

Spis treści

Wprowadzenie	2
1 Opis danych z testów koniunktury	2
2 Metoda transformacji danych	3
3 Transformacja danych z testu koniunktury	6
3.1 Opis przestrzeni znaków	6
3.2 Algorytm transformacji danych	8
4 Wnioski z transformacji dla modelowania	10
5 Dyskretny model Markowa dla struktury odpowiedzi	12
5.1 Konstrukcja przestrzeni znaków dla dyskretnego HMM	12
5.2 Postać modelu oraz wyniki estymacji	14
5.3 HMM dla struktur i MSM dla salda	17
6 Podsumowanie	18
Dodatek A Dane i wyniki obliczeń	20

Wprowadzenie

Zwykle reprezentacja graficzna danych z testów koniunktury ogranicza się do przedstawienia ich dynamiki w postaci wykresu salda. Obserwowanie salda udzielanych odpowiedzi pozbawia nas jednak pełnej informacji o tym, jaki procent respondentów wybierał poszczególne odpowiedzi. Dane z testów koniunktury są bowiem szeregami czasowymi złożonymi z wektorów trójwymiarowych, których elementami są właśnie frakcje respondentów wybierających każdą z trzech możliwości. Jak się okazuje, szczególna konstrukcja tych danych sprawia jednak, że można je przedstawić jako punkty na płaszczyźnie.

Głównym celem pracy będzie zaproponowanie metody transformacji danych z testów koniunktury. Jak się okaże, wiąże się to z istotnymi wnioskami dotyczącymi charakteru tych danych. Co ważne, dzięki zaproponowanej transformacji uzyskujemy również alternatywną metodę ich dyskretyzacji, a więc w dalszej kolejności segmentacji i klasyfikacji. Jest to szczególnie ważne w modelowaniu na podstawie danych z testów koniunktury wahań aktywności gospodarczej i cykli koniunkturalnych.

Po wstępnym opisie danych, w drugiej części formalnie wyprowadzimy sposób na ich transformację. Udowodnimy również, że transformacja ta posiada „dobre” własności. W części trzeciej pokażemy przydatność tej metody dla wnioskowania o charakterze danych z ankiet gospodarczych. W części czwartej zaprezentujemy potencjalną korzyść wynikającą z proponowanego przedstawienia danych dla modelowania przy użyciu ukrytych modeli Markowa. W celach prezentacji przedstawimy ostatecznie również prosty model – dyskretny ukryty model Markowa dla struktury odpowiedzi w ankietach gospodarczych. Dane wykorzystane w tej pracy, jak również wszystkie wyniki obliczeń znajdują się w dodatku.

1 Opis danych z testów koniunktury

Testy koniunktury IRG SGH są metodą badania ankietowego o charakterze jakościowym. Badania są przeprowadzane niezależnie dla szeregu branż. W tej pracy skoncentrujemy się na analizie danych z testów koniunktury w przemyśle. Testy te składają się z dwudziestu pytań dotyczących działalności przedsiębiorstw. Na większość z pytań możliwa jest jedna z trzech odpowiedzi. Wyniki zebranych ankiet są agregowane. W każdym okresie dostajemy zatem trójwymiarowy wektor, którego i -ty element odpowiada odsetkowi respondentów dających i -tą odpowiedź. Wektor taki nazywa się **strukturą odpowiedzi**. Różnicę między pierwszym, a trzecim elementem tego wektora nazywa się natomiast **saldem odpowiedzi**.

Jako przykład zostały wybrane wyniki odpowiedzi na pytanie dotyczące zmian wolumenu produkcji w przemyśle (pytanie 01). Respondent zakreśla jedną z trzech odpowiedzi. Wielkość produkcji w stosunku do poprzedniego miesiąca:

- **A** zwiększyła się
- **B** nie zmieniła się
- **C** zmniejszyła się.

Analizowane w tej pracy dane są danymi miesięcznymi. Zostały zebrane w okresie od marca 1997 do maja 2003 roku. Dysponujemy zatem szeregiem czasowym, będącym ciągiem trójwymiarowych wektorów (struktur odpowiedzi). Szczegółowe zestawienie zawiera dodatek A.

Zwykle prezentuje się wyniki testów w badanym okresie jako wykres salda odpowiedzi. Trzeba jednak zauważyć, że często dla różnych struktur odpowiedzi uzyskujemy to samo saldo (np. dla struktur postaci (20%, 50%, 30%) i (30%, 30%, 40%)). Zatem jeżeli chcemy przedstawić na wykresie szereg złożony ze struktur odpowiedzi musimy zastosować transformację inną niż do salda. Okazuje się, że dane z testów koniunktury mają na tyle szczególny charakter, że rzeczywiście istnieje przedstawienie na płaszczyźnie, które pozwala zachować wzajemne zależności między wszystkimi współrzędnymi struktury odpowiedzi.

2 Metoda transformacji danych

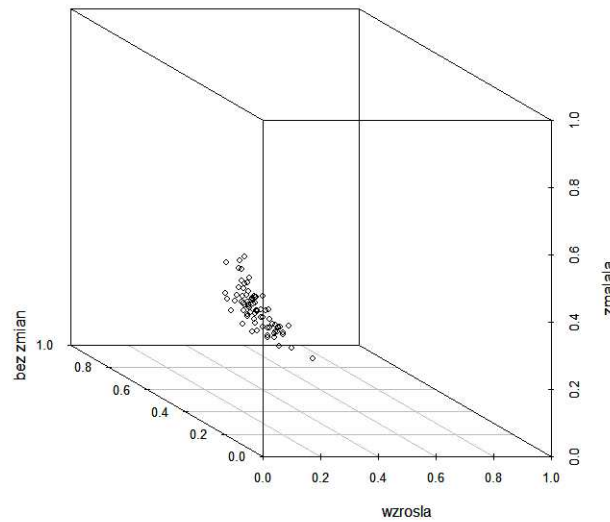
Rozpatrując strukturę odpowiedzi z testów koniunktury, dysponujemy w kolejnych okresach punktami w przestrzeni $x_t \in \mathbb{R}^3$, których kolejnymi współrzędnymi są odsetki odpowiedzi na pytanie o wielkość produkcji. $x_t = (x_{t1}, x_{t2}, x_{t3})$ oraz x_{ti} oznacza odsetek wybierający i -tą odpowiedź w okresie t . Analizowany szereg czasowy można przedstawić w poniższy sposób:

$$\left\{ \begin{pmatrix} x_{11} \\ x_{12} \\ x_{13} \end{pmatrix}, \begin{pmatrix} x_{21} \\ x_{22} \\ x_{23} \end{pmatrix}, \dots, \begin{pmatrix} x_{T1} \\ x_{T2} \\ x_{T3} \end{pmatrix} \right\},$$

gdzie T jest długością rozpatrywanego szeregu czasowego.

Zauważmy, że każdy respondent na każde pytanie w teście koniunktury IRG SGH mógł udzielić jednej i tylko jednej odpowiedzi. Suma elementów każdego wektora struktury odpowiedzi wynosi zatem 100%. Formalniej możemy zapisać, że dla każdej odpowiedzi i oraz dla każdego okresu t zachodzi $x_{ti} \geq 0$ oraz $\sum_{i=1}^3 x_{ti} = 1$. Jest to kluczowe dla konstrukcji transformacji danych z testów koniunktury zaprezentowanej w tej pracy.

Niech bowiem $x = (x_1, x_2, x_3)$ będzie punktem w trójwymiarowej przestrzeni euklidesowej. Jeżeli $\sum_{i=1}^3 x_i = 1$ oraz $x_i \geq 0$ dla każdego i , to x leży w dwuwymiarowym sympleksie w \mathbb{R}^3 będącym kombinacją wypukłą punktów $(1, 0, 0)$, $(0, 1, 0)$ i $(0, 0, 1)$. Łatwo bowiem można zauważyć, że punkt x ma nieujemne współrzędne, które spełniają równanie płaszczyzny P : $x_1 + x_2 + x_3 - 1 = 0$, zatem leży w zbiorze będącym przecięciem płaszczyzny P i \mathbb{R}_+^3 .



Wykres 1: Struktura odpowiedzi na pytanie dotyczące zmian wielkości produkcji w teście koniunktury dla przemysłu IRG SGH. Okres: marzej 1997r. – maj 2003 r. Źródło: opracowanie własne na podstawie danych IRG SGH.

