

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 427

**Taksonomia 27**

**Klasyfikacja i analiza danych –  
teoria i zastosowania**



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2016

Redaktor Wydawnictwa: Agnieszka Flasińska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Tytuł dofinansowany ze środków Narodowego Banku Polskiego  
oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Informacje o naborze artykułów i zasadach recenzowania  
znajdują się na stronach internetowych  
[www.pracnaukowe.ue.wroc.pl](http://www.pracnaukowe.ue.wroc.pl)  
[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Publikacja udostępniona na licencji Creative Commons  
Uznanie autorstwa-Użycie niekomercyjne-Bez utworów zależnych 3.0 Polska  
(CC BY-NC-ND 3.0 PL)



© Copyright by Uniwersytet Ekonomiczny we Wrocławiu  
Wrocław 2016

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)  
**e-ISSN 2392-0041**  
**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Zamówienia na opublikowane prace należy składać na adres:  
Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
ul. Komandorska 118/120, 53-345 Wrocław  
tel./fax 71 36 80 602; e-mail:[econbook@ue.wroc.pl](mailto:econbook@ue.wroc.pl)  
[www.ksiegarnia.ue.wroc.pl](http://www.ksiegarnia.ue.wroc.pl)

Druk i oprawa: TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Beata Bal-Domańska:</b> Propozycja procedury oceny zrównoważonego rozwoju w układzie <i>presja – stan – reakcja</i> w ujęciu przestrzennym / Proposal of the assessment of poviats sustainable development in the pressure – state – response system in spatial terms.....	11
<b>Tomasz Bartłomowicz:</b> Pomiar preferencji konsumentów z wykorzystaniem metody <i>Analytic Hierarchy Process</i> / Analytic Hierarchy Process as a method of measurement of consumers’ preferences.....	20
<b>Maciej Beręsewicz, Marcin Szymkowiak:</b> Analiza skupień wybranych lokalnych rynków nieruchomości w Polsce z wykorzystaniem internetowych źródeł danych / Cluster analysis of selected local real estate markets in Poland based on Internet data sources.....	30
<b>Beata Bieszk-Stolorz:</b> Wybrane modele przeciętnego efektu oddziaływania w analizie procesu wychodzenia z bezrobocia / Chosen average treatment effect models in the analysis of unemployment exit process.....	40
<b>Justyna Brzezińska:</b> Modele IRT i modele Rascha w badaniach testowych / IRT and Rasch models in test measurement.....	49
<b>Mariola Chrzanowska, Nina Drejerska:</b> Geograficznie ważona regresja jako narzędzie analizy poziomu rozwoju społeczno-gospodarczego na przykładzie regionów Unii Europejskiej / Geographically weighted regression as a tool of analysis of socio-economic development level of regions in the European Union.....	58
<b>Sabina Denkowska:</b> Zastosowanie analizy wrażliwości do oceny wpływu nieobserwowanej zmiennej w <i>Propensity Score Matching</i> / The application of sensitivity analysis in assessing the impact of an unobserved confounder in Propensity Score Matching.....	66
<b>Adam Depta:</b> Zastosowanie analizy czynnikowej do wyodrębnienia aspektów zdrowia wpływających na jakość życia osób jaskających się / The application of factor analysis to the identification of the health aspects affecting the quality of life of stuttering people.....	76
<b>Mariusz Doszyń, Sebastian Gnat:</b> Taksonomiczno-ekonometryczna procedura wyceny nieruchomości dla różnych miar porządkowania / Taxonomic and econometric method of real estate valuation for various classification measures.....	84

