

*Jan Acedański\**

## **OBCIĄŻENIE PROGNOZ STRUKTURY UZYSKANYCH PRZY POMOCY ŁAŃCUCHÓW MARKOWA**

**Streszczenie:** Celem artykułu jest ocena obciążenia prognoz struktury uzyskanych na podstawie łańcuchów Markowa w przypadku, gdy macierz przejścia łańcucha szacowana jest metodą najmniejszych kwadratów i jej odmianami przy użyciu makrodanych. Źródłem badanego obciążenia jest obciążenie estymatorów macierzy przejścia. Przy zastosowaniu metod symulacyjnych pokazano, że pomimo dużego obciążenia ocen macierzy przejścia uzyskane na jej podstawie prognozy cechują się niewielkim obciążeniem. Jednocześnie stwierdzono, że znany w literaturze wzór analityczny daje z reguły dobre oszacowania obciążenia macierzy przejścia, jednak nie nadaje się do oceny obciążenia prognoz. Rozważania zostały zilustrowane przykładem prognozowania struktury polskiego eksportu według głównych odbiorców.

### **1. WPROWADZENIE**

Prognozowanie struktur zjawisk jest jednym z ważnych zagadnień poruszanych przez ekonometrię przestrzenną. W pracy analizowane jest podejście, które wykorzystuje łańcuchy Markowa do modelowania dynamiki struktury procentowej badanego zjawiska. W omawianej metodzie struktura ta reprezentowana jest przez bezwarunkowy rozkład stacjonarnego, jednorodnego łańcucha Markowa. Podstawowym problemem w takim podejściu jest wybór takiej macierzy przejścia, by dynamika rozkładu bezwarunkowego łańcucha przy ustalonym zaobserwowanym rozkładzie początkowym była jak najbardziej zbliżona do obserwowanej dynamiki wektora udziałów opisującego badane zjawisko. Technicznie, szacowanie elementów macierzy przejścia jest tożsame z estymacją parametrów modelu VAR(1) z dodatkowymi ograniczeniami. W tym celu najczęściej stosowana jest metoda najmniejszych kwadratów i jej modyfikacje. Wiadomo jednak<sup>1</sup>, że uzyskane w ten sposób oceny przy małych próbach są obciążone. W efekcie obciążenie to przenosi się także na prognozy struktury uzyskane na podstawie modeli łańcuchów Markowa.

W pracy podjęto próbę oceny obciążenia prognoz struktury stawianych przy zastosowaniu łańcuchów Markowa, których macierz przejścia szacowana jest metodą najmniejszych kwadratów. W tym celu szczegółowo przeanalizowano obciążenie ocen elementów macierzy przejścia. Stosowano tutaj przede wszystkim metody symulacyjne. Przy okazji jednak poddano również ocenie dokładność oszacowania obciążenia estymatorów macierzy przejścia uzyskanych przybliżonym wzorem podanym przez Yamamoto

---

\* Doktor, Katedra Metod Statystyczno-Matematycznych w Ekonomii, Akademia Ekonomiczna im. Karola Adamiciego w Katowicach.

<sup>1</sup> M. Kendall, [1951], *Note on Bias in the Estimation of Autocorrelation*, *Biometrika*, Vol. 41; D. Tjostheim, J. Paulsen, [1983], *Bias of some Commonly-Used Time Series Estimates*, *Biometrika*, Vol. 70(2).

i Kunitomo (1984). Przedstawione metody zilustrowano w ostatniej części pracy na przykładzie prognozowania struktury polskiego eksportu według głównych odbiorców.

Powszechne stosowanie łańcuchów Markowa w prognozowaniu struktur zjawisk wynika z ich zalet<sup>2</sup>. Po pierwsze uzyskiwane tą metodą prognozy są zgodne, to znaczy suma prognozowanych udziałów jest równa 1. Po drugie, do wyznaczenia prognoz nie jest konieczna znajomość wartości zmiennych egzogenicznych w okresie prognozowanym. Podejście to jest także szczególnie przydatne w sytuacji, gdy dostępne są wiarygodne prognozy dotyczące łącznego poziomu badanego zjawiska. Wtedy do oceny poziomu zjawiska w poszczególnych grupach wystarczy wyznaczyć prognozy udziału danej grupy. Taka sytuacja ma miejsce na przykład przy prognozowaniu regionalnego PKB. Dane dotyczące tych zmiennych są publikowane ze znacznym opóźnieniem, mniej więcej dwóch lat. Tymczasem szacunki globalnego PKB są podawane znacznie szybciej. Dodatkowo jeszcze wcześniej znanych jest wiele wiarygodnych prognoz dotyczących tej wielkości.

## 2. METODOLOGIA

Struktura zjawiska w analizowanym modelu opisywana jest przez  $K$ -wymiarowy wektor  $\mathbf{D}_t = [D_{1t}, D_{2t}, \dots, D_{Kt}]$ , którego składowe spełniają warunki:

$$D_{it} \geq 0; \quad \sum_{i=1}^K D_{it} = 1. \quad (1-2)$$

Wektor  $\mathbf{D}_t$  utożsamiany jest z rozkładem bezwarunkowym łańcucha Markowa. Stacjonarny, nieprzywiedlny łańcuch Markowa reprezentowany jest przez kwadratową  $K$ -wymiarową macierz przejścia  $\mathbf{P}$ . Jej elementy  $p_{ij}$  są nieujemne i sumują się w wierszach do 1. Spełniają więc warunki:

$$p_{ij} \geq 0; \quad \sum_{i=1}^K p_{ij} = 1. \quad (3-4)$$

Dynamikę rozkładu bezwarunkowego opisuje równanie autoregresyjne:

$$\mathbf{D}_t = \mathbf{D}_{t-1} \mathbf{P}. \quad (5)$$

Równanie to jest podstawą wyznaczania prognoz na kolejne okresy. Przy próbie liczącej  $T$  obserwacji prognozę na  $h$  okresów do przodu wyznacza się jako:

$$\mathbf{D}_{T+h} = \mathbf{D}_T \mathbf{P}^h. \quad (6)$$

### 2.1. Estymacja macierzy przejścia $\mathbf{P}$

Równanie (5) stosowane jest także przy szacowaniu macierzy przejścia  $\mathbf{P}$ , gdyż dostępne są dane dotyczące empirycznych odpowiedników  $\mathbf{d}_t$ , wektorów  $\mathbf{D}_t$ . W takim przypadku przyjmuje ono postać:

$$\mathbf{d}_t = \mathbf{d}_{t-1} \mathbf{P} + \boldsymbol{\varepsilon}_t, \quad (7)$$

gdzie  $\boldsymbol{\varepsilon}_t$  jest wektorem składników losowych spełniających warunki  $E(\boldsymbol{\varepsilon}_t) = \mathbf{0}$ ,  $E(\boldsymbol{\varepsilon}_t' \boldsymbol{\varepsilon}_t) = \boldsymbol{\Omega}$ ,  $E(\boldsymbol{\varepsilon}_s' \boldsymbol{\varepsilon}_t) = \mathbf{0}$  dla  $s \neq t$ . Model wektorowej autoregresji VAR(1) postaci (7)

<sup>2</sup> Zob. I. Koźniewska (red.), [1980], *Prognozowanie struktury za pomocą łańcuchów Markowa*, SGPiS, Warszawa; J. Acedański, [2006], *O pewnej metodzie prognozowania przewozów w gospodarce*, [w:] P. Dittmann, J. Krupowicz (red.), *Prognozowanie w zarządzaniu firmą*, Prace Naukowe AE we Wrocławiu nr 1112, Wrocław.

jest modelem niestacjonarnym, gdyż ze względu na założenia (3-4) największa wartość własna macierzy  $\mathbf{P}$  jest co do wartości bezwzględnej równa 1.

Biorąc pod uwagę warunki (2) i (4) dla celów estymacji wystarczy rozpatrywać model składający się z  $K-1$  zmiennych:  $\mathbf{d}_t^* = [d_{1t} \ d_{2t} \ \dots \ d_{K-1t}]$  oraz odpowiedniej  $(K-1)$ -wymiarowej macierzy  $\mathbf{P}^*$  postaci:

$$\mathbf{P}^* = \begin{bmatrix} p_{11} - p_{1K} & p_{12} - p_{1K} & \dots & p_{1K-1} - p_{1K} \\ p_{21} - p_{2K} & p_{22} - p_{2K} & \dots & p_{2K-1} - p_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ p_{K-11} - p_{K-1K} & p_{K-12} - p_{K-1K} & \dots & p_{K-1K-1} - p_{K-1K} \end{bmatrix}.$$

Model zmodyfikowany dany jest równaniem:

$$\mathbf{d}_t^* = \mathbf{P}_K + \mathbf{d}_{t-1}^* \mathbf{P}^* + \boldsymbol{\varepsilon}_t^*, \tag{8}$$

gdzie  $\mathbf{P}_K = [p_{1K} \ p_{2K} \ \dots \ p_{K-1K}]$  oraz  $\boldsymbol{\varepsilon}_t^* = [\varepsilon_{1t} \ \varepsilon_{2t} \ \dots \ \varepsilon_{K-1t}]$ . W modelu (8) nieznanymi parametrami są wektor  $\mathbf{P}_K$  oraz macierz  $\mathbf{P}^*$ , przy czym teraz wszystkie wartości własne macierzy  $\mathbf{P}^*$  leżą wewnątrz koła jednostkowego. Przy oszacowanych wartościach  $\hat{\mathbf{P}}_K$  oraz  $\hat{\mathbf{P}}^*$  brakujące oceny  $\hat{p}_{1K}, \hat{p}_{2K}, \dots, \hat{p}_{KK}$  można wyznaczyć z warunku (2). W ten sposób uzyskuje się ocenę wyjściowej macierzy przejścia  $\hat{\mathbf{P}}$ .

Najczęściej stosowanymi metodami estymacji parametrów modelu (8) są metoda najmniejszych kwadratów (MNK) oraz warunkowa, uogólniona metoda najmniejszych kwadratów (WUMNK). W przypadku MNK oceny  $\hat{\mathbf{P}}_K$  oraz  $\hat{\mathbf{P}}^*$  uzyskiwane są przez rozwiązanie następującego problemu decyzyjnego:

$$\min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \tag{9}$$

gdzie:  $\mathbf{Y} = [d_{12} \ d_{13} \ \dots \ d_{1T} \ d_{22} \ d_{23} \ \dots \ d_{2T} \ \dots \ d_{K-12} \ d_{K-13} \ \dots \ d_{K-1T}]'$ ,

$$\mathbf{X} = \mathbf{I}_{K-1} \otimes \begin{bmatrix} \mathbf{d}_1^* & 1 \\ \mathbf{d}_2^* & 1 \\ \vdots & \vdots \\ \mathbf{d}_{T-1}^* & 1 \end{bmatrix}, \boldsymbol{\beta} = \text{vec} \begin{bmatrix} \mathbf{P}^* \\ \mathbf{P}_K \end{bmatrix}.$$

W powyższych wzorach  $\mathbf{I}_{K-1}$  oznacza  $(K-1)$ -wymiarową macierz jednostkową,  $\otimes$  jest iloczynem Kroneckera,  $\text{vec}$  jest operatorem, który przekształca macierz w wektor kolumnowy, natomiast  $T$  oznacza liczbę obserwacji. Wyznaczone w ten sposób estymatory są zgodne i asymptotycznie nieobciążone, jednak nie spełniają warunku (3) nieujemności elementów macierzy  $\mathbf{P}^3$ .

W metodzie WUMNK problem decyzyjny przyjmuje postać:

$$\min_{\boldsymbol{\beta}} (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\hat{\boldsymbol{\Omega}}^{*-1} \otimes \mathbf{I}_{T-1})(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}), \text{ p.w. } \mathbf{P}_K \geq \mathbf{0}, \quad \mathbf{0} \leq \mathbf{P}^* + \mathbf{P}_K \otimes \mathbf{1}'_{K-1} \leq \mathbf{1}. \tag{10}$$

<sup>3</sup> Zob. T.C. Lee, G. Judge, A. Zellner, [1970], *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*, North-Holland Publishing Company, Amsterdam-Londyn; M. Podgórska, P. Śliwka, M. Topolewski, M. Wrzosek, [2000], *Łańcuchy Markowa w teorii i zastosowaniach*, SGH, Warszawa.