Wykres 1 przedstawia elementy analizowanego szeregu struktury odpowiedzi. Choć z samego rysunku nie jest to bardzo dobrze widoczne, wszystkie punkty leżą na trójkącie o wierzchołkach $(1, 0, 0)$, $(0, 1, 0)$ i $(0, 0, 1)$. Wykorzystajmy teraz powyższe spostrzeżenia w konstrukcji przekształcenia tych punktów na płaszczyznę \mathbb{R}^2 . Polegać ono będzie w istocie na obróceniu rozpatrywanego trójkąta tak, by w całości leżał w \mathbb{R}^2 (rzutowanie na pierwsze dwie współrzędne nas nie interesuje). Formalny sposób konstrukcji takiego przekształcenia pokazuje poniższe twierdzenie.

Twierdzenie 1 *Niech P będzie płaszczyzną w \mathbb{R}^3 rozpinaną przez wektory v_1 i v_2 , a wektory $e_1 = [1, 0]$, $e_2 = [0, 1]$ tworzą bazę w \mathbb{R}^2 . Wtedy dowolne przekształcenie afiniczne $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ takie, że jego część liniowa spełnia układ $f(v_1) = e_1$, $f(v_2) = e_2$ przekształca różnowartościowo płaszczyznę P na \mathbb{R}^2 .*

Dowód: Jeżeli $v \in \text{lin}(v_1, v_2)$ to $f(v) = f(av_1 + bv_2) = af(v_1) + bf(v_2) = ae_1 + be_2 = [a, b]$, gdzie $a, b \in \mathbb{R}$. I na odwrót $[x, y] = xe_1 + ye_2 = xf(v_1) + yf(v_2) = f(xv_1 + yv_2)$ oraz

$xv_1 + yv_2 \in \text{lin}(v_1, v_2)$. Każdy wektor w danej przestrzeni ma jednoznaczne przedstawienie we współrzędnych bazowych. Zatem f przekształca różnowartościowo P na \mathbb{R}^2 . ■

Zwróćmy uwagę najpierw, że ograniczenie się do wektorów w \mathbb{R}^2 nie było potrzebne i służyło jedynie uproszczeniu zapisu przy dowodach. W ogólności możemy rozważać dowolną parę wektorów liniowo niezależnych. Zauważmy również, że płaszczyzna P z twierdzenia jest dana jako $p + \text{lin}(v_1, v_2)$, gdzie p jest dowolnym punktem należącym do P . Wtedy $f(p + \text{lin}(v_1, v_2)) = f(p) + \text{lin}(f(v_1), f(v_2))$, lecz teraz f jest przekształceniem afinicznym, które przy spełnieniu warunków $f(v_1) = e_1$, $f(v_2) = e_2$ zadane jest jednoznacznie przez wartość na jednym konkretnym punkcie p . Te dwie uwagi łączą się w poniższym twierdzeniu.

Twierdzenie 2 *Każde różnowartościowe przekształcenie afiniczne $f : P \rightarrow \mathbb{R}^2$, gdzie $P \subset \mathbb{R}^3$, jest jednoznacznie określone przez swoje wartości na trzech niewspółliniowych punktach.*

Dowód: Niech $p_1, p_2, p_3 \in P$ będą niewspółliniowe wtedy wektory $v_1 = p_2 - p_1$, $v_2 = p_3 - p_1$ są liniowo niezależne i rozpinają P . Znajomość $f(v_1)$, $f(v_2)$ oraz $f(p_1)$ gwarantuje nam jednoznaczność na mocy wcześniejszego rozumowania. ■

Chcemy zatem przekształcić trójkąt w \mathbb{R}^3 na trójkąt w \mathbb{R}^2 , tak by móc dane z testów koniunktury IRG SGH przedstawić jako punkty na płaszczyźnie. Nie interesuje nas jednak dowolne takie przekształcenie (jest ich nieskończenie wiele), ponieważ chcemy interpretować wynikowy wykres jako rzeczywisty obraz badanego zjawiska. Minimalnym, ale i wystarczającym warunkiem będzie to, by zachowywały się odległości między punktami. W ten sposób umożliwimy interpretację ułożenia się struktur względem siebie i wewnątrz rozpatrywanego trójkąta.

Bez trudu można sprawdzić, że przekształcenie, dla którego rozpatrywanemu trójkątowi odpowiada trójkąt w \mathbb{R}^2 o równym polu i odległościach między wierzchołkami, spełnia powyższy warunek. Przyjmijmy na przykład, że szukane przez nas przekształcenie f spełnia: $f(1, 0, 0) = (\sqrt{2}, 0)$, $f(0, 1, 0) = (0, 0)$ oraz $f(0, 0, 1) = (\frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2})$. Można pokazać, że pole obu trójkątów jest równe i wynosi $\frac{\sqrt{3}}{2}$. Wzór na to przekształcenie znajdujemy ad hoc, gdyż f jest afiniczne i spełnione jest: $f(x, y, z) = xf([1, 0, 0]) + yf([0, 1, 0]) + zf([0, 0, 1]) = (\sqrt{2}x + \frac{\sqrt{2}}{2}z, \frac{\sqrt{6}}{2}z)$.

Z algebry liniowej wiemy, że metryka na płaszczyźnie $x_1 + x_2 + x_3 - 1 = 0$ (a więc i na rozpatrywanym trójkącie) jest wyznaczona przez iloczyn skalarny dziedziczony z \mathbb{R}^3 . Jeżeli za wektory bazowe tej płaszczyzny przyjmiemy $v_1 = p_1 - p_2 = (1, -1, 0)$ oraz $v_2 = p_3 - p_2 = (0, -1, 1)$ (wówczas $(0, 1, 0)$ jest środkiem układu współrzędnych) to macierzą tego iloczynu

skalarne będzie:

$$\begin{pmatrix} \langle v_1, v_1 \rangle & \langle v_1, v_2 \rangle \\ \langle v_2, v_1 \rangle & \langle v_2, v_2 \rangle \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}, \quad (1)$$

gdzie $\langle \cdot, \cdot \rangle$ oznacza standardowy euklidesowy iloczyn skalarny w \mathbb{R}^3 . Można łatwo pokazać, że macierz tego iloczynu skalarnego odpowiada macierzy standardowego iloczynu skalarnego w \mathbb{R}^2 , co wystarczy by metryka była zachowana. Macierzą, która wyznacza pomiędzy nimi kongruencję jest macierz zamiany bazy $\{v_1, v_2\} \rightarrow \{e_1, e_2\}$.

Nie będziemy wchodzić głębiej w to zagadnienie, które należy raczej do teorii algebry liniowej. Poniższy przykład ilustruje przedstawione rozumowanie.

Przykład 1 Punktem $p = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$, który leży na płaszczyźnie odpowiada wektor $[\frac{1}{3}, -\frac{2}{3}, \frac{1}{3}]$, jako, że punkt $(0, 1, 0)$ jest tutaj środkiem układu. W bazie $\{v_1, v_2\}$ będzie to więc $(\frac{1}{3}, \frac{1}{3})$. Jego odległość od $(0, 1, 0)$ na płaszczyźnie P jest dana poprzez równość:

$$|p - p_2| = \sqrt{\langle p - p_2, p - p_2 \rangle} = \left(\left(\frac{1}{3}, \frac{1}{3} \right) \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} \left(\frac{1}{3}, \frac{1}{3} \right)^T \right)^{\frac{1}{2}} = \frac{\sqrt{6}}{3} \quad (2)$$

Wszystko to można pokazać mniej formalnie ale za to prościej. Jak już wspomnieliśmy, każdy punkt naszych trójkątów jest kombinacją wypukłą wierzchołków trójkąta (wagi nazywamy współzrędnymi barycentrycznymi). Można pokazać, że współzrędnne barycentryczne przy naszym przekształceniu się zachowują. Zatem punktem $p = \frac{1}{3}(1, 0, 0) + \frac{1}{3}(0, 1, 0) + \frac{1}{3}(0, 0, 1)$ z powyższego przykładu odpowiada punkt $f(p) = \frac{1}{3}(\sqrt{2}, 0) + \frac{1}{3}(0, 0) + \frac{1}{3}(\frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2}) = (\frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{6})$. Jego odległość od $f((0, 1, 0)) = (0, 0)$ na płaszczyźnie już bardzo łatwo obliczyć i jak widać wynik zgadza się z tym uzyskanym wcześniej.

