

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

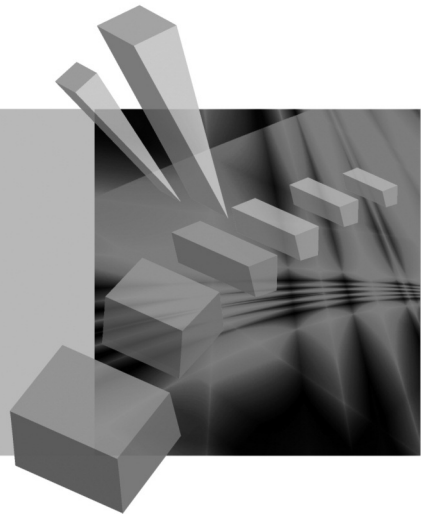
RESEARCH PAPERS

of Wrocław University of Economics

242

Taksonomia 19.

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak



Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2012

Recenzenci: Eugeniusz Gatnar, Elżbieta Gołata, Tadeusz Kufel, Józef Pocięcha,
Miroslaw Szreder, Feliks Wysocki

Redaktor Wydawnictwa: Aleksandra Śliwka

Redaktor techniczny: Barbara Łopusiewicz

Korektor: Barbara Cibis

Łamanie: Małgorzata Czupryńska

Projekt okładki: Beata Dębska

Tytuł sfinansowano ze środków Sekcji Klasyfikacji i Analizy Danych PTS
i Uniwersytetu Ekonomicznego we Wrocławiu

Publikacja jest dostępna na stronie www.ibuk.pl

Streszczenia opublikowanych artykułów są dostępne w międzynarodowej bazie danych
The Central European Journal of Social Sciences and Humanities <http://cejsh.icm.edu.pl>
oraz w The Central and Eastern European Online Library 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

Kopowanie i powielanie w jakiegokolwiek formie
wymaga pisemnej zgody Wydawcy

© Copyright by Uniwersytet Ekonomiczny we Wrocławiu
Wrocław 2012

ISSN 1899-3192 (Prace Naukowe Uniwersytetu Ekonomicznego we Wrocławiu)
ISSN 1505-9332 (Taksonomia)

Wersja pierwotna: publikacja drukowana

Druk: Drukarnia TOTEM
Nakład: 320 egz.

Spis treści

Wstęp	13
Stanisława Bartosiewicz , Jeszcze raz o skutkach subiektywizmu w analizie wielowymiarowej	17
Andrzej Sokolowski , Q uniwersalna miara odległości	22
Eugeniusz Gatnar , Jakość danych w systemach statystycznych banków centralnych (na przykładzie NBP)	31
Marek Walesiak , Pomiar odległości obiektów opisanych zmiennymi mierzonymi na skali porządkowej – strategię postępowania.....	39
Krzysztof Jajuga, Marek Walesiak , XXV lat konferencji taksonomicznych – fakty i refleksje	47
Józef Pocięcha, Barbara Pawelek , Model SEM w analizie zagrożenia bankructwem przedsiębiorstw w świetle koniunktury gospodarczej – problemy teoretyczne i praktyczne	50
Paweł Lula , Uczące się systemy pozyskiwania informacji z dokumentów tekstowych	58
Ewa Roszkowska , Zastosowanie metody TOPSIS do wspomaganie procesu negocjacji.....	68
Andrzej Młodak , Sąsiedztwo obszarów przestrzennych w ujęciu fizycznym oraz społeczno-ekonomicznym – podejście taksonomiczne	76
Andrzej Bąk , Modele kategorii nieuporządkowanych w badaniach preferencji	86
Jacek Kowalewski , Zintegrowany model optymalizacji badań statystycznych.....	96
Jan Paradysz, Karolina Paradysz , Obszary bezrobocia w Polsce – problem benchmarkowy.....	106
Tomasz Szubert , W co grać, aby jak najmniej przegrać? Próba klasyfikacji systemów gry w zakładach bukmacherskich.....	116
Izabela Szamrej-Baran , Klasyfikacja krajów UE ze względu na ubóstwo energetyczne	126
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , Analiza dojazdów do pracy za pomocą modelu grawitacji.....	135
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Minimum egzystencji a czynniki warunkujące skłonność do korzystania z pomocy społecznej. Klasyfikacja gospodarstw domowych	144
Hanna Dudek , Subiektywne skale ekwiwalentności – analiza na podstawie danych o satysfakcji z osiągniętych dochodów	153

