia, że rozkład składnika losowego jest asymetryczny, co jest główną przyczyną niezgodności estymatorów otrzymanych metodą najmniejszych kwadratów. W metodzie SCLS zakłada się[9], że u_i są niezależne o tym samym symetrycznym rozkładzie warunkowo od $\mathbf{X}_i$ (metoda dopuszcza dowolny symetryczny rozkład, niekoniecznie normalny). Korzystając z tego założenia sztucznie cenzuruje się dane: obserwacjom y_i takim, że $y_i > 2\mathbf{X}_i\boldsymbol{\beta}$ przypisuje się wartość $2\mathbf{X}_i\boldsymbol{\beta}$.

Estymator $\hat{\beta}_{SCLS}$ otrzymuje się wyznaczając argument minimalizujący wyrażenie:

$$\sum_{i=1}^{n}(y_i - \max\{\tfrac{1}{2}y_i; \mathbf{X}_i\boldsymbol{\beta}\})^2 + \sum_{i=1}^{n} 1\{y_i > 2\mathbf{X}_i\boldsymbol{\beta}\} \cdot (\tfrac{1}{4}y_i^2 - \max\{0; \mathbf{X}_i\boldsymbol{\beta}\}^2), \tag{5}$$

gdzie $1\{A\}$ jest funkcją charakterystyczną zbioru A.

Także ta metoda prowadzi do estymatorów zgodnych i asymptotycznie normalnych[10], nawet przy wystąpieniu heteroskedastyczności (o nieznanej formie). J.L. Powell (zob. [40], s. 1444–1445) wskazał zgodny estymator asymptotycznej macierzy kowariancji (por. także prace [18] oraz [33], s. 129).

METODA OPARTA NA RÓŻNICY PAR (ang. *pairwise difference method*). Metoda ta została zaproponowana przez B.E. Honoré'a i J.L. Powella w pracy [16] dla modeli regresji uciętej lub cenzurowanej, przy głównym założeniu, że u_i są niezależne o tym samym rozkładzie, a przy tym niezależne od zmiennych objaśniających $\mathbf{X}_i$. Dopuszczalny jest rozkład asymetryczny składnika losowego u_i, natomiast wykluczona jest jego heteroskedastyczność (por. także [5], s. 34). Idea metody polega na symetrycznym sztucznym cenzurowaniu par danych w taki sposób, aby po cenzurowaniu pary te miały identyczny rozkład[11]. Estymator parametrów modelu otrzymuje się wyznaczając argument minimalizujący wyrażenie (por. [4], s. 18):

$$\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} s(y_i, y_j, (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}), \tag{6}$$

[9] W pracy [33], s. 127, opisano sposoby weryfikacji założeń gwarantujących zgodność estymatorów otrzymanych metodą SCLS dla modelu tobitowego.

[10] Por. [40], gdzie zamieszczono szerszą dyskusję odnośnie istnienia globalnego minimum oraz omówiono własności otrzymanych estymatorów przy pewnych warunkach.

[11] Opis metody można znaleźć w [16], a także [4], s. 16, lub [15], s. 112.

gdzie:

$$s(y_1, y_2, \delta) = \begin{cases} \Xi(y_1) - (y_2 + \delta)\xi(y_1) & \text{gdy } \delta \leqslant -y_2 \\ \Xi(y_1 - y_2 - \delta) & \text{gdy } -y_2 < \delta < y_1 \\ \Xi(-y_2) + (y_1 - \delta)\xi(-y_2) & \text{gdy } y_1 \leqslant \delta \end{cases}, \tag{7}$$

$\Xi(.)$ – pewna parzysta i wypukła funkcja,

$\xi(.)$ – pewna nieparzysta funkcja, dla której powyższe wyrażenia mają sens.

Estymatory uzyskane tą metodą są zgodne i asymptotycznie normalne. Jednak za jej pomocą nie można wyznaczyć estymatora wyrazu wolnego bez nałożenia dodatkowych warunków (założenie o niezależności u_i od $\mathbf{X}_i$ nie nakłada restrykcji na położenie rozkładu składnika losowego)[12].

B.E. Honoré i J.L. Powell w pracy [16] przyjęli $\Xi(d) = |d|$ otrzymując estymatory o dobrych własnościach nawet dla małej próby. W literaturze przedmiotu (por. [5], s. 32) w tym wypadku metodę nazywa się metodą najmniejszych odchyleń bezwzględnych z identycznym cenzurowaniem (ang. *identically censored least absolute deviations*, ICLAD). W przypadku, gdy $\Xi(d) = d^2$ dla metody stosuje się nazwę metoda najmniejszych kwadratów z identycznym cenzurowaniem (ang. *identically censored least squares*, ICLS). Stosowanie estymatora ICLAD jest bardziej wskazane niż estymatora ICLS (por. [5], [16]).

Estymator $\hat{\boldsymbol{\beta}}_{ICLAD}$ otrzymuje się wyznaczając argument minimalizujący wyrażenie (por. [15], s. 113):

$$\sum_{i<j} \left(1 - 1\left\{y_i \leqslant \max\{0, (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\}; y_j \leqslant \max\{0, (\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta}\}\right\}\right) \cdot \left|y_i - y_j - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right|, \tag{8}$$

natomiast estymator $\hat{\boldsymbol{\beta}}_{ICLS}$ – minimalizujący wyrażenie (por. [15], s. 113):

$$\begin{aligned} &\sum_{i<j} \left(\max\{y_i, (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\} - \max\{y_j, (\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta}\} - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right)^2 + \\ &+2 \cdot 1\{y_i < (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\} \cdot ((\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta} - y_i)y_j + \\ &+2 \cdot 1\{y_j < (\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta}\} \cdot ((\mathbf{X}_j - \mathbf{X}_i)\boldsymbol{\beta} - y_j)y_i \end{aligned} . \tag{9}$$

Na zakończenie rozważań o metodach semiparametrycznych dla modelu tobitowego typu I warto wspomnieć o pewnej własności przydatnej przy wnioskowaniu o poprawności specyfikacji modelu, w tym poprawności założeń. Mianowicie wartości estymatorów otrzymanych metodą CLAD można traktować jako pewnego rodzaju nieformalne kryterium weryfikujące poprawność założeń modelu, gdyż estymator ten jest zgodny przy:

- założeniu normalności składnika losowego u_i – wymaganym do zgodności estymatorów otrzymanych metodą największej wiarygodności,

[12] Por. np. [4], s. 16.

- założeniu niezależności u_i – wymaganym do zgodności estymatorów otrzymanych metodą ICLAD lub ICLS,
- założeniu warunkowej symetryczności rozkładu u_i – wymaganym do zgodności estymatorów otrzymanych metodą SCLS.

Łączne porównanie[13] wartości estymatorów otrzymanych różnymi metodami pozwala wnioskować o ewentualnym naruszeniu założeń (por. [4], s. 21–24).