<b>Marta Dziechciarz-Duda, Anna Król:</b> Segmentacja konsumentów smartfonów na podstawie preferencji wyrażonych / Segmentation of smartphones' consumers on the basis of stated preferences .....	94
<b>Ewa Genge:</b> Zmienne towarzyszące w ukrytym modelu Markowa – analiza oszczędności polskich gospodarstw domowych / Latent Markov model with covariates – Polish households' saving behaviour .....	103
<b>Joanna Górna, Karolina Górna:</b> Modelowanie wzrostu gospodarczego z wykorzystaniem narzędzi ekonometrii przestrzennej / Economic growth modelling with the application of spatial econometrics tools .....	112
<b>Alicja Grześkowiak:</b> Wielowymiarowa analiza kompetencji zawodowych według grup wieku ludności / Multivariate analysis of professional competencies with respect to the age groups of the population .....	122
<b>Agnieszka Kozera, Feliks Wysocki:</b> Problem ustalania współrzędnych obiektów modelowych w metodach porządkowania liniowego obiektów / The problem of determining the coordinates of model objects in object linear ordering methods .....	131
<b>Mariusz Kubus:</b> Lokalna ocena mocy dyskryminacyjnej zmiennych / Local evaluation of a discrimination power of the variables.....	143
<b>Paweł Lula, Katarzyna Wójcik, Janusz Tuchowski:</b> Analiza wydźwięku polskojęzycznych opinii konsumenckich ukierunkowanych na cechy produktu / Feature-based sentiment analysis of opinions in Polish.....	153
<b>Aleksandra Łuczak, Agnieszka Kozera, Feliks Wysocki:</b> Ocena sytuacji finansowej jednostek samorządu terytorialnego z wykorzystaniem rozmytych metod klasyfikacji i programu R / Assessment of financial condition of local government units with the use of fuzzy classification methods and program R .....	165
<b>Dorota Rozmus:</b> Badanie stabilności taksonomicznej czynnikowej metody odległości probabilistycznej / Stability of the factor probability distance clustering method .....	176
<b>Adam Sagan, Aneta Rybicka, Justyna Brzezińska:</b> <i>Conjoint analysis</i> oparta na modelach IRT w zagadnieniu optymalizacji produktów bankowych / An IRT-approach for conjoint analysis for banking products preferences.....	184
<b>Michał Stachura:</b> O szacowaniu centrum populacji określonego obszaru na przykładzie Polski / On estimating centre of population of a given territory. Poland's case .....	195
<b>Michał Stachura, Barbara Wodecka:</b> Wybrane aspekty i zastosowania modeli zdarzeń ekstremalnych / Selected facets and application of models of extremal events .....	205
<b>Iwona Staniec, Jan Żółtowski:</b> Wykorzystanie analizy log-liniowej do wyboru czynników determinujących współpracę w przedsiębiorczości	

---

technologicznej / Use of log-linear analysis for the selection determinants of cooperation in technological entrepreneurship.....	215
<b>Marcin Szymkowiak, Wojciech Roszka:</b> Potencjał gospodarczy gmin aglomeracji poznańskiej w ujęciu taksonomicznym / The economic potential of municipalities of the Poznań agglomeration in the light of taxonomy analysis.....	224
<b>Lucyna Wojcieszka:</b> Zastosowanie modeli klas ukrytych w badaniu opinii respondentów na temat roli państwa w gospodarce / Implementation of latent class models in the respondents' survey on the role of the country in economy.....	234

## **Wstęp**

W dniach 14–16 września 2015 r. w Hotelu Novotel Gdańsk Marina w Gdańsku odbyła się XXIV Konferencja Naukowa Sekcji Klasyfikacji i Analizy Danych PTS (XXIX Konferencja Taksonomiczna) „Klasyfikacja i analiza danych – teoria i zastosowania”, zorganizowana przez Sekcję Klasyfikacji i Analizy Danych Polskiego Towarzystwa Statystycznego oraz Katedrę Statystyki Wydziału Zarządzania Uniwersytetu Gdańskiego.

W trakcie dwóch sesji plenarnych oraz 13 sesji równoległych wygłoszono 58 referatów poświęconych aspektom teoretycznym i aplikacyjnym zagadnienia klasyfikacji i analizy danych. Odbyła się również sesja plakatowa, na której zaprezentowano 14 plakatów.

Teksty 24 recenzowanych artykułów naukowych stanowią zawartość prezentowanej publikacji z serii Taksonomia nr 27. Teksty 25 recenzowanych artykułów naukowych znajdują się w Taksonomii nr 26.

*Krzysztof Jajuga, Marek Walesiak*

**Ewa Genge**

Uniwersytet Ekonomiczny w Katowicach  
e-mail: ewa.genge@ue.katowice.pl

---

**ZMIENNE TOWARZYSZĄCE  
W UKRYTYM MODELU MARKOWA –  
ANALIZA OSZCZĘDNOŚCI POLSKICH  
GOSPODARSTW DOMOWYCH**

---

**LATENT MARKOV MODEL WITH COVARIATES –  
POLISH HOUSEHOLDS’ SAVING BEHAVIOUR**

---

DOI: 10.15611/pn.2016.427.11

**Streszczenie:** Ukryte modele Markowa (modele LM) stosowane są najczęściej do analizy danych panelowych o charakterze jakościowym, gdzie celem jest nie tylko podział obserwacji na homogeniczne grupy, ale także pewna analiza zmian w czasie. Ponadto modele te pozwalają na badanie relacji pomiędzy zmiennymi ukrytymi i tzw. zmiennymi towarzyszącymi. W artykule przedstawiona zostanie rozbudowana wersja modelu LM, uwzględniająca zmienne towarzyszące, mające wpływ na prawdopodobieństwa początkowe oraz prawdopodobieństwa przejścia tego modelu. Celem badań będzie podział polskich gospodarstw domowych na klasy o podobnych skłonnościach do oszczędzania, a także zaobserwowanie zmian ich postaw na przestrzeni kilku lat, z wykorzystaniem wpływu zmiennych towarzyszących stałych i zmiennych w czasie  $t$ , na ukryte zmienne modelu. Badania przeprowadzone będą za pomocą ukrytych modeli Markowa (*latent Markov models*) z zastosowaniem pakietu LMest programu R.

