

Adam Sagan

Katedra Analizy Rynku i Badań Marketingowych
Uniwersytet Ekonomiczny w Krakowie

Zastosowanie programu Mplus w modelowaniu ukrytych przejść w segmentacji rynku

Streszczenie

W artykule omówiono zastosowanie modeli ukrytych przejść w segmentacji dynamicznej rynku. Modele te są uogólnieniem modeli klas ukrytych uwzględniających dynamiczny charakter danych. Pozwalają one na identyfikację ukrytych segmentów rynku, ocenę prawdopodobieństw warunkowych związanych z profilem ukrytych segmentów oraz prawdopodobieństw przejść w ramach klas (statusów) ukrytych w czasie. Model LTA został zastosowany do oceny segmentów dynamicznych w trzech falach analiz wyodrębnionych na podstawie wskaźników statusu ekonomicznego w danych z Polskiego Generalnego Sondażu Społecznego.

Słowa kluczowe: model ukrytych przejść, ukryte łańcuchy Markova, segmentacja rynku, Mplus.

1. Wprowadzenie

We współczesnych badaniach marketingowych, szczególnie w orientacji relacyjnej w marketingu, istotnym zagadnieniem staje się analiza zmiany w zachowaniach konsumentów. Jest to związane z oceną wartości dla konsumenta, wartości życiowej klienta, zmian w kształtowaniu zależności między zadowoleniem a lojalnością. Analizy dynamicznych aspektów zachowań rynkowych są również istotnym obszarem badań segmentacyjnych. W świetle zmieniających się

warunków otoczenia, identyfikacja i ocena segmentów rynku musi uwzględniać zarówno statyczny aspekt segmentacji, jak i zmiany w strukturze segmentów w czasie i czynniki je wyjaśniające. Czynniki te powodują, że tradycyjne podejścia w badaniach segmentacyjnych, takie jak analiza skupień czy analiza klas ukrytych, są uzupełniane o metody pozwalające na identyfikację zmian w strukturze wyodrębnianych segmentów w czasie.

Celem artykułu jest przedstawienie modelu ukrytych przejść (*latent transition model*) i jego zastosowań w badaniach segmentacyjnych na podstawie danych Polskiego Generalnego Sondażu Społecznego (PGSS). Jest ona mało znana w polskiej literaturze marketingowej, a w szczególności pracach dotyczących badań segmentacyjnych. Dodatkowo przedstawiona jest procedura estymacji i interpretacji parametrów za pomocą programu służącego do modelowania zmiennych ukrytych Mplus.

2. Modele ukrytych przejść w badaniach segmentacyjnych

W badaniach segmentacyjnych pozwalających na wyodrębnienie wewnętrznie spójnych grup nabywców różniących się między sobą wzorcami reakcji na oferty marketingowe firm można wyodrębnić kilka podstawowych wymiarów. Pierwszy odnosi się do sposobu wyodrębnienia kryteriów segmentacji, który dzieli je na kryteria *a priori* pozwalające na określenie konfiguracji zmiennych segmentacyjnych przed dokonywaniem klasyfikacji badanych (np. wg płci lub wieku) oraz *post hoc*, w których jest ona możliwa dopiero po przeprowadzeniu redukcji danych i oceny wymiarowości (np. postawy i wartości). Drugi wymiar odnosi się do charakteru kryteriów, które mogą być wyodrębniane na podstawie cech nabywców (np. wartości osobowe, wiek czy płeć) lub ich reakcji na instrumenty marketingowe oraz cechy produktu (np. częstość zakupu, preferencje lub postawy wobec marki). W wymiarze metodologicznym w badaniach segmentacyjnych są wykorzystywane metody opisowe (analiza skupień, analiza klas ukrytych czy analiza głównych składowych) lub predykcyjne (drzewa klasyfikacyjne, modele logitowe lub regresja czynnikowa). Stosowane są podejścia eksploracyjne (analiza skupień) lub modelowe (analiza klas ukrytych). Segmentacja jest dokonywana w sposób statyczny (analiza skupień i klas ukrytych) lub dynamiczny (modele krzywych rozwojowych, modele ukrytych przejść) [Wedel i Kamakura 2000, s. 17].

Modele ukrytych przejść w segmentacji rynku są związane z modelowym podejściem dynamicznym, w którym segmenty rynku wyłonione na podstawie analizy klas ukrytych są łączone poprzez prawdopodobieństwa przejścia w ujęciu dynamicznym. Pozwala to na ocenę stabilności i zmiany w strukturze ukrytych segmentów rynku.

Zastosowanie tego podejścia w segmentacyjnych badaniach marketingowych jest związane z segmentacją sekwencyjną wzorów poszukiwania informacji w sieci [Dias i Vermunt 2007], w której zastosowano ukryte łańcuchy Markova (*latent segment Markov chain model*) w celu dynamicznej segmentacji internautów, uwzględniającej zarówno niejednorodność populacji, jak i sekwencyjną zależność strategii poszukiwania informacji. Modele te pozwalają również na segmentację konsumentów z punktu widzenia wielkości efektów przestawienia i przesunięcia preferencji (*brand shifting/switching effects*) w badaniach panelowych [Poulsen 1990] i ewolucji struktury preferencji w czasie [Ramaswamy 1996]. Dynamiczna segmentacja jest jednak o wiele słabiej wykorzystywana w porównaniu z klasyczną segmentacją z wykorzystaniem analizy skupień lub klas ukrytych.