Macierz  $\hat{\Omega}^*$  oznacza ocenę macierzy kowariancji wektora reszt  $\varepsilon_t^*$ . Oszacowanie to uzyskuje się na podstawie reszt MNK postaci (9). Wektor  $\mathbf{1}'_{K-1}$  jest  $(K-1)$ -wymiarowym wierszowym wektorem jedynek. Ograniczenia występujące w problemie (10) są równoważne ograniczeniu (3) nieujemności elementów macierzy  $\mathbf{P}$ . Estymatory WUMNK są zgodne, asymptotycznie nieobciążone i cechują się większą efektywnością niż estymatory MNK<sup>4</sup>. Uzyskana na ich podstawie macierz  $\hat{\mathbf{P}}$  spełnia wszystkie warunki macierzy przejścia jednorodnego łańcucha Markowa.

## 2.2. Obciążenie estymatorów macierzy $\mathbf{P}$ oraz prognoz

W przypadku małych prób oba estymatory rozpatrywane powyżej są obciążone. Yamamoto i Kunitomo (1984) podali przybliżony wzór na obciążenie estymatorów MNK parametrów modelu (8). Przy przyjęciu następujących oznaczeń:

$$\mathbf{A} = \begin{bmatrix} \mathbf{P}^* & \mathbf{P}'_K \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{\Omega}_u = \begin{bmatrix} \mathbf{\Omega}^* & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}, \quad \mathbf{\Gamma} = \sum_{t=1}^{T-1} [\mathbf{d}_t^* \quad \mathbf{1}']' [\mathbf{d}_t^* \quad \mathbf{1}],$$

oraz pewnych warunków regularności obciążenie bias( $\mathbf{A}$ ) macierzy  $\mathbf{A}$  w przybliżeniu wyraża się wzorem<sup>5</sup>:

$$\text{bias}(\hat{\mathbf{A}}) = -T^{-1} \mathbf{\Omega}_u \sum_{k=0}^{\infty} (\mathbf{A}^k \text{tr}(\mathbf{A}^{k+1}) + \mathbf{A}^{2k+1}) \mathbf{\Gamma}^{-1}, \quad (11)$$

przy czym przez  $\text{tr}(\mathbf{A})$  oznaczono ślad macierzy  $\mathbf{A}$ . Oceny obciążenia uzyskuje się wstawiając zamiast macierzy  $\mathbf{\Omega}_u$  oraz  $\mathbf{A}$  oszacowania uzyskane na podstawie próby. Nieskończoną sumę przybliża się przyjmując dużą, ale skończoną liczbę składników. Jeżeli tylko wartości własne macierzy  $\hat{\mathbf{A}}$  leżą wewnątrz koła jednostkowego, szereg kolejnych potęg tej macierzy jest zbieżny. Powyższy wynik, a także analizy podobnych modeli<sup>6</sup>, wskazują, że przy wysokich wartościach elementów  $p_{ij}$  ich oceny MNK są średnio zbyt niskie, natomiast oszacowania elementów o wartościach bliskich 0 są zawyżone. Podobne wyniki dotyczą WUMNK.

Na podstawie wzoru (11) można również łatwo wyznaczyć obciążenie szacunków macierzy przejścia  $\mathbf{P}$ . Należy zwrócić tu uwagę, że w przypadku obu omawianych metod założenie sumowania się do 1 elementów w wierszach macierzy  $\mathbf{P}$  będzie spełnione. Tym samym suma obciążeń elementów w wierszach będzie równa 0:

$$\sum_{j=1}^K \text{bias}(p_{ij}) = 0. \quad (12)$$

Z uwagi na występowanie ograniczeń w postaci nierówności obciążenie estymatorów WUMNK jest praktycznie niemożliwe do oszacowania w sposób analityczny.

<sup>4</sup> Tamże.

<sup>5</sup> T. Yamamoto, N. Kunitomo, [1984], *Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models*, Annals of the Institute of Statistical Mathematics, Vol. 36(1), s. 421-422.

<sup>6</sup> Zob. M. Kendall, [1951]; K. Abadir, K. Hadri, E. Tzavalis, [1999], *The Influence of VAR Dimensions on Estimator Biases*, Econometrica, Vol. 67(1); S. Lawford., M. Stamatogiannis, [2004], *The Finite-Sample Effects of VAR Dimensions on OLS Bias, OLS Variance, and Minimum MSE Estimators*, Journal of Econometrics, Vol. 148(2); J. Kiviet, G. Phillips, [2005], *Moment Approximation for Least-Squares Estimators in Dynamic Regression Models with a Unit Root*, Econometrics Journal, Vol. 8; J. Acedański, [2008], *Dokładność prognoz struktury uzyskanych za pomocą modeli wektorowej autoregresji*, [w:] P. Dittmann, J. Szandula (red.), *Prognozowanie w zarządzaniu firmą*, AE Wrocław, Wydawnictwo Indygo Zahir Media, Wrocław.

W takim przypadku konieczne jest odwoływanie się do rezultatów analiz symulacyjnych.

Ze wzoru (6) wyraźnie wynika, że źródłem obciążenia prognoz w omawianym modelu jest obciążenie estymatorów macierzy  $\mathbf{P}^h$ :

$$\text{bias}(\hat{\mathbf{D}}_{T+h}) = \mathbf{D}_T \text{bias}(\hat{\mathbf{P}}^h). \tag{13}$$

W najprostszym przypadku, gdy  $h = 1$ , obciążenie tej macierzy można szacować korzystając z metod opisanych powyżej. Analityczna ocena obciążenia dla  $h > 1$  jest już problematyczna, gdyż wymaga znajomości wyższych momentów rozkładu estymatorów macierzy  $\mathbf{P}^h$ . Biorąc pod uwagę trudności związane z szacowaniem wartości oczekiwanej macierzy  $\mathbf{P}$ , analityczne oceny momentów wyższych rzędów nie będą zbyt dokładne. Z tego powodu przy szacowaniu obciążenia prognoz dla dłuższych horyzontów czasowych wskazane jest korzystanie z metod symulacyjnych.

