DYNAMICZNE MODELE EKONOMETRYCZNE
X Ogólnopolskie Seminarium Naukowe, 4–6 września 2007 w Toruniu
Katedra Ekonometrii i Statystyki, Uniwersytet Mikołaja Kopernika w Toruniu

Katarzyna Osiecka
Politechnika Warszawska
Józef Stawicki
Uniwersytet Mikołaja Kopernika w Toruniu

Markov Set - Chains jako narzędzie analizy zmian struktury wydatków gospodarstw domowych w Polsce w latach 1993-2005

Model łańcucha Markowa jest opisem procesu stochastycznego w dyskretnej przestrzeni stanów oraz w dyskretnym czasie. Podstawowa zależność określająca rozkład bezwarunkowy w chwili $t+1$ ma postać:

$$\bar{x}_{t+1} = \bar{x}_t \cdot P_t ,$$

gdzie $\bar{x}_t$ jest wektorem rozkładu bezwarunkowego, a P_t - macierzą prawdopodobieństw warunkowych w chwili t.

Jeśli dla każdego t macierz $P_t = P$ to proces nosi nazwę jednorodnego łańcucha Markowa. Macierz prawdopodobieństw warunkowych może zależeć od pewnych czynników w chwili t i wówczas podstawową zależność można przedstawić w postaci:

$$\bar{x}_{t+1} = \bar{x}_t \cdot P(\bar{z}_t) ,$$

gdzie $\bar{z}_t$ jest wektorem obserwowanych czynników.

W niniejszym referacie łańcuch Markowa stanowi narzędzie opisu zmiany struktury wydatków gospodarstw domowych rozumianej jako wektor:

$$\bar{x}_t = (x_{t1}, x_{t2}, \cdots, x_{tr}),$$

w którym x_{ti} określa udział wydatków w roku t na określoną i – tą grupę dóbr w całości wydatków gospodarstwa domowego spełniającego warunki:

$$\underset{t}{\exists}\underset{i}{\exists}\, x_{ti} \geq 0 \quad \text{oraz} \quad \underset{t}{\exists} \sum_{i=1}^{r} x_{ti} = 1 .$$

W tym przypadku zjawisko nie ma charakteru procesu stochastycznego jednowymiarowego, realizowanego w dyskretnej przestrzeni stanów. Można by traktować wektor $\overline{x}_t$ jako rozkład prawdopodobieństwa wydatkowania przez określone gospodarstwo domowe losowo wybranej jednostki pieniężnej na produkty z jednej z grup towarowych. Obserwacja dla pojedynczego gospodarstwa domowego mogłoby stanowić realizację wspomnianego procesu. W przypadku obserwowanych struktur wydatków dla ogółu gospodarstw domowych bądź wybranych grup ze względu na pewne kryterium (np. liczebności gospodarstwa domowego, źródła dochodów) prowadzi do posługiwania się średnim gospodarstwem bądź lepiej – typowym gospodarstwem domowym. Brak jasnej interpretacji procesu nie przeszkadza jednak w wykorzystaniu narzędzia, jakim są łańcuchy Markowa do badania zmian struktury. Proste miary podobieństwa czy zróżnicowania nie pozwalają na wyznaczenie prognoz. Łańcuchy Markowa pozwalają uzyskać prognozę struktury.

Przedziałowe łańcuchy Markowa są propozycją sprowadzenia poszukiwania niejednorodności macierzy przejść do wykorzystania przedziału macierzowego, z którego pochodzą stochastyczne macierze. Modele oparte o koncepcję Markov-Set Chains wprowadził D. J. Hartfiel[1]. Podstawą wspomnianej wyżej teorii jest przyjęcie założenia, że macierz przejścia w każdym kroku pochodzi ze zbioru macierzy określonego poprzez przedziały w jakich mogą zmieniać się poszczególne jej elementy. Struktura wyjściowa jest określona jako przedział wektorowy; struktury w kolejnych okresach są zbiorami wypukłymi i mogą być przybliżone przedziałami wektorowymi.