3. Model ukrytych przejść

Analiza ukrytych przejść należy do rodziny modeli klas ukrytych (*latent class analysis*), które wraz z modelami ukrytych profili należą do ogólnego typu modeli skończonych mieszanek (*finite mixture models*). Cechą charakterystyczną tych modeli jest próba identyfikacji klas ukrytych (dyskretnych zmiennych ukrytych) na podstawie kategoryalnych (np. binarnych) wskaźników. Podstawową zasadą identyfikacji klas ukrytych jest założenie lokalnej niezależności wskaźników, zgodnie z którym zależność między wskaźnikami wynika z ich związku z ukrytą zmienną. Po wprowadzeniu takiej zmiennej do modelu (jako zmiennej kontrolnej) występuje, dla jej ustalonych wartości lub klas, niezależność wskaźników – są one w stosunku do siebie lokalnie niezależne. Do podstawowych parametrów w modelach klas ukrytych należy prawdopodobieństwo klas ukrytych oraz prawdopodobieństwo warunkowe przynależności określonych przypadków do kategorii zmiennych obserwowalnych. Pierwsze z nich przedstawia rozkład klas zmiennej ukrytych, dla których występuje niezależność wskaźników. Rozkład ten jest funkcją liczby klas i ich wielkości. Prawdopodobieństwo warunkowe natomiast jest prawdopodobieństwem, z jakim dana jednostka należąca do klasy C zmiennej ukrytej L zostanie przyporządkowana do określonej kategorii zmiennej wskaźnikowej Y_k .

Analiza ukrytych przejść jest odmianą analizy klas ukrytych związaną z badaniem klas ukrytych w czasie. Stanowi ona rodzaj modeli mieszanych dla danych wzdłużnych, które w literaturze ekonometrycznej i marketingowej są również znane jako modele ukrytych łańcuchów Markova (*hidden Markov modeling*) [Everitt 2006, Frazer 2011]¹. Dla dwóch okresów i braku restrykcji nałożonych

¹ Dynamiczna wersja modeli regresji i modeli ścieżkowych jest związana z analizą ukrytych krzywych rozwojowych (*latent growth curves*).

na parametry modelu modele te są tożsame z modelem dwóch klas ukrytych dla powtarzanych pomiarów (*Repeated Measures LCA* – RMLCA), który jest właściwy dla analiz uwzględniających identyfikację klas charakteryzujących się swoistymi wzorcami zmian w poszczególnych badanych okresach. W modelach LTA klasyczne analizy klas ukrytych w poszczególnych okresach stanowią punkt wyjścia do budowy modelu LTA. Należy jednak podkreślić, że struktura klas i prawdopodobieństwa odpowiedzi mogą się istotnie różnić między modelami z powodu innej struktury danych i uwzględnienia dużej ilości informacji w modelach LTA (w porównaniu z LCA).

Analiza ukrytych przejść jest częściej wykorzystywana w celu oceny prawdopodobieństw przejść między klasami ukrytymi w kolejnych momentach. Stąd w przypadku, gdy zmiany charakteryzują się nielicznymi specyficznymi wzorami profili, RMLCA oferuje rozwiązania o mniejszej liczbie parametrów z LCA. W sytuacji wystąpienia przejść między parami klas ukrytych i spełnienia założenia procesu Markova I rzędu, model LTA jest modelem bardziej „oszczędny” w porównaniu z RMLCA. Pozwala on nie tylko na opis rozkładu przynależności do klas ukrytych, lecz także na ocenę prawdopodobieństw przejść z danej klasy do innej w czasie oraz wyjaśnienie tych prawdopodobieństw. Metodę LTA stosuje się w celu wyjaśnienia nie tylko struktury badanej populacji w danym momencie, ale również przejść różnych jednostek przynależących do różnych klas ukrytych. Podkreślając dynamiczny charakter klas ukrytych, są one z reguły nazywane ukrytymi statusami (*latent statuses*) [Collins i Lanza 2010].

Modele LTA są zazwyczaj identyfikowane dla bardzo dużych tabeli kontyngencji, ponieważ rozmiar w tabeli wynika z uwzględnienia wielu okresów pomiaru. Warunkiem koniecznym identyfikacji jest dodatnia liczba stopni swobody (*s.s.*), która jest różnicą między liczbą komórek tabeli kontyngencji (W) a liczbą szacowanych parametrów modelu ($s.s. = W - P - 1$). Zazwyczaj tabele danych w modelach klas ukrytych i ukrytych przejść są dość dużych rozmiarów. Dla przykładu: w przypadku czterech dwukategorialnych zmiennych występuje $2^4 = 16$ komórek tabeli dla analizy klas ukrytych. W modelu ukrytych przejść z trzema czasami pomiaru rozmiar tej samej tabeli wynosi już $W = 2^3 \times 4 = 2^{12} = 4,096$ komórek. Uwzględnianie kolejnych fal pomiaru powoduje wykładniczy wzrost rozmiarów tabeli danych. Stąd warunek konieczny identyfikacji modelu LTA wynikający z liczby stopni swobody jest zwykle spełniany. Drugim czynnikiem kształtującym identyfikację modelu jest liczebność próby (N) w relacji do liczby komórek (W) i związane z nią rozproszenie danych w tabeli. Wykładniczy wzrost liczby danych w stosunku do złożoności modelu powoduje, że modele LTA w celu otrzymania stabilnych oszacowań parametrów wymagają względnie dużych prób w procesie estymacji. Stopień rozproszenia jest mierzony stosunkiem liczebności próby do rozmiaru tabeli (N/W) i nie powinien być mniejszy od 5.