**Słowa kluczowe:** ukryty model Markowa, zmienne towarzyszące, dane panelowe.

**Summary:** Latent Markov (LM) models represent an important class of models for the analysis of longitudinal data, especially when response variables are categorical. Those models are specially tailored to study the evolution of an individual characteristic of interest that is not directly observable. We applied the extended version of LM to find groups of Polish households’ with similar saving behaviour. We focused especially on the different covariates having influence on the initial and transition probabilities of the model. We analyzed data collected as part of the Polish Social Diagnosis using LMest package of R.

**Keywords:** latent Markov model, covariates, longitudinal data.

## 1. Wstęp

Ukryty model Markowa (*latent Markov model*) można wpisać w nurt podejścia modelowego w taksonomii, wykorzystującego ideę mieszanek rozkładów. Modele te wykorzystywane są do analizy panelowych zbiorów danych o niejednorodnej strukturze, w których liczba klas (stanów) jest nieznana (tzw. zmienna ukryta). Celem badań jest już nie tylko podział obserwacji na homogeniczne grupy, ale również pewna analiza zmian w czasie  $t$ .

Ukryty model Markowa został zaproponowany przez L.M. Wigginsa [1973]. Pierwsza znacząca pozycja tego autora ukazała się po wcześniejszej publikacji jego pracy doktorskiej w roku 1955 (zob. [Wiggins 1955]). Do jednej z najważniejszych modyfikacji modelu zaliczyć należy z pewnością możliwość uwzględnienia zmiennych towarzyszących mających wpływ na ukrytą zmienną modelu (zob. np. [Vermunt, Langeheine, Böckenholt 1999; Bartolucci, Pennoni, Francis 2007; Bartolucci, Montanari, Pandolfi 2015]). Innym wariantem modelu LM jest możliwość badania wpływu zmiennych towarzyszących na rozkłady warunkowe zmiennych obserwowanych (zob. np. [Bartolucci, Farcomeni 2009; Bartolucci, Farcomeni, Pennoni 2013]). Bardziej szczegółowe informacje na temat różnych wariantów modeli LM można znaleźć w pracach [Bartolucci, Farcomeni, Pennoni 2013; Vermunt 2010].

Artykuł jest kontynuacją prac autora poświęconą ukrytym modelom Markowa i ich zastosowaniom w badaniach społeczno-ekonomicznych. Ukryty model Markowa w badaniu skłonności Polaków do oszczędzania (różnych form oszczędności) przedstawiono w pracy [Genge 2014]. Model Markowa z uwzględnieniem zmiennych towarzyszących mających wpływ na prawdopodobieństwa przejścia wykorzystano w badaniu zaufania do instytucji publicznych i finansowych w polskim społeczeństwie oraz nastawienia Polaków do emigracji [Genge 2015a, b]. Celem niniejszego referatu jest podział polskich gospodarstw domowych na klasy o podobnych skłonnościach do oszczędzania (o podobnych celach oszczędzania), a także zaobserwowanie zmian ich postaw na przestrzeni kilku lat w zależności od wybranych zmiennych demograficznych (stałych i zmiennych w czasie  $t$ ), mających wpływ zarówno na prawdopodobieństwa początkowe, jak i prawdopodobieństwa przejścia modelu LM.

## 2. Ukryty model Markowa

W ukrytym modelu Markowa dany jest wektor  $m$  jakościowych zmiennych  $\mathbf{Y}^{(t)}$ ,  $\mathbf{Y}^{(t)} = Y_1^{(t)}, \dots, Y_m^{(t)}$ , obserwowanych w  $t$  okresach ( $t = 1, \dots, T$ ). Liczba kategorii każdej ze zmiennych określana jest jako  $l_j$  przybierających wartości  $0, \dots, l_j - 1$



( $j = 1, \dots, m$ ). Niech  $\tilde{\mathbf{Y}}$  oznacza wektor wszystkich zmiennych  $\mathbf{Y}^{(t)}$ , obserwowanych w każdym okresie (wektor o wymiarze  $mT$ ). Zaś  $\tilde{\mathbf{Z}}$  oznacza dodatkowy wektor wszystkich analizowanych zmiennych towarzyszących  $\mathbf{Z}^{(t)}$  obserwowanych w czasie  $t = 1, \dots, T$ .

Rozkład zmiennej ukrytej definiowany jest za pomocą łańcucha Markowa, którego stany  $\mathbf{S} = (S^{(1)}, \dots, S^{(T)})$  nie są bezpośrednio mierzalne ( $\mathbf{S}$  jest ukrytą zmienną modelu)<sup>1</sup>. Można zatem wyróżnić obserwowaną i ukrytą składową modelu.