Przez stochastyczny przedział wektorowy $[\overline{\mathbf{p}}, \overline{\mathbf{q}}]$ w przestrzeni R^n, gdzie $\overline{\mathbf{q}} \geq \overline{\mathbf{p}} \geq \overline{\mathbf{0}}$ rozumie się zbiór wektorów:

$$[\overline{\mathbf{p}}, \overline{\mathbf{q}}] = \{\overline{\mathbf{x}} \in R^n ; \overline{\mathbf{p}} \leq \overline{\mathbf{x}} \leq \overline{\mathbf{q}}\},$$

przy czym $\sum_i x_i = 1$.

Stochastyczny przedział wektorowy jest wielościanem wypukłym. W szczególnym przypadku, gdy $\overline{\mathbf{q}} = \overline{\mathbf{p}}$ przedział wektorowy sprowadza się do punktu. W przypadku wektorów z przestrzeni R^2 jest to odcinek. Przedział

[1] Teoria ta została przedstawiona w pracy Hartfiela (1999). W Polsce zagadnieniami tymi zajmowali się A. Decewicz i A. Gyczew (2001). Autorzy ci nie wprowadzili polskojęzycznego odpowiednika Markov-Set Chains. Termin przedziałowe łańcuchy Markowa Stawicki (2002). Wcześniej terminem „łańcuch przedziałami Markowa” posługiwał się w swoich pracach A. Tabeau. Słowo „przedział” odnosiło się jednak do osi czasu.

wektorowy w przestrzeni R^3 jest wielokątem wypukłym o maksymalnie sześciu kątach. Może on być zredukowany do odcinka lub w szczególnym przypadku do punktu. Każdy stochastyczny przedział wektorowy może być przedstawiony w postaci tak zwanego przedziału $[\overline{\mathbf{p}}, \overline{\mathbf{q}}]$ napiętego. Przedział napięty to taki przedział wektorowy, w którym wszystkie współrzędne wektora $\overline{\mathbf{p}}$ oraz wektora $\overline{\mathbf{q}}$ są napięte. Dla współrzędnych tych wektorów spełnione są następujące warunki:

$$p_i = \min_{\bar{x}\in[\bar{p},\bar{q}]} x_i ,$$

$$q_j = \max_{\bar{x}\in[\bar{p},\bar{q}]} x_j .$$

Dowolnemu przedziałowi stochastycznemu odpowiada równoważny mu przedział stochastyczny napięty.

W podobny sposób określa się stochastyczne przedziały macierzowe $[\mathbf{P}, \mathbf{Q}]$, gdzie $\mathbf{P} = \lfloor p_{ij} \rfloor$, $\mathbf{Q} = \lfloor q_{ij} \rfloor$ są macierzami kwadratowymi stopnia n oraz $\mathbf{Q} \geq \mathbf{P} \geq \mathbf{0}$. Stochastycznym przedziałem macierzowym nazywa się zbiór macierzy:

$$[\mathbf{P},\mathbf{Q}] = \{\mathbf{A}; \mathbf{P} \leq \mathbf{A} \leq \mathbf{Q}\},$$

gdzie $\mathbf{A} = \lfloor a_{ij} \rfloor$ jest macierzą stochastyczną.

Dla stochastycznych przedziałów macierzowych określa się w podobny sposób jak dla przedziałów wektorowych napięcie przedziału.