Ostatnim warunkiem identyfikacji jest „informacyjność” analizy określanej przez siłę relacji między zmienną ukrytą a jej wskaźnikami. Ilość informacji zawartych w danych jest związana z wielkością prawdopodobieństw warunkowych dla poprawnie specyfikowanego modelu, które są efektem homogeniczności wskaźników i odpowiedniej separacji klas ukrytych.

Ocena absolutnego dopasowania modelu jest dokonywana na podstawie statystyki G^2 [Collins i Lanza 2010, s. 83]:

$$G^2 = 2 \sum_{w=1}^W f_w \log\left(\frac{f_w}{f_e}\right), \quad (1)$$

gdzie f_w oznacza liczebności empiryczne, a f_e – liczebności oczekiwane.

Należy podkreślić, że występująca duża liczba stopni swobody i rozproszenie tabeli powoduje bardzo wysokie prawdopodobieństwo odrzucenia modelu i tym samym niewielką użyteczność tego testu.

Względne dopasowanie modelu LTA jest dokonywane na podstawie wartości ilorazu wiarygodności i informacyjnych wskaźników dopasowania. Pozwalają na bardziej optymalne określenie równowagi między dopasowaniem a prostotą analizowanego modelu. W modelach przejścia szacowane są trzy zestawy parametrów, z których dwa mają bezpośrednie odpowiedniki w analizie klas ukrytych: 1) prawdopodobieństwa odpowiedzi dla każdego okresu, 2) liczby przypadków w statusach ukrytych oraz 3) prawdopodobieństwa przejść między statusami [Guo *et al.* 2009].

Prawdopodobieństwa przynależności do ukrytych statusów (*latent status prevalences*) pełnią taką samą rolę, jak w analizie klas ukrytych. Określają one prawdopodobieństwa przynależności do statusu w danym czasie. Prawdopodobieństwo w czasie t_1 jest szacowane niezależnie, natomiast w pozostałych okresach są estymowane na podstawie prawdopodobieństw przynależności do ukrytych klas w poprzednim okresie (δ_{t-1}) oraz warunkowych prawdopodobieństw przejść ($\tau_{S_t|S_{t-1}}$) [Collins i Lanza 2010, s. 199]:

$$\delta_{S_t} = \sum_{S_{t-1}=1}^S \delta_{S_{t-1}} \tau_{S_t|S_{t-1}}. \quad (2)$$

Prawdopodobieństwa odpowiedzi na kategorie dla danego okresu (*item-response probabilities*) pozwalają na nazwanie ukrytych statusów. Stanowią warunkowe prawdopodobieństwa odpowiedzi na kategorie przy założeniu przynależności do określonego statusu (klasy ukrytej w czasie).

Prawdopodobieństwa przejść (*transition probabilities*) nie mają odpowiednika w analizie klas ukrytych i stanowią podstawowy typ informacji w modelach LTA. Są to warunkowe prawdopodobieństwa zmiany ukrytego statusu w czasie t przy założeniu przynależności do statusu w czasie $t - 1$. Wartości na głównej

przekątnej określają prawdopodobieństwo pozostanie w tej samej klasie ukrytej pod warunkiem przynależności do tej klasy w okresie poprzednim. Prawdopodobieństwa te są najczęściej wyrażone w postaci macierzy przejść [Collins i Lanza 2010, s. 198]:

$$\begin{array}{cccc} \tau_{1_{t+1}|1_t} & \tau_{2_{t+1}|1_t} & \dots & \tau_{S_{t+1}|1_t} \\ \tau_{1_{t+1}|2_t} & \tau_{2_{t+1}|2_t} & \dots & \tau_{S_{t+1}|2_t} \\ \dots & \dots & \dots & \dots \\ \tau_{1_{t+1}|S_t} & \tau_{2_{t+1}|S_t} & \dots & \tau_{S_{t+1}|S_t} \end{array} \quad (3)$$

Należy ostrożnie interpretować te prawdopodobieństwa, uwzględniając ilość czasu, jaki upłynął między pomiarami jego wpływ na wskaźniki zmiennych ukrytych.

Podstawowy model ukrytych przejść wyjaśnia dany profil odpowiedzi na pytania wskaźnikowe jako funkcję prawdopodobieństwa przynależności do ukrytego statusu w danym czasie, prawdopodobieństwa przejścia z jednego statusu do drugiego (uwzględniając przynależność do określonej klasy w czasie) oraz prawdopodobieństwa odpowiedzi na poszczególną kategorię wskaźnika przy założeniu przynależności do określonego statusu ukrytego. Jest to wyrażone za pomocą wzoru [Collins i Lanza 2010, s. 198]:

$$P(Y=y) = \sum_{s_1=1}^S \dots \sum_{S_T=1}^S \delta_{s_1} \tau_{s_2|s_1} \dots \tau_{S_T|S_{T-1}} \prod_{t=1}^T \prod_{j=1}^J \prod_{r_{j,t}=1}^{R_j} \rho_{j,r_{j,t}|S_t}^{I(y_{j,t}=r_{j,t})} \quad (4)$$

Interpretacja modelu jest związana z oceną interwału pomiaru, charakteru wskaźników (typ pytań kwestionariuszowych, słownictwo itp.) Szczególne znaczenie dla interpretacji parametrów ma czas i interwał obserwacji. W przypadku długich przerw między okresami pomiaru i jednocześnie silnych i zmiennych przejść między statusami można domniemywać występowanie braków danych w okresach pomiaru lub istnienie istotnych czynników moderujących prawdopodobieństwa przejścia.

