

MODELE UKRYTYCH ŁAŃCUCHÓW MARKOWA JAKO METODA ANALIZY DANYCH PANELOWYCH

1. Wprowadzenie

Celem niniejszego artykułu jest prezentacja modelu ukrytego łańcucha Markowa (*latent Markov model*) oraz jego rozszerzeń jako narzędzia analizy danych panelowych zarówno od strony teoretycznej, jak i od strony praktycznego zastosowania w dziedzinie mikroekonomii.

Pierwszą część artykułu poświęcono teorii modelu ukrytego łańcucha Markowa. Opisano w niej podstawową postać modelu oraz jego rozszerzenie – model mieszanych ukrytych łańcuchów Markowa (*mixed latent Markov model*) ze współzmiennymi. Omówiono parametryzację tych modeli, metodę estymacji i ocenę jakości. W drugiej części artykułu zaprezentowano model mieszanych ukrytych łańcuchów Markowa ze współzmiennymi dla zagadnienia zachowań finansowych polskich gospodarstw domowych. W dotychczasowych ekonometrycznych badaniach zachowań finansowych koncentrowano się na poszukiwaniu determinant wybranego aspektu zachowań finansowych, najczęściej stopy oszczędności. Używano w tym celu przede wszystkim metod regresji, również panelowej. Modele ukrytych łańcuchów Markowa mają tę przewagę, że pozwalają analizować jednocześnie i w powiązaniu wiele aspektów zachowań finansowych, dając pełniejszy obraz tego zagadnienia. Ponadto, umożliwiają uwzględnienie motywów oszczędzania i zadłużania się, które zajmują istotne

miejsce w teoriach zachowań finansowych, a rzadko były brane pod uwagę w badaniach ekonometrycznych.

2. Podstawowy model ukrytego łańcucha Markowa

Model ukrytego łańcucha Markowa przedstawił Wiggins w 1955 r.¹ Łączył on model łańcucha Markowa z modelem klas ukrytych. Niech $\{S^t\}$, $t = 0, 1, \dots, T$ będzie skończonym łańcuchem Markowa. W łańcuchu Markowa $\{S^t\}$ rozkład zmiennej losowej S^t zależy wyłącznie od wartości przyjętej przez S^{t-1} , a nie zależy bezpośrednio od wartości przyjmowanych wcześniej². W modelu ukrytego łańcucha Markowa dyskretna zmienna losowa S^t nie jest bezpośrednio mierzalna³, a stany łańcucha $s^t = 0, 1, \dots, L_S$ nazywa się ukrytymi.

Pomiar ukrytego stanu odbywa się za pośrednictwem zmiennych wskaźnikowych Y_k^t , $k = 1, 2, \dots, K$, przyjmujących wartości y_k^t . W każdym okresie t dla zmiennej S^t przyjmuje się model klas ukrytych⁴. W modelu tym zakłada się, że cała zależność pomiędzy zmiennymi wskaźnikowymi wynika wyłącznie z ich zależności od ukrytej klasy (tzw. założenie lojalnej niezależności)⁵. Interpretacji stanów ukrytych dokonuje się na podstawie parametrów rozkładów zmiennych wskaźnikowych, warunkowych względem stanów. Aby zachować stałą interpretację stanów, zakłada się niezmiennosc pomiaru, która oznacza, że w każdym okresie występują te same zmienne wskaźnikowe, o takich samych rozkładach warunkowych.

Model ukrytego łańcucha Markowa można zapisać w postaci⁶:

$$P(\mathbf{S} = \mathbf{s}, \mathbf{Y} = \mathbf{y}) = P(S^0 = s^0) \cdot \prod_{k=1}^K P(Y_k^0 = y_k^0 | S^0 = s^0) \cdot \prod_{t=1}^T \left[P(S^t = s^t | S^{t-1} = s^{t-1}) \cdot \prod_{k=1}^K P(Y_k^t = y_k^t | S^t = s^t) \right], \quad (1)$$

gdzie \mathbf{S} jest wektorem S^t , \mathbf{s} jest wektorem s^t , \mathbf{Y} jest macierzą Y_k^t , a \mathbf{y} jest macierzą y_k^t .

¹ R. Langeheine, F. van de Pol, *Latent Markov Chains*, w: *Applied Latent Class Analysis*, red. J.A. Hagenaars, A.L. McCutcheon, Cambridge University Press, New York 2002, s. 323.

² M. Podgórska, P. Śliwka, M. Topolewski, M. Wrzosek, *Łańcuchy Markowa w teorii i zastosowaniach*, Oficyna Wydawnicza SGH, Warszawa 2002, s. 11–12.

³ R. Langeheine, F. van de Pol, op.cit., s. 323.

⁴ A.L. McCutcheon, *Basic Concepts and Procedures In Single and Multiple-Group Latent Class Analysis*, w: *Applied Latent Class Analysis*, op.cit., s. 56.

⁵ Ibidem, s. 58.

⁶ R. Langeheine, F. van de Pol, op.cit., s. 323; J.K. Vermunt, B. Tran, J. Magidson, *Latent Class Models in Longitudinal Research*, w: *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, red. S. Menard, Elsevier, Burlington 2008, s. 374, 378.