Przez **przedziałowy łańcuch Markowa** rozumie się ciąg zbiorów macierzowych:

$$\mathbf{M}, \mathbf{M}^2, \mathbf{M}^3, ..., \mathbf{M}^k, ...$$

gdzie $\mathbf{M}$ jest zbiorem macierzy stochastycznych, natomiast $\mathbf{M}^k$ jest zbiorem określonym następująco:

$$\mathbf{M}^k = \{\mathbf{A};\ \mathbf{A} = \mathbf{A}_1 \cdot \mathbf{A}_2 \cdot ... \cdot \mathbf{A}_k\},\ \text{gdzie}\ \underset{i}{\forall}\ \mathbf{A}_i \in \mathbf{M} .$$

Symbolicznie możemy to zapisać następującą formułę:

$$\mathbf{S}_1 = \mathbf{S}_0 \cdot \mathbf{M},$$

gdzie $\mathbf{S}_1$ oznacza zbiór wektorów stochastycznych otrzymanych z dowolnego wektora stochastycznego ze zbioru początkowego $\mathbf{S}_0$ pomnożonego przez dowolną macierz stochastyczną ze zbioru $\mathbf{M}$.

Przedziałowy łańcuch Markowa jest zatem niejednorodnym łańcuchem, w którym macierze prawdopodobieństw przejść w kolejnych krokach są dowolnymi macierzami ze zbioru $\mathbf{M}$. W szczególności zbiorem $\mathbf{M}$ może być stochastyczny przedział macierzowy. Twierdzenia dotyczące tak rozumianych przedziałowych łańcuchów Markowa pozwalają określić zbiory $\mathbf{S}_1, \mathbf{S}_2, \mathbf{S}_3, \ldots$ w kolejnych okresach, jeśli znany jest początkowy stochastyczny przedział wektorowy $\mathbf{S}_0$ i zbiór macierzy $\mathbf{M}$ w postaci przedziału macierzowego. Proponowane są trzy metody wyznaczania zbioru $\mathbf{S}_1$: metoda wyznaczania wierzchołków, metoda Hi-Lo oraz symulacyjna metoda Monte Carlo[2]. Metoda Hi-Lo pozwala w każdym kroku określić przedział wektorowy.

Praktyczna strona zagadnienia sprowadza się do określenia zbioru macierzy $\mathbf{M}$ w postaci przedziału macierzowego $[\mathbf{P}, \mathbf{Q}]$, oraz przedziału wektorowego $\mathbf{S}_0$. Jako przedział $\mathbf{S}_0$ można przyjąć wektor struktury charakteryzujący określone gospodarstwo domowe. Pozwoli to na prognozę struktury w kolejnych latach w postaci przedziału wektorowego.

W omawianym zagadnieniu analizy struktury wydatków gospodarstw domowych wyodrębniono cztery grupy wydatków:

1. wydatki na żywność,
2. wydatki na artykuły nieżywnościowe (w tym: odzież i obuwie, mieszkanie, higiena osobista),
3. wydatki na usługi (w tym: transport i łączność, rekreacja i kultura, edukacja, zdrowie),
4. pozostałe wydatki.

Struktura wydatków konsumpcyjnych gospodarstw domowych zależy od wielu czynników o różnym charakterze. Na podstawie danych z budżetów gospodarstw domowych bada się strukturę wydatków gospodarstw między innymi według grup społeczno-ekonomicznych, do których te gospodarstwa należą, ale również według grup dochodowych czy liczby osób w rodzinie.

W latach 90-tych miały miejsce istotne zmiany w strukturze wydatków gospodarstw domowych. Tabela 1 zawiera dane dotyczące ogólnej struktury wydatków gospodarstw domowych z podziałem na 4 grupy wydatków. Największy udział w ogóle wydatków miały wydatki na zaspokojenie potrzeb podstawowych takich jak żywność. Chociaż udział wydatków na żywność był najwyższy, to jednak w badanym okresie, czyli w latach 1993-2005, systematycznie się zmniejszał. Przyczyną tego była podwyżka cen podstawowych towarów, a także komercjalizacja usług w sferze utrzymania mieszkania, ochrony zdrowia, kształcenia oraz zwolnienie tempa wzrostu cen żywności w ostatnich latach. Na skutek wzrostu obciążenia budżetów gospodarstw domowych stałymi opłatami za użytkowanie mieszkania w tym podwyżka czynszu oraz cen nośników energii, następował systematyczny, powolny wzrost udziału wydatków na mieszkanie, a co za tym idzie wzrost udziału wydatków na artykuły nieżywnościowe w