W budowie modeli LTA z wykorzystaniem jakościowych wskaźników zmiennych ukrytych występuje problem inwariancji pomiaru. Przyjmuje się założenie o niezmienniczości narzędzia pomiarowego w czasie (*measurement invariance*). Założenie to prowadzi do budowy modeli LTA z ograniczonymi parametrami odpowiadania na pozycje wynikającymi z przyjęcia równości wartości prawdopodobieństw odpowiedzi na poszczególne pozycje w czasie. Ograniczenie to powoduje łatwiejszą interpretację modelu oraz możliwość porównywalnej interpretacji znaczenia ukrytych statusów w czasie i tym samym jednoznaczną interpretację prawdopodobieństw klas ukrytych (samo znaczenie ukrytych statusów jest takie same w poszczególnych falach badania). Również interpretacja prawdopodobieństw przejść jest taka sama dla kolejnych okresów pomiaru. W sytuacji braku

inwariancji pomiarowej interpretacja prawdopodobieństw przejścia musiałaby uwzględniać zmianę w czasie również i samych znaczeń ukrytych statusów identyfikowanych na podstawie warunkowych prawdopodobieństw odpowiadania na poszczególne wskaźniki tych statusów. Model ograniczony (z dużo mniejszą liczbą parametrów) pozwala również na uzyskanie bardziej stabilnych oszacowań parametrów i zmniejsza prawdopodobieństwo braku identyfikacji i uzyskania osobliwych rozwiązań [Chung i Walls 2007].

4. Estymacja modelu LTA w programie Mplus

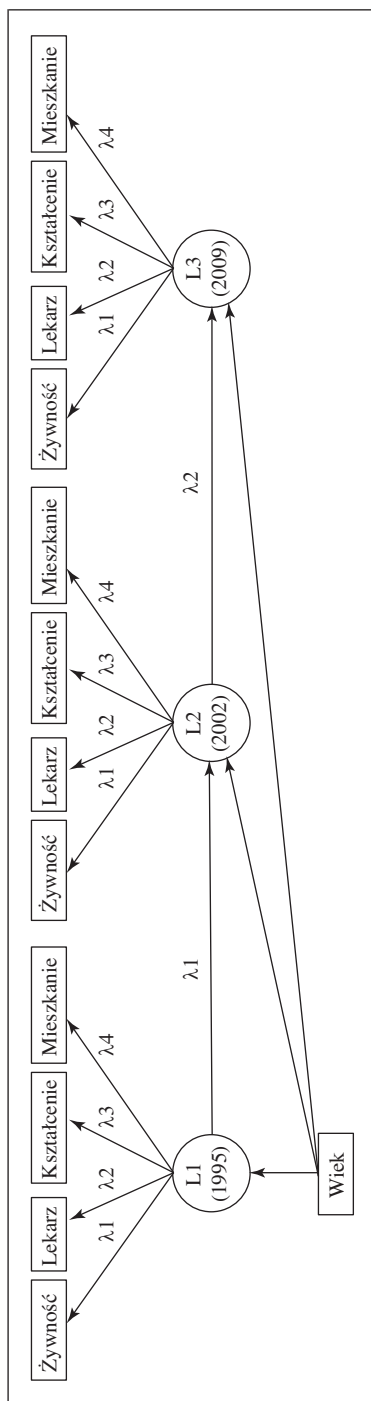
Ilustracją modelu LTA jest model ukrytych przejść zbudowany na podstawie danych z Polskiego Generalnego Sondażu Społecznego z lat 1995–2008. Baza danych obejmuje 10 fal badań w latach 1992–2008 obejmujących w sumie 16 234 respondentów i 975 zmiennych obejmujących ocenę problemów społecznych i ekonomicznych przez dorosłych mieszkańców Polski. Z bazy danych wyodrębnione zostały cztery pytania wskaźnikowe dotyczące kondycji ekonomicznej obejmujące:

- brak pieniędzy na żywność i ubranie (zmienna „żywność”),
- brak pieniędzy na lekarza (zmienna „lekarz”),
- brak pieniędzy na kształcenie (zmienna „kształcenie”),
- brak pieniędzy na świadczenia mieszkaniowe (zmienna „mieszkanie”).

Dla wszystkich pytań występowały binarne kategorie odpowiedzi („tak” – „nie”)².

Ponieważ PGSS ma postać powtarzanych sondaży prowadzonych na próbach niezależnych, w celu uzyskania efektu pomiaru panelowego dokonano w pierwszym etapie agregacji przypadków w mikrosegmenty *a priori*. Jako kryterium mikrosegmentacji przyjęto następujące zmienne: 1) płeć, 2) wiek (w latach), 3) liczba osób w gospodarstwie domowym, 4) miejsce zamieszkania. Czynniki te kształtują profile gospodarstw domowych w Polsce. Krzyżując powyższe kombinacje cech społeczno-demograficznych, uzyskano 4981 mikrosegmentów (pozostałe stanowiły segmenty puste). Uwzględniając zmienne opisujące segmenty, można (w uproszczeniu) przyjąć, że stanowią one reprezentację gospodarstw domowych w Polsce. Dane z 10 fal badań zostały przekodowane na trzy fale: 1) dane 1992–1995, 2) dane 1997–2002 i 3) dane 2005–2009. Porównując wzorce odpowiedzi na powyższe cztery pytania, wśród 4981 jednostek analizy uzyskano możliwość porównania fal badań w ujęciu panelowym (porównanie wzorców odpowiedzi wśród tych samych mikrosegmentów).

