

# PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

# RESEARCH PAPERS

of Wrocław University of Economics

Nr 327

**Taksonomia 22**

**Klasyfikacja i analiza danych –  
teoria i zastosowania**

Redaktorzy naukowci

Krzysztof Jajuga, Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu  
Wrocław 2014

Redaktor Wydawnictwa: Barbara Majewska

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Beata Mazur

Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:

[www.ibuk.pl](http://www.ibuk.pl), [www.ebscohost.com](http://www.ebscohost.com),

w Dolnośląskiej Bibliotece Cyfrowej [www.dbc.wroc.pl](http://www.dbc.wroc.pl),

The Central and Eastern European Online Library [www.ceeol.com](http://www.ceeol.com),

a także w adnotowanej bibliografii zagadnień ekonomicznych BazEkon

[http://kangur.uek.krakow.pl/bazy\\_ae/bazekon/nowy/index.php](http://kangur.uek.krakow.pl/bazy_ae/bazekon/nowy/index.php)

Informacje o naborze artykułów i zasadach recenzowania znajdują się

na stronie internetowej Wydawnictwa

[www.wydawnictwo.ue.wroc.pl](http://www.wydawnictwo.ue.wroc.pl)

Tytuł dofinansowany ze środków Narodowego Banku Polskiego

oraz ze środków Sekcji Klasyfikacji i Analizy Danych PTS

Kopiowanie i powielanie w jakiegokolwiek formie

wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu

Wrocław 2014

**ISSN 1899-3192** (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)

**ISSN 1505-9332** (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM

## Spis treści

<b>Wstęp</b> .....	9
<b>Eugeniusz Gatnar</b> , Balance of payments statistics and external competitiveness of Poland.....	15
<b>Andrzej Sokolowski, Magdalena Czaja</b> , Efektywność metody $k$ -średnich w zależności od separowalności grup.....	23
<b>Barbara Pawelek, Józef Pocięcha, Adam Sagan</b> , Wielosektorowa analiza ukrytych przejść w modelowaniu zagrożenia upadłością polskich przedsiębiorstw .....	30
<b>Elżbieta Gołata</b> , Zróżnicowanie procesu starzenia i struktur demograficznych w Poznaniu i aglomeracji poznańskiej na tle wybranych dużych miast Polski w latach 2002-2011.....	39
<b>Aleksandra Łuczak, Feliks Wysocki</b> , Ustalanie systemu wag dla cech w zagadnieniach porządkowania liniowego obiektów .....	49
<b>Marek Walesiak</b> , Wzmacnianie skali pomiaru dla danych porządkowych w statystycznej analizie wielowymiarowej .....	60
<b>Paweł Lula</b> , Identyfikacja słów i fraz kluczowych w tekstach polskojęzycznych za pomocą algorytmu <i>RAKE</i> .....	69
<b>Mariusz Kubus</b> , Propozycja modyfikacji metody złagodzonego LASSO.....	77
<b>Andrzej Bąk, Tomasz Bartłomowicz</b> , Wielomianowe modele logitowe wyborów dyskretnych i ich implementacja w pakiecie <i>DiscreteChoice</i> programu R.....	85
<b>Justyna Brzezińska</b> , Wykorzystanie modeli logarytmiczno-liniowych do analizy bezrobocia w Polsce w latach 2004-2012.....	95
<b>Andrzej Bąk, Marcin Pelka, Aneta Rybicka</b> , Zastosowanie pakietu <i>dcMNM</i> programu R w badaniach preferencji konsumentów wódki .....	104
<b>Barbara Batóg, Jacek Batóg</b> , Analiza stabilności klasyfikacji polskich województw według sektorowej wydajności pracy w latach 2002-2010 .....	113
<b>Małgorzata Markowska, Danuta Strahl</b> , Klasyfikacja europejskiej przestrzeni regionalnej ze względu na filary inteligentnego rozwoju z wykorzystaniem referencyjnego systemu granicznego.....	121
<b>Kamila Migdał-Najman, Krzysztof Najman</b> , Formalna ocena jakości odwzorowania struktury grupowej na mapie Kohonena .....	131
<b>Kamila Migdał-Najman, Krzysztof Najman</b> , Graficzna ocena jakości odwzorowania struktury grupowej na mapie Kohonena .....	139
<b>Beata Basiura, Anna Czapkiewicz</b> , Badanie jakości klasyfikacji szeregów czasowych .....	148
<b>Michał Trzęsiok</b> , Wybrane metody identyfikacji obserwacji oddalonych.....	157