4. ESTYMACJA SEMIPARAMETRYCZNA W MODELU TOBITOWYM TYPU II

Metody semiparametryczne dla modeli, w których równanie selekcji jest opisane za pomocą zmiennej binarnej I, w tym dla modelu tobitowego typu II danego wzorem (2), na ogół wymagają, aby co najmniej jedna zmienna należąca do zbioru zmiennych W opisującego równanie selekcji, nie należała do zbioru zmiennych X objaśniającego równanie regresji (por. [5], s. 35, [30], s. 24, [46],[47], s. 219). W zastosowaniach założenie to nie jest wygodne i prawdopodobnie z tego względu semiparametryczne modele tobitowe typu II są znacznie rzadziej stosowane niż np. modele tobitowe typu I lub III. Ponadto całkowity brak znajomości wartości zmiennej Z^* w modelu tobitowym typu II sprawia, że nie jest możliwe w szczególności zastosowanie metod analogicznych do tych, odpowiednich dla modelu tobitowego typu I. Częściej natomiast stosuje się metody nieparametryczne.

Tak jak poprzednio metody semiparametryczne pozwalają na osłabienie założenia o rozkładzie normalnym składnika losowego.

Większość z metod dla modelu tobitowego typu II przebiega w dwóch etapach, podobnie jak dwustopniowa metoda Heckmana[14]. W każdym z etapów stosuje się metody semiparametryczne lub nieparametryczne. Podstawą estymacji jest zapisanie warunkowej wartości oczekiwanej zmiennej zależnej w następujący sposób:

$$E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = \mathbf{X}_i\boldsymbol{\beta} + E(u_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i). \tag{10}$$

Przy tym zazwyczaj zakłada się, że warunkowa wartość oczekiwana składnika losowego u_i jest funkcją $\mathbf{W}_i\boldsymbol{\gamma}$ i tylko w ten sposób zależy od $\mathbf{W}_i$, tzn. (por. [46], s. 139 i n.):

$$E(u_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = g(\mathbf{W}_i\boldsymbol{\gamma}), \tag{11}$$

gdzie g jest nieznaną funkcją[15].

W pierwszym kroku estymuje się wektor parametrów $\boldsymbol{\gamma}$ w równaniu selekcji bez narzucania konkretnego rozkładu na składnik losowy v_i równania selekcji. Można za-

[13] Porównania te przeprowadza się w sposób podobny do testu Hausmana.

[14] Por. [11], [12].

[15] Warunkową wartość oczekiwaną można zapisać w ogólny sposób $E(y_i|I_i = 1, \mathbf{X}_i, \mathbf{W}_i) = g_1(\mathbf{X}_i) + g_2(\mathbf{W}_i\boldsymbol{\gamma})$, gdzie g_1 oraz g_2 – nieznane funkcje, przy czym zazwyczaj określa się postać funkcji g_1 np. liniową. Szerszą dyskusję odnośnie odpowiednich założeń i postaci funkcji można znaleźć np. w pracy [29], s. 353–358.

stosować jedną z metod semiparametrycznych lub nieparametrycznych dla modeli binarnych np. metodę R.W. Kleina i R.H. Spady (zob. [19])[16] prowadzącą do zgodnych i asymptotycznie normalnych estymatorów.

Również w drugim kroku estymacji semiparametrycznej modelu tobitowego typu II można postąpić na wiele sposobów[17]. Jedna z metod przebiega następująco (por. [31]). Korzystając z estymatora $\mathbf{W}_i\hat{\boldsymbol{\gamma}}$ oraz przybliżając funkcję g za pomocą szeregu aproksymacji opartego na odwrotności współczynnika Millsa równanie regresji modelu tobitowego typu II należy zapisać w postaci[18]:

$$y_i = \mathbf{X}_i\boldsymbol{\beta} + \zeta_0 + \zeta_1\lambda(\mathbf{W}_i\hat{\boldsymbol{\gamma}}) + \zeta_2\lambda^2(\mathbf{W}_i\hat{\boldsymbol{\gamma}}) + \ldots + \zeta_L\lambda^L(\mathbf{W}_i\hat{\boldsymbol{\gamma}}) + \varepsilon_i, \tag{12}$$

gdzie:

$\lambda(\alpha) = \dfrac{\varphi(\alpha)}{1-\Phi(\alpha)}$ – odwrotność współczynnika Millsa dla cenzurowania lewostronnego,

$\beta_0 + \zeta_0, \quad \beta_1, \ldots, \beta_k, \quad \zeta_1, \ldots, \zeta_L$ – nieznane parametry.

Ten sposób aproksymacji ma tę zaletę, że przy $L = 1$ model (12) jest tożsamy z modelem tobitowym typu II z założeniem rozkładu normalnego. Jeżeli oszacowane parametry $\zeta_2, \ldots, \zeta_L$ modelu (12) nie są łącznie statystycznie nieistotne, można przypuszczać, że założenie o rozkładzie normalnym nie jest poprawne (por. [25], s. 333), a także [37]). Postać asymptotycznej wariancji podano w pracy [34] autorstwa W.K. Neweya.

Estymację parametrów równania (12) przeprowadza się na podstawie obserwacji, dla których $I_i = 1$. Należy zastosować jedną z metod odpornych na heteroskedastyczność składnika losowego ε_i (np. metodę P.M. Robinsona [43], por. [31], s. 70).

Przy zastosowaniu aproksymacji (12) nie ma jednoznacznej reguły określającej wartość L. Jedna z propozycji ustala kryterium doboru L w taki sposób, by dodanie następnego elementu w szeregu aproksymacji nie zmieniało zasadniczo wartości wyrazu wolnego (por. [25]).

Wyraz wolny modelu (12) jest sumą dwóch wyrazów wolnych: $\beta_0 + \zeta_0$. W celu estymacji oddzielnie β_0 i ζ_0 można skorzystać z propozycji J. Heckmana w pracy [13] lub D. Andrewsa i M. Schafgans w pracy [2][19].

5. ESTYMACJA SEMIPARAMETRYCZNA W MODELU TOBITOWYM TYPU III

Sposoby estymacji semiparametrycznej w wypadku modelu tobitowego typu III danego wzorem (3) są odmienne od tych mających zastosowanie w modelu tobitowym

[16] Opis metody można znaleźć np. w [30], s. 25, [36], s. 283. Inną metodę zaproponowano np. w pracy [42]. W pracy [30], Appendix A, przedstawiono także testy umożliwiające porównanie równania selekcji opisanego za pomocą modelu probitowego z tym otrzymanym metodą semi- lub nieparametryczną.

[17] W artykule zaprezentowano tylko jedno z wielu możliwych podejść. Szeroki opis stosowanych metod można znaleźć np. w [41] lub [46].

[18] Inne postaci aproksymacji zaprezentowano np. w pracy [8], [25], [34], [37].

[19] Por. prace [25] lub [44] odnośnie dyskusji na temat własności tego estymatora.

typu II. Różnice wynikają przede wszystkim z faktu, że obecnie dostępnych jest więcej informacji: znane są wartości zmiennej Z. Oczywiście w celu estymacji modelu tobitowego typu III można zastosować również metody semiparametryczne odpowiednie do estymacji modelu tobitowego typu II wykorzystując tylko informację o znaku zmiennej Z^*. Jednakże ze względu na utratę informacji (nie korzysta się ze znanych dokładnych wartości zmiennej cenzurowanej Z), otrzymane estymatory zazwyczaj będą mniej efektywne.