² Dane można pobrać ze strony Polskiego Generalnego Sondażu Społecznego: <http://pgss.iss.uw.edu.pl/>.



Rys. 1. Model ukrytych przejść z ograniczonymi parametrami

Źródło: opracowanie własne.

W estymacji modelu wykorzystana została metoda odporna największej wiarygodności (*robust maximum likelihood*) z parametryzacją probabilistyczną (zamiast logitowej). Pozwala ona na bezpośrednie szacowanie prawdopodobieństw przejść³.

Model ukrytych przejść z kowariantą jest zaprezentowany na rys. 1. W modelu tym prawdopodobieństwa przejść są wyjaśniane za pomocą zmiennej wiek „głowy rodziny”. Rys. 1 przedstawia układ trzech kategoryalnych zmiennych ukrytych (L1–L3). Każda ze zmiennych ukrytych składa się z trzech klas ukrytych⁴. Daje to w sumie $3 \times 3 \times 3 = 27$ różnych kombinacji zmiennych i klas w przekroju fal pomiaru. Zakładając inwariancję pomiaru, model ten został estymowany z ograniczonymi parametrami zakładającymi takie samo znaczenie klas ukrytych w przekroju poszczególnych fal badania. Wskaźniki dopasowania modelu przedstawione są w tabeli 1.

Tabela 1. Wskaźniki dopasowania modelu

Wskaźnik	Wartość
Liczba wolnych parametrów	34
Czynnik skalujący	0,9615
Kryterium Akaike	40911,217
Kryterium Bayesowskie	41132,632
Statystyka chi-kwadrat dla wskaźników kategoryalnych	4146,287, (s.s. = 4048, poziom $p = 0,1376$)

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Wskaźniki sugerują również odpowiednie dopasowanie modelu. Liczebności klas i względne proporcje ukrytych statusów dla poszczególnych zmiennych ukrytych w tym modelu przedstawiono w tabeli 2.

Tabela 2. Wzory ukrytych klas (statusów)

Wzór klasy ukrytej	Liczebność oszacowana	Frakcja
1 1 1	1415,8	0,28
1 1 2	727,3	0,14

³ W parametryzacji logitowej prawdopodobieństwa przejść są obliczane na podstawie estymowanych szans.

⁴ Liczba ukrytych statusów została wyodrębniona na podstawie wielu modeli klasycznej analizy klas ukrytych, z których jako najlepszy model wybrano model o optymalnej relacji między dopasowaniem (entropią) a prostotą (oszczędnością). Syntetycznymi wskaźnikami informacyjnymi, które wykorzystano przy wyborze modelu były wskaźniki informacyjne Akaike (AIC) oraz bayesowskie kryterium informacyjne (BIC).

cd. tabeli 2

Wzór klasy ukrytej	Liczebność oszacowana	Fracja
1 1 3	249,2	0,05
1 2 1	0,0	0,00
1 2 2	17,5	0,00
1 2 3	258,1	0,05
1 3 1	0,0	0,00
1 3 2	602,7	0,12
1 3 3	155,5	0,03
2 1 1	149,2	0,02
2 1 2	0,0	0,00
2 1 3	0,0	0,00
2 2 1	0,0	0,00
2 2 2	171,0	0,03
2 2 3	95,0	0,01
2 3 1	0,00	0,00
2 3 2	160,2	0,03
2 3 3	0,3	0,00
3 1 1	431,9	0,08
3 1 2	0,0	0,00
3 1 3	0,0	0,00
3 2 1	0,0	0,00
3 2 2	231,3	0,04
3 2 3	128,8	0,02
3 3 1	0,0	0,00
3 3 2	185,4	0,03
3 3 3	0,3	0,00

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Struktura liczebności wskazuje, że dominująca liczba przypadków należy do pierwszej klasy ukrytej w przekroju wszystkich trzech zmiennych ukrytych (1, 1, 1) oraz pierwszej klasy dla zmiennej L1 i L2 oraz drugiej klasy dla zmiennej L3 (1, 1, 2). W modelu tym występują już puste klasy ukryte (1, 2, 1). Liczebności i proporcje poszczególnych statusów ukrytych przedstawiono w tabeli 3.

Tabela 3. Proporcje ukrytych statusów

Zmienna ukryta	Status ukryty	Liczebność	Frakcja
C1	1	3426,56	0,688
	2	576,04	0,115
	3	978,39	0,196
C2	1	2973,66	0,597
	2	902,53	0,181
	3	1104,80	0,221
C3	1	1996,98	0,400
	2	2096,33	0,420
	3	887,68	0,178

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Z tabeli 3 wynika, że najliczniejszą grupę przypadków stanowią przedstawiciele pierwszych klas ukrytych w poszczególnych falach badań. Najmniej liczna jest trzecia klasa ukryta w przekroju zmiennych.

5. Profilowanie ukrytych statusów

Prawdopodobieństwa odpowiedzi dla poszczególnych kategorii w klasach ukrytych (z ograniczonymi parametrami) są podane poniżej (podano jedynie odpowiedzi „tak”).

Tabela 4. Profil pierwszej klasy ukrytej

Pierwsza klasa ukryta	Parametr	Błąd standardowy	Iloraz krytyczny	Poziom p
Żywność	0,038	0,003	12,129	0,00
Lekarz	0,010	0,003	3,460	0,00
Kształcenie	0,036	0,002	15,391	0,00
Mieszkanie	0,031	0,002	13,835	0,00

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Pierwszą klasę ukrytą stanowią przypadki nieujawniające braku pieniędzy na realizację podstawowych potrzeb życiowych. Są to gospodarstwa „zamożne”⁵.