Gdy $P(S^0 = s^0)$, wówczas prawdopodobieństwa początkowe przynależności do ukrytych stanów $P(S^t = s^t | S^{t-1} = s^{t-1})$ to prawdopodobieństwa przejścia.

Dla okresu $t > 0$ rozkład S^t zależy tylko od wartości S^{t-1} , a nie zależy od wartości S^0, \dots, S^{t-2} na podstawie własności Markowa. Rozkład łączny zmiennych wskaźnikowych jest opisany iloczynem ich rozkładów na podstawie założenia o lokalnej niezależności, rozszerzonego na wiele okresów.

3. Model mieszanych ukrytych łańcuchów Markowa ze współzmiennymi

Model ukrytego łańcucha Markowa można rozszerzyć poprzez uwzględnienie obserwowalnych oraz nieobserwowalnych współzmiennych. Współzmiennie determinują przynależność do ukrytych stanów poprzez wpływ na prawdopodobieństwa początkowe oraz prawdopodobieństwa przejścia. Nieobserwowalna współzmienna pełni rolę klasy ukrytej, stałej w całym badanym okresie. Uwzględniając ją, otrzymuje się model mieszany, w którym prawdopodobieństwa początkowe i macierze przejścia są różne, w zależności od klasy ukrytej. Prawdopodobieństwo przynależności do klasy ukrytej jest przedmiotem estymacji i może zależeć od obserwowalnych współzmiennych.

Niech $X_m^t, m = 1, \dots, M$ przyjmujące wartości $x_m^t \in R$ będą obserwowalnymi współzmiennymi determinującymi przynależność do stanów, $Z_q, q = 1, \dots, Q$ przyjmujące wartości $z_q \in R$ będą obserwowalnymi współzmiennymi determinującymi przynależność do klasy ukrytej, a U o wartościach $u = 1, \dots, L_U$ oznacza klasę ukrytą. Model mieszanych ukrytych łańcuchów Markowa można zapisać w postaci⁷:

$$\begin{aligned}
 &P(U = u, \mathbf{S} = \mathbf{s}, \mathbf{Y} = \mathbf{y} | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}) = \\
 &= P(U = u | \mathbf{Z} = \mathbf{z})P(S^0 = s^0 | U = u, \mathbf{X}^0 = \mathbf{x}^0) \cdot \prod_{k=1}^K P(Y_k^0 = y_k^0 | S^0 = s^0) \\
 &\cdot \prod_{t=1}^T \left[P(S^t = s^t | S^{t-1} = s^{t-1}, U = u, \mathbf{X}^t = \mathbf{x}^t) \cdot \prod_{k=1}^K P(Y_k^t = y_k^t | S^t = s^t) \right],
 \end{aligned} \tag{2}$$

gdzie \mathbf{X} jest macierzą X_m^t , \mathbf{x} jest macierzą x_m^t , \mathbf{Z} jest wektorem Z_q , a \mathbf{z} jest wektorem z_q .

⁷ J.K. Vermunt, B. Tran, J. Magidson, op.cit., s. 374.

4. Estymacja i ocena jakości modeli ukrytego łańcucha Markowa

Model ukrytego łańcucha Markowa można zapisać, używając 3 alternatywnych parametryzacji: probabilistycznej, log-liniowej oraz logistycznej⁸. Parametryzacje te dla najprostszych modeli, w szczególności modelu ukrytego łańcucha Markowa z wyłącznie nominalnymi wskaźnikami, są równoważne⁹. W bardziej skomplikowanych modelach przewagę zyskuje najbardziej elastyczna z nich parametryzacja logistyczna, która pozwala m.in. użyć ciągłych współzmiennych i zmiennych wskaźnikowych. W parametryzacji logistycznej rozkłady stanów i klas ukrytych oraz nominalnych zmiennych wskaźnikowych opisują modele wielomianowej regresji logistycznej, a rozkłady ciągłych zmiennych wskaźnikowych można opisać modelem regresji odpowiednim dla jej rozkładu¹⁰.

Do szacowania modelu wykorzystuje się metodę największej wiarygodności (*Maximum Likelihood* – ML)¹¹. Podstawiając we wzorze (1) lub (2) za poszczególne prawdopodobieństwa warunkowe wzory wynikające z przyjętego modelu wielomianowej regresji logistycznej i ewentualnie regresji odpowiedniej dla wskaźników ciągłych, otrzymuje się wzór na wiarygodność pojedynczej obserwacji, warunkową względem wektora parametrów. Na potrzeby estymacji wystarczy, aby obiekt występował w próbie w dwóch okresach. Nie jest więc konieczne posiadanie danych ze wszystkich okresów dla wszystkich obiektów. Źródłem problemów w estymacji, w szczególności występowania nieidentyfikowalnych lub brzegowych parametrów, mogą być wskaźniki i współzmiennie o rzadko występujących kategoriach. Z tego powodu niepożądane są również stany i klasy ukryte o niskich prawdopodobieństwach przynależności.