Prawdopodobieństwa warunkowe zmiennych obserwowanych można zapisać jako:

$$\phi_{j|y|sz}^{(t)} = p(Y_j^{(t)} = y | S^{(t)} = s, \mathbf{Z}^{(t)} = \mathbf{z}), \quad j = 1, \dots, m, \quad y = 0, \dots, l_j - 1. \quad (1)$$

Parametry ukrytej składowej modelu to: prawdopodobieństwa początkowe dane równaniem (2):

$$\pi_{s|z} = p(S_1 = s | \mathbf{Z}^{(1)} = \mathbf{z}), \quad s = 1, \dots, u, \quad (2)$$

oraz prawdopodobieństwa przejścia definiowane za pomocą równania (3):

$$\pi_{s|\bar{s}z}^{(t)} = p(S^{(t)} = s | S^{(t-1)} = \bar{s}, \mathbf{Z}^{(t)} = \mathbf{z}), \quad t = 2, \dots, T, \quad s, \bar{s} = 1, \dots, u. \quad (3)$$

Ukryty model Markowa z uwzględnieniem zmiennych towarzyszących, mających wpływ na prawdopodobieństwa początkowe oraz prawdopodobieństwa przejścia tego modelu można zapisać wtedy jako:

$$p(\tilde{\mathbf{Y}} = \tilde{\mathbf{y}} | \tilde{\mathbf{Z}} = \tilde{\mathbf{z}}) = \sum_s \pi_{s^{(1)}|z^{(1)}} \pi_{s^{(2)}|s^{(1)}z^{(2)}} \dots \pi_{s^{(T)}|s^{(T-1)}z^{(T)}} \times \phi_{y^{(1)}|s^{(1)}}^{(1)} \dots \phi_{y^{(T)}|s^{(T)}}^{(T)}. \quad (4)$$

W powyższym równaniu  $\tilde{\mathbf{y}}$  jest realizacją zmiennej  $\tilde{\mathbf{Y}}$  złożonej z podwektorów  $\mathbf{y}^{(t)} = (y_1^{(t)}, \dots, y_m^{(t)})$ , zaś  $\mathbf{y}$  jest realizacją zmiennej  $\mathbf{Y}$  o elementach  $y_j, j = 1, \dots, m$ .

Wpływ zmiennych towarzyszących na ukrytą składową modelu, tj. prawdopodobieństwa początkowe oraz prawdopodobieństwa przejścia modelu LM [Bartolucci, Lupporelli, Montanari 2009], wyrażany jest za pomocą wielomianowych modeli logitowych danych równaniami (5) oraz (6):

$$\log \frac{p(S^{(1)} = s | \mathbf{Z}^{(1)} = \mathbf{z})}{p(S^{(1)} = 1 | \mathbf{Z}^{(1)} = \mathbf{z})} = \log \frac{\pi_{s|z}}{\pi_{1|z}} = \beta_{0s} + \mathbf{z}' \boldsymbol{\beta}_{1s}, \quad s \geq 2, \quad (5)$$

$$\log \frac{p(S^{(t)} = s | S^{(t-1)} = \bar{s}, \mathbf{Z}^{(t)} = \mathbf{z})}{p(S^{(t)} = \bar{s} | S^{(t-1)} = \bar{s}, \mathbf{Z}^{(t)} = \mathbf{z})} = \log \frac{\pi_{s|\bar{s}z}^{(t)}}{\pi_{\bar{s}|\bar{s}z}^{(t)}} = \alpha_{0s\bar{s}} + \mathbf{z}' \boldsymbol{\alpha}_{1s\bar{s}}, \quad t \geq 2, \quad \bar{u} \neq u. \quad (6)$$

<sup>1</sup> Analizowane obserwacje są zmienne w czasie  $t$ , dlatego w literaturze z zakresu łańcuchów Markowa często spotykanym określeniem jest stan. Z racji zastosowania teorii łańcuchów Markowa na gruncie taksonomii w dalszej części pracy autorka posługiwać się będzie również terminami „klasa” lub „grupa”. W literaturze polskiej wśród prac z zakresu teorii łańcuchów Markowa wymienić należy [Podgórska i in. 2000; Stawicki 2004].

W powyższych równaniach  $\beta_s = (\beta_{0s}, \beta'_{1s})'$  oraz  $\alpha_{ss} = (\alpha_{0ss}, \alpha'_{1ss})'$  są szacowanymi parametrami modeli logitowych.

Parametry ukrytych modeli Markowa najczęściej szacowane są za pomocą algorytmu EM [Dempster, Laird, Rubin 1977], a wybór modelu optymalnego dokonywany jest na podstawie kryteriów informacyjnych tj. BIC oraz AIC [Akaike 1974; Schwarz 1978].

### 3. Analiza empiryczna

Analizę empiryczną przeprowadzono na podstawie danych panelowych dotyczących polskich gospodarstw domowych, pochodzących z projektu Diagnoza Społeczna [Czapiński, Panek (red.) 2013]. Analizie podano 969 gospodarstw, które wzięły udział w trzech edycjach badania panelowego, przeprowadzonego w latach 2009, 2011 i 2013. Szacowano więc modele dla 2907 obserwacji łącznie. Analiza została przeprowadzona z uwzględnieniem 13 zmiennych obserwowanych, charakteryzujących cele oszczędności gospodarstw domowych<sup>2</sup>.