<b>Grażyna Dehnel, Tomasz Klimanek</b> , Taksonomiczne aspekty estymacji pośredniej uwzględniającej autokorelację przestrzenną w statystyce gospodarczej.....	167
<b>Michał Bernard Pietrzak, Justyna Wilk</b> , Odległość ekonomiczna w modelowaniu zjawisk przestrzennych z wykorzystaniem modelu grawitacji.....	177
<b>Maciej Beręsewicz</b> , Próba zastosowania różnych miar odległości w uogólnionym estymatorze Petersena .....	186
<b>Marcin Szymkowiak, Tomasz Józefowski</b> , Konstrukcja i praktyczne wykorzystanie estymatorów typu SPREE na przykładzie dwuwymiarowych tabel kontyngencji .....	195
<b>Marcin Pelka</b> , Klasyfikacja pojęciowa danych symbolicznych w podejściu wielomodelowym .....	202
<b>Małgorzata Machowska-Szewczyk</b> , Ocena klas w rozmytej klasyfikacji obiektów symbolicznych.....	210
<b>Justyna Wilk</b> , Problem wyboru liczby klas w taksonomicznej analizie danych symbolicznych.....	220
<b>Andrzej Dudek</b> , Metody analizy skupień w klasyfikacji markerów map Google .....	229
<b>Ewa Roszkowska</b> , Ocena ofert negocjacyjnych w słabo ustrukturyzowanych problemach negocjacyjnych z wykorzystaniem rozmytej procedury SAW	237
<b>Marcin Szymkowiak, Marek Witkowski</b> , Zastosowanie analizy korespondencji do badania kondycji finansowej banków spółdzielczych.....	248
<b>Bartłomiej Jefmański</b> , Budowa rozmytych indeksów satysfakcji klientów z zastosowaniem programu R.....	257
<b>Karolina Bartos</b> , Odkrywanie wzorców zachowań konsumentów za pomocą analizy koszykowej danych transakcyjnych .....	266
<b>Joanna Trzęsiok</b> , Taksonomiczna analiza krajów pod względem dzietności kobiet oraz innych czynników demograficznych .....	275
<b>Beata Bal-Domańska</b> , Próba identyfikacji większych skupisk regionalnych oraz ich konwergencja.....	285
<b>Beata Bieszk-Stolorz, Iwona Markowicz</b> , Wpływ zasiłku na proces poszukiwania pracy .....	294
<b>Marta Dziechciarz-Duda, Klaudia Przybysz</b> , Wykształcenie a potrzeby rynku pracy. Klasyfikacja absolwentów wyższych uczelni.....	303
<b>Tomasz Klimanek</b> , Problem pomiaru procesu dezagrarnizacji wsi polskiej w świetle wielowymiarowych metod statystycznych.....	313
<b>Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska</b> , Wybrane metody analizy danych wzdłużnych.....	321
<b>Artur Zaborski</b> , Zastosowanie miar odległości dla danych porządkowych do agregacji preferencji indywidualnych .....	330
<b>Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek</b> , Zastosowanie analizy korespondencji do badania sytuacji mieszkańców strefy podmiejskiej Warszawy na rynku pracy.....	338

<b>Katarzyna Wawrzyniak</b> , Klasyfikacja województw według stopnia realizacji priorytetów Strategii Rozwoju Kraju 2007-2015 z wykorzystaniem wartości centrum wierszowego .....	346
---	-----

## Summaries

<b>Eugeniusz Gatnar</b> , Statystyka bilansu płatniczego a konkurencyjność gospodarki Polski .....	22
<b>Andrzej Sokółowski, Magdalena Czaja</b> , Cluster separability and the effectiveness of $k$ -means method .....	29
<b>Barbara Pawelek, Józef Pocięcha, Adam Sagan</b> , Multisectoral analysis of latent transitions in bankruptcy prediction models.....	38
<b>Elżbieta Golata</b> , Differences in the process of aging and demographic structures in Poznań and the agglomeration compared to selected Polish cities in the years 2002-2011 .....	48
<b>Aleksandra Łuczak, Feliks Wysocki</b> , Determination of weights for features in problems of linear ordering of objects .....	59
<b>Marek Walesiak</b> , Reinforcing measurement scale for ordinal data in multivariate statistical analysis .....	68
<b>Paweł Lula</b> , Automatic identification of keywords and keyphrases in documents written in Polish.....	76
<b>Mariusz Kubus</b> , The proposition of modification of the relaxed LASSO method.....	84
<b>Andrzej Bąk, Tomasz Bartłomowicz</b> , Microeconomic multinomial logit models and their implementation in the <code>DiscreteChoice</code> R package .	94
<b>Justyna Brzezińska</b> , The analysis of unemployment data in Poland in 2004-2012 with application of log-linear models .....	103
<b>Andrzej Bąk, Marcin Pelka, Aneta Rybicka</b> , Application of the MMLM package of R software for vodka consumers preference analysis.....	112
<b>Barbara Batóg, Jacek Batóg</b> , Analysis of the stability of classification of Polish voivodeships in 2002-2010 according to the sectoral labour productivity .....	120
<b>Małgorzata Markowska, Danuta Strahl</b> , Classification of the European regional space in terms of smart growth pillars using the reference limit system.....	130
<b>Kamila Migdał Najman, Krzysztof Najman</b> , Formal quality assessment of group structure mapping on the Kohonen's map .....	138
<b>Kamila Migdał Najman, Krzysztof Najman</b> , Graphical quality assessment of group structure mapping on the Kohonen's map .....	147
<b>Beata Basiura, Anna Czapkiewicz</b> , Validation of time series clustering .....	156
<b>Michał Trzęsiok</b> , Selected methods for outlier detection.....	166