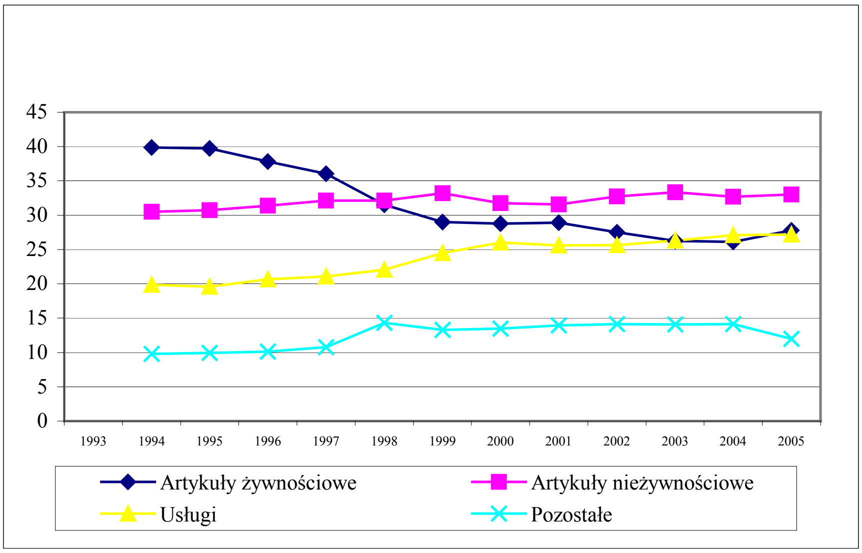
[2] Charakterystykę metody wierzchołków oraz metody Hi-Lo przedstawić Hartfiel (op.cit), Metodę symulacyjną jak również wcześniej wspomniane przedstawiła Samuels (2001).

ogóle wydatków gospodarstw. Również udział wydatków na usługi (takie jak: zdrowie oraz transport i łączność) ulegał ciągłemu wzrostowi.

Tabela 1. Struktura wydatków gospodarstw domowych w wydatkach ogółem w Polsce w latach 1993 – 2005 (w %)

Rok	Artykuły żywno-ściowe	Artykuły nieżywno-ściowe	Usługi	Pozostałe
1993	41,50	29,90	19,24	9,36
1994	39,86	30,50	19,85	9,80
1995	39,71	30,75	19,60	9,95
1996	37,80	31,39	20,68	10,13
1997	36,04	32,13	21,09	10,75
1998	31,53	32,11	22,06	14,30
1999	29,02	33,20	24,46	13,31
2000	28,75	31,77	26,01	13,46
2001	28,89	31,57	25,60	13,94
2002	27,49	32,73	25,63	14,15
2003	26,24	33,36	26,32	14,09
2004	26,11	32,67	27,08	14,14
2005	27,80	33,00	27,20	12,00

Źródło: Opracowanie własne na podstawie danych GUS.



Rys.1. Struktura wydatków gospodarstw domowych w Polsce w latach 1993–2005 (udział wydatków w %)
Źródło: Opracowanie własne na podstawie Tabeli 1.

W dalszej części posłużono się obserwacjami dotyczącymi gospodarstw domowych z podziałem na gospodarstwa pracownicze, gospodarstwa pracownicze użytkujących gospodarstwa rolne, gospodarstwa rolników, gospodarstwa emerytów i rencistów, gospodarstwa pracujących na własny rachunek. Obserwowane struktury dotyczą wskazanych grup ogółem oraz z podziałem na wielkość gospodarstw domowych. Podział ten uwzględnia gospodarstwa jednoosobowe aż do gospodarstw sześcioosobowych i większych.