Należy również mieć na uwadze, że jeżeli elementy w kolumnach macierzy  $\text{bias}(\hat{\mathbf{P}}^h)$  będą posiadać przeciwne znaki, wtedy ich obciążenie będzie się częściowo znosić przy wyznaczaniu obciążenia prognoz. Dokładna wielkość tego efektu zależy także od wektora  $\mathbf{D}_T$ .

### 3. WYNIKI BADAŃ SYMULACYJNYCH

W badaniach symulacyjnych analizowano dwie różne macierze przejścia:

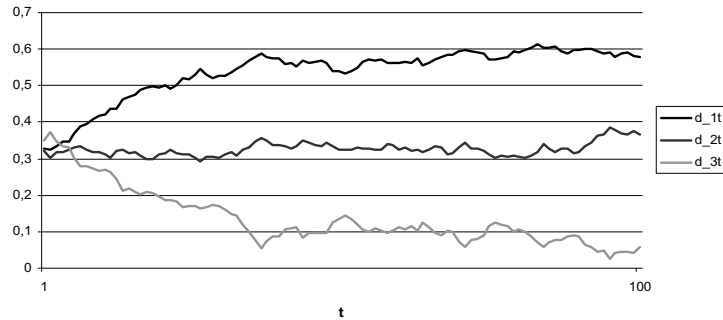
$$\mathbf{P}_1 = \begin{bmatrix} 0,99 & 0,01 & 0 \\ 0,01 & 0,98 & 0,01 \\ 0,03 & 0,02 & 0,95 \end{bmatrix}, \quad \mathbf{P}_2 = \begin{bmatrix} 0,6 & 0,2 & 0,2 \\ 0,3 & 0,5 & 0,2 \\ 0,3 & 0,3 & 0,4 \end{bmatrix}.$$

W pierwszej macierzy wszystkie wartości własne są bliskie 1, co wskazuje, że zbieżność rozkładu łańcucha o takiej macierzy przejścia do rozkładu stacjonarnego jest bardzo powolna. W drugiej macierzy natomiast druga z kolei największa wartość własna jest równa 0,3, a więc zbieżność do rozkładu granicznego jest bardzo szybka. Przykładowe trajektorie rozkładów bezwarunkowych omawianych łańcuchów dla 100 obserwacji prezentują rysunki 1 i 2.

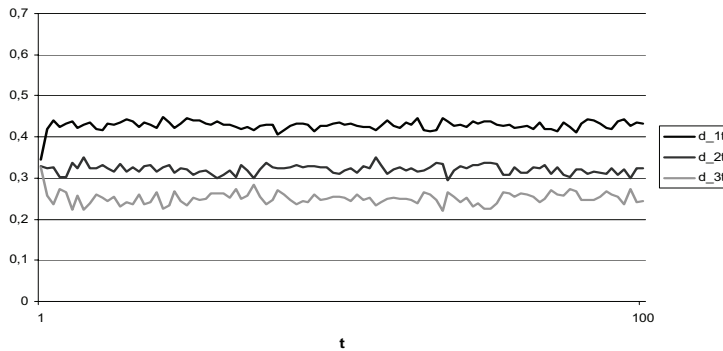
Dla macierzy  $\mathbf{P}_1$  badano również wpływ różnych rozkładów początkowych na wyniki estymacji i prognozowania. Analizowano trzy rozkłady startowe:

$$\mathbf{d}_{s1} = [0 \ 0 \ 1], \quad \mathbf{d}_{s2} = [1/3 \ 1/3 \ 1/3], \quad \mathbf{d}_{s3} = [0,5714 \ 0,3571 \ 0,0714].$$

W pierwszym przypadku liczba realizacji łańcucha potrzebna do uzyskania zbieżności jest największa, gdyż rozkład  $\mathbf{d}_{s1}$  znacząco różni się od rozkładu granicznego. Trzeci rozkład początkowy równy jest natomiast rozkładowi granicznemu, a więc w tym przypadku czas zbieżności jest najkrótszy. Dla macierzy  $\mathbf{P}_2$  rozpatrywano tylko jeden rozkład początkowy  $\mathbf{d}_{s2}$ , gdyż zbieżność uzyskiwana jest zawsze bardzo szybko. W symulacjach przyjęto, że zakłócenia losowe mają rozkład normalny z osobliwą macierz kowariancji. Odnośnie pierwszych  $K - 1$  zmiennych była ona macierzą diagonalną z elementami na głównej przekątnej o wartościach równych 0,0001. Zaburzenie ostatniej zmiennej równe było sumie zaburzeń dla pierwszych  $K - 1$  zmiennych, co gwarantowało spełnienie warunku (2) sumowania się do jedności składowych wektora rozkładu bezwarunkowego.

**Rys. 1** Przykładowa trajektoria rozkładów bezwarunkowych dla macierzy  $P_1$ 

Źródło: opracowanie własne.

**Rys. 2** Przykładowa trajektoria rozkładów bezwarunkowych dla macierzy  $P_2$ 

Źródło: opracowanie własne.

### 3.1. Wyniki badania obciążenia macierzy przejścia

W tabeli 1 przedstawiono obciążenie ocen macierzy przejścia  $P_1$  przy różnych rozkładach początkowych oraz różnej liczbie obserwacji. Kolumna YK zawiera oceny obciążenia obliczone wzorem (11) Yamamoto-Kunitomo. Aby zapewnić zbieżność szeregów występujących w omawianym wzorze, przed jego obliczeniem badano wartości własne uzyskanej macierzy przejścia. Jeżeli największa przekraczała 1, wtedy zmniejszono proporcjonalnie elementy leżące na głównej przekątnej tak, aby druga co do wielkości wartość własna była mniejsza niż 0,99. Taką skorygowaną macierz przejścia stosowano jedynie przy obliczaniu obciążenia zgodnie ze wzorem (11).