<b>Grażyna Dehnel, Tomasz Klimanek</b> , Taxonomic aspects of indirect estimation accounting for spatial correlation in enterprise statistics .....	176
<b>Michał Bernard Pietrzak, Justyna Wilk</b> , Economic distance in modeling spatial phenomena with the application of gravity model.....	185
<b>Maciej Beręsewicz</b> , An attempt to use different distance measures in the Generalized Petersen estimator .....	194
<b>Marcin Szymkowiak, Tomasz Józefowski</b> , Construction and practical using of SPREE estimators for two-dimensional contingency tables.....	201
<b>Marcin Pelka</b> , The ensemble conceptual clustering for symbolic data.....	209
<b>Małgorzata Machowska-Szewczyk</b> , Evaluation of clusters obtained by fuzzy classification methods for symbolic objects.....	219
<b>Justyna Wilk</b> , Problem of determining the number of clusters in taxonomic analysis of symbolic data .....	228
<b>Andrzej Dudek</b> , Clustering techniques for Google maps markers.....	236
<b>Ewa Roszkowska</b> , The evaluation of negotiation offers in ill structure negotiation problems with the application of fuzzy SAW procedure .....	247
<b>Marcin Szymkowiak, Marek Witkowski</b> , The use of correspondence analysis in analysing the financial situation of cooperative banks.....	256
<b>Bartłomiej Jefmański</b> , The construction of fuzzy customer satisfaction indexes using R program.....	265
<b>Karolina Bartos</b> , Discovering patterns of consumer behaviour by market basket analysis of the transactional data.....	274
<b>Joanna Trzęsiok</b> , Cluster analysis of countries with respect to fertility rate and other demographic factors .....	284
<b>Beata Bal-Domańska</b> , An attempt to identify major regional clusters and their convergence .....	293
<b>Beata Bieszk-Stolorz, Iwona Markowicz</b> , The influence of benefit on the job finding process .....	302
<b>Marta Dziechciarz-Duda, Klaudia Przybysz</b> , Education and labor market needs. Classification of university graduates .....	312
<b>Tomasz Klimanek</b> , The problem of measuring deagrarianisation process in rural areas in Poland using multivariate statistical methods.....	320
<b>Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska</b> , Selected methods for an analysis of longitudinal data.....	329
<b>Artur Zaborski</b> , The application of distance measures for ordinal data for aggregation individual preferences .....	337
<b>Mariola Chrzanowska, Nina Drejerska, Iwona Pomianek</b> , Application of correspondence analysis to examine the situation of the inhabitants of Warsaw suburban area in the labour market .....	345
<b>Katarzyna Wawrzyniak</b> , Classification of voivodeships according to the level of the realization of priorities of <i>the National Development Strategy 2007-2015</i> with using the values of centroid of the rows .....	355

**Małgorzata Sej-Kolasa, Mirosława Sztemberg-Lewandowska**

Uniwersytet Ekonomiczny we Wrocławiu

---

## WYBRANE METODY ANALIZY DANYCH WZDŁUŻNYCH

---

**Streszczenie:** W artykule przedstawiono możliwości wykorzystania wybranych metod wielowymiarowych do analizy zjawisk opisanych danymi wzdluznymi (*longitudinal data*). Omówiono dwa podejścia wykorzystujące modele ze zmiennymi ukrytymi. Pierwsze podejście bazuje na modelach równań strukturalnych, a dokładniej na modelu latentnych krzywych rozwojowych. Model ten zakłada, że zmiana jest procesem ciągłym scharakteryzowanym przez zmienne, których realizacje różnią się między obiektami. Drugie podejście bazuje na analizie danych funkcjonalnych. Jedną z metod analizy danych funkcjonalnych jest funkcjonalna analiza głównych składowych, która posiada zalety klasycznej analizy głównych składowych, dodatkowo umożliwia analizę danych o charakterze dynamicznym.

**Słowa kluczowe:** dane wzdluzne, funkcjonalna analiza czynnikowa, SEM, krzywe rozwojowe.

### 1. Wstęp

Badania wzdluzne są badaniami czasowo-przekrojowymi, powstają z połączenia w jeden zbiór szeregów czasowych obserwacji pochodzących z różnych obiektów. Obiektami, czyli jednostkami przekroju, mogą być np. regiony administracyjno-gospodarcze w obrębie jednego lub kilku państw czy powiązane ze sobą makroregiony gospodarcze. Na zachowanie każdego obiektu wpływają zarówno czynniki indywidualne, wynikające ze specyfiki obiektu, jak i czynniki determinujące działalność wszystkich obiektów w jednakowy sposób i w jednakowym stopniu.