Dane statystyczne dostępne są w postaci macierzy $X_g = \left[x_{ti}^g\right]_{13\times 4}$ przedstawiającej struktury wydatków dla poszczególnych grup społeczno-ekonomicznych. Wykorzystując metodę estymacji macierzy przejścia dla danych zagregowanych z kryterium sumy bezwzględnych odchyleń[3] otrzymano macierze przejścia dla poszczególnych grup w postaci $\mathbf{P}_g = \left[p_{ij}^g\right]_{4\times 4}$.

Macierze $\mathbf{P}$ oraz $\mathbf{Q}$ określające przedział $[\mathbf{P}, \mathbf{Q}]$ definiuje się następująco:

$$p_{ij} = \min_g \left\{p_{ij}^g\right\},$$

$$q_{ij} = \max_g \left\{p_{ij}^g\right\}.$$

W podobny sposób określony został przedział wektorowy $\mathbf{S}_0$ dla ostatniego okresu obserwacji.

Zarówno przedział wektorowy $\mathbf{S}_0$ jak i przedział macierzowy $[\mathbf{P}, \mathbf{Q}]$ poddane zostały procedurze napinania. Prognozy wyznaczono na dwa kolejne lata wykorzystując metodę Hi-Lo. Poniżej przedstawione są zarówno przedział wektorowy, przedział macierzowy oraz przedziały macierzowe będące prognozą struktury. Prognozy te należy odnieść do dowolnego gospodarstwa domowego, które może zmienić swoją liczebność oraz źródło dochodów.

Macierz dolna dla badanych zmian struktury ma postać następującą:

$$P = \begin{bmatrix} 0{,}7924 & 0{,}0036 & 0{,}0130 & 0{,}0000 \\ 0{,}0000 & 0{,}7166 & 0{,}0037 & 0{,}0000 \\ 0{,}0000 & 0{,}0000 & 0{,}6363 & 0{,}0000 \\ 0{,}0211 & 0{,}0059 & 0{,}0038 & 0{,}5807 \end{bmatrix}$$

Macierz górna ma postać:

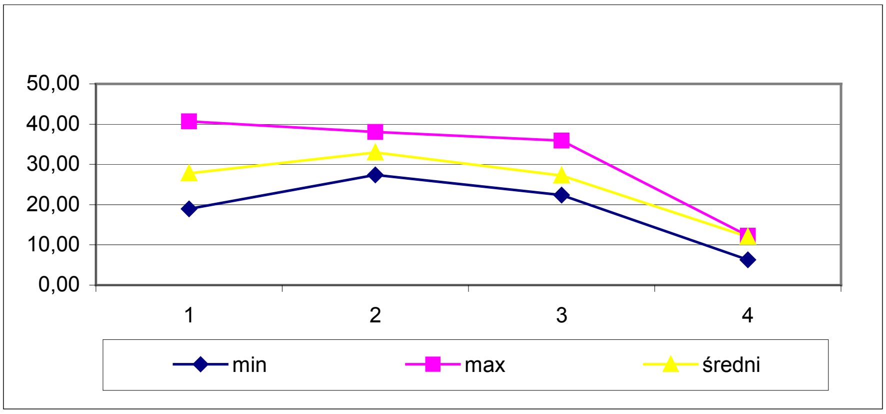
[3] Patrz: Lee, Judge, Zelner (1970).

$$Q=\begin{bmatrix}0{,}9562 & 0{,}1163 & 0{,}0913 & 0{,}0000\\ 0{,}0302 & 0{,}9837 & 0{,}2034 & 0{,}1309\\ 0{,}1127 & 0{,}2578 & 0{,}9753 & 0{,}1058\\ 0{,}2079 & 0{,}2233 & 0{,}1676 & 0{,}9579\end{bmatrix}$$

Przedział wektorowy wyznaczony dla ostatniego roku obserwacji w postaci minimalnych i maksymalnych udziałów wydatków w badanych grupach towarowych obliczanych po wszystkich grupach społeczno-ekonomicznych ma postać następującą:

$$S_T=[\overline{\mathbf{p}},\overline{\mathbf{q}}]=[(18{,}91\quad 27{,}38\quad 22{,}38\quad 6{,}26)\ \ (40{,}68\quad 38{,}06\quad 35{,}87\quad 11{,}29)]$$