⁵ Etykiety klas ukrytych mają charakter względny. Brak problemów w zaspokajaniu podstawowych potrzeb życiowych (żywność, opieka zdrowotna, kształcenie i mieszkanie) nie jest wskaźnikiem zamożności społeczeństwa, a jedynie zachowywania *status quo*. Jednakże

Tabela 5. Profil drugiej klasy ukrytej

Druga klasa ukryta	Parametr	Błąd standardowy	Iloraz krytyczny	Poziom p
Żywność	0,027	0,006	4,973	0,00
Lekarz	0,322	0,011	30,053	0,00
Kształcenie	0,024	0,004	6,875	0,00
Mieszkanie	0,011	0,003	3,623	0,00

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Do drugiej klasy ukrytej należą przypadki, które wskazują na względnie zauważalne problemy finansowe w obszarze wydatków na leczenie. W pozostałych obszarach nie brakuje im pieniędzy na zaspokajanie potrzeb życiowych. Są to gospodarstwa „względnie zamożne”.

Tabela 6. Profil trzeciej klasy ukrytej

Trzecia klasa ukryta	Parametr	Błąd standardowy	Iloraz krytyczny	Poziom p
Żywność	0,916	0,009	106,629	0,00
Lekarz	0,754	0,010	73,556	0,00
Kształcenie	0,279	0,009	31,935	0,00
Mieszkanie	0,519	0,010	51,451	0,00

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Trzecią klasę ukrytą stanowią przypadki, które cechują się wysokim prawdopodobieństwem braku pieniędzy na zaspokojenie potrzeb życiowych (z wyjątkiem wydatków na kształcenie) Są to gospodarstwa „biedne”.

Ocena dynamiki przejść pomiędzy stanami ukrytymi w czasie jest przedstawiona za pomocą prawdopodobieństw warunkowych w przekrojach lat 1992–1995 do 2005–2009 (tabele 7 i 8).

Analiza regresji czynników dyskretnych (ukrytych statusów) z uwzględnieniem wpływu zmiennej kowariancyjnej wiek wskazuje na zależność między statusami ukrytymi w pierwszej fali pomiaru na statusy występujące w drugiej fali oraz wpływ tych ostatnich na statusy w trzeciej fali badań.

w przypadku kraju tak biednego, jak Polska, w okresie transformacji lat 1995–2009, gospodarstwa domowe, które były w stanie na podstawowym poziomie realizować potrzeby fizjologiczne mogły uważać się za „zamożne”.

Tabela 7. Macierz przejść pierwszego okresu

1992–1995 względem 1997–2002	1	2	3
1	0,698	0,080	0,221
2	0,259	0,462	0,279
3	0,441	0,369	0,190

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Tabela 8. Macierz przejść drugiego okresu

1997–2002 względem 2005–2009	1	2	3
1	0,672	0,245	0,084
2	0,000	0,466	0,534
3	0,000	0,859	0,141

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

W przedstawionych analizach wpływ wieku na przynależność do określonych statusów jest najsilniejszy w przypadku przynależności do klasy gospodarstw „zamożnych” (w stosunku do „biednych”) w drugiej i trzeciej fali badań. Dla pierwszej fali badań (1992–1995) współczynnik regresji odzwierciedla zmianę w szansie przynależności do klasy ukrytej w relacji do klasy odniesienia związaną z jednostkowym wzrostem wartości zmiennej kowariancyjnej.

Dla kolejnych fal badań (1997–2002 i 2005–2009) współczynniki regresji odzwierciedlają warunkową zmianę w szansie przejścia z ukrytego statusu S_t do S_{t+1} w relacji do statusu odniesienia (pod warunkiem przynależności do statusu S_t) wraz z jednostkowym wzrostem wartości zmiennej kowariancyjnej. Dla drugiej fali wraz z wzrostem wieku głowy gospodarstwa domowego o rok następuje spadek przynależności do klasy „zamożnych” w stosunku do „biednych” o 49% (iloraz szans wynosi $e^{-0,49} = 0,61$). W ostatniej fali pomiaru wpływ wieku daje efekt netto 43% (*i.s.* wynosi $e^{-0,56} = 0,57$). Dla pozostałych klas i fal badań wpływ wieku na prawdopodobieństwa przynależności do ukrytych statusów jest niewielki (ilorazy szans znajdują się na poziomie 1,0).

Na podstawie modelu można oszacować warunkowe prawdopodobieństwa dla ukrytych statusów i przejść dla określonego poziomu wartości zmiennej kowariancyjnej (wiek). W tabeli 6 przedstawione zostały one dla dwóch poziomów wieku głowy rodziny – 25 lat (młode małżeństwa) i 50 lat (dojrzałe małżeństwa z samodzielnymi dziećmi).