W celu znalezienia wartości parametrów maksymalizujących funkcję wiarygodności najczęściej używa się algorytmu maksymalizacji wartości oczekiwanej (*expectation-maximization* – EM)¹² oraz metody Newtona–Rapsona (NR)¹³. Każdy krok algorytmu EM składa się z dwóch etapów¹⁴. W etapie wartości oczekiwanej (E), podstawiając za parametry ich oszacowania z poprzedniego kroku, oblicza się wartość oczekiwaną stanów i klas ukrytych. W etapie maksymalizacji (M) znajduje się nowe parametry maksymalizujące, a przynajmniej zwiększające, wartość

⁸ A.L. McCutcheon, op.cit., s. 57–63.

⁹ Ibidem, s. 63.

¹⁰ J.K. Vermunt, J. Magidson, *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Statistical Innovations, Belmont 2005, s. 9.

¹¹ A.L. McCutcheon, op.cit., s. 64–65.

¹² A.P. Dempster, N.M. Laird, D.B. Rubin, *Maximum Likelihood from incomplete data via EM algorithm*, „Journal of the Royal Statistical Society” (series B) 1977, vol. 39, no. 1, s. 1–38.

¹³ A.L. McCutcheon, op.cit., s. 64.

¹⁴ A.P. Dempster, N.M. Laird, D.B. Rubin, op.cit., s. 3.

funkcji wiarygodności, przyjmując za wartości stanów i klas ukrytych ich wartości oczekiwane obliczone w etapie E. Zaletą algorytmu EM w porównaniu do NR jest niewielka wrażliwość na wybór punktu startowego¹⁵. Algorytm NR natomiast, o ile wystartuje w dobrym punkcie, szybko osiąga zbieżność, a ponadto generuje macierz wariancji-kowariancji oszacowań. Aby wykorzystać zalety obu algorytmów, rozpoczyna się estymację od algorytmu EM, a po pewnej liczbie kroków przechodzi się na algorytm NR¹⁶.

W celu wyboru najlepszego z porównywanych modeli ocenia się ich dopasowanie oraz trafność zdefiniowania stanów i klas ukrytych. Ocena dopasowania bazuje na wartości logarytmu funkcji wiarygodności LL. Na jej podstawie oblicza się kryteria informacyjne, m.in. Akaike (AIC i CAIC) i Bayesa–Schwarza (BIC). Im niższa jest ich wartość, tym lepszy jest model ze względu na to, że jest dobrze dopasowany przy relatywnie niewielkiej liczbie parametrów. W praktycznych zastosowaniach najczęściej wybierany jest model, dla którego BIC osiąga minimum¹⁷. Na podstawie wartości LL można przeprowadzić test ilorazu wiarygodności dla modeli zagnieżdżonych – np. o różnej liczbie stanów i klas. W szczególności można w ten sposób przetestować założenie jednorodności łańcucha, mówiące o stałości macierzy przejścia w czasie.

Trafność zdefiniowania stanów i klas ukrytych bada się poprzez sprawdzanie, na ile precyzyjnie można przyporządkować do nich obserwowane obiekty, mając do dyspozycji zmienne wskaźnikowe¹⁸. W tym celu oblicza się prawdopodobieństwa przynależności do poszczególnych stanów warunkowe względem wartości zmiennych wskaźnikowych i oszacowanych parametrów. Na ich podstawie oblicza się procent błędnych klasyfikacji oraz kilka rodzajów pseudo- R^2 , opisujących względny przyrost jakości klasyfikacji do stanów, uzyskiwany dzięki wykorzystaniu informacji o zmiennych wskaźnikowych.

¹⁵ A.L. McCutcheon, op.cit., s. 64–65.

¹⁶ J.K. Vermunt, J. Magidson, op.cit., s. 52.

¹⁷ Przykłady: L.J. Paas, J.K. Vermunt, T.H.A. Bijmolt, *Discrete time discrete state latent Markov modeling for assessing and predicting household acquisition of financial products*, „Journal of the Royal Statistical Society” (series A) 2007, vol. 170, no. 4, s. 955–974; J. Magidson, J.K. Vermunt, B. Tran, *Using a Mixture Latent Markov Model to Analyze Longitudinal U.S. Employment Data Involving Measurement Error*, w: *New Trends in Psychometrics*, red. K. Shigemasu, A. Okada, T. Imaizumi, T. Hoshino, „Frontiers Science Series” 2007, no. 55, Universal Academy Press Inc., s. 235–242; R. Du, W.A. Kamakura, *Household Lifecycles and Life Styles in America*, „Journal of Marketing Research” 2006, vol. 43, s. 121–132.

¹⁸ J.K. Vermunt, J. Magidson, op.cit., s. 62.