Przedział ten można przedstawić na rysunku w postaci następującej:



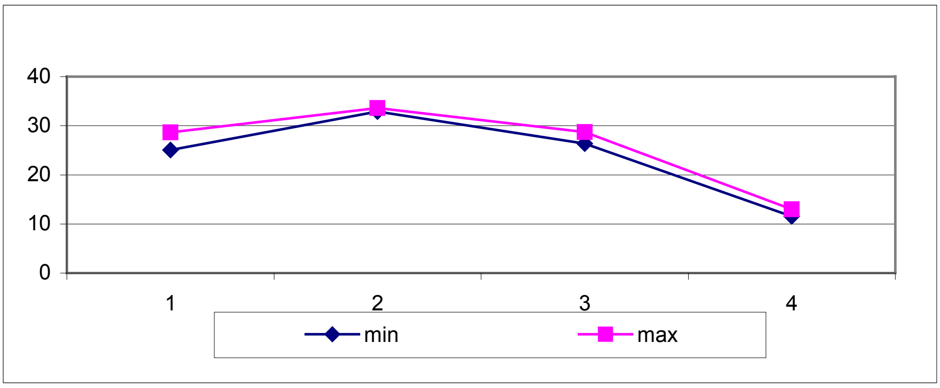
Rys.2. Przedział wektorowy struktury wydatków gospodarstw domowych w Polsce w roku 2005
Źródło: opracowanie własne.

Jako podstawę prognozy przyjęto wektor średni, będący strukturą dla gospodarstw ogółem. Prognozę na rok 2006 otrzymano metodą symulacji. Najpierw generowano macierz stochastyczną P_s z przedziału macierzowego $[\mathbf{P},\mathbf{Q}]$, a następnie wyznaczano prognozę ze wzoru:

$$\overline{x}^{p}{}_{2006}=\overline{x}_{2005}\cdot P_s$$

Z tak otrzymanych prognoz, wyznaczając elementy minimalne i maksymalne otrzymano przedział wektorowy.

$$S_{T+1}=[\overline{\mathbf{p}},\overline{\mathbf{q}}]^{p}=[(25{,}05\quad 32{,}89\quad 26{,}39\quad 11{,}56)\ \ (28{,}65\quad 33{,}58\quad 28{,}71\quad 12{,}96)]$$



Rys.3. Przedział prognozy dla struktury wydatków gospodarstw domowych w 2006 r.
Źródło: opracowanie własne.

Należy zauważyć, że duża rozpiętość udziału wydatków jest dla wydatków na żywność wynosząca 3,6, natomiast udział wydatków na artykuły nieżywnościowe jest bardzo stabilny. Dużą niestabilnością charakteryzuje się też udział wydatków na usługi (2,32). Otrzymane prognozy potwierdzają również ogólną tendencję zmian w strukturze wydatków, a mianowicie: spadek udziału wydatków na żywność i wzrost udziału wydatków na artykuły nieżywnościowe i usługi w ogólnej strukturze wydatków.

Literatura

Decewicz, A., Gyczew, A. (2001), *Wprowadzenie do teorii i zastosowań Markov-Set Chains*, praca niepublikowana, napisana w ramach badań statutowych SGH pod kierunkiem prof. dr hab.. M. Podgórskiej, Warszawa.

Hartfiel, D.J. (1999), *Markov Set-Chains*, Springer, New York.

Lee T.C., Judge, G.G., Zelner, A. (1970), *Estimating the Parameters of Markov Probability Model from Aggregate Time Series Date*, Nort-Holland Publ. Co.,Amsterdam.

Samuels, C. L. (2001), *Markov Set-Chains as Models of Plant Succession, University of Tennessee*, Knoxville.

Stawicki, J. (2002), *Łańcuchy Markowa w analizie rynku kapitałowego*, Toruń.