Celem badań wzdluznych jest obserwacja wybranych obiektów i zjawisk w dłuższym okresie, np. kilkumiesięcznym, kilkuletnim i dłuższym. Badania te umożliwiają systematyczne gromadzenie danych dotyczących badanego zjawiska, obserwację zachodzących zmian oraz wyjaśnianie ich przyczyn. Dane wzdluzne (*longitudinal data*) pokazują stan obiektów w czasie; są wynikiem wielokrotnych obserwacji tych samych zmiennych w długich okresach.

W artykule pokazano możliwości wykorzystania metod wielowymiarowych do analizy zjawisk opisanych danymi wzdluznymi zaprezentowano dwa podejścia

wykorzystujące modele ze zmiennymi ukrytymi. Pierwsze podejście bazuje na modelach równań strukturalnych, drugie na analizie danych funkcjonalnych. Celem artykułu jest porównanie obu metod z uwzględnieniem ich odmienności oraz wskazaniem ich zalet i wad.

W celu zobrazowania tych metod przytoczono przykład empiryczny, którego przedmiotem badania był poziom edukacji w państwach europejskich, scharakteryzowany przez współczynnik skolaryzacji<sup>1</sup>, dotyczący studentów uczelni wyższych studiów I i II stopnia (ISCED 5, 6). Zakres przestrzenny badania to 30 państw europejskich, natomiast zakres czasowy to lata 1999-2011. Dane pochodzą z baz Eurostatu.

## 2. Funkcjonalna analiza głównych składowych

Analiza głównych składowych polega na transformacji zmiennych pierwotnych w zbiór nowych wzajemnie ortogonalnych zmiennych, zwanych głównymi składowymi. Funkcjonalna analiza głównych składowych posiada zalety klasycznej analizy głównych składowych, dodatkowo umożliwia analizę danych o charakterze dynamicznym. Podstawową różnicą tych dwóch metod jest rodzaj danych: PCA bazuje na danych wielowymiarowych, natomiast FPCA na danych funkcjonalnych. Danymi funkcjonalnymi są krzywe i trajektorie, czyli ciąg indywidualnych obserwacji, a nie jak zwykle pojedyncza obserwacja.

W przypadku funkcjonalnej analizy głównych składowych (FPCA) każda główna składowa wyrażona jest przez funkcję wagową głównych składowych (*principal component weight function*), inaczej nazwaną funkcją własną (*eigenfunction*)  $\xi_j(t)$  zależną od czasu [Daniele 2006; Ramsay, Silverman 2005]. Funkcja własna maksymalizuje wariancję funkcji głównych składowych:

$$v(t, s) \stackrel{\text{def}}{=} \frac{1}{n-1} \sum_{i=1}^n \{x_i(t) - \bar{x}(t)\} \{x_i(s) - \bar{x}(s)\}. \quad (1)$$

Analogicznie do klasycznej PCA, problemem w funkcjonalnej jest rozkład wariancji funkcji:

$$v(t, s) = \sum_j \lambda_j \xi_j(t) \xi_j(s), \quad (2)$$

gdzie  $\lambda_j, \xi_j(t)$  spełniają równanie własne:

$$\langle v(u, \cdot), \xi_j \rangle = \lambda_j \xi_j(s) \quad (3)$$

<sup>1</sup> Współczynnik skolaryzacji wyraża liczbę studentów na 100 osób należących do tej samej grupy wiekowej.



oraz wartości własne są dodatnie i niemalejące:

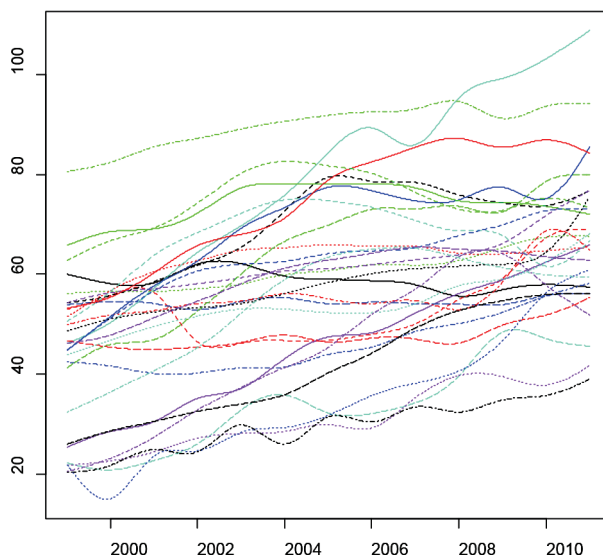
$$\lambda_j \stackrel{\text{def}}{=} \int_T \xi_j(t) v(t,s) \xi_j(s) dt ds. \quad (4)$$

Funkcje własne spełniają warunek:

$$\int_T \xi_j^2(t) dt = 1 \text{ oraz } \int_T \xi_j(t) \xi_i(t) dt = 0 \quad (i < j). \quad (5)$$

Funkcje własne określają główne składowe zmienności między próbkowymi funkcjami  $x_t$  [Ingrassia, Costanzo 2005; Hall i in. 2006].