W literaturze przedmiotu metody semiparametryczne dla modelu tobitowego typu III rozważano m.in. w pracach: [15], a także [6], [26]. W artykule zaprezentowano metody opisane w pierwszej z tych prac.

Pierwsza metoda zaproponowana w pracy[20] [15] opiera się głównie na założeniu, że dwuwymiarowy rozkład składników losowych (u_i, v_i) warunkowo od $(\mathbf{X}_i, \mathbf{W}_i)$ jest rozkładem symetrycznym w takim sensie, że (u_i, v_i) pochodzą z tego samego rozkładu co $(-u_i, -v_i)$. Ponadto, aby otrzymane estymatory parametrów $\boldsymbol{\beta}$ oraz $\boldsymbol{\gamma}$ były zgodne i asymptotycznie normalne, muszą być spełnione pewne dodatkowe warunki regularności[21]. Warto podkreślić, że metoda dopuszcza heteroskedastyczność składnika losowego.

W pierwszym kroku omawianej metody należy zastosować metodę CLAD lub SCLS do równania selekcji (będącego modelem tobitowym typu I) w celu wyznaczenia estymatora wektora parametrów $\boldsymbol{\gamma}$ (por. wzór (4) lub, odpowiednio, (5)). W drugim kroku wartości $\mathbf{W}_i\hat{\boldsymbol{\gamma}}_{CLAD}$ (odpowiednio $\mathbf{W}_i\hat{\boldsymbol{\gamma}}_{SCLS}$) wykorzystuje się do symetrycznego przycięcia (cenzurowania) rozkładu składników losowych u_i. Następnie wystarczy zastosować np. metodę LAD lub, odpowiednio, metodę najmniejszych kwadratów do równania regresji po przycięciu, czyli na nowej, przyciętej próbie. W rezultacie estymator wektora parametrów $\boldsymbol{\beta}$ otrzymuje się wyznaczając argument $\tilde{\tilde{\boldsymbol{\beta}}}_{CLAD}$ minimalizujący wyrażenie:

$$\frac{1}{n}\sum_{i=1}^{n} 1\{0 < z_i < 2\mathbf{W}_i\hat{\boldsymbol{\gamma}}_{CLAD}\} \cdot |y_i - \mathbf{X}_i\boldsymbol{\beta}|, \tag{13}$$

przy zastosowaniu metody LAD lub argument $\tilde{\tilde{\boldsymbol{\beta}}}_{SCLS}$ minimalizujący wyrażenie:

$$\frac{1}{n}\sum_{i=1}^{n} 1\{0 < z_i < 2\mathbf{W}_i\hat{\boldsymbol{\gamma}}\} \cdot (y_i - \mathbf{X}_i\boldsymbol{\beta})^2, \tag{14}$$

przy zastosowaniu metody najmniejszych kwadratów.

Druga z proponowanych w pracy [15] metod bazuje głównie na założeniu, że zmienne objaśniające X oraz W są niezależne od składników losowych u oraz v, przy czym nie narzuca się założenia odnośnie symetryczności rozkładu składników

[20] W pracy [15] podano także sposób wyznaczania estymatorów w wypadku modeli selekcji uciętej.
[21] Szczegółowe założenia można znaleźć w [15], s. 115.

losowych[22]. W celu uzyskania estymatorów zgodnych i asymptotycznie normalnych wymagane są pewne dodatkowe warunki regularności[23].

Tym razem w pierwszym kroku metody następuje estymacja równania selekcji metodą semiparametryczną służącą do estymacji modelu tobitowego typu I przy założeniu niezależności zmiennych objaśniających i składników losowych. W pracy [15] stosowano metody ICLAD lub ICLS zaproponowane przez B.E. Honoré'a i J.L. Powella (zob. [16]), tzn. wyznaczono estymator $\hat{\boldsymbol{\gamma}}_{ICLAD}$ (por. wzór (8)) lub odpowiednio $\hat{\boldsymbol{\gamma}}_{ICLS}$ (por. wzór (9)) wektora parametrów $\boldsymbol{\gamma}$, traktując równanie selekcji jako model tobitowy typu I.

W drugim kroku w celu eliminacji efektu selekcji zastosowano podejście bazujące na rozkładzie różnicy par[24], analogicznie jak w opisanej wcześniej semiparametrycznej metodzie B.E. Honoré'a i J.L. Powella (por. [16]) dla modelu tobitowego typu I (por. wzory (8) i (9)). Estymatory wektora parametrów otrzymuje się wyznaczając argument $\tilde{\tilde{\boldsymbol{\beta}}}_{ICLAD}$ minimalizujący wyrażenie:

$$\sum_{i<j} 1\left\{z_i > \max\{0, (\mathbf{W}_i - \mathbf{W}_j)\hat{\boldsymbol{\gamma}}_{ICLAD}\}; z_j > \max\{0, (\mathbf{W}_j - \mathbf{W}_i)\hat{\boldsymbol{\gamma}}_{ICLAD}\}\right\} \cdot \left|y_i - y_j - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right|$$

lub odpowiednio $\tilde{\tilde{\boldsymbol{\beta}}}_{ICLS}$ minimalizujący wyrażenie:

$$\sum_{i<j} 1\left\{z_i > \max\{0, (\mathbf{W}_i - \mathbf{W}_j)\hat{\boldsymbol{\gamma}}_{ICLS}\}; z_j > \max\{0, (\mathbf{W}_j - \mathbf{W}_i)\hat{\boldsymbol{\gamma}}_{ICLS}\}\right\} \cdot \left(y_i - y_j - (\mathbf{X}_i - \mathbf{X}_j)\boldsymbol{\beta}\right)^2.$$

Przy tym, tak jak poprzednio, bardziej wskazane jest stosowanie metody bazującej na ICLAD.

6. WERYFIKACJA HIPOTEZ STATYSTYCZNYCH

Przydatna jest możliwość porównania różnych specyfikacji modeli tobitowych np. postaci parametrycznej z modelem semiparametrycznym. W literaturze przedmiotu stosuje się np. test J.A. Hausmana (zob. [10]). Idea tego testu jest następująca[25]. Niech $M-$elementowe wektory $\hat{\boldsymbol{\pi}}_0$ oraz $\hat{\boldsymbol{\pi}}_1$ będą dwoma estymatorami wektora parametrów $\boldsymbol{\pi}$, które przy założeniu hipotezy zerowej są zgodne i asymptotycznie normalne o asymptotycznych macierzach wariancji odpowiednio: V_0 oraz V_1. Ponadto przy założeniu hipotezy zerowej $\hat{\boldsymbol{\pi}}_0$ jest estymatorem asymptotycznie efektywnym oraz $V_1 - V_0$ jest nieujemnie określona. Przy założeniu hipotezy alternatywnej estymator $\hat{\boldsymbol{\pi}}_1$ pozostaje

[22] W przypadku przyjęcia wspomnianego założenia o niezależności należy rozpatrywać model bez wyrazu wolnego w równaniu regresji i w równaniu selekcji (por. [15], s. 112).