Tabela 9. Analiza regresji statusów i predyktorów

Statusy ukryte	Predyktory	Parametr	Błąd standardowy	Iloraz krytyczny	Poziom p
Zamożni 1997–2002	Zamożni 1992–1995	-4,328	0,578	-7,556	0,00
	Średnio zamożni 1992–1995	16,001	8,917	1,794	0,07
Średnio zamożni 1997–2002	Zamożni 1992–1995	2,086	0,709	2,942	0,00
	Średnio zamożni 1992–1995	-0,053	1,182	-0,045	0,96
Zamożni 2005–2009	Zamożni 97–92	62,869	0,000	–	–
	Średnio zamożni 97–92	-166,984	0,000	–	–
Średnio zamożni (2005–2009)	Zamożni 97–92	0,899	0,227	3,958	0,000
	Średnio zamożni 97–92	-5,555	0,859	-6,467	0,000
Zamożni 1997–2002	Wiek	-0,003	0,002	-1,166	0,243
Średnio zamożni (1992–1995)		0,025	0,003	7,806	0,000
Zamożni (1997–2002)		-0,494	0,228	-2,164	0,030
Średnio zamożni (1997–2002)		0,057	0,015	3,837	0,000
Zamożni (2005–2009)		-0,560	0,068	-8,244	0,000
Średnio zamożni (2005–2009)		-0,089	0,014	-6,454	0,000

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Tabela 10. Warunkowe prawdopodobieństwa przynależności do klas ukrytych w zależności od wieku

Wiek 25	Wiek 50
$P(C1 = 1) = 0,741$	$P(C1 = 1) = 0,691$
$P(C1 = 2) = 0,059$	$P(C1 = 2) = 0,111$
$P(C1 = 3) = 0,200$	$P(C1 = 3) = 0,199$
$P(C2 = 1 C1 = 1) = 0,698$	$P(C2 = 1 C1 = 1) = 0,698$
$P(C2 = 2 C1 = 1) = 0,080$	$P(C2 = 2 C1 = 1) = 0,080$
$P(C2 = 3 C1 = 1) = 0,221$	$P(C2 = 3 C1 = 1) = 0,221$

cd. tabeli 10

Wiek 25	Wiek 50
$P(C2 = 1 C1 = 2) = 1,000$	$P(C2 = 1 C1 = 2) = 0,022$
$P(C2 = 2 C1 = 2) = 0,000$	$P(C2 = 2 C1 = 2) = 0,419$
$P(C2 = 3 C1 = 2) = 0,000$	$P(C2 = 3 C1 = 2) = 0,559$
$P(C2 = 1 C1 = 3) = 0,920$	$P(C2 = 1 C1 = 3) = 0,277$
$P(C2 = 2 C1 = 3) = 0,015$	$P(C2 = 2 C1 = 3) = 0,399$
$P(C2 = 3 C1 = 3) = 0,064$	$P(C2 = 3 C1 = 3) = 0,324$
$P(C3 = 1 C1 = 1, C2 = 1) = 1,000$	$P(C3 = 1 C1 = 1, C2 = 1) = 0,512$
$P(C3 = 2 C1 = 1, C2 = 1) = 0,000$	$P(C3 = 2 C1 = 1, C2 = 1) = 0,454$
$P(C3 = 3 C1 = 1, C2 = 1) = 0,000$	$P(C3 = 3 C1 = 1, C2 = 1) = 0,034$
$P(C3 = 1 C1 = 1, C2 = 2) = 0,000$	$P(C3 = 1 C1 = 1, C2 = 2) = 0,000$
$P(C3 = 2 C1 = 1, C2 = 2) = 0,164$	$P(C3 = 2 C1 = 1, C2 = 2) = 0,021$
$P(C3 = 3 C1 = 1, C2 = 2) = 0,836$	$P(C3 = 3 C1 = 1, C2 = 2) = 0,979$
$P(C3 = 1 C1 = 1, C2 = 3) = 0,000$	$P(C3 = 1 C1 = 1, C2 = 3) = 0,000$
$P(C3 = 2 C1 = 1, C2 = 3) = 0,981$	$P(C3 = 2 C1 = 1, C2 = 3) = 0,846$
$P(C3 = 3 C1 = 1, C2 = 3) = 0,019$	$P(C3 = 3 C1 = 1, C2 = 3) = 0,154$
$P(C3 = 1 C1 = 2, C2 = 1) = 1,000$	$P(C3 = 1 C1 = 2, C2 = 1) = 1,000$
$P(C3 = 2 C1 = 2, C2 = 1) = 0,000$	$P(C3 = 2 C1 = 2, C2 = 1) = 0,000$
$P(C3 = 3 C1 = 2, C2 = 1) = 0,000$	$P(C3 = 3 C1 = 2, C2 = 1) = 0,000$
$P(C3 = 1 C1 = 2, C2 = 2) = 0,000$	$P(C3 = 1 C1 = 2, C2 = 2) = 0,000$
$P(C3 = 2 C1 = 2, C2 = 2) = 0,643$	$P(C3 = 2 C1 = 2, C2 = 2) = 0,643$
$P(C3 = 3 C1 = 2, C2 = 2) = 0,357$	$P(C3 = 3 C1 = 2, C2 = 2) = 0,357$
$P(C3 = 1 C1 = 2, C2 = 3) = 0,000$	$P(C3 = 1 C1 = 2, C2 = 3) = 0,000$
$P(C3 = 2 C1 = 2, C2 = 3) = 0,998$	$P(C3 = 2 C1 = 2, C2 = 3) = 0,998$
$P(C3 = 3 C1 = 2, C2 = 3) = 0,002$	$P(C3 = 3 C1 = 2, C2 = 3) = 0,002$
$P(C3 = 1 C1 = 3, C2 = 1) = 1,000$	$P(C3 = 1 C1 = 3, C2 = 1) = 1,000$
$P(C3 = 2 C1 = 3, C2 = 1) = 0,000$	$P(C3 = 2 C1 = 3, C2 = 1) = 0,000$
$P(C3 = 3 C1 = 3, C2 = 1) = 0,000$	$P(C3 = 3 C1 = 3, C2 = 1) = 0,000$
$P(C3 = 1 C1 = 3, C2 = 2) = 0,000$	$P(C3 = 1 C1 = 3, C2 = 2) = 0,000$
$P(C3 = 2 C1 = 3, C2 = 2) = 0,643$	$P(C3 = 2 C1 = 3, C2 = 2) = 0,643$
$P(C3 = 3 C1 = 3, C2 = 2) = 0,357$	$P(C3 = 3 C1 = 3, C2 = 2) = 0,357$
$P(C3 = 1 C1 = 3, C2 = 3) = 0,000$	$P(C3 = 1 C1 = 3, C2 = 3) = 0,000$
$P(C3 = 2 C1 = 3, C2 = 3) = 0,998$	$P(C3 = 2 C1 = 3, C2 = 3) = 0,998$
$P(C3 = 3 C1 = 3, C2 = 3) = 0,002$	$P(C3 = 3 C1 = 3, C2 = 3) = 0,002$