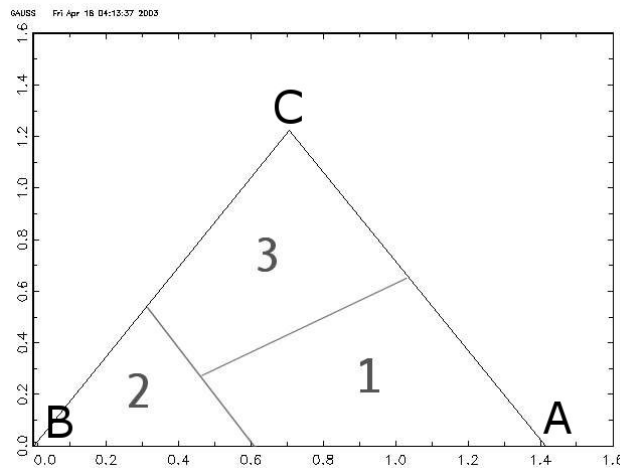
3 Transformacja danych z testu koniunktury

3.1 Opis przestrzeni znaków

Zaproponowana w poprzednim punkcie transformacja danych zapewni, że będziemy mogli przedstawić obserwacje w \mathbb{R}^2 i wszystkie one leżeć będą w trójkącie o wierzchołkach $A(\sqrt{2}, 0)$, $B(0, 0)$ oraz $C(\frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2})$. Nie uwzględniając specyfiki badanego zjawiska możemy podzielić rozpatrywany trójkąt na łatwo interpretowalne podobszary (patrz rysunek 2). Obszary te zostały wydzielone na bazie środka ciężkości badanego trójkąta czyli punktu $(\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. Odcinek oddzielający obszary 1 i 3 stanowi część dwusiecznej, a odcinek wydzielający obszar 2 jest równoległy do boku trójkąta. Prześledzenie wcześniejszych obliczeń prowadzi następujących wniosków:

- punkt $A(\sqrt{2}, 0)$ odpowiada sytuacji, kiedy 100% respondentów odpowiedziało, że ich produkcja się zwiększyła,
- punkt $B(0, 0)$ oznacza, że 100% odpowiada, że produkcja się nie zmieniła,
- punkt $C(\frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2})$ odpowiada sytuacji, gdy 100% respondentów odpowiedziało, że ich produkcja zmalała,

Wszystkie punkty wewnątrz rozpatrywanego trójkąta, są kombinacjami wypukłymi jego wierzchołków. Odcinek łączący wierzchołki B i C będziemy zatem nazywać „**bokiem zerowego wzrostu**”, gdyż opisują go kombinacje wierzchołków typu $(0, t, 1 - t)$, gdzie $t \in [0, 1]$. Można również łatwo zauważyć, że, że dwusieczna trójkąta odpowiada kombinacjom typu $(t, 1 - 2t, t)$, gdzie $t \in [0, \frac{1}{2}]$. Każdy odcinek zawarty w dwusiecznej będziemy zatem nazywać **linią zerowego salda**. Widać teraz, że obserwacje leżące odpowiednio w obszarach 1, 2 i 3 możemy interpretować w poniższy sposób:



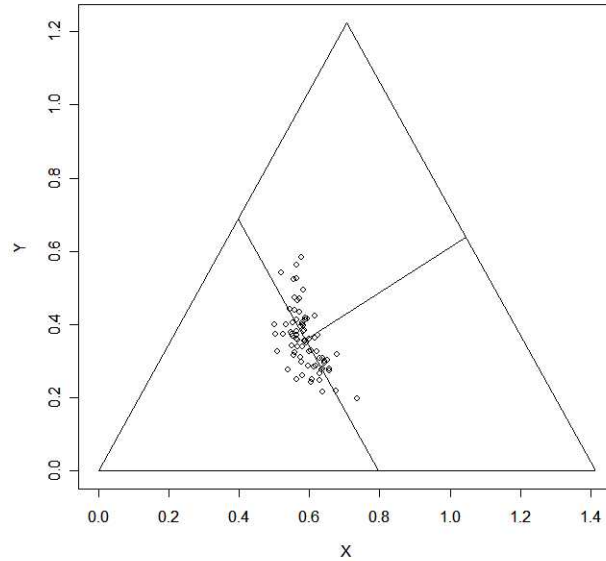
Rysunek 2: Podział rozpatrywanego trójkąta na trzy podobszary. Źródło: opracowanie własne.

obszar(1) Większa liczba respondentów deklaruje, że ich produkcja wzrosła.

obszar(2) Struktura odpowiedzi wykazuje stabilizację wielkości produkcji przemysłowej

obszar(3) Przedsiębiorcy w większości stwierdzają, że ich produkcja spadła.

Pokażemy niżej, że dokonanie tego podziału a priori i bez znajomości wykresu struktury, jest nieuzasadnione i może prowadzić do mylnych wniosków.



Wykres 3: Transformowana struktura odpowiedzi na pytanie dotyczące zmian wielkości produkcji w teście koniunktury dla przemysłu IRG SGH. Okres: marzej 1997r. – maj 2003 r.. Źródło: opracowanie własne.

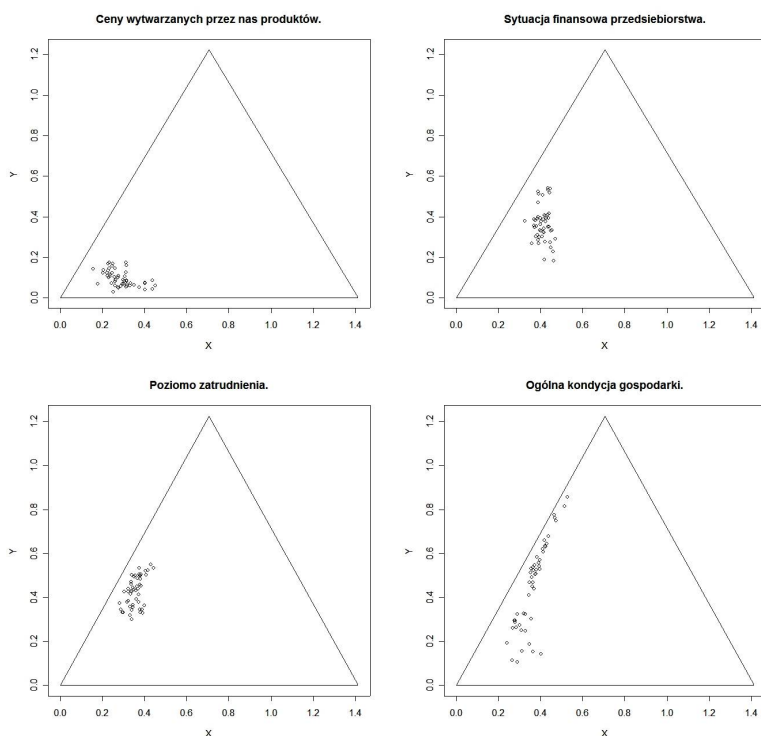
3.2 Algorytm transformacji danych

W poprzedniej części zaproponowaliśmy wzór na przekształcenie, które pozwoli nam dokonać transformacji danych z testów koniunktury. Istotą algorytmu transformacji danych jest, by dla każdego wektora struktury z naszego szeregu czasowego postaci (x_{t1}, x_{t2}, x_{t3}) obliczyć wartość współrzędnych nowego punktu $(y_{t1}, y_{t2}) = f(x_{t1}, x_{t2}, x_{t3}) = (\sqrt{2}x_{t1} + \frac{\sqrt{2}}{2}x_{t3}, \frac{\sqrt{6}}{2}x_{t3})$. I tak na przykład strukturze odpowiedzi w marcu 1997 roku czyli $(0, 336; 0, 438; 0, 226)$ odpowiada dwuwymiarowy punkt $(\sqrt{2} \cdot 0, 336 + \frac{\sqrt{2}}{2} \cdot 0, 226; \frac{\sqrt{6}}{2} \cdot 0, 226) = (0, 635; 0, 277)$. Dane dotyczące struktur odpowiedzi w teście IRG SGH w przemyśle oraz ich dwuwymiarowe odpowiedniki zawiera tabela w dodatku A.