[23] Por. [15], s. 117.

[24] Idea metody bazuje na fakcie, że rozkład różnicy niezależnych zmiennych losowych pochodzących z tego samego rozkładu jest symetryczny wokół zera.

[25] Por. np. [28], s. 558–559, [32], s. 1321, [33].

zgodny, ale estymator $\hat{\boldsymbol{\pi}}_0$ traci tę własność (nie jest zgodny). Statystyka testowa jest postaci:

$$N \cdot (\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_0)'(\hat{V}_1 - \hat{V}_0)^{-1}(\hat{\boldsymbol{\pi}}_1 - \hat{\boldsymbol{\pi}}_0) \sim \chi^2[M], \tag{15}$$

gdzie N jest liczbą elementów próby.

W.K. Newey (zob. [33]) podał sposób zastosowania testu Hausmana w celu porównania estymatorów parametrów modelu tobitowego typu I otrzymanych metodą największej wiarygodności przy założeniu rozkładu normalnego i homoskedastyczności składnika losowego z estymatorami otrzymanymi metodą semiparametryczną SCLS odporną na naruszenie tych założeń. Dzięki odporności tego ostatniego estymatora test Hausmana można zastosować jako narzędzie sprawdzające podstawowe założenia parametrycznego modelu tobitowego typu I. W przypadku odrzucenia hipotezy zerowej należy uznać, że co najmniej jedno z założeń (o rozkładzie normalnym lub o homoskedastyczności) nie jest spełnione. Do przeprowadzenia dalszego wnioskowania należy posłużyć się estymatorami otrzymanymi metodą semiparametryczną.

Na podobnej zasadzie można skonstruować test porównujący estymatory otrzymane metodą największej wiarygodności z estymatorami otrzymanymi dowolną inną metodą semiparametryczną (por. [17], s. S67, [39]). Przy tym należy zwrócić uwagę, że macierze wariancji estymatorów $\hat{\boldsymbol{\pi}}_0$ oraz $\hat{\boldsymbol{\pi}}_1$ wyznacza się przy założeniu hipotezy zerowej. Jest to szczególnie przydatne, gdyż macierze wariancji estymatorów otrzymanych metodami semiparametrycznymi są zazwyczaj proste do wyznaczenie przy założeniu rozkładu normalnego (por. [4]).

B. Melenberg i A. van Soest w pracy [31] zastosowali test Hausmana w celu porównania modelu tobitowego z rozkładem normalnym z modelem otrzymanym metodą CLAD, jak również przy porównaniu semiparametrycznych specyfikacji modelu. Także w pracy [4] przeprowadzono analizę różnic wartości semiparametrycznych estymatorów powołując się na podobieństwo do testu Hausmana. Głównym celem takiej analizy jest wskazanie, które z założeń modelu zostało naruszone, gdyż każdy z omówionych estymatorów semiparametrycznych modelu tobitowego wymaga innych założeń.

Test Hausmana w odpowiedniej postaci można również zastosować do porównania specyfikacji parametrycznych (model probitowy) i semiparametrycznych równania selekcji (por. np. [35]).

Także analiza reszt modelu w semiparametrycznych modelach tobitowych może pomóc w określeniu, czy któreś z założeń o rozkładzie składników losowych jest naruszone (por. np. [4]). W literaturze przedmiotu zazwyczaj stosuje się mniej formalne testy, w tym w dużej mierze graficzne (por. [7], [17]).

7. UWAGI KOŃCOWE

Zaprezentowane metody estymacji semiparametrycznej znajdują zastosowanie w modelach tobitowych typu I–III, a także innych modelach zmiennej zależnej ograniczonej. Metody te pozwalają w szczególności na rozluźnienie założenia o rozkładzie

normalnym składników losowych, narzucając jednocześnie inne, na ogół mniej krępujące warunki. Ponadto sięgnięcie po metody semiparametryczne umożliwia weryfikację, które z założeń o rozkładzie składnika losowego modelu zostało naruszone. Dzięki temu można zweryfikować, czy można zastosować np. metodę największej wiarygodności, umożliwiającą m.in. poprawną interpretację wyników.

Omawiane metody nie są jednak pozbawione pewnych wad. Jedną z nich jest występujący w niektórych metodach problem niejednoznacznej identyfikacji wyrazu wolnego modelu. Ponadto niektóre metody, np. CLAD, charakteryzują się dużą złożonością obliczeniową. Kolejną niedogodnością jest brak metod pozwalających na poprawną interpretację uzyskanych wyników. Specyfika modeli tobitowych typu I–III, oraz innych modeli zmiennej zależnej ograniczonej, sprawia, że wpływy krańcowe zmiennych objaśniających nie są równe parametrom modelu[26]. Ze względu na brak określenia parametrycznej postaci rozkładu składnika losowego występują trudności z wyznaczeniem wpływów krańcowych zmiennych objaśniających na analizowane zjawisko. Brak możliwości wyznaczenia wpływów krańcowych równoznaczny jest z brakiem poprawnej interpretacji otrzymanych wyników na podstawie modeli tobitowych. Jest to duże utrudnienie w stosowaniu metod semiparametrycznych w wypadku rozważanych modeli tobitowych.

Uniwersytet Ekonomiczny w Krakowie

LITERATURA

[1] Amemiya T. (1984), *Tobit Models: A survey*, Journal of Econometrics, vol. 24, s. 3-61.
[2] Andrews D., Schafgans M. (1998), *Semiparametric Estimation of a Sample Selection Model*, Review of Economic Studies, vol. 65, nr 224, s. 497-518.
[3] Bloomfield P., Steiger W.L. (1983), *Least Absolute Deviations: Theory, Applications, and Algorithms*, Progress in Probability and Statistics, vol. 6, Birkhäuser Boston, Boston, Mass, USA.
[4] Chay K.Y., Honoré B.E. (1998), *Estimation of Semiparametric Regression Models*, Journal of Human Resources, vol. 33, nr 1, s. 4-38.
[5] Chay K.Y., Powell J.L. (2001), *Semiparametric Censored Regression Models*, Journal of Economic Perspectives, vol. 15, nr 4, s. 29-42.
[6] Chen S. (1997), *Semiparametric Estimation of the Type-3 Tobit Model*, Journal of Econometrics, vol. 80, s. 1-34
[7] Chesher A., Irish M. (1987), *Residual Analysis on the Grouped and Censored Normal Linear Model*, Journal of Econometrics, vol. 34, s. 33-61.
[8] Gallant R., Nychka G. (1987), *Semi-Nonparametric Maximum Likelihood Estimatnion*, Econometrica, vol. 55, s. 363-390.
[9] Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, wyd. 2, Monografie i Opracowania 490, Oficyna Wydawnicza SGH, Warszawa.
[10] Hausman J.A. (1978), *Specification Tests in Econometrics*, Ecoonmetrica, vol. 46, s. 1251-1272.