Źródło: opracowanie własne na podstawie obliczeń w programie Mplus 7.

Dla przykładu, „młode małżeństwa” z pierwszej fali badań należą do pierwszej klasy ukrytej („zamożni”) z prawdopodobieństwem 0,741, a „dojrzałe małżeństwa” należą do niej z prawdopodobieństwem 0,691. Szansa przejścia „młodych małżeństw” z klasy „średnio zamożnych” z lat 1992–1995, do klasy „zamożnych” w latach 1997–2002 wynosi 1,00, a szansa takiego przejścia dla „dojrzałych małżeństw” wynosi tylko 0,022. Prawdopodobieństwo, że „młode małżeństwa” w latach 2005–2009 znajdą się w klasie „zamożnych”, jeżeli w okresie 1992–1995 byli w klasie „zamożnych” oraz w latach 1997–2002 również byli w tej klasie wynosi 1,00. Prawdopodobieństwo takiego przejścia dla „małżeństw dojrzałych” wynosi 0,512. W ostatnim wyróżnionym w tabeli przykładzie prawdopodobieństwo przejścia „młodych małżeństw” z klasy „średnio zamożnych” w latach 1992–1995 do statusu „biednych” w okresie 1997–2002 wynosi 0,00, lecz dla małżeństw „dojrzałych” prawdopodobieństwo to wynosi już 0,559.

6. Zakończenie

Modele ukrytych przejść stanowią istotne wzbogacenia tradycyjnych analiz segmentacyjnych rynku. Pozwalają na wyodrębnienie segmentów rynkowych w układzie longitudinalnym oraz określenie prawdopodobieństw przejść z jednego stanu do drugiego.

Analiza ukrytych przejść może być również wzbogacona o uwzględnienie przynależności do segmentów *a priori* (np. regionów geograficznych). W takich sytuacjach modele LTA są przekształcane w wielogrupowe modele LTA z uwzględnieniem znanych klas mających charakter jawny. Inne istotne zastosowanie tych modeli w analizie rynku stanowi połączenie ukrytych przejść z kontrolą efektów kohortowych i identyfikacja prawdopodobieństw przejścia między segmentami w przekroju kohort wiekowych. W przedstawionej analizie kontroli podlegał tylko efekt wieku, wprowadzenie informacji o roku urodzenia umożliwiłoby dokonanie ocen prawdopodobieństw przejść w poszczególnych kohortach wiekowych. Obie te analizy można łączyć w celu identyfikacji układów przejść międzysegmentowych dla kohort wiekowych w przekrojach kulturowych czy regionalnych.

Literatura

- Collins L.M., Lanza S.T. [2010], *Latent Class and Latent Transition Analysis*, Wiley, Hoboken.
- Chung W., Walls T.A. [2007], *A Latent Transition Model with Logistic Regression*, „Psychometrika” vol. 72, nr 3.

- Dias J.G., Vermunt J.K. [2007], *Latent Class Modeling of Website Users' Search Patterns: Implications for Online Market Segmentation*, „Journal of Retailing and Consumer Services”, vol. 14, nr 6.
- Everitt B. [2006], *The Cambridge Dictionary of Statistics*, Cambridge University Press, Cambridge.
- Frazer A.M. [2011], *Hidden Markov Models and Dynamic Systems*, Society for Industrial and Applied Mathematics, Philadelphia.
- Guo B. *et al.* [2009], *Using Latent Class and Latent Transition Analysis to Examine the Transtheoretical Model Staging Algorithm and Sequential Stage Transition in Adolescent Smoking*, „Substance Use & Misuse”, vol. 44, nr 14.
- Poulsen C.S. [1990], *Mixed Markov and Latent Markov Modeling Applied to Brand Choice Behavior*, „International Journal of Research in Marketing”, vol. 7, nr 11.
- Ramaswamy V. [1997], *Evolutionary Preference Segmentation with Panel Survey Data: An Application to New Products*, „International Journal of Research in Marketing”, vol. 14, nr 1.
- Wedel M., Kamakura W.A. [2000], *Market Segmentation. Conceptual and Methodological Foundation*, Kluwer Academic, Boston.

Mplus in Latent Transition Model Estimations in Market Segmentation

The article is devoted to the applications of latent transition models (LTA) in dynamic market segmentation. A generalisation of the latent class models, these models take into account the dynamic nature of the data. They allow for the identification of unobserved market segments, the assessment of conditional probabilities and transition patterns within latent statuses in time. LTA model was used to evaluate the dynamic segments in three separate waves of research based on economic status indicators in data from the Polish General Social Survey.

Keywords: latent transition model, market segmentation, hidden Markov chains, Mplus software.