Na rysunku 3, który zawiera transferowane struktury odpowiedzi, dostrzegamy, że punkty odpowiadające obserwacjom w poszczególnych okresach skupiają się na dość małym obszarze¹. Oczywiście teoretycznie punkty mogą pojawić się w obrębie całego trójkąta. Obserwowane zagęszczenie jest istotną cechą badanego zjawiska. Wiąże się również z małymi odchyleniami punktów od średniej.

Powstaje teraz dodatkowe pytanie, co sprawia, że ułożenie struktur odpowiedzi jest takie a nie inne. Czy jest to ogólnie cecha danych z testów koniunktury, czy też „trójkąty” będą się różniły pomiędzy poszczególnymi pytaniami.

¹Warto przypomnieć, że rozpatrywane przekształcenie z przestrzeni trójwymiarowej nie zmieniało odległości między punktami.



Wykres 4: Wykres struktury odpowiedzi na pytania z testów koniunktury w przemyśle IRG SGH w okresie marzec 1997 r. – maj 2001 r. Kolejno: (1) Ceny wytwarzanych przez nas wyrobów. (2) Sytuacja finansowa przedsiębiorstwa. (3) Poziomo zatrudnienia. (4) Ogólna sytuacja gospodarki polskiej. Źródło: opracowanie własne.

Pierwszą hipotezą, która się nasuwa jest stwierdzenie, że taki szczególny wygląd wykresu struktury w przypadku pytania o wielkość produkcji odpowiada psychologii respondentów. Analiza struktury odpowiadającej innym pytaniom z ankiet gospodarczych (patrz wykres 4) dotyczących przemysłu, pozwala sądzić, że zagęszczenie, jakie obserwowaliśmy w pytaniu o wielkość produkcji, obserwujemy także w ogólnym przypadku. Jedyne znacząco innym „trójkątem” jest ten przedstawiający odpowiedzi na pytanie o zmiany w ogólnej kondycji gospodarki. Jest to o tyle charakterystyczne, że z przedstawionych pytań jest to jedyne, które dotyczy otoczenia przedsiębiorstwa, a nie jego bezpośrednio.

Warto również zauważyć, że względnie duże rozstrzelenie punktów, które obserwujemy w tym przypadku przebiega w okolicach odcinka łączącego punkty $(0, 0)$ i $(\frac{\sqrt{2}}{2}, \frac{\sqrt{6}}{2})$. Większe przyciąganie od strony „boku zerowego wzrostu”, któremu podlegają obserwowane punkty, jest również cechą innych wykresów na rysunku 4.

Są zatem dwie ważne cechy przedstawionych wykresów struktur, które nasuwają się od razu. Jedną jest obserwowane w większości przypadków duże zagęszczenie punktów, a drugą to, że znajdują się stosunkowo blisko „boku zerowego wzrostu”. Wydaje się, że można te cechy tłumaczyć zachowawczym zachowaniem przedsiębiorców, może nawet niechętnym przyznawaniem się do pozytywnych zmian w firmie oraz ogólnie odczuwalnej w polskiej gospodarce obniżce aktyw-

ności gospodarczej. Wyjaśnienie zagęszczenie punktów można szukać także w samej strukturze przemysłu, który jest mocno niejednorodny. Oznaczać to może na przykład, że zmiany pozytywne i negatywne w pewnym stopniu się niwelują, w momencie gdy jedna branża notuje pozytywne zmiany, druga może odczuwać zastój.

4 Wnioski z transformacji dla modelowania koniunktury przy pomocy modeli Markowa

Modelując wahania gospodarcze przy użyciu danych z testów koniunktury, często korzysta się z klasycznych łańcuchów Markowa (patrz: Podgórska, Decewicz[4]). Przestrzeń stanów ŁM tworzy się wtedy poprzez podzielenie możliwych sald na przedziały o zadanej długości (dyskretyzacja salda).

Zauważmy, że dwusieczna trójkąta w przestrzeni przetransformowanych danych, wychodząca z początku układu współrzędnych jest w istocie linią zerowego salda (wagi wierzchołka A i C pozostają równe). Zatem jeżeli podzielimy nasz trójkąt na dwa obszary względem dwusiecznej, to tak jakbyśmy zbudowali dwustanowy ŁM, gdzie proces znajduje się w pierwszym stanie przy ujemnym saldzie, a w drugim w przeciwnym przypadku. Dyskretyzacja salda odpowiada w związku z tym po prostu podzieleniu rozpatrywanego trójkąta na podobszary względem większej liczby siecznych kąta przy wierzchołku $(0,0)$. W tym miejscu łatwo zauważyć, co dała nam opisana wyżej transformacja danych. Mianowicie jesteśmy teraz w stanie bardziej świadomie dzielić trójkąt na podobszary i budować przestrzeń stanów ŁM.

Wyniki testu koniunktury gospodarczej bywają analizowane także przy pomocy nieklasycznych modeli Markowa. Najczęściej w tym kontekście pojawiają się ukryte modele Markowa (HMM) lub szersza klasa przełącznikowych modeli Markowa (MSM). W pracy tej nie będziemy formalizować wymienionych metod. Są one opisane w szeregu publikacji ekonometrycznych, inżynierskich czy statystycznych. Najbardziej kompletne omówienie tego typu procesów znajduje się w artykule Ephraima i Merhava[1]. Wprowadzenie do przełącznikowych modeli Markowa znajduje się w podręczniku Hamiltona[2].

W literaturze ekonometrycznej HMM z ciągłą przestrzenią znaków traktuje się po prostu jako modele typu MSM. Przedstawimy krótko istotę tej szerokiej klasy modeli na przykładzie dyskretnego ukrytego modelu Markowa. W skrótowy sposób można opisać go jako dwuwymia-

rowy proces stochastyczny

$$\left\{ \left(\begin{array}{c} Y_1 \\ X_1 \end{array} \right), \left(\begin{array}{c} Y_2 \\ X_2 \end{array} \right), \left(\begin{array}{c} Y_3 \\ X_3 \end{array} \right), \dots \right\},$$

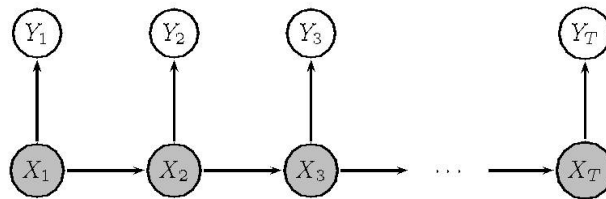
gdzie:

- (X_t) jest ukrytym łańcuchem Markowa o skończonej przestrzeni stanów $S_X = \{i_1, i_2, \dots, i_N\}$.
- (Y_t) jest obserwowanym ciągiem generowanych przez układ „znaków” z alfabetu $S_Y = \{s_1, s_2, \dots, s_M\}$.

Zakładamy ponadto, że prawdopodobieństwo emisji znaku s zależy wyłącznie od stanu, w którym znajduje się ŁM i jest niezmiennicze w czasie:

$$\forall_{t \in \mathbb{N}} P(Y_t = s | X_1, \dots, X_t, Y_1, \dots, Y_{t-1}) = P(Y_t = s | X_t) = P(Y_1 = s | X_1)$$

Dobrze jest utożsamiać ukryty proces Markowa z odpowiadającym mu zorientowanym acyklicznym grafem stochastycznym. Na rysunku 5 zaciemnione zostały zmienne ukryte, a strzałki mówią o bezpośredniej zależności probabilistycznej między zmiennymi.



Rysunek 5: Ukryty proces Markowa jako acykliczny zorientowany graf stochastyczny. Źródło: opracowanie własne.

Duża popularność wymienionych dwóch metod w modelowaniu koniunktury wynika głównie z dość interesujących walorów interpretacyjnych. Aktywność gospodarcza nie jest bowiem zmienną bezpośrednio obserwowalną. Poszczególne agregaty, jak PKB czy produkcja przemysłowa nie oddają w pełni kształtowania się stanu koniunktury. Dlatego właśnie traktuje się w tych modelach aktywność gospodarczą jako zmienną ukrytą² i szacuje taki dwustanowy ukryty ŁM, gdzie stany odpowiadają poziomowi aktywności gospodarczej (niski, wysoki). Jako zmienne obserwowalne bierze się często zmiany logarytmów kwartalnego PKB.