[26] Zob. szerszą dyskusję na temat interpretacji w modelach tobitowych w pracy [24].

[11] Heckman J. (1976), *The Common Structure of Statistical Models of Truncation, Sample Selection and Limited Dependent Variables and a Simple Estimator for Such Models*, Annals of Economic and Social Measurement, nr 5/4, 1976, s. 475-492.
[12] Heckman J. (1979), *Sample Selection Bias as a Specification Error*, Econometrica, vol. 47, nr 1, s. 153-161.
[13] Heckman J. (1990), *Varieties of Selection Bias*, American Economic Review, 80, s. 313-318.
[14] Heckman J. (2003), *Microdata, Heterogeneity and the Evaluation of Public Policy* [w:] *Nobel Lectures in Economic Sciences 1996 – 2000*, T. Persson (ed.), World Scientific Publishing Co., Singapore, s. 255-322.
[15] Honoré B.E., Kyriazidou E., Udry C. (1997), *Estimation of Type 3 Tobit Models Using Symmetric Trimming and Pairwise Comperisions*, Journal of Econometrics, vol. 76, s. 107-128.
[16] Honoré B.E., Powell J.L. (1994), *Pairwise Difference Estimators of Censored and Truncated Regression Models*, Journal of Econometrics, vol. 64, nr 1–2, s. 241-278.
[17] Horowitz J.L., Neumann G.R. (1989), *Specification Testing in Censored Regression Models: Parametric and Semiparametric Methods*, Journal of Applied Econometrics, vol. 4, s. S61-S86.
[18] Johnston J. i DiNardo J. (1997), *Econometric Methods*, McGraw-Hill Companies, Inc., New York.
[19] Klein R.W., Spady R.H. (1993), *An Efficient Semiparametric Estimator of Binary Response Models*, Econometrica, vol. 61, nr 2, s. 387-421.
[20] Kostrzewska J. (2003), *Model tobitowy jako szczególny przypadek cenzurowanego modelu regresji* [w:] *Przestrzenno-czasowe modelowanie i prognozowanie zjawisk gospodarczych*, red. A. Zeliaś, Wyd. AE w Krakowie, Kraków, s. 397-404.
[21] Kostrzewska J. (2008), *Zastosowanie wybranych modeli tobitowych do opisu tygodniowej liczby godzin pracy* [w:] *Taksonomia 15. Klasyfikacja i analiza danych – teoria i zastosowania*, pod red. K. Jajugi i M. Walesiaka, Wyd. AE we Wrocławiu, Wrocław.
[22] Kostrzewska J. (2009), *Truncated and Censored Dependent Variables and Their Models*, referat wygłoszony na 15th Polish-Slovak-Ukrainian Scientific Seminar "Statistical Analysis of Socio-Economic Consequences of Transition Processes in Central-East European Countries", Kraków, 21–24 października 2008 r.
[23] Kostrzewska J. (2011), *Analiza porównawcza tygodniowego czasu pracy zamężnych kobiet pracujących na własny rachunek lub najemnie*, [w:] *Osiągnięcia i perspektywy modelowania i prognozowania zjawisk społeczno-gospodarczych*, pod red. B. Pawełek, Wyd. UEK, Kraków, w druku.
[24] Kostrzewska J. (2011), *Interpretacja w modelach tobitowych*, Przegląd Statystyczny, tom 58, z. 3-4, s. 256-280.
[25] Lanot G., Walker I. (1998), *The union/non-union wage differential: An application of semiparametric methods*, Journal of Econometrics, vol. 84, nr 2, s. 327-349.
[26] Lee L.-F. (1994), *Semiparametric Two-Stage Estimation of Sample-Selection Models Subject to Tobit-Type Selection Rules*, Journal of Econometrics, vol. 61, s. 305-344.
[27] Maddala G.S. (1994), *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.
[28] Maddala G.S. (2006), *Ekonometria*, Wyd. Nauk. PWN, Warszawa.
[29] Manski C.F. (1989), *Anatomy of the Selection Problem*, The Journal of Human Resources, vol. 24, nr 3, s. 343-360.
[30] Martins M.F. (2001), *Parametric and Semiparametric Estimation of Sample Selection Models: An Empirical Application to the Female Labour Force in Portugal*, Journal of Applied Econometrics, vol. 16, nr 1, s. 23-39.
[31] Melenberg B., van Soest A. (1996), *Parametric and Semi-Parametric Modelling on Vacation Expenditures*, Journal of Applied Econometrics, vol. 11, s. 59-76.
[32] Nelson F.D. (1981), *A Test for Misspecification in the Censored Normal Model*, Econometrica, vol. 49, s. 1317-1329.

[33] Newey W.K. (1987), *Specification Tests for Distributional Assumption in the Tobit Model*, Journal of Econometrics, vol. 34, s. 125-145.
[34] Newey W.K. (1991), *Efficient Estimation of Tobit Models under Conditional Symmetry* [w:] *Nonparametric and Semiparametric Estimation Methods in Econometrics and Statistics*, pod red. W.A. Barnetta, J. Powella, G.E. Tauchena, Cambridge University Press, Cambridge.
[35] Newey W.K., Powell J.L., Walker J.R. (1990), *Semiparametric Estimation of Selection Models: Some Empirical Results*, The American Economic Review, vol.80, nr 2, s. 324-328.
[36] Pagan A.R., Ullah A. (1999), *Nonparametric Econometrics*, Cambridge University Press, Cambridge.
[37] Pagan A.R., Vella F. (1989), *Diagnostic Tests for Models Based on Individual Data: A Survey*, Journal of Applied Econometrics, vol. 4, s. S29-S59.
[38] Powell J.L. (1984), *Least Absolute Deviations Estimation for the Censored Regression Model*, Journal of Econometrics, vol. 25, nr 3, s. 303-325.
[39] Powell J.L. (1986), *Censored Regression Quantiles*, Journal of Econometrics, vol. 32, s. 143-155.
[40] Powell J.L. (1986), *Symmetrically Trimmed Least Squares Estimation for Tobit Models*, Econometrica, vol. 54, nr 6, s. 1435-1460.
[41] Powell J.L. (1994), *Estimation of Semiparametric Models* [w:] *Handbook of Econometrics*, vol. IV, red. R.F. Engle, D.L. McFadden, Elsevier Science Publishers B.V., s. 2444-2514.
[42] Powell J.L., Stock J.H., Stoker T.M. (1989), *Semiparametric Estimation of Index Coefficients*, Econometrica, vol. 57, nr 6, s. 1403-1430.
[43] Robinson P.M. (1987), *Asymptotically Efficient Estimation in the Presence of Heteroskedasticity of Unknown Form*, Econometrica, vol. 55, s. 875-891.
[44] Schafgans M. (2004), *Finite Sample Properties for the Semiparametric Estimation of the Intercept of a Censored Regression Model*, Statistica Neerlandica, vol. 58, nr. 1, s. 35-56.
[45] Tobin J. (1958), *Estimation of Relationships for Limited Dependent Variables*, Econometrica, vol. 26, s. 24-36.
[46] Vella F. (1998), *Estimating Models with Sample Selection Bias: A Survey*, Journal of Human Resources, vol. 33, nr 1, s. 127-169.
[47] Verbeek M. (2001), *A Guide to Modern Econometrics*, John Wiley & Sons, Chichester.