W nawiasach podano oryginalne nazwy zmiennych analizowanego zbioru, udostępnianego na stronie internetowej [www.diagnoza.com](http://www.diagnoza.com). Litera *e* poprzedzająca symbol zmiennej (np. *ef4\_1*) dotyczy badania w 2009 r., litera *f* – badania w 2011 r., zaś litera *G* – badania w 2013 r.

$Y_1$  (*ef4\_1, fF4\_1, GF9\_01*): rezerwa na bieżące wydatki konsumpcyjne (np. żywność, odzież i ubranie);  $Y_2$  (*ef4\_2, fF4\_2, GF9\_02*): stałe opłaty (np. mieszkaniowe);  $Y_3$  (*ef4\_3, fF4\_3, GF9\_03*): zakup dóbr trwałego użytku;  $Y_4$  (*ef4\_4, fF4\_4, GF9\_04*): zakup domu, zakup mieszkania, wkład do spółdzielni mieszkaniowej;  $Y_5$  (*ef4\_5, fF4\_5, GF9\_05*): remont domu, mieszkania;  $Y_6$  (*ef4\_6, fF4\_6, GF9\_06*): leczenie;  $Y_7$  (*ef4\_7, fF4\_7, GF9\_07*): rehabilitacja;  $Y_8$  (*ef4\_8, fF4\_8, GF9\_08*): wypoczynek;  $Y_9$  (*ef4\_9, fF4\_9, GF9\_09*): rezerwa na sytuacje losowe, tj. na „czarną godzinę”;  $Y_{10}$  (*ef4\_10, fF4\_10, GF9\_10*): zabezpieczenie przyszłości dzieci;  $Y_{11}$  (*ef4\_11, fF4\_11, GF9\_11*): zabezpieczenie na starość;  $Y_{12}$  (*ef4\_12, fF4\_12, GF9\_13*): inne cele;  $Y_{13}$  (*ef4\_13, fF4\_13, GF9\_14*): bez specjalnego przeznaczenia.

W badaniu uwzględniono również następujące zmienne towarzyszące:  $Z_1$ : źródło utrzymania gospodarstwa<sup>3</sup>;  $Z_2$ : łączna wartość oszczędności posiadanych przez

<sup>2</sup> Pytanie w kwestionariuszu brzmiało: „W jakim celu Pana(i) gospodarstwo domowe gromadzi oszczędności?”. Należało udzielić odpowiedzi (tak lub nie) dla każdego z poniższych celów oszczędzania (zmienne  $Y_1$ – $Y_{13}$ ).

<sup>3</sup> 1: gospodarstwa pracowników, 2: gospodarstwa rolników, 3: gospodarstwa pracujących na własny rachunek, 4: gospodarstwa emerytów, 5: gospodarstwa rencistów, 6: gospodarstwa utrzymujące się z niezarobkowych źródeł.

gospodarstwo<sup>4</sup>;  $Z_3$ : typ biologiczny rodziny<sup>5</sup>;  $Z_4$ : liczba osób w gospodarstwie domowym;  $Z_5$ : klasa miejscowości zamieszkania<sup>6</sup>;  $Z_6$ : województwo.

W badaniach wykorzystano pakiet LMest programu R. Ukrytą liczbę klas wybrano na podstawie wartości kryteriów informacyjnych AIC oraz BIC. Kryteria BIC oraz AIC wskazały minimalną wartość dla liczby klas równej 3. Do klasy pierwszej, stanowiącej 59% wszystkich gospodarstw, należą gospodarstwa najmniej skłonne do oszczędzania, które prawie w ogóle nie oszczędzają na leczenie, rehabilitację oraz opłaty stałe (poniżej 5% gospodarstw). Gospodarstwa te cechuje głównie przezornościowy motyw oszczędzania.

Klasa druga jest klasą najmniej liczną – należy do niej 20% gospodarstw domowych. W klasie tej również największy odsetek gospodarstw swe oszczędności gromadzi na tzw. czarną godzinę (77%). W odróżnieniu od klasy pierwszej duży odsetek gospodarstw swe oszczędności planuje przeznaczyć na leczenie (67%) oraz starość (59%). Członkowie tych gospodarstw w ogóle nie gromadzą środków na mieszkanie czy nowy dom, co może świadczyć o tym, że motywem skłaniającym tę grupę gospodarstw do oszczędzania jest również godny poziom życia w okresie starości, po przejściu na emeryturę.