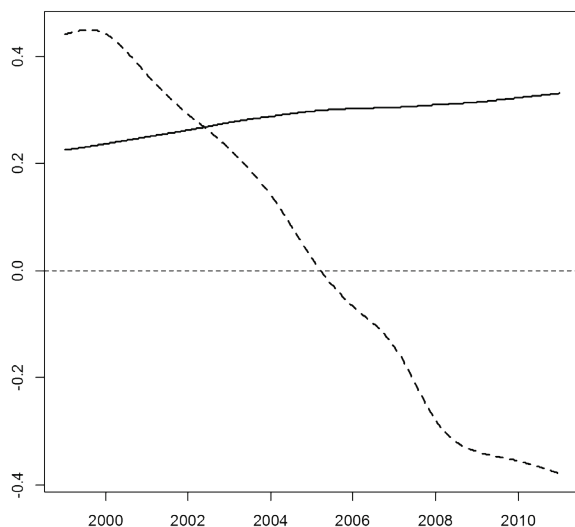
Empiryczne wartości współczynnika skolaryzacji dla 30 państw europejskich w latach 1999-2011 przedstawia rysunek 1. Dane wielowymiarowe przekształcono na dane funkcjonalne metodą B-spline, a następnie za pomocą funkcjonalnej analizy głównych składowych wyodrębniono dwie główne składowe (rys. 2).



**Rys. 1.** Skolaryzacja państw europejskich

Źródło: opracowanie własne z wykorzystaniem programu R.

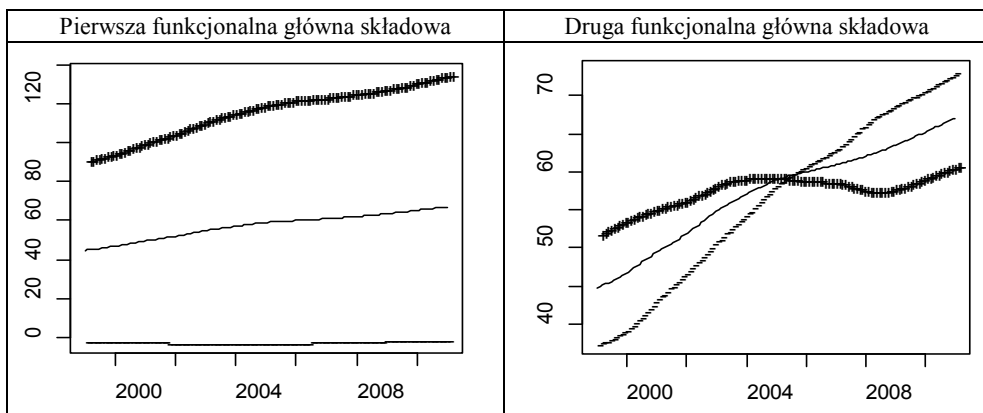
Pierwsza funkcjonalna główna składowa, oznaczona linią ciągłą, wyjaśnia 59% zmienności wspólnej, natomiast druga, oznaczona linią przerywaną – 40%. Podobnie jak w klasycznej analizie główne składowe należy zinterpretować, jednak w przypadku danych funkcjonalnych interpretacja jest trudniejsza. Praktyczne wyjaśnienie funkcjonalnych głównych składowych ułatwiają wykresy odchylenia każdej ze składowych od średniej (rys. 3). Linia ciągła (leżąca pośrodku) przedstawia



**Rys. 2.** Funkcjonalne główne składowe

Źródło: opracowanie własne z wykorzystaniem programu R.

średnią wielkość zmian współczynnika skolaryzacji, linia zbudowana ze znaków „+” (najgrubsza) oznacza średnią powiększoną o tę część składowej, która wyjaśnia zmienność wspólną. Linia zbudowana ze znaków „-” oznacza średnią pomniejszoną o tę wartość.



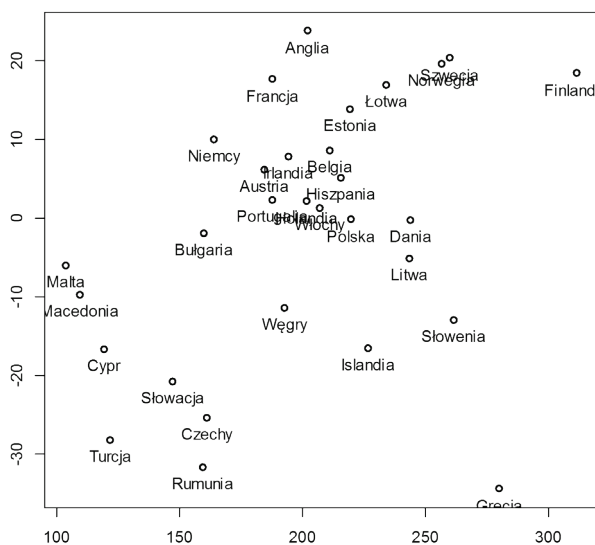
**Rys. 3.** Odchylenia funkcjonalnych głównych składowych od średniej

Źródło: opracowanie własne z wykorzystaniem programu R.