5. Postać modelu mieszanych ukrytych łańcuchów Markowa w kontekście teorii zachowań finansowych

Gospodarstwa domowe przejawiają różnego rodzaju zachowania finansowe. Antonides i van Raaij¹⁹ wyróżniają 11 rodzajów zachowań, z których najczęściej przedmiotami badań są oszczędzanie i pożyczanie. Opisując zachowania finansowe, można analizować przejawianie konkretnych zachowań lub stany finansów, do których te zachowania prowadzą. Stany mają tę przewagę, że są bardziej trwałe i dają pogląd o zachowaniach przejawianych w dłuższym okresie. Hipoteza cyklu życia głosi, że zachowania finansowe, a więc i stany finansów, zmieniają się w ciągu życia²⁰. Wobec tego adekwatnym podejściem do analizy zachowań finansowych wydaje się model łańcucha Markowa, w którym stany łańcucha obrazują stan finansów gospodarstwa domowego. Ponieważ nie sposób opisać tego stanu jedną zmienną, należy użyć szeregu wskaźników, co prowadzi do modelu ukrytego łańcucha Markowa. Wskaźnikami stanów ukrytych powinny być przede wszystkim zmienne opisujące stan oszczędności i zadłużenia.

W teoriach zachowań finansowych ważną rolę zajmują motywy ich podejmowania, głównie – motywy oszczędzania. Osiem motywów oszczędzania wyróżnił już Keynes²¹, kolejny dodali Browning i Lusardi²². Antonides i van Raaij²³ wyróżnili cztery poziomy zachowań finansowych, bazując właśnie na motywach oszczędzania i zadłużania się. Shefrin i Thaler²⁴ w teorii księgowania mentalnego postulowali, że jednostki dzielą posiadane środki na wyobrażone konta, według ich przeznaczenia. Wobec tego przeznaczenie oszczędności i kredytów, odzwierciedlające motywy ich zgromadzenia czy zaciągnięcia, powinno się znaleźć wśród wskaźników stanu finansów.

Przejawianie poszczególnych zachowań finansowych jest szeregiem uwarunkowane czynników. Najważniejszym z nich w świetle niemal wszystkich teorii zachowań finansowych jest dochód. W różnych postaciach występuje on u Keynesa²⁵,

¹⁹ G. Antonides, W.F. van Raaij, *Consumer Behaviour. A European Perspective*, John Wiley & Sons, New York 1998, s. 432.

²⁰ F. Modigliani, R. Brumberg, *Utility analysis and the consumption function: an interpretation of cross-section data*, w: *Post-Keynesian economics*, red. K. Kurihara, NJ: Rutgers University Press, New Brunswick 1954.

²¹ J.M. Keynes, *The general theory of employment, interest and money*, MacMillan, London 1936, s. 56–99.

²² M. Browning, A. Lusardi, *Household saving: micro theories and micro facts*, „Journal of Economic Literature” 1996, vol. 34, s. 1797.

²³ G. Antonides, W.F. van Raaij, op.cit., s. 433–436.

²⁴ H.M. Shefrin, R.H. Thaler, *The behavioral life-cycle hypothesis*, „Economic Inquiry” 1988, vol. 26 (4), Oxford University Press, s. 609–643.

²⁵ J.M. Keynes, op.cit.

Dunsberry'ego²⁶, w psychologicznej teorii konsumpcji Katony²⁷ i oczywiście w hipotezie dochodu permanentnego²⁸. Dochód powinien więc być uwzględniony w modelu jako najważniejsza współzmienna. W świetle hipotezy cyklu życia bardzo istotną determinantą zachowań finansowych jest wiek²⁹, a szerzej – faza przebiegu życia. Wpływa ona na możliwości dochodowe oraz potrzeby, przekładające się na motyw. Wobec tego również aspekty związane z przebiegiem życia powinny być uwzględnione jako współzmiennie.

Według teorii samokontroli decyzje finansowe są wypadkową działania dwóch części osobowości o różnych preferencjach międzyokresowych³⁰. Strotz dzielił ludzi na naiwnych i wysublimowanych – w zależności od tego, czy zdawali sobie sprawę z dwoistości swoich preferencji i próbowali sprawować samokontrolę³¹. Nieobserwowalny podział na „naiwnych” i „wysublimowanych” można uchwycić w modelu mieszanych ukrytych łańcuchów Markowa w postaci klasy ukrytej.

Zdolność do sprawowania samokontroli zależy od szeregu czynników, głównie subiektywnych, np. postaw. Badania empiryczne związane z teorią społecznych reprezentacji wykazały, że w społeczeństwie dominują dwie przeciwstawne postawy wobec zachowań finansowych: prooszczędnościowa i prokonsumpcyjna³². Zaobserwowano, że orientacja prooszczędnościowa jest powszechniejsza w wyższych klasach społecznych, a prokonsumpcyjna w niższych. Wobec tego wskaźniki klasy społecznej powinny zostać w modelu uwzględnione jako determinanty przynależności do klas ukrytych.

6. Model zachowań finansowych polskich gospodarstw domowych

W badaniu wykorzystano dane dotyczące oszczędności i kredytów z *Diagnozy społecznej*³³ – badania prowadzonego na reprezentatywnej próbie polskich gospodarstw

²⁶ J. Dunsberry, *Income, Saving and the theory of consumer behavior*, Harvard University Press, Cambridge 1949.

²⁷ Ibidem, s. 177–195.

²⁸ M. Freidman, *A theory of the consumption function*, Princeton University Press, Princeton 1957.

²⁹ F. Modigliani, R. Brumberg, op.cit., s. 4, 15–16.