W kolejnych kolumnach zestawiono obciążenia ocen MNK i WUMNK, które uzyskano przy zastosowaniu symulacji. W tym celu symulowano wielokrotnie realizacje łańcucha o danej macierzy przejścia i korzystając z tych realizacji szacowano macierze przejścia oraz prognozy. Rezultaty tych oszacowań porównywano następnie z rzeczywistymi wartościami macierzy przejścia i realizacjami łańcucha. Wyniki w tabeli 1 są średnimi wartościami obliczonymi na podstawie 10000 symulacji. W tabeli 2 przedstawiono rezultaty takiego samego badania dla łańcucha z macierzą przejścia  $P_2$  i rozkładem początkowym  $\mathbf{d}_{s2}$ .

Tab. 1 Obciążenie ocen macierzy przejścia  $P_1$ 

Liczba obserwacji	Metoda YK	MNK	WUMNK
Rozkład początkowy $d_{s1}$			
$T = 25$	$\begin{bmatrix} -0,088 & 0,114 & -0,026 \\ 0,122 & -0,182 & 0,060 \\ 0,001 & 0,001 & -0,002 \end{bmatrix}$	$\begin{bmatrix} -0,105 & 0,152 & -0,047 \\ 0,155 & -0,235 & 0,080 \\ 0,001 & 0,001 & -0,001 \end{bmatrix}$	$\begin{bmatrix} -0,131 & 0,127 & -0,004 \\ 0,195 & -0,197 & 0,002 \\ 0,001 & 0,001 & -0,002 \end{bmatrix}$
$T = 50$	$\begin{bmatrix} -0,037 & 0,052 & -0,016 \\ 0,054 & -0,082 & 0,028 \\ 0,001 & 0,001 & -0,001 \end{bmatrix}$	$\begin{bmatrix} -0,049 & 0,072 & -0,024 \\ 0,073 & -0,113 & 0,040 \\ 0,001 & 0,001 & -0,001 \end{bmatrix}$	$\begin{bmatrix} -0,062 & 0,059 & 0,003 \\ 0,095 & -0,093 & -0,002 \\ 0,001 & 0,001 & -0,001 \end{bmatrix}$
$T = 100$	$\begin{bmatrix} -0,016 & 0,024 & -0,008 \\ 0,025 & -0,038 & 0,013 \\ 0,001 & 0,001 & -0,001 \end{bmatrix}$	$\begin{bmatrix} -0,021 & 0,032 & -0,011 \\ 0,032 & -0,051 & 0,019 \\ 0,001 & 0,001 & -0,001 \end{bmatrix}$	$\begin{bmatrix} -0,028 & 0,025 & 0,002 \\ 0,043 & -0,040 & -0,003 \\ 0 & 0,001 & -0,001 \end{bmatrix}$
Rozkład początkowy $d_{s2}$			
$T = 25$	$\begin{bmatrix} -0,206 & 0,143 & 0,064 \\ 0,180 & -0,248 & 0,068 \\ 0,050 & 0,116 & -0,165 \end{bmatrix}$	$\begin{bmatrix} -0,067 & 0,141 & -0,075 \\ 0,072 & -0,216 & 0,144 \\ -0,002 & 0,077 & -0,075 \end{bmatrix}$	$\begin{bmatrix} -0,093 & 0,087 & -0,006 \\ 0,112 & -0,134 & 0,022 \\ -0,008 & 0,052 & -0,044 \end{bmatrix}$
$T = 50$	$\begin{bmatrix} -0,034 & 0,050 & -0,017 \\ 0,039 & -0,079 & 0,040 \\ 0,003 & 0,032 & -0,035 \end{bmatrix}$	$\begin{bmatrix} -0,026 & 0,066 & -0,040 \\ 0,032 & -0,105 & 0,073 \\ -0,001 & 0,041 & -0,041 \end{bmatrix}$	$\begin{bmatrix} -0,044 & 0,041 & 0,003 \\ 0,060 & -0,066 & 0,006 \\ -0,008 & 0,028 & -0,020 \end{bmatrix}$
$T = 100$	$\begin{bmatrix} -0,011 & 0,021 & -0,011 \\ 0,013 & -0,035 & 0,022 \\ 0,003 & 0,016 & -0,019 \end{bmatrix}$	$\begin{bmatrix} -0,011 & 0,028 & -0,018 \\ 0,014 & -0,047 & 0,033 \\ -0,002 & 0,021 & -0,024 \end{bmatrix}$	$\begin{bmatrix} -0,021 & 0,018 & 0,002 \\ 0,030 & -0,031 & 0,001 \\ -0,003 & 0,015 & -0,012 \end{bmatrix}$
Rozkład początkowy $d_{s3}$			
$T = 25$	$\begin{bmatrix} -0,356 & 0,163 & 0,193 \\ 0,469 & -0,305 & -0,164 \\ 0,484 & 0,247 & -0,731 \end{bmatrix}$	$\begin{bmatrix} -0,105 & 0,090 & 0,015 \\ 0,137 & -0,163 & 0,025 \\ 0,142 & 0,087 & -0,229 \end{bmatrix}$	$\begin{bmatrix} -0,102 & 0,086 & 0,017 \\ 0,139 & -0,154 & 0,016 \\ 0,125 & 0,083 & -0,208 \end{bmatrix}$
$T = 50$	$\begin{bmatrix} -0,045 & 0,038 & 0,007 \\ 0,059 & -0,067 & 0,007 \\ 0,064 & 0,032 & -0,096 \end{bmatrix}$	$\begin{bmatrix} -0,053 & 0,048 & 0,005 \\ 0,069 & -0,086 & 0,017 \\ 0,074 & 0,042 & -0,115 \end{bmatrix}$	$\begin{bmatrix} -0,053 & 0,044 & 0,010 \\ 0,072 & -0,078 & 0,006 \\ 0,064 & 0,040 & -0,104 \end{bmatrix}$
$T = 100$	$\begin{bmatrix} -0,019 & 0,018 & 0,001 \\ 0,025 & -0,032 & 0,007 \\ 0,028 & 0,015 & -0,043 \end{bmatrix}$	$\begin{bmatrix} -0,025 & 0,024 & 0,001 \\ 0,032 & -0,042 & 0,010 \\ 0,036 & 0,020 & -0,055 \end{bmatrix}$	$\begin{bmatrix} -0,026 & 0,021 & 0,006 \\ 0,036 & -0,037 & 0,001 \\ 0,031 & 0,018 & -0,049 \end{bmatrix}$

Źródło: opracowanie własne.