W przypadku modelowania z użyciem danych z testów koniunktury, modele możemy podzielić ze względu na przestrzeń znaków (wartości procesu widzialnego) na ciągłe i dyskretne

²W ogólności jako zmienne ukryte możemy modelować nie tylko zmienne nieobserwowalne bezpośrednio, ale także zmienne, których dokładne zaobserwowanie wiąże się z dużymi nakładami finansowymi, czasowymi itp.

HMM, a ze względu na transformację danych na modele dla struktury i modele dla sald. Najczęściej wykorzystywane są ciągłe HMM dla sald. Są to głównie modele przełączanych rozkładów normalnych bądź przełącznikowej autoregresji. Są to sposoby proste i zapewniają niską liczbę parametrów do oszacowania.

W modelowaniu struktury odpowiedzi wykorzystuje się natomiast niemal wyłącznie dyskretne HMM, ponieważ nie ma oczywistego sposobu zdefiniowania ciągłej funkcji gęstości na przestrzeni znaków (czyli na rozpatrywanym trójkącie), która zależna by była od małej liczby parametrów. Najczęściej zatem w modelowaniu struktury wykorzystuje się przełączanie rozkładów dwu- i trójmianowych (patrz: Podgórska, Decewicz[4]).

Istnieją ważne powody, dla których możemy chcieć sięgać po dyskretne HMM. Ciągłe jeszcze sprawa weryfikacji statystycznej ukrytych modeli Markowa jest sprawą otwartą. Odczuwa się duży brak wiarygodnych testów na istotność parametrów modelu, na jego specyfikację. Jeżeli próbujemy modelować ukryty proces zarówno w przypadku ciągłej, jak i dyskretnej przestrzeni znaków i w obu przypadkach wyniki są podobne, może to być silny argument za tym, że rzeczywiście ukryty proces istnieje.

5 Dyskretny model Markowa dla struktury odpowiedzi

5.1 Konstrukcja przestrzeni znaków dla dyskretnego HMM

Jako przykład zaprezentujemy prosty model – dyskretne HMM dla struktur. Na mocy tego co zostało wcześniej powiedziane, zależy nam na tym by podzielić ciągłą przestrzeń znaków na podobszary, które stworzą alfabet dla dyskretnego HMM. Ciągłą przestrzenią znaków w przypadku struktury jest oczywiście rozpatrywany trójkąt³.

Powstaje pytanie o to, jakimi przesłankami kierować się przy wyborze podobszarów. Zagęszczanie się danych powoduje, że obszary te powinny być odpowiednio „duże”, tak by prawdopodobieństwo trafienia w nie (na podstawie danych historycznych) nie było zerowe. Żadnego miejsca na trójkącie *a priori* nie możemy odrzucić, chyba, że zależy nam jedynie na modelu diagnostycznym. Dodatkowym ograniczeniem na wybór alfabetu w modelowaniu ekonometrycznym powinna być łatwa interpretacja ekonomiczna wyznaczonych podobszarów, bo tylko w ten sposób HMM zyska walory poznawcze.

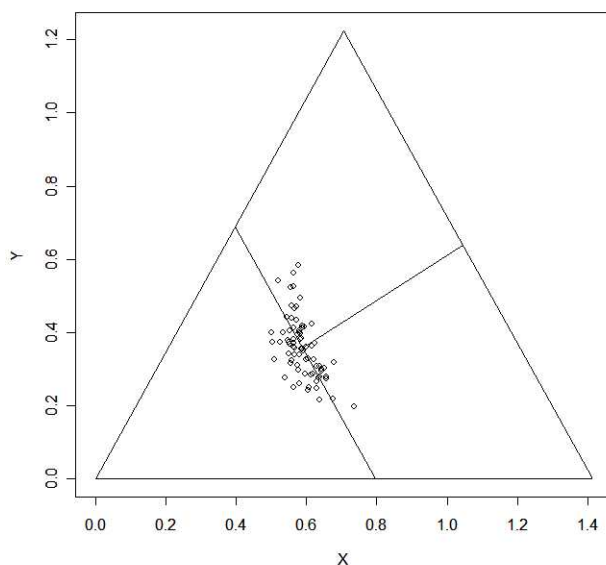
Przedstawiliśmy już wcześniej intuicyjny sposób podzielenia trójkąta na podobszary (patrz rysunek 2). Nie jest to jednak sposób, który aplikuje się w ogólnych przypadkach. Powierzchow-

³W przypadku modelowania sald jest to odcinek $[-1, 1]$.

na nawet inspekcja wykresów na rysunku 4 powinna nas przekonać, że dla niektórych pytań z ankiet istnieją podobszary, które pozostaną puste. Poza tym jest to sposób podziału, który w żadnym wypadku nie oddaje charakteru danych.

Co zatem należy zrobić, by podział oddawał charakter danych? Odpowiednim rozwiązaniem wydaje się przesunięcie punktu odniesienia, którym wcześniej był środek ciężkości trójkąta, na środek ciężkości rozpatrywanych danych – czyli na średnią. W ten sposób punkty leżące bliżej wierzchołka $A(\sqrt{2}, 0)$ niż średnia charakteryzują okresy o względnie większej aktywności gospodarczej. Podobną argumentację możemy powtórzyć dla dwóch pozostałych wierzchołków.

Prostą wydzielającą obszar 2 poprowadzimy przez punkt będący średnią, a więc punkt $(0, 269; 0, 438; 0, 293)$, któremu w dwuwymiarowej przestrzeni będzie odpowiadał punkt $(x_0, y_0) = (0, 588; 0, 358)$. Prosta ta opisana jest wzorem $y = -\sqrt{3}x + y_0 + \sqrt{3}x_0$. Druga interesująca nas prosta, która rozdziela obszary 1 i 3, również przechodzi przez punkt (x_0, y_0) wychodząc z wierzchołka B i ma oczywiście wzór $y = \frac{y_0}{x_0}x$.



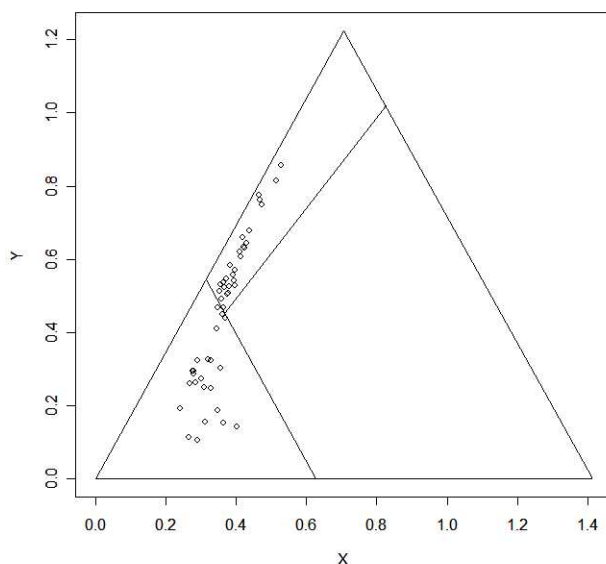
Wykres 6: Struktura odpowiedzi na pytanie o zmiany w wielkości produkcji wraz z naniesionymi liniami podziału na podobszary według metody poprawki na średnią. Źródło: opracowanie własne.

W celu przyporządkowania znaku dla każdej obserwacji, musimy dla każdego punktu $(y_{t1}, y_{t2}) = f(x_{t1}, x_{t2}, x_{t3})$ sprawdzić znak wyrażenia $y_{t2} + \sqrt{3}y_{t1} + y_0 + \sqrt{3}x_0$. Jeżeli znak jest ujemny to badany punkt znajduje się poniżej prostej $y = -\sqrt{3}x + y_0 + \sqrt{3}x_0$, a więc w obszarze 2. Jeżeli znak jest dodatni to w dalszej kolejności badamy znak wyrażenia $y_{t2} - \frac{y_0}{x_0}y_{t1}$. Jeżeli jest dodatni to punkt (y_{t1}, y_{t2}) znajduje się w obszarze 3. W pozostałych przypadkach znajduje się w

obszarze 1⁴.