³⁰ R.H. Thaler, H.M. Shefrin, *An Economic Theory of Self-Control*, „Journal of Political Economy” 1981, vol. 89 (2), University of Chicago Press, s. 392–406; H.M. Shefrin, R.H. Thaler, op.cit., s. 610–612.

³¹ P. Bańbuła, *Oszczędności i wybór międzyokresowy: podejście behawioralne*, „Materiały i Studia” 2006, nr 208, Narodowy Bank Polski, Warszawa 2006, s. 29, 40.

³² P. Webley, E.K. Nyhus, *Representations of Saving and Saving Behaviour*, w: *Everyday Representations of the Economy*, red. Ch. Roland-Lévy, E. Kirchler, E. Penz, C. Gray, WUV Universitätsverlag, Wien 2001, s. 96–99.

³³ Rada Monitoringu Społecznego, *Diagnoza społeczna: zintegrowana baza danych, 2000–2011*, www.diagnoza.com [dostęp 17.02.2012]; Rada Monitoringu Społecznego, *Diagnoza społeczna* (kwestionariusze:

domowych z próbą panelową. Jako wskaźników stanu finansów gospodarstwa domowego użyto zmiennych z części dotyczącej zasobności materialnej, przy czym niektóre kategorie pogrupowano ze względu na niską liczebność, która jest niepożądana w estymacji. Wskaźnikami stanu finansów były: posiadanie oszczędności, wysokość oszczędności w stosunku do dochodu gospodarstwa, forma oszczędności (gotówka, lokaty, formy zaawansowane), przeznaczenie oszczędności (bieżące wydatki, wypoczynek, wypadki losowe, leczenie, remont, zakup dóbr trwałego użytku, zabezpieczenie przyszłości dzieci i zabezpieczenie własnej starości) oraz posiadanie kredytów, ich wysokość w stosunku do dochodu, źródło kredytu (banki, źródła pozabankowe) i przeznaczenie kredytu (bieżące wydatki; spłata wcześniejszych zobowiązań; leczenie; zakup dóbr trwałego użytku, w tym nieruchomości; rozwój – np. kształcenie, rozwój działalności gospodarczej). Z wyjątkiem zmiennych opisujących wysokość zgromadzonych oszczędności i zaciągniętych zobowiązań wszystkie mierzone są na skali – tak/nie.

W charakterze współzmiennych wpływających na przynależność do stanów uwzględniono: dochód na jednostkę konsumpcyjną w minionym roku i ocenę zmiany dochodu w ostatnich 2 latach oraz wiek osoby będącej głową gospodarstwa domowego i typ rodziny biologicznej. Jako współzmiennie determinujące przynależność do klasy ukrytej wzięto pod uwagę: klasę wielkości miejscowości, wykształcenie głowy gospodarstwa domowego, przedział miesięcznego dochodu z ostatniego roku oraz źródło utrzymania.

Analizie poddano 3796 gospodarstw domowych, które brały udział przynajmniej w trzech edycjach badania, spośród tych przeprowadzonych w latach: 2003, 2005, 2007, 2009, 2011. Edycję z 2000 r. pominięto, ponieważ nie uwzględniono w niej ważnych motywów oszczędzania (zabezpieczenie starości) i zadłużania się (spłata wcześniejszych zobowiązań). Kolejne edycje wyznaczały kolejne analizowane momenty czasu. W sumie model oszacowano na 14 342 obserwacjach gospodarstw w czasie.

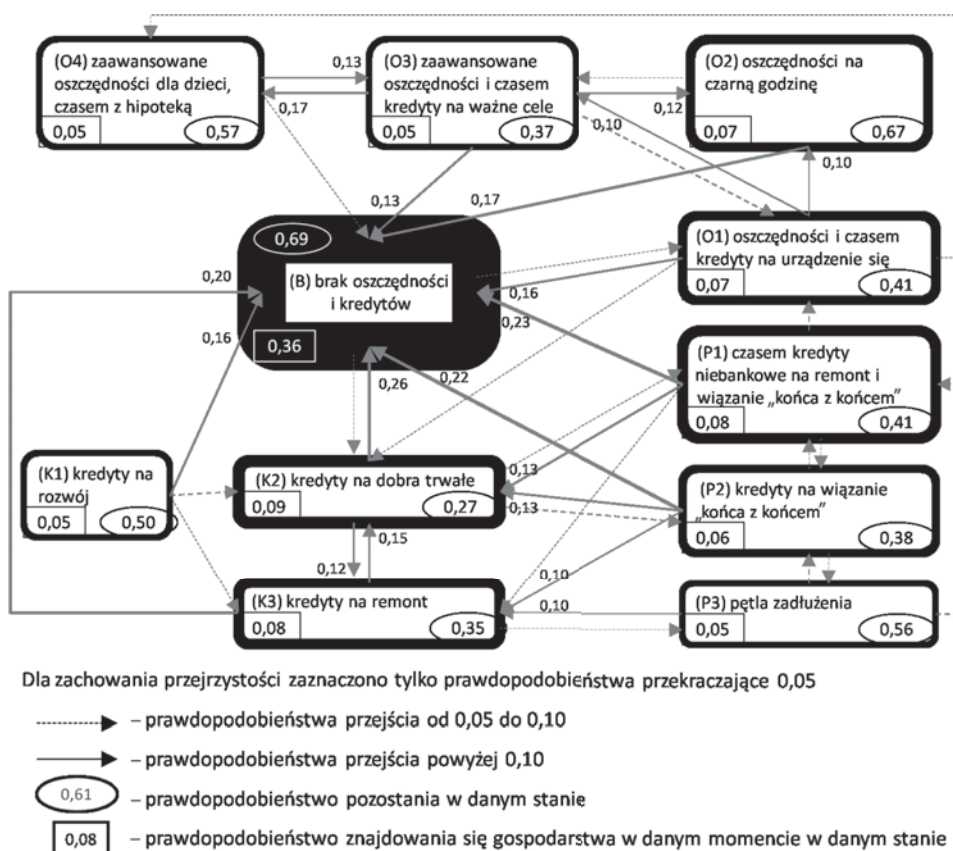
Analizę wykonano w module Syntax programu LatentGOLD 4.5³⁴. Założono niezmiennosć pomiaru. Oszacowano model z jedną oraz dwoma klasami ukrytymi dla liczby stanów od 4 do 14, zakładając homogeniczność macierzy przejścia. Najniższą wartość kryteriów informacyjnych BIC i CAIC osiągnięto dla modelu z 11 stanami, zarówno przy jednej, jak i przy dwóch klasach ukrytych. Dla obu tych modeli zbadano słuszność założenia homogeniczności macierzy przejścia, poprzez oszacowanie ich odpowiedników z heterogenicznymi macierzami przejścia i przeprowadzenie testu ilorazu wiarygodności. W obu przypadkach prawdopodobieństwo krytyczne wyniosło 1, nie dając podstaw do odrzucenia założenia homogeniczności łańcucha.