PRZEGLĄD SEMIPARAMETRYCZNYCH METOD ESTYMACJI MODELI TOBITOWYCH TYPU I–III

S t r e s z c z e n i e

Estymacja parametrów modeli tobitowych, tzn. modeli opisujących zmienną zależną cenzurowaną, za pomocą metody największej wiarygodności wymaga m.in. założenia o normalności rozkładu składników losowych. Sięgnięcie po metody semiparametryczne pozwalające na osłabienie założeń jest szczególnie przydatne. W kolejnych punktach krótko przedstawiono modele tobitowe typu I, II oraz III. Następnie zaprezentowano wybrane metody estymacji semiparametrycznej tych modeli oraz wskazano przydatność przedstawionych metod przy weryfikacji założeń narzuconych na rozkład składnika losowego.

THE REVIEW OF THE SEMI-PARAMETRIC METHODS FOR THE TOBIT TYPE I–III MODELS

S u m m a r y

The estimation of parameters of the tobit type models, that is models for a censored dependent variable, by the MLE method requires an assumption of the normal distribution of disturbances. Semi-parametric methods are very useful, because allow to weaken these assumptions. In the paper, at first the tobit type I, II and III models are introduced, next some known in the literature semi-parametric methods for these models are presented. At the end of the paper there is shown the usefulness of semi-parametric methods to a verification of assumptions imposed on the distribution of disturbances.

PRZEGLĄD STATYSTYCZNY
R. LVIII – ZESZYT 3-4 – 2011

CZESŁAW DOMAŃSKI

„LOSY PRZEGLĄDU STATYSTYCZNEGO”

„Statystyka: narzędzie nieodzowne w poszukiwaniu prawdy”.
C. Radhakrishua Rao

1. WPROWADZENIE

Głównym celem działalności Polskiego Towarzystwa Statystycznego, które zostało powołane w Krakowie w 1912 r. było opracowanie publikacji statystycznych obejmujących swym zasięgiem terytorialnym wszystkie trzy zabory. Wykonania tego zamierzenia podjął się ówczesny kierownik krakowskiego Miejskiego Biura Statystycznego doc. dr hab. Kazimierz Władysław Komaniecki we współpracy z Władysławem Stadnickim i poparciu prezydenta m. Krakowa prof. dr Juliusza Leo, pierwszego prezesa PTS.

W ramach zaprojektowanej działalności publikacyjnej PTS wydało „Statystykę Polski” wydrukowaną w 1915 r. w drukarni Uniwersyteckiej. „Statystyka Polski” była pierwszym polskim rocznikiem statystycznym. (por. strona tytułowa).

A. KRZYŻANOWSKI I K. KUMANIECKI

STATYSTYKA POLSKI

HANDBUCH DER POLNISCHEN STATISTIK

TABLEAU STATISTIQUE DE LA POLOGNE

WYDANA STARANIEM I NAKŁADEM
POLSKIEGO TOW. STATYSTYCZNEGO

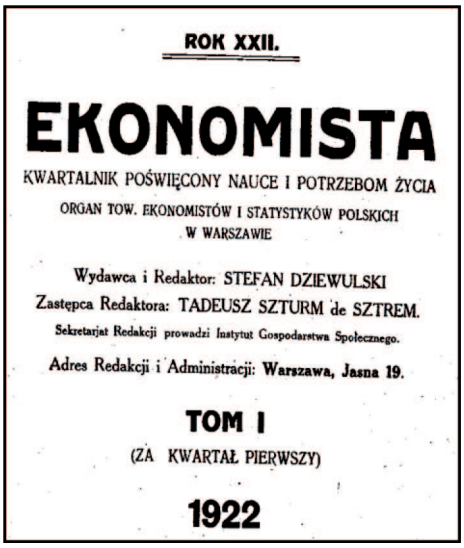
KRAKÓW 1915
SKŁAD GŁÓWNY W KSIĘGARNI G. GEBETHNERA I SP. W KRAKOWIE.

Warto wspomnieć o najstarszych obecnie działających towarzystwach statystycznych:

1. Królewskie Towarzystwo Statystyczne , powstało w 1834r. i liczy obecnie około 7200 członków,
2. Amerykańskie Towarzystwo Statystyczne, powstało w 1839r. i liczy obecnie ponad 17000 członków,
3. Paryskie Towarzystwo Statystyczne , powstało w 1860r. (obecnie Francuskie Towarzystwo Statystyczne),
4. Niemieckie Towarzystwo Statystyczne, powstało w 1911r. i liczy obecnie 800 członków,
5. Polskie Towarzystwo Statystyczne , powstało w 1912r. i liczy obecnie około 700 członków.

Aktywność statystyków polskich przejawiała się na wielu płaszczyznach, o czym świadczy powołanie Towarzystwa Ekonomistów i Statystyków Polskich w Warszawie w 1917 r., a o jego genezie mówi następujący zapis: „ Wśród grona ekonomistów i statystyków zamieszkujących w Warszawie i skupiających się wokół redakcji „Ekonomisty”, który od roku 1916 redagował Stefan Dziewulski i był kwartalnikiem poświęconym sprawom gospodarczym i społecznym, powstała myśl powołania do życia organizacji naukowej, specjalistycznej pod nazwą ≪Towarzystwa Ekonomistów i Statystyków Polskich≫”.

W 1921 r. Rada Towarzystwa uznała kwartalnik „Ekonomistę” za organ Towarzystwa Ekonomistów i Statystyków Polskich. Można też przyjąć, że „Ekonomista” był jednym z pierwszych czasopism statystycznych o profilu ekonomiczno-statystycznym (por. strona tytułowa).


ROK XXII.

EKONOMISTA

KWARTALNIK POŚWIĘCONY NAUCE I POTRZEBOM ŻYCIA

ORGAN TOW. EKONOMISTÓW I STATYSTYKÓW POLSKICH W WARSZAWIE

Wydawca i Redaktor: STEFAN DZIEWULSKI

Zastępca Redaktora: TADEUSZ SZTURM de SZTREM.

Sekretarjat Redakcji prowadzi Instytut Gospodarstwa Społecznego.

Adres Redakcji i Administracji: Warszawa, Jasna 19.

TOM I

(ZA KWARTAŁ PIERWSZY)

1922

Jeszcze w okresie niewoli przed I wojną światową zaczął się rozwój statystki na ziemiach polskich. Zwracali na to uwagę m.in. prof. Zygmunt Szweykowski i dr Wacław Skrzywan. Statystyka była wykładana na polskich uniwersytetach w Krakowie i we Lwowie. Publikowane były też prace oparte na danych statystycznych, jak np. Witolda Załęskiego pt. „Z statystyki porównawczej Królestwa Polskiego (1908r.)”, Jó-

zefa Kończyńskiego pt. „Ludność Warszawy (1913r.), czy Jana Czekanowskiego, który wydał swój podręcznik pt. „Zarys metod statystycznych w zastosowaniu do antropologii (1913r.)".