I tak na przykład w marcu 1997 roku mieliśmy $(y_{t1}, y_{t2}) = (0,635; 0,277)$ (patrz dodatek A). Zachodzi zatem nierówność $y_{t2} + \sqrt{3}y_{t1} + y_0 + \sqrt{3}x_0 = 0,277 + \sqrt{3} \cdot 0,635 + 0,358 + \sqrt{3} \cdot 0,588 = 2,75 > 0$. Musimy więc zbadać znak kolejnego wyrażenia. Mamy $y_{t2} - \frac{y_0}{x_0}y_{t1} = -0,11 < 0$, zatem rozpatrywany punkt znajduje się w obszarze 1.

Wykres struktury odpowiedzi na pytanie o zmiany wielkości produkcji w teście koniunktury w przemyśle IRG SGH wraz z podziałem na obszary zawiera wykres 6. Trzeba jednak zauważyć, że identyczne podejście może prowadzić do zaskakujących rezultatów w przypadku innych pytań z testu koniunktury. W szczególności dla pytania o ogólną kondycję gospodarki (wykres 7) mimo zastosowania poprawki na średnią, ciągle jeden podobszar pozostaje pusty. W przypadku zadania diagnostycznego dla naszego modelu powoduje to całkowite wykluczenie obszaru 1 z szacowanego modelu.



Wykres 7: Struktura odpowiedzi na pytanie o zmiany w ogólnej kondycji gospodarki wraz z naniesionymi liniami podziału na podobszary według metody poprawki na średnią. Źródło: opracowanie własne.

5.2 Postać modelu oraz wyniki estymacji

Ukrytymi stanami zaproponowanego w tej części HMM będą dwa stany $S_X = \{1, 2\}$. Alfabet dany jest przez podział trójkąta $S_Y = \{znak(1), znak(2), znak(3)\}$. Ukryte stany chcielibyśmy

⁴Pozostaje tylko zdecydować się, co jeżeli wyrażenia będą zerowe. W badanej próbie takie zdarzenie nie nastąpiło, wydaje się jednak, że w takim przypadku wystarczy arbitralnie przyjąć którąś z dwóch możliwości, sprawdzając wszakże, czy nie ma to istotnego wpływu na wynik oszacowań.

interpretować, odpowiednio, jako stan niższej i wyższej aktywności gospodarczej. Interpretacja znaków jest taka sama jak w poprzednim przypadku. Znak 1 będzie świadczył o rosnącej aktywności gospodarczej w przemyśle, znak 3 o malejącej aktywności, a punkty z obszaru 2 będziemy interpretować jako świadczące o stabilizacji produkcji.

Stawiamy więc hipotezę, że stan aktywności gospodarczej wpływa na strukturę odpowiedzi w testach koniunktury. Wpływ ten ma oczywiście charakter stochastyczny, a stan drugi będzie odpowiadał okresom, gdy przedsiębiorcy lepiej oceniają kondycję gospodarki.

Problem szacowania parametrów ukrytego modelu Markowa powyższego przypadku jest rozwiązywany przez algorytm Bauma–Welcha, który dokładnie opisują na przykład MacDonald i Zucchini[3]. Naszą uwagę skupiają przede wszystkim wartości parametrów macierzy emisji i przejść. Chcielibyśmy mianowicie, by macierz emisji miała jasną interpretację. Macierz przejść powinna natomiast odzwierciedlać cechy poszczególnych faz cykli koniunkturalnych. Jeżeli rzeczywiście chcemy mówić o okresach większej i mniejszej aktywności gospodarczej, musimy oczekiwać, że prawdopodobieństwa pozostania w stanach (p_{11} i p_{22}) będą odpowiednio duże (w praktyce powyżej 0,5).

Zgodnie z przewidywaniami pierwszy znak z największym prawdopodobieństwem jest emitowany w stanie charakteryzującym się wyższymi ocenami stanu gospodarki (stan drugi). Jest to jednak stan najmniej prawdopodobny⁵, mimo poprawki na średnią. Oznacza to, że respondenci znacznie częściej dostrzegają negatywne tendencje w wolumenie produkcji.

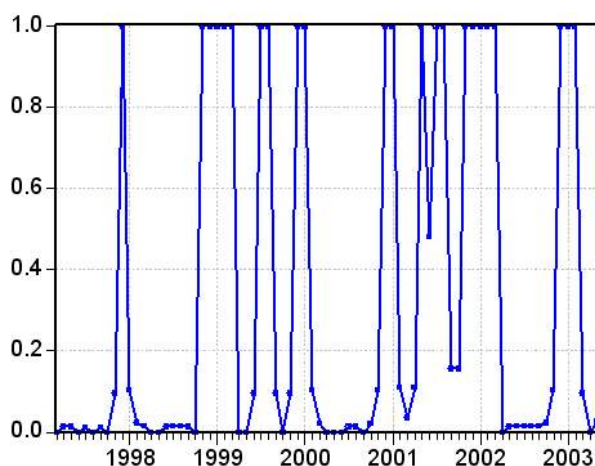
Tabela 1: Oszacowanie macierzy emisji, macierzy przejścia oraz rozkładu granicznego rozpatrywanego HMM.

$\hat{\pi}_{ij}$	Obszar 1	Obszar 2	Obszar 3	Macierz przejścia		δ_i
Stan pierwszy	0	0,09	0,91	0,656	0,344	0,342
Stan drugi	0,322	0,678	0	0,179	0,821	0,658
$P(Y = i)$	0,212	0,477	0,311			

Zwracają jeszcze uwagę zerowe prawdopodobieństwa wyemitowania znaku 1 w stanie pierwszym i znaku 3 w stanie drugim. Oznacza to, że model traktuje wszystkie obserwacje leżące w pierwszym obszarze jako wyemitowane ze stanu drugiego (wyższa aktywność gospodarcza), a te w obszarze trzecim jako wyemitowane w stanie pierwszym. Co więcej wyemitowanie znaku 2 ze stanu pierwszego wiąże się z bardzo niskim prawdopodobieństwem ($P(Y_t = 2|X_t = 1) = 0,09$).

⁵Bezwzględne prawdopodobieństwa obliczamy według wzoru $P(Y = j) = \sum_i P(Y = j, X = i) = \sum_i \pi_{ij} \cdot \delta_i$. Wyniki zawiera tabela 1.

Wysokie prawdopodobieństwa pozostania w stanie ($p_{11} = 0,656$ oraz $p_{22} = 0,821$) oraz zgodna z intuicją postać szacowanej macierzy emisji wskazują, że HMM dla struktury może mieć dobre walory interpretacyjne. Na podstawie prawdopodobieństw pozostania w stanie jesteśmy w stanie obliczyć średni czas przebywania w stanie, który dla pierwszego stanu wynosi 2,9, a dla drugiego 5,6 miesiąca. Potwierdza to obserwację dotyczącą cykli koniunkturalnych, która stwierdza przewagę długości okresów wyższej, nad długością okresów niższej aktywności gospodarczej. Wydaje się zatem, że ukryte stany rozważanego HMM, mogą być interpretowane, jako stany krótkookresowych wahań w aktywności gospodarczej.



Wykres 8: Wyglądzone prawdopodobieństwa przebywania w pierwszym stanie na podstawie HMM dla struktury odpowiedzi. Źródło: opracowanie własne.

Wykres 8 przedstawia wygładzone prawdopodobieństwa przebywania HMM w pierwszym stanie, który charakteryzuje się niższą oceną aktywności gospodarczej przez przedsiębiorców. W okresach, gdy wykres przebiega blisko jedynki, przemysł z dużym prawdopodobieństwem znajdował się w stanie, gdy przedsiębiorcy gorzej oceniają stan aktywności w przemyśle. Zauważyć należy, że w okresie zimowym, a dokładniej w pierwszym kwartale każdego roku, występuje zwiększone prawdopodobieństwo przebywania w drugim stanie. Jest to oczywiście wpływ czynnika sezonowego. Wydaje się więc, że o ile wahania krótkookresowe są bardzo dobrze odzwierciedlone na wykresie wygładzonych prawdopodobieństw, to wahania średniookresowe powinny być badane na wstępie przerobionym szeregu, które powinno obejmować między innymi pozabawienie szeregu wahań sezonowych. Głębsza recesja, jak wynika w wykresie, dotknęła przemysł w 1999 oraz w 2001 roku.