Tab. 2 Obciążenie ocen macierzy przejścia  $P_2$ 

Liczba obserwacji	Metoda YK	MNK	WUMNK
$T = 25$	$\begin{bmatrix} -0,012 & 0,019 & -0,007 \\ 0,010 & -0,046 & 0,037 \\ 0,007 & 0,028 & -0,035 \end{bmatrix}$	$\begin{bmatrix} -0,011 & 0,019 & -0,008 \\ 0,008 & -0,048 & 0,039 \\ 0,008 & 0,028 & -0,036 \end{bmatrix}$	$\begin{bmatrix} -0,010 & 0,020 & -0,010 \\ 0,006 & -0,050 & 0,044 \\ 0,009 & 0,027 & -0,039 \end{bmatrix}$
$T = 50$	$\begin{bmatrix} -0,009 & 0,010 & -0,001 \\ 0,007 & -0,023 & 0,016 \\ 0,006 & 0,014 & -0,019 \end{bmatrix}$	$\begin{bmatrix} -0,009 & 0,010 & -0,001 \\ 0,006 & -0,024 & 0,018 \\ 0,007 & 0,014 & -0,021 \end{bmatrix}$	$\begin{bmatrix} -0,008 & 0,011 & 0,002 \\ 0,005 & -0,025 & 0,020 \\ 0,007 & 0,015 & -0,022 \end{bmatrix}$
$T = 100$	$\begin{bmatrix} -0,007 & 0,005 & 0,002 \\ 0,005 & -0,012 & 0,007 \\ 0,004 & 0,007 & -0,011 \end{bmatrix}$	$\begin{bmatrix} -0,007 & 0,004 & 0,002 \\ 0,006 & -0,010 & 0,005 \\ 0,004 & 0,006 & -0,010 \end{bmatrix}$	$\begin{bmatrix} -0,007 & 0,005 & 0,002 \\ 0,005 & -0,011 & 0,006 \\ 0,004 & 0,006 & -0,010 \end{bmatrix}$

Źródło: opracowanie własne.

Analizując wielkości zestawione w omawianych tabelach można wysnuć kilka wniosków. Po pierwsze potwierdzają się znane w literaturze własności analizowanych estymatorów: wysokie wartości elementów na głównej przekątnej są niedoszacowane, natomiast wartości niskie, poza główną przekątną – przeszacowane – przy czym w małych próbach obciążenie może być bardzo duże (w kilku przypadkach ponad 0,1 przy 50 obserwacjach, także dla elementów leżących poza główną przekątną); obciążenie ocen maleje wraz ze wzrostem liczby obserwacji oraz im mniejsze jest tempo zbieżności łańcucha do rozkładu stacjonarnego, tym większe jest obciążenie szacunków macierzy przejścia.

Ponadto widać wyraźnie, że nie ma większych różnic pomiędzy obciążeniem ocen MNK oraz WUMNK. Nie można też wskazać żadnej metody, która zawsze dawałaby oszacowania cechujące się mniejszym obciążeniem.

Odnośnie wzoru YK można sformułować dwa stwierdzenia. Po pierwsze, daje on dobre oszacowania obciążenia w przypadku dużych prób (w analizowanym przypadku było to  $T=50$  oraz więcej) lub łańcuchów cechujących się szybką zbieżnością do rozkładu stacjonarnego, a więc zawsze gdy obciążenie szacunków jest niewielkie; w przypadku małych prób ( $T=25$ ) i łańcucha o powolnej zbieżności oceny obciążenia uzyskane omawianym wzorem mogą się znacząco różnić od rzeczywistego obciążenia.

Po drugie wzór (11) daje szczególnie duże błędy w sytuacji, gdy większość obserwacji pochodzi z okolic rozkładu stacjonarnego. Błędy oszacowania wynikające ze stosowania tego wzoru były bowiem szczególnie duże dla rozkładu początkowego  $\mathbf{d}_{s,3}$  równego rozkładowi granicznemu. W przypadku rozkładu początkowego  $\mathbf{d}_{s,1}$  błędy były znacznie mniejsze.

### 3.2. Wyniki badania obciążenia prognoz

W tabeli 3 przedstawiono wyniki badania obciążenia prognoz konstruowanych dla łańcucha o macierzy przejścia  $\mathbf{P}_1$ , analogicznie jak to było w tabeli 1 dotyczącej obciążenia ocen macierzy przejścia. Rozpatrywano dwa horyzonty prognoz:  $h=1$  oraz  $h=5$ . W przypadku wzoru YK obciążenie obliczano tylko dla horyzontu  $h=1$ . W pozostałych przypadkach w pierwszym wierszu podawano zawsze obciążenie dla  $h=1$ , a w drugim dla  $h=5$ . Wartości równe 0 w omawianej tabeli oznaczają, że obciążenie co do wartości bezwzględnej było mniejsze niż 0,001. Nie podano rezultatów dla łańcucha o macierzy przejścia  $\mathbf{P}_2$ , gdyż we wszystkich analizowanych przypadkach obciążenie było mniejsze od wartości granicznej podanej powyżej.

Pomimo znacznego obciążenia oszacowań macierzy przejścia, obciążenie prognoz jest raczej niewielkie. W przypadku prognoz na jeden okres do przodu nigdy nie przekracza ono poziomu 0,003 dla  $T=25$  oraz 0,001 dla  $T=50$  obserwacji i więcej. Dla  $h=5$  błędy w niektórych przypadkach są już wyraźnie wyższe i przekraczają 0,01 dla  $\mathbf{d}_{s,2}$  oraz  $T=25$ .