Rozwój polskiej statystyki staje się zorganizowany i usystematyzowany dopiero po powstaniu Głównego Urzędu Statystycznego w 1918r. W dniu 13 lipca 1918r. utworzono GUS, a kierowanie pracami powierzono prof. dr Józefowi Buzkowi. Jego zasługi w zakresie organizacji GUS są ogromne. Skupił on m.in. w tej instytucji najwybitniejszych statystyków tamtych czasów i tak zorganizował wszystkie prace, że były one od początku na najwyższym poziomie.

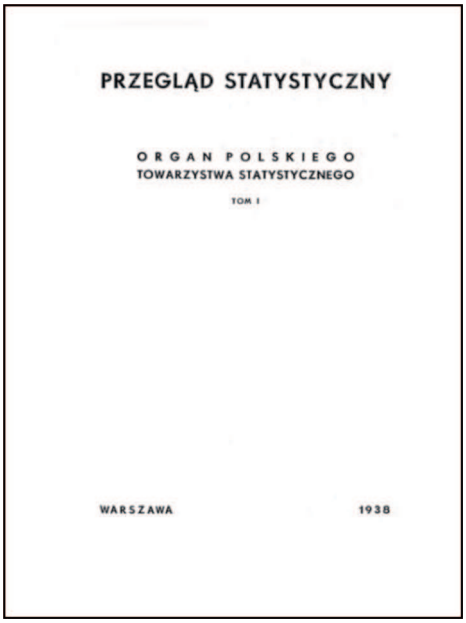
W 1919 r. zaczął się ukazywać „Miesięcznik Statystyczny", a od 1924r. w „Kwartalnik Statystyczny". Miał on za zadanie podawanie informacji liczbowych, interpretacji danych, formułowanie wniosków, a także rozwój teorii statystyki. Autorami artykułów byli wszyscy wybitni statystycy polscy.

W połowie lat trzydziestych, kiedy pojęcie zawodu statystyka było już ukształtowane, statystycy polscy widzieli potrzebę stworzenia odrębnej organizacji społecznej o charakterze naukowym.

Dnia 14 maja 1937 r. komisarz rządu na m. ST. Warszawę zarejestrował statut Polskiego Towarzystwa Statystycznego oparty na statucie z 1912 r. Walne Zgromadzenie odbyło się w dniach 31 października i 1 listopada 1937 r. w lokalu Szkoły Głównej Handlowej, któremu przewodniczył prof. Jan Czekanowski. Na zgromadzeniu tym omawiano plany pracy, działalność wydawniczą i organizacyjną zebrań naukowych. Powołano 4 sekcje tematyczne, a mianowicie: Sekcję Statystyki Matematycznej, Sekcję Statystyki Ludności, Sekcję Statystyki Gospodarczej i Społecznej oraz Sekcję Statystki w Przedsiębiorstwie. Postanowiono przystąpić do wydawania własnego czasopisma, poświęconego zagadnieniom badań statystycznych. Na zebraniu stwierdzono, że Towarzystwo musi znaleźć poparcie wśród społeczeństwa, a popularyzacja statystyki musi stać się jednym z zasadniczych celów Towarzystwa, aby społeczeństwo zrozumiało potrzebę prowadzenia tego rodzaju badań. Już wówczas zdawano sobie sprawę z tego, że zadania PTS realizowane powinny być w aspekcie ogólnospołecznym.

W dniu 1 listopada 1937 r. odbyło się posiedzenie Rady Polskiego Towarzystwa Statystycznego, na którym wybrano Komitet Redakcyjny Przeglądu Statystycznego w składzie: *Z. Limanowski (przewodniczący), S. Szulc (wiceprzewodniczący), J. Wiśniewski (sekretarz) oraz A. Łomnicki i J. Piekałkiewicz.*

Na szczególne podkreślenie zasług statystyków zrzeszonych w Polskim Towarzystwie Statystycznym zasługuje utworzenie nowego czasopisma naukowego pt.: „Przegląd Statystyczny” i uznanie tego czasopisma jako organu Polskiego Towarzystwa Statystycznego. (por. strona tytułowa I tomu).



PRZEGLĄD STATYSTYCZNY

ORGAN POLSKIEGO
TOWARZYSTWA STATYSTYCZNEGO

TOM I

WARSZAWA 1938

Warto przytoczyć „Słowo wstępne” do I tomu oraz strukturę spisu treści:

Inicjatorem reaktywowania po II wojnie światowej Polskiego Towarzystwa Statystycznego był 1947r, prof. Stefan Szulc, ówczesny prezes Głównego Urzędu Statystycznego.

Już 15 czerwca 1947 r. odbyło się w Warszawie walne zgromadzenie PTS. Podjęto szereg uchwał dotyczących m.in. wznowienia prac działających przed wojną sekcji statystyki: matematycznej, ludności, społecznej i gospodarczej, w przedsiębiorstwie, utworzenia nowych sekcji – statystyki w gospodarce planowej i statystyki miejskiej oraz wznowienia wydawania „Przeglądu Statystycznego”.

Załączono stronę tytułową III tomu i Redakcję „Przeglądu Statystycznego” i spis treści.

SŁOWO WSTĘPNE

Polskie Towarzystwo Statystyczne postawiło sobie jako jedno z naczelnych zadań wydawanie czasopisma poświęconego teorii i praktyce statystycznej.

Statystyka administracyjna, demografia, życie gospodarcze i społeczne, nauki przyrodnicze korzystają ze wzajemnych doświadczeń i wspólnej praktyki statystycznej, korzystają z tych samych metod, doskonalą je i rozwijają. Zainteresowania statystyków sięgają więc daleko poza wąski nieraz zakres badanych przez nich zagadnień.

Zrozumiała jest w tych warunkach potrzeba czasopisma, które by informowało o pracach statystycznych, podejmowanych w tych różnych dziedzinach, ułatwiając wymianę myśli i koordynację wysiłków.

Potrzeba takiego czasopisma w Polsce zaznaczyła się szczególnie silnie z chwilą zawieszenia Kwartalnika Statystycznego, który choć poświęcony przede wszystkim statystyce administracyjnej, starał się zaspakajać i szersze potrzeby.

Przegląd Statystyczny w miarę możności ma się ukazywać 4 razy do roku.

Przegląd Statystyczny ogłaszać będzie: 1) oryginalne prace teoretyczne autorów polskich i obcych, 2) oryginalne prace z zastosowań statystyki, ważne ze względów metodologicznych, 3) referaty przedstawiające stan lub postępy pewnej gałęzi statystyki, 4) recenzje i krytyki dochodzeń statystycznych, 5) kronikę informującą o życiu organizacyjnym i naukowym P. T. S. oraz jego sekcyj i oddziałów, jak również pokrewnych instytucyj zagranicznych i międzynarodowych, 6) szczegółową bibliografię wydawnictw polskich.