5.3 HMM dla struktur i MSM dla salda

Wymieniając potencjalne korzyści ze stosowania dyskretnych HMM, zwróciliśmy uwagę na fakt niewielkiej liczby testów będących w stanie weryfikować hipotezy związane z modelami tej klasy. Porównywanie HMM ciągłych i dyskretnych dla tego samego ukrytego procesu, jest jedną z możliwości badania takich modeli. W tej części porównamy przedstawiony w poprzedniej części prosty model dla struktur z ciągłym HMM dla salda odpowiedzi.

HMM dla struktur ma jedynie trzejelementową przestrzeń znaków. W przypadku MSM dla salda, przestrzeń znaków jest mocy continuum. Są to więc modele w znaczącym stopniu różne. Gdyby jednak ich estymacja dawała podobne wyniki, mogłoby to potwierdzać tezę, że istnieje ukryty proces generujący krótkookresowe wahania w ocenach aktywności gospodarczej.

MSM dla sald jest to model przełącznikowy Markowa w najprostszej z wersji, a więc markowska mieszanina dwóch rozkładów normalnych. Model zapisujemy jako równanie postaci:

$$y_{s_t} = \mu_{s_t} + \epsilon_{s_t,t}, \quad (3)$$

gdzie $s_t = \{1, 2\}$ oraz $\epsilon_{s_t,t} \sim N(0, \sigma_{s_t})$.

W pierwotnej wersji zaprezentowanej w artykule Podgórskiej, Decewicz[4], model nie zakładał identyczności wariancji między stanami, jednak ich oszacowania dla badanego okresu praktycznie się nie różniły. Szacowane wartości parametrów modelu zawiera tabela 2.

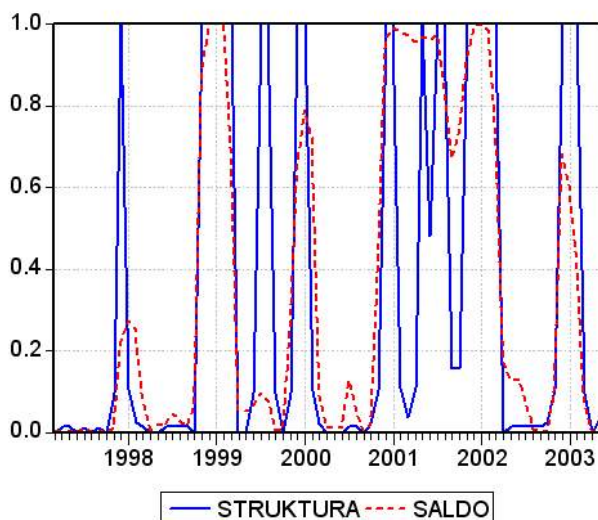
Tabela 2: Wyestymowana macierz parametrów MSM dla salda, okres marzec 1997–maj 2003.

Reżim 1	Reżim2	Macierz przejścia	
$\hat{\mu}_1 = -13,27$	$\hat{\mu}_2 = 3,557$	0,812	0,188
$\hat{\sigma} = 8,82$	$\hat{\sigma} = 8,82$	0,103	0,897

Wykres 9 przedstawia wykres wygładzonych prawdopodobieństw przebywania modelu dla salda w pierwszym stanie skonfrontowany z odpowiednimi prawdopodobieństwami HMM dla struktury.

Choć oszacowane macierze przejścia dla obu modeli różnią się znacząco, to z wykresu wynika, że może istnieć ukryty proces generujący wahania. Trzeba podkreślić, że HMM dla struktur ma tylko trzy znaki, a więc jest to granicznie prosty model — zbieżność wykresów wygładzonych prawdopodobieństw jest więc zaskakująca. Jedyna rzucająca się w oczy różnica jest taka, że wykres wygładzonych prawdopodobieństw modelu dla struktury odpowiedzi charakteryzuje się dużo większą zmiennością. Wahania wykresów występują w tych samych momentach jednak w przypadku modelu dla struktury są one dużo gwałtowniejsze. Fakt ten nie powinien dziwić,

gdyż wydaje się, że modele dla struktury odpowiedzi w ogólności powinny się charakteryzować większą zmiennością, gdyż często dla tych samych sald mamy do czynienia z różną strukturą, nigdy odwrotnie.



Wykres 9: Wygładzone prawdopodobieństwa przebywania w pierwszym stanie na podstawie HMM dla struktury odpowiedzi zestawione z wygładzonymi prawdopodobieństwami dla modelu przełącznikowego dla sald. Źródło: opracowanie własne.

Można przypuszczać, że wraz z dodawaniem kolejnych podobszarów trójkąta wydzielanych za pomocą siecznych, powinniśmy uzyskiwać wyniki coraz bliższe wynikom uzyskanym w modelu dla sald. Mimo wszystko tak duże podobieństwo uzyskane nawet w tak prostym przypadku pozwala sądzić, że istnieje pewien ukryty proces rządzący zmianami stanów aktywności gospodarki. Drugim, istotnym dla tej pracy wnioskiem, może być stwierdzenie, że podobieństwo to potwierdza słuszność zaproponowanej formuły podziału ciągłej przestrzeni znaków (trójkąta).

6 Podsumowanie

Przedstawiona w tej pracy metoda zapewnia alternatywne spojrzenie na dane z testów koniunktury i metody ich modelowania. Możliwość przedstawienia danych trójwymiarowych w dwóch wymiarach, daje nam dużo większe możliwości doboru narzędzi do ich analizy. Analizowanie wykresów jest często pierwszym etapem modelowania ekonometrycznego. Do tego, jak staraliśmy się pokazać, istnieją również inne potencjalne zastosowania transformacji zmiennych trójwymiarowych, jak na przykład możliwość dyskretyzacji ciągłej przestrzeni obserwacji w sposób bardziej świadomy. Może to mieć wykorzystanie na przykład w modelowaniu zjawiska za pomocą klasycznych i nieklasycznych modeli Markowa.

Z możliwością przedstawieniu danych z ankiet gospodarczych w dwóch wymiarach wiąże

się istotne kwestie dotyczące ich umieszczenia wewnątrz trójkąta. Interesujące jest obserwowane zagęszczenie punktów odpowiadających strukturom odpowiedzi. Obserwowaliśmy je w przypadku wszystkich pytań dotyczących wewnętrznej sytuacji przedsiębiorstwa. Drugą istotną cechą danych, jest ich „ciążenie” w stronę boku „zerowego wzrostu”, czyli boku, gdy żaden z respondentów nie zanotował poprawy zmiennej, której dotyczyło pytanie.

Przedstawiona transformacja pozwoliła nam także zdyskretyzować ciągłą przestrzeń możliwych struktur odpowiedzi. Możliwe jest dokonanie tego podziału z uwzględnieniem charakterystyki badanego procesu. Transformacja pozwala dostrzec tę charakterystykę i w łatwy sposób (za pomocą funkcji w \mathbb{R}^2) przydzielać obserwacje do znaków.

Na bazie takiego podziału została stworzona przestrzeń znaków dla ukrytego modelu Markowa. Jego estymacja daje podobne wyniki, jak w przypadku modelu dla sald z ciągłą przestrzenią znaków. Zbieżność wyników może potwierdzać, że istnieje ukryty proces wpływający na odpowiedzi w testach koniunktury w przemyśle IRG SGH. Proces taki interpretujemy, jako wahania w aktywności gospodarczej.