Pierwsza składowa odpowiada za ogólną tendencję. Przyjmuje wartości dodatnie, zatem dla danego kraju dodatni ładunek na tej składowej oznacza, że krzywa opisująca wielkość zmian współczynnika skolaryzacji leży powyżej średniej. Dru-

ga składowa pokazuje tendencje w pierwszych latach w odniesieniu do ostatnich („początek kontra koniec”) – porównuje okres do roku 2006 z okresem po 2006 roku. Dodatni ładunek na drugiej składowej oznacza, że tempo zmian na początku badanego okresu było większe od średniej, natomiast na końcu tempo było mniejsze od średniej.

Funkcjonalna analiza czynnikowa pozwala na wizualizację danych, umożliwiającą porównanie badanych obiektów. Rysunek 4 przedstawia rzut danych na płaszczyznę wyznaczoną przez dwie funkcjonalne główne składowe.



**Rys. 4.** Rzut obiektów na płaszczyznę wyznaczoną przez dwie funkcje składowe

Źródło: opracowanie własne z wykorzystaniem programu R.

### 3. Model krzywych rozwojowych

Modelowanie równań strukturalnych (SEM – *Structural Equation Modeling*), inaczej nazywane LISREL (*Linear Structural RELationships*), jest ogólną, głównie liniową, wielowymiarową techniką statystyczną. Jest ona bardziej confirmacyjna niż eksploracyjna, czyli wykorzystuje się ją do sprawdzania dopasowania określonego modelu do danych, a nie do budowy pasującego modelu. Jednym ze złożonych modeli strukturalnych jest model krzywych rozwojowych, który służy do analizy zmiany wartości poszczególnych zmiennych w czasie. Model ten zakłada, że zmiana jest procesem ciągłym scharakteryzowanym przez zmienne, których realizacje różnią się między obiektami [Acock, Li 2002; Konarski 2004].

Model pomiarowy opisuje zmienne latentne, tzn. parametry krzywych rozwojowych. Natomiast model strukturalny zawiera informacje o zróżnicowaniu tych

parametrów. Model krzywych rozwojowych pozwala na estymację parametrów nie tylko funkcji liniowej i wielomianowej, ale także np. logarytmicznej, wykładniczej. Takie podejście do analizy zmiany pozwala na uwzględnienie korelacji między statusem początkowym i tempem zmiany krzywej rozwojowej oraz testowanie relacji między błędami pomiarowymi [Sej-Kolasa, Sztemberg-Lewandowska 2009].

Przy założeniu, że chwile pomiaru nie różnią się między obiektami, model pomiarowy ma postać:

$$\mathbf{Y} = \mathbf{\Lambda}\boldsymbol{\eta} + \mathbf{E}, \quad (6)$$

gdzie:  $\mathbf{Y}$  – wektor obserwowalnych zmiennych,

$\mathbf{\Lambda}$  – macierz ładunków czynnikowych,

$\boldsymbol{\eta}$  – wektor zmiennych latentnych,

$\mathbf{E}$  – wektor błędów pomiaru zmiennych obserwowalnych.

Natomiast model strukturalny jest wyrażony wzorem:

$$\boldsymbol{\eta} = \boldsymbol{\alpha} + \boldsymbol{\xi}, \quad (7)$$

gdzie  $\boldsymbol{\alpha}$  zawiera średnie wartości indywidualnych parametrów regresji  $\mu_\alpha, \mu_\beta$ ;  $\boldsymbol{\xi}$  zawiera pozostałości regresji.

Opisany model SEM z założeniami:  $E(\boldsymbol{\eta}) = \boldsymbol{\alpha}$ ,  $E(\boldsymbol{\xi}) = 0$ ,  $Var(\boldsymbol{\xi}) = \Psi$  oraz  $E(\mathbf{E}) = 0$ ,  $Var(\mathbf{E}) = \Theta$ ,  $Cov(\mathbf{E}, \boldsymbol{\eta}) = 0$  bazuje na następujących równaniach:

$$\boldsymbol{\mu}_y = \mathbf{\Lambda}\boldsymbol{\alpha}, \quad (8)$$

$$\boldsymbol{\Sigma} = \mathbf{\Lambda}\boldsymbol{\Phi}\mathbf{\Lambda}^T + \boldsymbol{\Theta}. \quad (9)$$

Obiektami w badaniu są państwa europejskie, zmienną zależną – wartości współczynnika solaryzacji, natomiast  $t$  jest numerem pomiaru, badanie obejmuje dwanaście lat.

Zmiennymi latentnymi w modelu są: stan początkowy i tempo wzrostu (szybkość zmian)  $i$ -tego obiektu. W większości modeli równań strukturalnych ładunki czynnikowe są estymowane, tutaj są znane. Stan początkowy  $i$ -tego obiektu jest taki sam w każdej chwili pomiaru, dlatego ładunki zmiennych na tym czynniku są równe 1. Natomiast ładunki zmiennych na drugim czynniku odzwierciedlają zróżnicowanie czasowe kolejnych pomiarów w odniesieniu do 1999 roku.