Do klasy trzeciej zaliczono 21% gospodarstw, które cechują się największą skłonnością do oszczędzania (o najbardziej różnorodnych motywach oszczędzania). Największy odsetek gospodarstw tej klasy stanowią ponownie gospodarstwa odkładające swe środki na sytuacje losowe (95%). Zarówno 75% gospodarstw oszczędza na starość, jak i na wypoczynek. Nieco mniej, bo 70% swe oszczędności planuje przeznaczyć na remont domu lub zakup dóbr trwałego użytku (69%) czy też na leczenie (63% gospodarstw). Klasę tę wyróżnia najwyższy odsetek gospodarstw oszczędzających również z pozostałych powodów. Cechują się więc najwyższą potrzebą niezależności i dążeniem do stopniowej poprawy jakości życia.

Oszacowane prawdopodobieństwa przejścia ukrytego modelu Markowa, obrazujące stabilność pozostania w danej klasie w kolejnych okresach, przedstawione zostały w tab. 1.

Największe prawdopodobieństwo pozostania w tej samej klasie w kolejnym okresie odpowiada klasie pierwszej, kolejno mniejsze prawdopodobieństwa, równe 0,75 i 0,58 odpowiadają klasie drugiej i trzeciej. Niestety gospodarstwa, które ce-

---

<sup>4</sup> 1: o wysokości miesięcznych dochodów gospodarstwa, 2: powyżej miesięcznych do 3-miesięcznych dochodów, 3: powyżej 3-miesięcznych do półrocznych dochodów, 4: powyżej półrocznych do rocznych dochodów, 5: powyżej rocznych dochodów gospodarstwa domowego, 6: trudno powiedzieć.

<sup>5</sup> 1: małżeństwa bez dzieci, 2: małżeństwa z 1 dzieckiem, 3: małżeństwa z 2 dziećmi, 4: małżeństwa z 3 i więcej dziećmi, 5: rodziny niepełne, 6: wielorodzinne, 7: nierodzinne jednoosobowe, 8: nierodzinne wieloosobowe.

<sup>6</sup> 1: miasta o liczbie mieszkańców 500 tys. i więcej, 2: miasta o liczbie mieszkańców 200-500 tys., 3: miasta o liczbie mieszkańców 100-200 tys., 4: miasta o liczbie mieszkańców 20-100 tys., 5: miasta o liczbie mieszkańców poniżej 20 tys., 6: wieś.

chują się najmniejszą skłonnością do oszczędności (klasa pierwsza), najprawdopodobniej nie zmieniają swego podejścia. Jeżeli jednak uda im się zmienić swoje nastawienie czy pokonać różnego rodzaju bariery, to bardziej prawdopodobne okazuje się przejście do klasy trzeciej ( $S_{13} = 0,11$ )<sup>7</sup> aniżeli do klasy drugiej ( $S_{12} = 0,07$ )<sup>8</sup>. Mniejsze szanse na pozostanie w danej klasie w następnym okresie mają respondenci należący do klasy drugiej ( $S_{22} = 0,75$ ). Ankietowani należący do grupy respondentów o największej skłonności do oszczędzania cechują się najniższą tendencją pozostania w tej grupie ( $S_{33} = 0,58$ ) w następnym okresie. Dla tej grupy osób istnieje większe prawdopodobieństwo przejścia do klasy pierwszej aniżeli do klasy drugiej.

**Tabela 1.** Prawdopodobieństwa przejścia dla trzech klas

Klasa S/Klasa R	Klasa 1	Klasa 2	Klasa 3
Klasa 1	<b>0,82</b>	0,07	0,11
Klasa 2	0,20	<b>0,75</b>	0,05
Klasa 3	0,34	0,08	<b>0,58</b>

Źródło: opracowanie własne.

W kolejnej części pracy dokonano analizy wpływu zmiennych towarzyszących na prawdopodobieństwa początkowe oraz przejścia (zmiany nastawienia) respondentów do poszczególnych klas. Spośród analizowanych zmiennych towarzyszących istotny wpływ na prawdopodobieństwa początkowe wykazały zmienne „źródło utrzymania” oraz „wartość posiadanych oszczędności”.

Na początku badanego okresu największą skłonnością do oszczędzania cechują się gospodarstwa, których źródłem utrzymania jest praca na własny rachunek, oraz gospodarstwa, których członkowie nie zdradzają wysokości swych dochodów lub gospodarstwa o oszczędnościach wyższych niż wartość rocznych dochodów. Najniższą skłonnością do oszczędzania na początku badanego okresu (najwyższe prawdopodobieństwo przynależności do  $S_1$ ) cechują się gospodarstwa pracowników, o oszczędnościach niższych niż wartość miesięcznego dochodu.

Istotny wpływ na prawdopodobieństwa przejścia miały zmienne: „źródło utrzymania”, „wartość posiadanych oszczędności” oraz „typ rodziny biologicznej”.