2000, 2003, 2005, 2007, 2009, 2011), www.diagnoza.com [dostęp 20.10.2011].

³⁴ J.K. Vermunt, J. Magidson, *LG-Syntax User's Guide: Manual for Latent GOLD 4.5 Syntax Module*, Statistical Innovations, Belmont 2008.

Test ilorazu wiarygodności przeprowadzony dla porównania modelu z jedną i dwoma klasami ukrytymi pokazał, że ten z dwoma jest istotnie lepiej dopasowany. Ponadto, dodanie klasy ukrytej pozwoliło uzyskać jakość klasyfikacji, mierzoną błędem klasyfikacji i pseudo- R^2 , porównywalną z najlepszymi pod tym względem modelami z jedną klasą. Dodatkowo model z dwoma klasami ma również uzasadnienie w teorii zachowań finansowych. Wobec tego do analizy przyjęto model z dwoma klasami. Wszystkie współzmiennne były istotne statystycznie.

Na rysunku 1 zaprezentowano oszacowany model zachowań finansowych polskich gospodarstw domowych. Prostokąty symbolizują zidentyfikowane stany. W każdym z nich umieszczono krótki opis danego stanu, sporządzony na podstawie interpretacji oszacowanych parametrów rozkładów zmiennych wskaźnikowych. Każdemu stanowi nadano też skróconą nazwę, np. O2. Strzałki odzwierciedlają wartości oszacowanych prawdopodobieństw przejścia.



Rysunek 1. Schemat przejść pomiędzy stanami finansów

Źródło: opracowanie własne.

W zachowaniach finansowych polskich gospodarstw domowych dominuje stan B, w którym brakuje zarówno oszczędności, jak i kredytów. Znajduje się w nim większość gospodarstw w większości momentów czasu, a co za tym idzie – prawdopodobieństwa przynależności do innych stanów są niewielkie. Wysokie prawdopodobieństwa pozostawania w danym stanie i niskie prawdopodobieństwa przejścia pokazują, że okres 2 lat jest relatywnie krótki w odniesieniu do dynamiki zmian stanu finansów gospodarstw domowych.

W pierwszej klasie ukrytej prawdopodobieństwo przynależności do stanu B, w której brakuje oszczędności i kredytów, przekracza 0,4, w porównaniu do 0,2 dla klasy drugiej. Ponadto, w klasie pierwszej jest wyższe niż w klasie drugiej prawdopodobieństwo przynależności do stanów K2, K3, w których dominuje zadłużenie z co najwyżej niewielkim buforem oszczędności, oraz do stanów P1–P3, charakteryzujących się finansowaniem bieżących wydatków z pożyczek. W klasie drugiej zaś wyższe są prawdopodobieństwa przynależności do stanów O2–O4, w których dominują oszczędności, a kredyty są zaciągane jedynie na istotne wydatki, oraz do stanu K1, w którym występuje zadłużenie na cele rozwojowe z pewnym buforem oszczędności. Można więc powiedzieć, że klasy ukryte faktycznie pozwoliły rozróżnić „naiwnych” od „wysublimowanych”.