Krótki czas, którym przy przygotowaniu niniejszego numeru rozporządzała Redakcja, nie pozwolił rozwinąć działu recenzji. Z tego samego powodu nie zdołano wykończyć opracowania szczegółowej bibliografii za 1937 r. Braki te zostaną wyrównane w następnych numerach.

Oddając ten pierwszy numer Przeglądu Statystycznego do rąk czytelników, Redakcja sądzi, iż mimo swych usterek będzie on przez nich życzliwie przyjęty i znajdzie czynne poparcie wśród wszystkich osób pracujących na niwie statystycznej.

REDAKCJA

PRZEGLĄD STATYSTYCZNY

Spis rzeczy tomu I, 1938

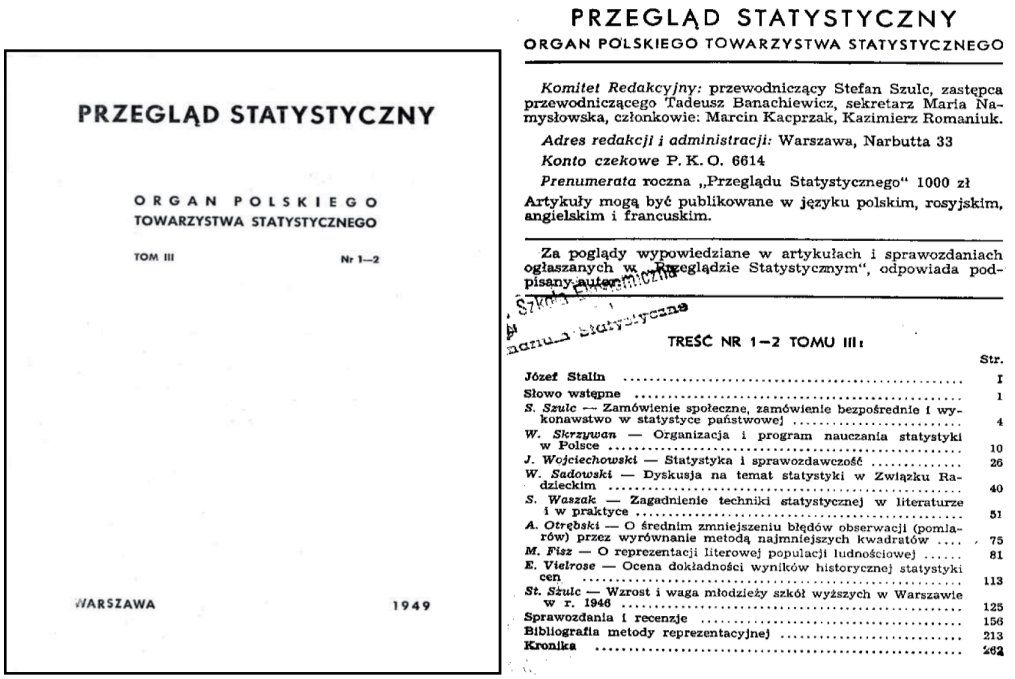
STATISTICAL REVIEW

Index to Vol. I, 1938

ARTYKUŁY — ARTICLES

SPRAWOZDANIA — REVIEWS

5



PRZEGLĄD STATYSTYCZNY

ORGAN POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

TOM III Nr 1—2

WARSZAWA 1949

PRZEGLĄD STATYSTYCZNY

ORGAN POLSKIEGO TOWARZYSTWA STATYSTYCZNEGO

Komitet Redakcyjny: przewodniczący Stefan Szulc, zastępca przewodniczącego Tadeusz Banachiewicz, sekretarz Maria Namysłowska, członkowie: Marcin Kacprzak, Kazimierz Romaniuk.

Adres redakcji i administracji: Warszawa, Narbutta 33

Konto czekowe P. K. O. 6614

Prenumerata roczna „Przeglądu Statystycznego" 1000 zł

Artykuły mogą być publikowane w języku polskim, rosyjskim, angielskim i francuskim.

Za poglądy wypowiedziane w artykułach i sprawozdaniach ogłaszanych w „Przeglądzie Statystycznym", odpowiada podpisany autor.

TREŚĆ NR 1—2 TOMU III:

W latach 1950 – 1956 totalna kontrola polityczna aparatu partyjno-państwowego nad życiem społeczno-gospodarczym przyczyniła się do stopniowego zamierania prac różnych organizacji, w tym Polskiego Towarzystwa Statystycznego, których działalność kolidowała z kierunkiem polityki państw komunistycznych. Na łamach radzieckich czasopism statystycznych (Waprosy Ekonomiki, 1949r., Z.4) zaatakowano statystykę za uleganie burżuazyjnemu obiektywizmowi i formalizmowi. Takie tendencje występowały oczywiście i w PRL – na I Kongresie Nauki Polskiej (czerwiec/lipiec 1951r.) oraz na konferencji katedr statystyki szkół wyższych (grudzień 1952r.) gdzie proponowano utworzenie sekcji statystyki w ramach Polskiego Towarzystwa Ekonomicznego.

Losy Towarzystwa rozstrzygnęło ostatnie Walne Zgromadzenie w dniu 12 grudnia 1953r. W obecności 10% członków podjęto uchwałę o likwidacji PTS i przekazaniu majątku Polskiemu Towarzystwu Ekonomicznemu. Z formalnego punktu widzenia za ostateczną datę uznaje się ostatnie posiedzenie Komisji Likwidacyjnej PTS w dniu 4 kwietnia 1955r. Organ PTS „Przegląd Statystyczny" został przejęty przez sekcję statystyki PTE i w latach 1955 – 1973 wychodził pod redakcją prof. Kazimierza Romaniuka, od 1974r. jest organem Komitetu Ekonometrii i Statystyki Polskiej Akademii Nauk.

Działalność Polskiego Towarzystwa Statystycznego w okresie blisko 100 lat przyczyniła się niewątpliwie do rozwoju polskiej myśli statystycznej, w tym m. in. do rozwiązywania problemów teoretycznych i praktycznych badań statystycznych. Ponadto wskazała kierunki dalszych badań umożliwiających tworzenie w zakresie polskiej nauki statystycznej zwłaszcza oryginalnych metod wśród których na szczególne wy-

różnienie zasługują prace Jerzego Spławy-Neymana z zakresu estymacji, weryfikacji hipotez i metody reprezentacyjnej.

Ze względu na inicjujący charakter prezentowania i inspirowania wyników badań statystycznych najwybitniejszych statystyków polskich szczególnie w „Przeglądzie Statystycznym” postuluję *po pierwsze:* o nową numerację „Przeglądu Statystycznego” od bieżącego roku uwzględniając 3 tomy „Przeglądu Statystycznego” z lat 1938, 1939 i 1949; *po drugie:* o zamieszczenie nazwy i logo Polskiego Towarzystwa Statystycznego na stronie tytułowej tego czasopisma.

„Nie trzeba kłaniać się okolicznością
A prawdom kazać, by za drzwiami stały”
Cyprian Kamil Norwid

Czesław Domański

SPIS TREŚCI