Jeżeli obciążenie prognoz nie jest znikomo małe, wtedy oszacowania uzyskane ze wzoru YK wyraźnie różnią się od rzeczywistych obciążeń wyznaczonych na podstawie symulacji. We wszystkich omawianych sytuacjach oszacowania analityczne przekraczają faktyczne obciążenia. Z tego punktu widzenia omawiany wzór może być stosowany jedynie do oceny górnej granicy obciążenia. Prawdopodobnie jednak będzie ono w rzeczywistości wyraźnie niższe, niż to wynika z takich oszacowań.



Tab. 3 Obciążenie prognoz dla łańcucha z macierzą przejścia  $P_1$ 

Liczba obserwacji	Metoda YK	MNK	WUMNK
Rozkład początkowy $d_{s,1}$			
$T = 25$	$[-0,001 \ 0 \ 0,001]$	$\begin{bmatrix} 0 & 0 & 0 \\ -0,002 & 0 & 0,002 \end{bmatrix}$	$\begin{bmatrix} -0,001 & 0 & 0,001 \\ -0,004 & 0,002 & 0,006 \end{bmatrix}$
$T = 50$	$[-0,001 \ 0 \ 0,001]$	$\begin{bmatrix} 0 & 0 & 0 \\ -0,002 & 0 & 0,001 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ -0,002 & 0 & 0,002 \end{bmatrix}$
$T = 100$	$[0 \ 0 \ 0]$	$\begin{bmatrix} 0 & 0 & 0,001 \\ -0,001 & 0 & 0,001 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0,001 \\ -0,001 & 0 & 0,001 \end{bmatrix}$
Rozkład początkowy $d_{s,2}$			
$T = 25$	$[-0,029 \ 0,010 \ 0,039]$	$\begin{bmatrix} -0,003 & 0 & 0,003 \\ -0,013 & -0,002 & 0,015 \end{bmatrix}$	$\begin{bmatrix} -0,002 & -0,001 & 0,003 \\ -0,010 & -0,002 & 0,012 \end{bmatrix}$
$T = 50$	$[-0,004 \ 0,001 \ 0,003]$	$\begin{bmatrix} -0,001 & 0 & 0,001 \\ -0,005 & 0 & 0,004 \end{bmatrix}$	$\begin{bmatrix} -0,001 & 0 & 0,001 \\ -0,005 & 0 & 0,005 \end{bmatrix}$
$T = 100$	$[-0,001 \ 0,001 \ 0,001]$	$\begin{bmatrix} 0 & 0 & 0 \\ -0,002 & 0 & 0,002 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ -0,002 & 0 & 0,002 \end{bmatrix}$
Rozkład początkowy $d_{s,3}$			
$T = 25$	$[-0,005 \ 0,005 \ -0,009]$	$\begin{bmatrix} -0,001 & -0,001 & 0,002 \\ -0,004 & -0,003 & 0,007 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0,001 \\ -0,001 & 0 & 0,001 \end{bmatrix}$
$T = 50$	$[0 \ 0 \ 0]$	$\begin{bmatrix} 0 & 0 & 0 \\ -0,002 & 0 & 0,002 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$
$T = 100$	$[0 \ 0 \ 0]$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

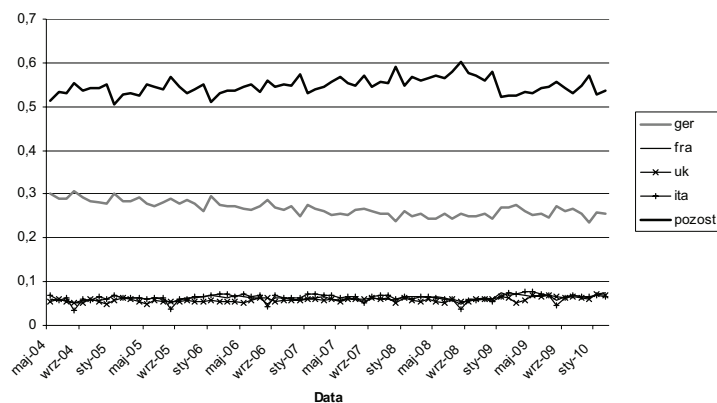
Źródło: opracowanie własne.

#### 4. PRZYKŁAD EMPIRYCZNY

Przedstawione powyżej metody prognozowania struktury oraz obliczania obciążenia szacunków macierzy przejścia, a także prognoz zastosowano do prognozowania struktury polskiego eksportu według głównych odbiorców. Dynamikę tej struktury przedstawiono na rysunku 3. Miesięczne dane obejmują okres 05.2004-02.2010 i pochodzą z bazy danych Eurostatu. Szereg liczy więc 70 obserwacji.

W badanym okresie największa zmiana dotyczyła eksportu do Niemiec, którego udział spadł z około 30% do około 25%. Jednocześnie zwiększeniu uległ eksport do Francji, mniej więcej z 6% do 7,5% oraz do Wielkiej Brytanii, z 5,5% do 7%. Większym zmianom nie uległ eksport do Włoch oraz do pozostałych państw, które wyniosły odpowiednio około 6,5% oraz 53%.

Rys. 3 Struktura polskiego eksportu według głównych partnerów w okresie 05.2004-02.2010



Źródło: opracowanie własne.

Macierz przejścia łańcucha oszacowana na podstawie omawianych danych przyjęła postać:

$$\hat{\mathbf{P}}_{WUMNK} = \begin{bmatrix} 0,756 & 0 & 0 & 0,043 & 0,201 \\ 0 & 0,498 & 0,075 & 0,261 & 0,166 \\ 0 & 0,093 & 0,452 & 0 & 0,455 \\ 0,123 & 0 & 0,109 & 0,253 & 0,515 \\ 0,103 & 0,047 & 0,033 & 0,036 & 0,755 \end{bmatrix},$$

przy czym kolejne kolumny odpowiadają następującym krajom: Niemcy, Francja, Wielka Brytania, Włochy oraz pozostałe kraje. Oszacowana macierz kowariancji składników losowych równa była:

$$\hat{\mathbf{\Omega}}_{WUMNK} = 10^{-4} \begin{bmatrix} 1,372 & 0,003 & 0,013 & -0,240 & -1,147 \\ 0,003 & 0,232 & 0,042 & 0,300 & -0,577 \\ 0,013 & 0,042 & 0,157 & 0,058 & -0,270 \\ -0,240 & 0,300 & 0,058 & 0,674 & -0,792 \\ -1,147 & -0,577 & -0,270 & -0,792 & 2,796 \end{bmatrix}.$$