Prawdopodobieństwo przynależności do klasy pierwszej, czyli do „naiwnych”, wynosi 0,7. Należą do niej zwłaszcza gospodarstwa pracowników, emerytów i rencistów oraz utrzymujące się ze źródeł niezarobkowych. Łączne miesięczne dochody prawie połowy z nich nie przekraczają 1,5 tys. zł. Gospodarstwa te relatywnie rzadko są w największych miastach, a osoby będące głowami tych gospodarstw mają zwykle wykształcenie podstawowe lub zawodowe. Do klasy drugiej, czyli „wysublimowanych”, należą zwłaszcza gospodarstwa rolników i przedsiębiorców, ale również część gospodarstw pracowników, emerytów i rencistów. Prawie połowa tych gospodarstw osiąga miesięczne dochody przekraczające 2 tys. zł. W ponad połowie przypadków osoby będące głowami tych gospodarstw mają wykształcenie średnie lub wyższe. Gospodarstwa te również nieco częściej są w większych miastach.

Rozkład dochodu w poszczególnych stanach przemawia na korzyść hipotezy o nieproporcjonalności. Najwyższe dochody na jednostkę konsumpcyjną uzyskują gospodarstwa ze stanów O1–O4, w których posiadanie oszczędności jest powszechne, a ich wysokość jest najwyższa w stosunku do dochodu. W tych też stanach dominujące jest poczucie, że sytuacja dochodowa nie uległa zmianie, lub nawet się poprawiła. Dalej plasują się gospodarstwa ze stanów K1–K3, w których oszczędności występują rzadko, a i wtedy są niskie w stosunku do dochodu. Najniższe dochody osiągają gospodarstwa w stanach B i P1–P4, w których oszczędności praktycznie nie występują. W stanach P1–P3 dominuje też poczucie, że sytuacja dochodowa się pogorszyła. Podobne obserwacje dotyczą również dochodu łącznego.

Na podstawie uzyskanych wyników trudno jednoznacznie odnieść się do hipotezy cyklu życia. Z pewnością nie ma jednej sekwencji stanów finansowych. Stany charakterystyczne dla klasy drugiej, „wysublimowanych”, układają się w łańcuch, w którym na każdym etapie występują oszczędności, a ważne przedsięwzięcia – zakup domu, założenie własnej działalności – mają miejsce wcześniej w przebiegu życia. Najniższym wiekiem osoby będącej głową gospodarstwa domowego charakteryzują się stany O4 i K1. Gospodarstwa w stanie O4 należą najczęściej do małżeństwa z jednym dzieckiem lub dwojgiem dzieci. Małżeństwa w stanie K1 zwykle mają dwoje lub troje dzieci i dość często prowadzą gospodarstwa wielorodzinne. Gospodarstwa z klasy drugiej na środkowym etapie życia należą do stanów O3 i O1 – rodziny posiadają oszczędności i niekiedy kredyty na ważne cele. Ostatni w przestrzeni stanów jest stan O2, w którym występują gospodarstwa osób najstarszych i bez dzieci.

Dla stanów charakterystycznych dla klasy pierwszej, „naiwnych”, obraz nie jest już tak prosty. W początkowym etapie życia mogą występować stany K2, P1 i P3, a więc oznacza to: zakup mieszkania, pętla zadłużenia i sporadyczne pożyczki. Do stanów P1 i P3, najbardziej obciążonych finansowo, relatywnie często należą rodziny niepełne. W środkowych etapach życia mogą wystąpić stany K3 lub O1, a więc remont lub drobne oszczędności. W stanie K3 relatywnie często znajdują się gospodarstwa wielorodzinne. W końcowych etapach życia występują zaś stany P2 i B, a więc brak oszczędności i kredytów lub wiązanie „końca z końcem” dzięki pożyczkom. W stanie B, podobnie jak w klasie drugiej w stanie O2, często występują gospodarstwa bez dzieci. Natomiast dla stanu P2 charakterystyczne są rodziny niepełne, podobnie jak na wcześniejszych etapach dla stanów P1 i P3.

7. Podsumowanie

W niniejszym artykule zaprezentowano model mieszanych ukrytych łańcuchów Markowa ze współzmiennymi. Opisano postać modelu, metodę estymacji i sposoby oceny jakości modelu. W charakterze ilustracji praktycznego zastosowania stworzono model leżących u podstaw mikro- i makroekonomicznej sytuacji w kraju zachowań finansowych polskich gospodarstw domowych. Zaprezentowano również podejście do modelowania. Na podstawie teorii badanego zjawiska ustalono, że model mieszanych ukrytych łańcuchów Markowa jest adekwatny do jego opisu, oraz wybrano wskaźniki stanów ukrytych i współzmiennie. Następnie oszacowano modele dla liczby stanów od kilku do kilkunastu i dokonano wyboru ostatecznej postaci modelu na podstawie kryteriów informacyjnych i miar jakości klasyfikacji, przetestowawszy uprzednio założenie jednorodności łańcucha. Wreszcie scharakteryzowano wyróżnione stany, bazując na oszacowanych rozkładach zmiennych wskaźnikowych,

przeanalizowano ścieżki przejść pomiędzy nimi, zinterpretowano klasy ukryte oraz omówiono zależności pomiędzy współzmiennymi a stanami i klasami ukrytymi.

Uzyskany model zachowań finansowych pozwolił odnieść się do wybranych teorii zachowań finansowych, wskazać stymulanty i destymulanty oszczędzania i zadłużania się oraz wyróżnić grupy społeczne szczególnie narażone na problemy finansowe.