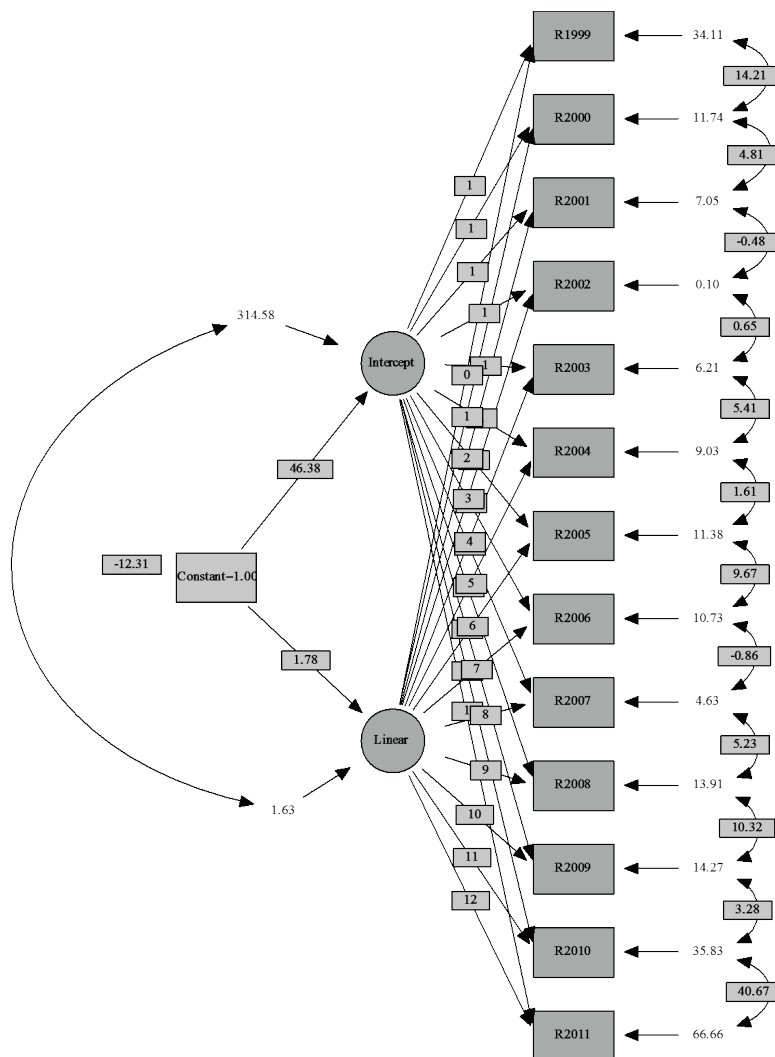
W tabeli 1 zamieszczono miary dopasowania opisanego modelu liniowego do danych. Otrzymany model przedstawiono na rysunku 5.

Zadowolająca wartość RMSEA wynosi poniżej 0,05, natomiast dla wszystkich pozostałych miar im wartość bliższa jest jedności, tym lepsze dopasowanie. IFI i MFI wykazują dobre dopasowanie (wartości powyżej 0,8), CFI i GFI wykazują gorsze dopasowanie (wartości powyżej 0,6), natomiast wartość RMSEA jest graniczna.

**Tabela 1.** Miary dopasowania otrzymanego modelu

Statystyka chi-kwadrat	$p = 0,1$
Indeks Bollena IFI ( <i>Incremental Fix Index</i> )	0,810
Indeks Bentlera I Muellera CFI ( <i>Normed Comparative Fit Index</i> )	0,611
Indeks Joreskoga i Sörboma GFI ( <i>Goodness-of-Fit Index</i> )	0,636
Indeks McDonalda MFI ( <i>McDonald Fit Index</i> )	0,911
Wskaźnik RMSEA ( <i>Root Mean Square Error of Approximation</i> )	0,055

Źródło: opracowanie własne z wykorzystaniem programu R.

**Rys. 5.** Estymowane parametry modelu

Źródło: opracowanie własne z wykorzystaniem programu EQS.

Wartość *constant* potrzebna jest w przypadku modelowania zarówno średniej, jak i kowariancji modelu. Strzałka charakteryzująca zależność *constant* z *Intercept* reprezentuje średnią dla stałej modelu liniowego, natomiast strzałka łącząca *constant* z *Linear* reprezentuje średnią dla współczynnika kierunkowego modelu. Średnie te są wolnymi parametrami modelu.

Współczynnik kierunkowy otrzymanego modelu liniowego wynosi 1,8, zatem tendencja zmian współczynnika solaryzacji jest rosnąca.

Model krzywych rozwojowych pozwala dodatkowo na ocenę modelu w każdym badanym roku, co umożliwi określenie, w którym roku tendencja zmian współczynnika solaryzacji dla wszystkich obiektów była największa.