Stąd wyznaczono prognozy udziału danego partnera w eksporcie dla  $h = 1, 2, \dots, 6$  miesięcy do przodu. Prognozy te kształtowały się następująco:

$$\begin{aligned} \hat{\mathbf{d}}_{mar} &= [0,256 \quad 0,069 \quad 0,064 \quad 0,067 \quad 0,544], \\ \hat{\mathbf{d}}_{kwi} &= [0,258 \quad 0,066 \quad 0,062 \quad 0,065 \quad 0,548], \\ \hat{\mathbf{d}}_{maj} &= [0,260 \quad 0,065 \quad 0,061 \quad 0,065 \quad 0,549], \\ \hat{\mathbf{d}}_{cze} &= [0,262 \quad 0,064 \quad 0,060 \quad 0,064 \quad 0,550], \\ \hat{\mathbf{d}}_{lip} &= [0,263 \quad 0,064 \quad 0,060 \quad 0,064 \quad 0,550], \\ \hat{\mathbf{d}}_{sie} &= [0,264 \quad 0,064 \quad 0,059 \quad 0,064 \quad 0,549]. \end{aligned}$$

W następnej kolejności badano obciążenia szacunków. Stosując metody symulacyjne stwierdzono, że obciążenie macierzy  $\hat{\mathbf{P}}_{WUMNK}$  przy zaburzeniach losowych o rozkładzie normalnym z macierzą kowariancji  $\hat{\mathbf{\Omega}}_{WUMNK}$  równe jest:

$$\text{bias}(\hat{\mathbf{P}}_{WUMNK}) = \begin{bmatrix} -0,041 & 0,012 & 0,011 & 0,019 & -0,002 \\ 0,083 & -0,110 & 0,008 & -0,102 & 0,122 \\ 0,083 & 0,036 & -0,049 & 0,058 & -0,129 \\ -0,016 & 0,041 & 0,001 & 0,009 & -0,035 \\ 0,003 & -0,002 & -0,001 & -0,005 & 0,005 \end{bmatrix}.$$

Przy tych samych założeniach obciążenie prognoz w każdym przypadku było mniejsze niż 0,001 i dlatego nie zostało dokładnie podane.

Rezultaty zaprezentowane w tej części potwierdzają analizy z poprzednich rozdziałów. Mimo znacznego obciążenia oszacowań macierzy przejścia, obciążenie prognoz jest bardzo niewielkie.

## 5. PODSUMOWANIE

Przedstawione w pracy wyniki wskazują, że pomimo znacznego obciążenia oszacowań macierzy przejścia łańcucha Markowa konstruowane na jej podstawie prognozy cechuje niewielkie obciążenie. W związku z tym przy ocenie jakości prognoz większe praktyczne znaczenie będzie mieć ich odchylenie standardowe.

Jednocześnie pokazano, że stosowanie wzoru analitycznego podanego przez Yamamoto i Kunitomo daje dobre przybliżenia obciążenia estymatorów MNK i WUMNK macierzy przejścia. Jednak oceny obciążenia prognoz uzyskane z tego wzoru są zdecydowanie zawyżone i do tym celu wzór ten nie można uznać za przydatny.

**LITERATURA**

- Abadir K., Hadri K., Tzavalis E., [1999], *The Influence of VAR Dimensions on Estimator Biases*, *Econometrica*, Vol. 67(1).
- Acedański J., [2006], *O pewnej metodzie prognozowania przewozów w gospodarce*, [w:] Dittmann P., Krupowicz J. (red.), *Prognozowanie w zarządzaniu firmą*, Prace Naukowe AE we Wrocławiu nr 1112, Wrocław.
- Acedański J., [2008], *Dokładność prognoz struktury uzyskanych za pomocą modeli wektorowej autoregresji*, [w:] Dittmann P., Szandula J. (red.), *Prognozowanie w zarządzaniu firmą*, AE Wrocław, Wydawnictwo Indygo Zahir Media, Wrocław.
- Kendall M., [1951], *Note on Bias in the Estimation of Autocorrelation*, *Biometrika*, Vol. 41.
- Kiviet J., Phillips G., [2005], *Moment Approximation for Least-Squares Estimators in Dynamic Regression Models with a Unit Root*, *Econometrics Journal*, Vol. 8.
- Koźniewska I. (red.), [1980], *Prognozowanie struktury za pomocą łańcuchów Markowa*, SGPiS, Warszawa.
- Lawford S., Stamatiogiannis M., [2004], *The Finite-Sample Effects of VAR Dimensions on OLS Bias, OLS Variance, and Minimum MSE Estimators*, *Journal of Econometrics*, Vol. 148(2).
- Lee T. C., Judge G., Zellner A., [1970], *Estimating the Parameters of the Markov Probability Model from Aggregate Time Series Data*, North-Holland Publishing Company, Amsterdam-Londyn.
- Podgórska M., Śliwka P., Topolewski M., Wrzosek M., [2000], *Łańcuchy Markowa w teorii i zastosowaniach*, SGH, Warszawa.
- Tjostheim D., Paulsen J., [1983], *Bias of some Commonly-Used Time Series Estimates*, *Biometrika*, Vol. 70(2).
- Yamamoto T., Kunitomo N., [1984], *Asymptotic Bias of the Least Squares Estimator for Multivariate Autoregressive Models*, *Annals of the Institute of Statistical Mathematics*, Vol. 36(1).

**A BIAS OF STRUCTURE FORECASTS FROM MARKOV CHAIN MODELS**

The paper investigates a bias of forecasts of structure obtained from Markov chain models when the transition matrix of the chain is estimated using least squares methods and macrodata. The bias comes from a bias of the transition matrix estimators. Based on simulation methods it is shown that despite significant bias in the transition matrix estimates, bias of the forecasts is rather small. It is also acknowledged that analytical equation known from literature gives accurate estimates of the transition matrix bias, but it is not useful for assessing bias of the forecasts. The analysis is illustrated with forecasting of structure of Polish export by the main partners.