Bibliografia

- Antonides G., Raaij W.F. van, *Consumer Behaviour. A European Perspective*, John Wiley & Sons, New York 1998.
- Bańbuła P., *Oszczędności i wybór międzyokresowy: podejście behawioralne*, „Materiały i Studia” 2006, nr 208, Narodowy Bank Polski, Warszawa 2006.
- Browning M., Lusardi A., *Household saving: micro theories and micro facts*, „Journal of Economic Literature” 1996, vol. 34, s. 1797–1855.
- Dempster A.P., Laird N.M., Rubin D.B., *Maximum Likelihood from incomplete data via EM algorithm*, „Journal of the Royal Statistical Society” (series B) 1977, vol. 39, no. 1, s. 1–38.
- Du R., Kamakura W.A., *Household Lifecycles and Life Styles in America*, „Journal of Marketing Research” 2006, vol. 43, s. 121–132.
- Dunsberry J., *Income, Saving and the theory of consumer behavior*, Harvard University Press, Cambridge 1949.
- Freidman M., *A theory of the consumption function*, Princeton University Press, Princeton 1957.
- Keynes J.M., *The general theory of employment, interest and money*, MacMillan, London 1936.
- Langeheine R., Pol F. van de, *Latent Markov Chains*, w: *Applied Latent Class Analysis*, red. J.A. Hagenaars, A.L. McCutcheon, Cambridge University Press, New York 2002, s. 304–341.
- Magdison J., Vermunt J.K., Tran B., *Using a Mixture Latent Markov Model to Analyze Longitudinal U.S. Employment Data Involving Measurement Error*, w: *New Trends in Psychometrics*, red. K. Shigemasu, A. Okada, T. Imaizum, T. Hoshino, „Frontiers Science Series” 2007, no. 55, Universal Academy Press Inc., s. 235–242.
- McCutcheon A.L., *Basic Concepts and Procedures In Single and Multiple-Group Latent Class Analysis*, w: *Applied Latent Class Analysis*, red. J.A. Hagenaars, A.L. McCutcheon, Cambridge University Press, New York 2002, s. 56–85.
- Modigliani F., Brumberg R., *Utility analysis and the consumption function: an interpretation of cross-section data*, w: *Post-Keynesian economics*, red. K. Kurihara, NJ: Rutgers University Press, New Brunswick 1954.
- Paas L.J., Vermunt J.K., Bijmolt T.H.A., *Discrete time discrete state latent Markov modeling for assessing and predicting household acquisition of financial products*, „Journal of the Royal Statistical Society” (series) 2007, vol. 170, no. 4, s. 955–974.

- Podgórska M., Śliwka P., Topolewski M., Wrzosek M., *Łańcuchy Markowa w teorii i zastosowaniach*, Oficyna Wydawnicza SGH, Warszawa 2002.
- Rada Monitoringu Społecznego, *Diagnoza społeczna* (kwestionariusze 2000, 2003, 2005, 2007, 2009, 2011), www.diagnoza.com [dostęp 20.10.2011].
- Rada Monitoringu Społecznego, *Diagnoza społeczna: zintegrowana baza danych, 2000–2011*, www.diagnoza.com [dostęp 17.02.2012].
- Shefrin H.M., Thaler R.H., *The behavioral life-cycle hypothesis*, „Economic Inquiry” 1988, vol. 26 (4), Oxford University Press, s. 609–643.
- Thaler R.H., Shefrin H.M., *An Economic Theory of Self-Control*, „Journal of Political Economy” 1981, vol. 89 (2), University of Chicago Press, s. 392–406.
- Vermunt J.K., Magidson J., *LG-Syntax User’s Guide: Manual for Latent GOLD 4.5 Syntax Module*, Statistical Innovations, Belmont 2008.
- Vermunt J.K., Magidson J., *Technical Guide for Latent GOLD 4.0: Basic and Advanced*, Statistical Innovations, Belmont 2005.
- Vermunt J.K., Tran B., Magidson J., *Latent Class Models in Longitudinal Research*, w: *Handbook of Longitudinal Research: Design, Measurement, and Analysis*, red. S. Menard, Elsevier, Burlington 2008, s. 373–385.
- Webley P., Nyhus E.K., *Representations of Saving and Saving Behaviour*, w: *Everyday Representations of the Economy*, red. Ch. Roland-Lévy, E. Kirchler, E. Penz, C. Gray, WUV Universitätsverlag, Wien 2001, s. 93–110.

Summary

Latent Markov models in panel data analysis

In the present article the mixed latent Markov model with covariates was presented. Its formulation, estimation and assessment were presented. As an example of practical use a model was created for financial behaviour of polish households, for which the mixed latent Markov model is adequate in the light of financial behaviour theories. Having chosen the best model, the latent classes, latent states and the transitions between them were characterized and their dependencies on covariates were described. The model created allowed to assess some of financial behaviour theories, point out stimulants of saving and indebteding and characterize social groups most exposed to financial problems.

Keywords: latent Markov models, mixed latent Markov models, households financial behaviour, Social Diagnosis

JEL classification: C33, C35, C38, D14