Największe prawdopodobieństwo pozostania w  $S_1$  mają rolnicy i pracownicy, o oszczędnościach niższych niż wartość miesięcznego dochodu. Są to przede wszystkim małżeństwa z trójką lub większą liczbą dzieci czy też gospodarstwa wielorodzinne. Największe prawdopodobieństwo pozostania w  $S_2$  mają emeryci i renciści. Są to więc głównie małżeństwa bez dzieci o najwyższych oszczędno-

<sup>7</sup> Autorka posługuje się skrótami  $S_1$ ,  $S_2$  oraz  $S_3$  na określenie klas (stanów) pierwszej, drugiej i trzeciej. Symbolem  $S_{13}$  oznaczono prawdopodobieństwo przejścia z  $S_1$  do  $S_3$ .

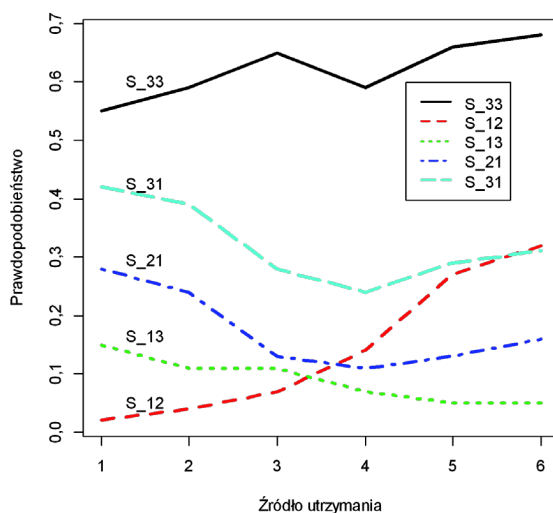
<sup>8</sup> Wyższe prawdopodobieństwo przejścia z  $S_{13}$  aniżeli z  $S_{12}$  można również tłumaczyć analizą wpływu zmiennych towarzyszących ( $S_2$  tworzą głównie gospodarstwa emerytów i rencistów).

ściach lub niechcący określić ich wysokości. Największe prawdopodobieństwo pozostania w  $S_3$  mają gospodarstwa nierodzinne wieloosobowe i jednoosobowe, pracujące na własny rachunek, lub gospodarstwa o niezarobkowych źródła utrzymania. Gospodarstwa te cechuje również wartość oszczędności przekraczająca wysokość rocznych dochodów lub niechcący określić ich wysokości.

Zaobserwowano również istotny wpływ zmiennych towarzyszących na prawdopodobieństwa przejścia do poszczególnych stanów.

Prawdopodobieństwo przejścia z  $S_1$  do  $S_2$  jest wyraźnie wyższe dla gospodarstw innych niż gospodarstwa pracowników (najwyższe dla gospodarstw o niezarobkowych źródłach utrzymania (0,32), rencistów (0,27) i emerytów (0,13)). Z kolei prawdopodobieństwo przejścia z  $S_2$  do  $S_1$  jest najwyższe dla gospodarstw pracowników (0,28), rolników (0,24) oraz o niezarobkowych źródłach utrzymania (0,16).

Prawdopodobieństwo przejścia z  $S_1$  do  $S_3$  jest najwyższe dla gospodarstw pracowników (0,15), rolników (0,11) i pracujących na własny rachunek (0,11). Największe prawdopodobieństwo przejścia z  $S_3$  do  $S_1$  mają pracownicy (0,42) i rolnicy (0,39) oraz gospodarstwa o niezarobkowych źródłach utrzymania (0,31). Ilustrację graficzną wybranych prawdopodobieństw przejścia dla zmiennej „źródło utrzymania” pokazano na rys. 1.



**Rys. 1.** Prawdopodobieństwa przejścia do poszczególnych klas dla zmiennej „źródło utrzymania”  
Źródło: opracowanie własne.

Dla gospodarstw o łącznej wartości oszczędności wyższej niż wysokość miesięcznych dochodów wzrasta prawdopodobieństwo przejścia z  $S_1$  do  $S_3$ , a spada z  $S_3$  do  $S_1$ . Dla rodzin z dziećmi wzrasta prawdopodobieństwo przejścia z  $S_2$  do  $S_3$ , a spada z  $S_3$  do  $S_2$ .

## 4. Zakończenie

W artykule przedstawiono przykład zastosowania ukrytego modelu Markowa z uwzględnieniem zmiennych towarzyszących w analizie oszczędności polskich gospodarstw domowych. Podobnie jak w pracy [Genge 2014], na podstawie kryteriów informacyjnych wyodrębniono trzy klasy o podobnych motywach oszczędzania wśród Polaków. Zbadano także wpływ zmiennych towarzyszących na oszacowane prawdopodobieństwa początkowe oraz prawdopodobieństwa przejścia (prawdopodobieństwa zmiany nastawienia do oszczędzania). Liczne badania pokazują, iż posiadanie jasno określonego celu oszczędzania zwiększa szanse na skuteczne zgromadzenie oszczędności. Uzyskane wyniki mogą zatem stać się motorem do działań podejmowanych zarówno przez podmioty gospodarcze, np. instytucje finansowe, jak i przez państwo (inicjatywy zwiększające wiedzę i świadomość finansową Polaków, wzmacniające edukację społeczeństwa w zakresie zasad ustalania wysokości emerytury czy form gromadzenia oszczędności, zapewniające atrakcyjność form gromadzenia oszczędności). W dalszych badaniach analizie poddaną zostaną również ukryte modele Markowa z uwzględnieniem teorii efektów losowych [Bartolucci i in. 2015].