#### 4. Podsumowanie

Zarówno funkcjonalna analiza głównych składowych, jak i model krzywych rozwojowych bazują na zmiennych ukrytych. Obie metody pozwalają na analizę danych wzdluznych (czasowo-przekrojowych). Każda z metod zakłada, że zmiana jest procesem ciągłym scharakteryzowanym przez zmienne, których realizacje różnią się między obiektami.

Podstawową różnicą obu metod jest rodzaj wykorzystywanych danych: dane funkcjonalne lub wielowymiarowe.

Funkcjonalną analizę głównych składowych stosujemy w celu wyodrębnienia funkcjonalnych głównych składowych, natomiast model krzywych rozwojowych – w celu estymacji parametrów wybranego modelu, np. liniowego. Zaletą pierwszej z metod jest możliwość pokazania nie tylko ogólnej tendencji zmian, ale także

**Tabela 2.** Funkcjonalna analiza głównych składowych i model krzywych rozwojowych - porównanie

Wyszczególnienie	Funkcjonalna analiza głównych składowych	Model krzywych rozwojowych
Podstawowe założenie	Zmiana jest procesem ciągłym scharakteryzowanym przez zmienne, których realizacje różnią się między obiektami	
Rodzaj danych	Funkcjonalne	Wielowymiarowe
Funkcje przybliżające dane	B-spline, stała, wykładnicza, funkcja Fouriera, wielomianowa, potęgowa	Liniowa, wielomianowa, logarytmiczna, wykładnicza
Zmienne latentne	Ogólna tendencja „Początek kontra koniec”	Stan początkowy modelu liniowego (współczynnik kierunkowy prostej) Tempo zmian modelu liniowego (przecięcie z osią)
Cel procedury	Wyodrębnienie funkcjonalnych głównych składowych	Estymacja parametrów wybranego modelu
Istotne zalety metody	Wizualizacja obiektów	Miary dobroci dopasowania parametrów modelu Ocena modelu w każdym badanym okresie dla wszystkich obiektów

Źródło: opracowanie własne.

tendencji w początkowym okresie w odniesieniu do okresu ostatniego. Dodatkowo metoda pozwala na wizualizację danych. Zaletą drugiej metody jest możliwość dokonania oceny modelu w każdym badanym okresie dla wszystkich obiektów oraz oceny dobroci dopasowania poszczególnych parametrów modelu. Wadą funkcjonalnej analizy głównych składowych jest utrata informacji spowodowana koniecznością przekształcenia danych empirycznych na dane funkcjonalne. Utrata informacji występuje również w modelu krzywych rozwojowych na etapie estymacji parametrów modelu.

Syntetyczne porównanie obu zaprezentowanych metod zawiera tabela 2.

## Literatura

- Acock A.C., Li F. (2002), *Latent growth curve analysis: A gentle introduction*, <http://oregonstate.edu/dept/hdfs/papers/lgcgeneral.pdf> (24.04.2014).
- Daniele M. (2006), *Functional principal components analysis to study environmental data*, [http://www.sis-statistica.it/files/pdf/atti/Spontanee%202006\\_677-680.pdf](http://www.sis-statistica.it/files/pdf/atti/Spontanee%202006_677-680.pdf) (2.06.2013).
- Hall P., Müller H.G., Wang J.L. (2006), *Properties of Principal Component Methods for Functional and Longitudinal Data Analysis*, „The Annals of Statistics” Vol. 34, No. 3.
- Ingrassia S., Costanzo G.D. (2005), *Functional principal component analysis of financial time series*, [w:] Vichi M., Monari P., Mignani S., Montanari A. (red.), *New Developments in Classification and Data Analysis*, Springer-Verlag, Berlin.
- Konarski R. (2004), *Analiza zmiany z zastosowaniem analizy latentnych krzywych rozwojowych*, „ASK. Społeczeństwo. Badania. Metody”, nr 13.
- Ramsay J.O., Silverman B.W. (2005), *Functional Data Analysis*, Springer, USA.
- Sej-Kolasa M., Sztemberg-Lewandowska M. (2009), *Antropopresja na środowisko państw europejskich*, [w:] Walesiak M. (red.), *Wizualizacja wyników badań marketingowych-podejścia, metody i zastosowania*, Wydawnictwo Uniwersytetu Ekonomicznego, Wrocław.

## SELECTED METHODS FOR AN ANALYSIS OF LONGITUDINAL DATA

**Summary:** This paper presents the possibilities of using selected multivariate methods to analyse phenomena described with longitudinal data. Two approaches based on models with latent variables are described. The first approach is based on the structural equation modelling and, more precisely, on the latent growth modelling. This model assumes that changes are continuous processes characterised by variables the realizations of which differ between objects. The other approach is based on the functional data analysis. One of the methods for analysing functional data is the functional principal components analysis which has the advantages of the principal components analysis and enables analysing data with dynamic nature.

**Keywords:** longitudinal data, functional factor analysis, SEM, growth curves models.