Dodatek A Dane i wyniki obliczeń

Dane, z których korzystaliśmy w tej pracy zawarte są w poniższej tabeli. Kolejne kolumny oznaczają:

- **okres** — okres, którego dotyczą obserwacje
- **struktura odpowiedzi** — dla każdego okresu wektor złożony z frakcji respondentów odpowiadających: WW1 produkcja wzrosła, WW2 produkcja nie zmieniła się, WW3 produkcja spadła
- **saldo** — różnica między WW1 a WW2
- **transformowane** — dla każdego okresu dwuwymiarowy wektor złożony ze współrzędnych dla struktury odpowiedzi po przetransformowaniu
- **obszar** — numer obszaru trójkąta, w którym znalazł się przetransformowany punkt
- **prawdopodobieństwo** — prawdopodobieństwo przebywania w stanie charakteryzującym się niższą aktywnością gospodarczą dla dyskretnego HMM dla struktury oraz dla ciągłego HMM dla salda

Okres	Struktura odpowiedzi			Saldo	Transformowane		Obszar	Prawdopodobieństwo	
	WW1	WW2	WW3		X	Y		wg struktury	wg salda
199703	0,336	0,438	0,226	0,11	0,635	0,277	1	0,000	0,015
199704	0,343	0,456	0,202	0,141	0,628	0,247	2	0,013	0,002
199705	0,329	0,472	0,198	0,131	0,605	0,242	2	0,013	0,002
199706	0,337	0,417	0,246	0,091	0,651	0,301	1	0,000	0,005
199707	0,274	0,441	0,286	-0,012	0,590	0,350	2	0,012	0,051
199708	0,337	0,434	0,228	0,109	0,638	0,279	1	0,000	0,005
199709	0,363	0,46	0,176	0,187	0,638	0,216	2	0,012	0,001
199710	0,44	0,4	0,16	0,28	0,735	0,196	1	0,000	0,000
199711	0,317	0,451	0,232	0,085	0,612	0,284	2	0,096	0,006
199712	0,246	0,423	0,33	-0,084	0,581	0,404	3	1,000	0,210
199801	0,242	0,446	0,311	-0,069	0,562	0,381	2	0,105	0,346
199802	0,219	0,476	0,305	-0,086	0,525	0,374	2	0,023	0,550

Okres	Struktura odpowiedzi			Saldo	Transformowane		obszar	prawdopodobieństwo	
	WW1	WW2	WW3		X	Y		struktura	saldo
199803	0,262	0,472	0,265	-0,003	0,558	0,325	2	0,014	0,260
199804	0,335	0,424	0,241	0,094	0,644	0,295	1	0,000	0,016
199805	0,277	0,428	0,294	-0,017	0,600	0,360	1	0,000	0,061
199806	0,28	0,466	0,254	0,026	0,576	0,311	2	0,013	0,032
199807	0,248	0,45	0,302	-0,054	0,564	0,370	2	0,014	0,140
199808	0,272	0,451	0,276	-0,004	0,580	0,338	2	0,014	0,087
199809	0,288	0,471	0,242	0,046	0,578	0,296	2	0,013	0,023
199810	0,324	0,424	0,252	0,072	0,636	0,309	1	0,000	0,009
199811	0,215	0,426	0,359	-0,144	0,558	0,440	3	1,000	0,504
199812	0,171	0,354	0,476	-0,305	0,578	0,583	3	1,000	0,996
199901	0,168	0,372	0,46	-0,292	0,563	0,563	3	1,000	0,999
199902	0,178	0,395	0,428	-0,25	0,554	0,524	3	1,000	0,997
199903	0,262	0,392	0,346	-0,084	0,615	0,424	3	1,000	0,902
199904	0,35	0,391	0,259	0,091	0,678	0,317	1	0,000	0,111
199905	0,29	0,407	0,303	-0,013	0,624	0,371	1	0,000	0,092
199906	0,27	0,44	0,289	-0,019	0,586	0,354	2	0,096	0,095
199907	0,257	0,429	0,314	-0,057	0,585	0,385	3	1,000	0,200
199908	0,256	0,433	0,311	-0,055	0,582	0,381	3	1,000	0,272
199909	0,32	0,446	0,234	0,086	0,618	0,287	2	0,096	0,020
199910	0,35	0,422	0,228	0,122	0,656	0,279	1	0,000	0,003
199911	0,249	0,473	0,278	-0,029	0,549	0,340	2	0,096	0,072
199912	0,245	0,414	0,341	-0,096	0,588	0,418	3	1,000	0,343
200001	0,228	0,418	0,354	-0,126	0,573	0,434	3	1,000	0,744
200002	0,191	0,483	0,326	-0,135	0,501	0,399	2	0,104	0,919
200003	0,268	0,508	0,225	0,043	0,538	0,276	2	0,021	0,283
200004	0,319	0,43	0,251	0,068	0,629	0,307	1	0,000	0,031
200005	0,286	0,417	0,297	-0,011	0,614	0,364	1	0,000	0,059
200006	0,333	0,423	0,244	0,089	0,643	0,299	1	0,000	0,008
200007	0,214	0,459	0,327	-0,113	0,534	0,400	2	0,013	0,338
200008	0,263	0,461	0,276	-0,013	0,567	0,338	2	0,013	0,191
200009	0,352	0,424	0,223	0,129	0,655	0,273	1	0,000	0,006
200010	0,303	0,484	0,213	0,09	0,579	0,261	2	0,021	0,005

Okres	Struktura odpowiedzi			Saldo	Transformowane		obszar	prawdopodobieństwo	
	WW1	WW2	WW3		X	Y		struktura	saldo
200011	0,227	0,507	0,266	-0,039	0,509	0,326	2	0,104	0,089
200012	0,21	0,386	0,404	-0,194	0,583	0,495	3	1,000	0,829
200101	0,205	0,435	0,36	-0,155	0,544	0,441	3	1,000	0,958
200102	0,225	0,445	0,33	-0,105	0,552	0,404	2	0,110	0,925
200103	0,203	0,492	0,305	-0,102	0,503	0,374	2	0,035	0,910
200104	0,237	0,459	0,305	-0,068	0,551	0,374	2	0,110	0,819
200105	0,231	0,433	0,336	-0,105	0,564	0,412	3	1,000	0,882
200106	0,233	0,459	0,308	-0,075	0,547	0,377	2	0,481	0,826
200107	0,21	0,41	0,379	-0,169	0,565	0,464	3	1,000	0,969
200108	0,25	0,426	0,324	-0,074	0,583	0,397	3	1,000	0,867
200109	0,264	0,479	0,257	0,007	0,555	0,315	2	0,156	0,422
200110	0,255	0,455	0,291	-0,036	0,566	0,356	2	0,156	0,337
200111	0,248	0,415	0,337	-0,089	0,589	0,413	3	1,000	0,559
200112	0,147	0,411	0,442	-0,295	0,520	0,541	3	1,000	0,995
200201	0,184	0,386	0,429	-0,245	0,564	0,525	3	1,000	0,997
200202	0,202	0,412	0,386	-0,184	0,559	0,473	3	1,000	0,988
200203	0,25	0,411	0,339	-0,089	0,593	0,415	3	1,000	0,908
200204	0,306	0,426	0,267	0,039	0,622	0,327	1	0,000	0,293
200205	0,272	0,44	0,289	-0,017	0,589	0,354	2	0,013	0,183
200206	0,241	0,459	0,3	-0,059	0,553	0,367	2	0,014	0,277
200207	0,251	0,454	0,295	-0,044	0,564	0,361	2	0,014	0,279
200208	0,306	0,461	0,233	0,073	0,598	0,285	2	0,015	0,028
200209	0,328	0,469	0,203	0,125	0,607	0,249	2	0,015	0,003
200210	0,337	0,446	0,217	0,12	0,630	0,266	2	0,023	0,003
200211	0,293	0,439	0,268	0,025	0,604	0,328	2	0,105	0,023
200212	0,213	0,402	0,384	-0,171	0,573	0,470	3	1,000	0,667
200301	0,247	0,427	0,327	-0,08	0,581	0,400	3	1,000	0,730
200302	0,246	0,431	0,322	-0,076	0,576	0,394	3	1,000	0,749
200303	0,29	0,444	0,267	0,023	0,599	0,327	2	0,096	0,259
200304	0,388	0,433	0,179	0,209	0,675	0,219	1	0,000	0,001
200305	0,295	0,5	0,205	0,09	0,562	0,251	2	0,028	0,005

Literatura

- [1] Y. Ephraim, N. Merhav. Hidden markov processes. *IEEE Transactions on Information Theory*, 48(6):1518–1569, Czerwiec 2002.
- [2] J.D. Hamilton. *Time series analysis*. Princeton University Press, Princeton, New Jersey, 1994.
- [3] I.L. MacDonald, W. Zucchini. *Hidden Markov and other models for discrete-valued time series*. Chapman & Hall, London, 1997.
- [4] M. Podgórska, A. Decewicz. *Analiza tendencji rozwojowych w polskiej gospodarce na podstawie testu koniunktury. Metody i wyniki*, wolumen 70 serii *Prace i Materiały IRG*, rozdział Modele Markowa w analizie wyników testu koniunktury. SGH, Warszawa, 2001.