## Literatura

- Akaike H., 1974, *A new look at statistical model identification*, IEEE Transactions on Automatic Control, vol. 19, no. 6, s. 716–723.
- Bartolucci F., Farcomeni A., 2009, *A multivariate extension of the dynamic logit model for longitudinal data based on latent Markov heterogeneity structure*, Journal of the American Statistical Association, vol. 104, no. 486, s. 816–831.
- Bartolucci F., Farcomeni A., Pandolfi S., Pennoni, F. 2015, *LMest: An R package for latent Markov models for categorical longitudinal data*, <http://arxiv.org/abs/1501.04448> (19.03.2015).
- Bartolucci F., Farcomeni A., Pennoni F., 2013, *Latent Markov Models for Longitudinal Data*, Chapman and Hall/CRC Press, Boca Raton.
- Bartolucci F., Lupporelli M., Montanari G.E., 2009, *Latent Markov model for binary longitudinal data: An application to the performance evaluation of nursing homes*, Annals of Applied Statistics, vol. 3, no. 1, s. 611–636.
- Bartolucci F., Montanari G. E., Pandolfi S., 2015, *Three-step estimation of latent Markov models with covariates*, Computational Statistics & Data Analysis, vol. 83, no. 1 s. 287–301.
- Bartolucci F., Pennoni F., Francis B., 2007, *A latent Markov model for detecting patterns of criminal activity*, Journal of the Royal Statistical Society, vol. 170, no. 1, ser. A, s. 151–132.
- Czapiński J., Panek T. (red.), *Diagnoza społeczna 2013. Warunki i jakość życia Polaków (raport)*, Rada Monitoringu Społecznego, Warszawa.
- Dempster A.P., Laird N.P., Rubin D.B., 1977, *Maximum likelihood for incomplete data via the EM algorithm (with discussion)*, Journal of the Royal Statistical Society, vol. 39, no. 1, ser. B, s. 1–38.
- Genge E., 2014, *Zastosowanie ukrytych modeli Markowa w analizie oszczędności wśród Polaków*, [w:] J. Kolonko, G. Kończak (red.), *Metody wnioskowania statystycznego w badaniach ekono-*

- micznych*, Studia Ekonomiczne, nr 189, Wydawnictwo Uniwersytetu Ekonomicznego w Katowicach, Katowice, s. 58–66.
- Genge E., 2015a, *A longitudinal study of Polish emigration attitudes: A latent Markov model approach*, [w:] M. Papież, S. Śmiech (red.), *The 9th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-Economic Phenomena. Conference Proceedings*, Foundation of the Cracow University of Economics, Cracow, s. 49–57 (CD-ROM).
- Genge E., 2015b, *Zaufanie do instytucji publicznych i finansowych w polskim społeczeństwie – analiza empiryczna z wykorzystaniem ukrytych modeli Markowa*, Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu, nr 384, *Taksonomia 24: Klasyfikacja i analiza danych – teoria i zastosowania*, s. 100–107.
- Podgórska M., Śliwka P., Topolewski M., Wrzosek M., 2000, *Łańcuchy Markowa w teorii i zastosowaniach*, Oficyna Wydawnicza SGH, Warszawa.
- Schwarz G., 1978, *Estimating the dimension of a model*, *Annals of Statistics*, vol. 6, no. 2, s. 461–464.
- Stawicki J., 2004, *Wykorzystanie łańcuchów Markowa w analizie rynku kapitałowego*, Wydawnictwo Uniwersytetu Mikołaja Kopernika, Toruń.
- Vermunt J.K., 2010, *Longitudinal research using mixture models*, [w:] K. van Montfort, J.H.L. Oud, A. Satorra (red.), *Longitudinal Research with Latent Variables*, Springer, Berlin–Heidelberg, s. 119–152.
- Vermunt J.K., Langeheine R., Böckenholt U., 1999, *Discrete-time discrete-state latent Markov models with time-constant and time-varying covariates*, *Journal of Educational and Behavioral Statistics*, vol. 24, no. 1, s. 178–205.
- Wiggins L.M., 1955, *Mathematical Models for the Analysis of Multiwave Panels*, Ph.D. Dissertation, Columbia University, Ann Arbor, MI.
- Wiggins L.M., 1973, *Panel Analysis: Latent Probability Models for Attitude and Behaviour Processes*, Elsevier, Amsterdam.