Joanicjusz Nazarko, Ewa Chodakowska, Marta Jaročka , Segmentacja szkół wyższych metodą analizy skupień <i>versus</i> konkurencja technologiczna ustalona metodą DEA – studium komparatywne.....	163
Ewa Chodakowska , Wybrane metody klasyfikacji w konstrukcji ratingu szkół.....	173
Bartosz Soliński , Sektor energetyki odnawialnej w krajach Unii Europejskiej – klasyfikacja w świetle strategii zarządzania zmianą.....	182
Krzysztof Szwarz , Klasyfikacja powiatów województwa wielkopolskiego ze względu na sytuację demograficzną.....	192
Elżbieta Gołata, Grażyna Dehnel , Rejestry administracyjne w analizie przedsiębiorczości.....	202
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Wykorzystanie metod taksonomicznych w prognozowaniu wskaźników rentowności banków giełdowych w Polsce.....	212
Katarzyna Dębowska , Modelowanie upadłości przedsiębiorstw przy wykorzystaniu metod dyskryminacji i regresji.....	222
Alina Bojan , Wykorzystanie metod wielowymiarowej analizy danych do identyfikacji zmiennych wpływających na atrakcyjność wybranych inwestycji.....	231
Justyna Brzezińska , Analiza logarytmiczno-liniowa w badaniu przyczyn umieralności w krajach UE.....	240
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Analiza klas ukrytych w badaniach satysfakcji studentów.....	247
Bartłomiej Jefmański , Pomiar opinii respondentów z wykorzystaniem elementów teorii zbiorów rozmytych i środowiska R.....	256
Julita Stańczuk , Porównanie rezultatów wielostanowej klasyfikacji obiektów ekonomicznych z wykorzystaniem analizy dyskryminacyjnej oraz sieci neuronowych.....	265
Jerzy Krawczuk , Skuteczność metod klasyfikacji w prognozowaniu kierunku zmian indeksu giełdowego S&P500.....	275
Anna Czapkiewicz, Beata Basiura , Symulacyjne badanie wpływu zaburzeń na grupowanie szeregów czasowych na podstawie modelu Copula-GARCH.....	283
Radosław Pietrzyk , Ocena efektywności inwestycji funduszy inwestycyjnych z tytułu doboru papierów wartościowych i umiejętności wykorzystania trendów rynkowych.....	291
Aleksandra Witkowska, Marek Witkowski , Zastosowanie metody Panzara-Rosse’a do pomiaru poziomu konkurencji w sektorze banków spółdzielczych.....	306
Marcin Pelka , Podejście wielomodelowe z wykorzystaniem metody <i>boosting</i> w analizie danych symbolicznych.....	315
Justyna Wilk , Analiza porównawcza oprogramowania komputerowego w klasyfikacji danych symbolicznych.....	323

Tomasz Bartłomowicz, Justyna Wilk , Zastosowanie metod analizy danych symbolicznych w przeszukiwaniu dziedzinowych baz danych.....	333
Kamila Migdał-Najman , Propozycja hybrydowej metody grupowania opartej na sieciach samouczących	342
Dorota Rozmus , Porównanie dokładności taksonomii spektralnej oraz zagregowanych algorytmów taksonomicznych opartych na idei metody <i>bagging</i>	352
Krzysztof Najman , Grupowanie dynamiczne z wykorzystaniem samouczących się sieci GNG	361
Małgorzata Misztal , Wpływ wybranych metod uzupełniania brakujących danych na wyniki klasyfikacji obiektów z wykorzystaniem drzew klasyfikacyjnych w przypadku zbiorów danych o niewielkiej liczebności – ocena symulacyjna	370
Mariusz Kubus , Zastosowanie wstępnego uwarunkowania zmiennej objaśnianej do selekcji zmiennych.....	380
Barbara Batóg, Jacek Batóg , Wykorzystanie analizy dyskryminacyjnej do identyfikacji czynników determinujących stopę zwrotu z inwestycji na rynku kapitałowym	387
Katarzyna Wójcik, Janusz Tuchowski , Analiza porównawcza miar podobieństwa tekstów opartych na macierzy częstości i tekstów opartych na wiedzy dziedzinowej	396
Iwona Staniec , Analiza czynnikowa w identyfikacji obszarów determinujących doskonalenie systemów zarządzania w polskich organizacjach	406
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Analiza porównawcza wybranych technik eksploracji danych do klasyfikacji danych medycznych z brakującymi obserwacjami	416
Iwona Foryś , Wykorzystanie analizy log-liniowej do wyboru czynników determinujących atrakcyjność cenową mieszkań w obrocie wtórnym na przykładzie lokalnego rynku mieszkaniowego.....	426
Ewa Genge , Analiza skupień oparta na mieszankach uciętych rozkładów normalnych.....	436
Jerzy Korzeniewski , Ocena efektywności metody uśredniania zmiennych i metody Ichino selekcji zmiennych w analizie skupień	444
Andrzej Dudek , SMS – propozycja nowego algorytmu analizy skupień	451
Artur Mikulec , Metody oceny wyniku grupowania w analizie skupień.....	460
Małgorzata Machowska-Szewczyk , Algorytm klasyfikacji rozmytej dla obiektów opisanych za pomocą zmiennych symbolicznych oraz rozmytych	469
Artur Zaborski , Analiza PROFIT i jej wykorzystanie w badaniu preferencji	479
Karolina Bartos , Analiza skupień wybranych państw ze względu na strukturę wydatków konsumpcyjnych obywateli – zastosowanie sieci Kohonena	488

Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Klasyfikacja gospodarstw domowych ze względu na bodźce do zawierania umowy o ubezpieczenie z wykorzystaniem modeli zmiennych jakościowych .	496
Izabela Kurzawa , Zastosowanie modelu LA/AIDS do badania elastyczności cenowych popytu konsumpcyjnego w gospodarstwach domowych w relacji miasto–wieś	505
Aleksandra Łuczak, Feliks Wysocki , Metody porządkowania liniowego obiektów opisanych za pomocą cech metrycznych i porządkowych	513
Agnieszka Sompolska-Rzechuła , Porównanie klasycznej i pozycyjnej taksonomicznej analizy zróżnicowania jakości życia w województwie zachodniopomorskim	523
Joanna Banaś, Małgorzata Machowska-Szewczyk , Ocena intensywności wykorzystania skrzynek poczty elektronicznej za pomocą uporządkowanego modelu probitowego	532
Iwona Bąk , Segmentacja gospodarstw domowych emerytów i rencistów pod względem wydatków na rekreację i kulturę	541
Aneta Becker , Zastosowanie metody ANP do porządkowania województw Polski pod względem dynamiki wykorzystania ICT w latach 2008-2010	552
Katarzyna Dębowska , Klasyfikacja sektorów ze względu na ich kondycję finansową przy użyciu metod wielowymiarowej analizy statystycznej	562
Anna Domagała , Propozycja metody doboru zmiennych do modeli DEA (procedura kombinowanego doboru w przód).....	571
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Analiza statystyczna w badaniach zapotrzebowania na usługi teleinformatyczne sieci łączności ruchomej	580
Hanna Gruchociak , Konstrukcja estymatora regresyjnego dla danych o strukturze dwupoziomowej.....	590
Tomasz Klimanek, Marcin Szymkowiak , Zastosowanie estymacji pośredniej uwzględniającej korelację przestrzenną w opisie niektórych charakterystyk rynku pracy	601
Jarosław Lira , Prognozowanie opłacalności produkcji żywca wieprzowego w Polsce	610
Christian Lis , Wykorzystanie metody klasyfikacji w ocenie konkurencyjności portów południowego Bałtyku	619
Beata Bieszk-Stolorz, Iwona Markowicz , Wykorzystanie wielomianowego modelu logitowego do oceny szansy podjęcia pracy przez bezrobotnych .	628
Lucyna Przezbórska-Skobiej, Jarosław Lira , Przestrzeń agroturystyczna Polski i ocena jej atrakcyjności.....	637
Paweł Ulman , Model rozkładu wydatków a funkcje popytu.....	646
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Zastosowanie metod analizy statystycznej w badaniach mięczaków	655

Summaries

Stanisława Bartosiewicz , The effects of subjectivism in multivariate analysis revisited.....	21
Andrzej Sokółowski , Q universal distance measure	30
Eugeniusz Gatnar , Data quality in central banks' statistical systems (NBP example)	38
Marek Walesiak , Distance measures for ordinal data – strategies of proceedings.....	46
Krzysztof Jajuga, Marek Walesiak , XXV years of taxonomic conferences – some facts and remarks.....	49
Józef Pocięcha, Barbara Pawelek , General SEM model in researching corporate bankruptcy and business cycles – theoretical and practical problems.....	57
Paweł Lula , Learning-based systems of information extraction from textual resources	67
Ewa Roszkowska , The application of the TOPSIS method to support the negotiation process	75
Andrzej Młodak , Neighborhood of spatial areas in the physical and socio-economic context – a taxonomic approach.....	85
Andrzej Bąk , Models for unordered categories in preference analysis.....	95
Kowalewski Jacek , An integrated model of optimizing statistical surveys	105
Jan Paradysz, Karolina Paradysz , Areas of unemployment in Poland – benchmark problem	115
Tomasz Szubert , How to play to lose the least? Classification of systems in sports bets	125
Izabela Szamrej-Baran , Classification of EU member states in view of fuel poverty	134
Sylvia Filas-Przybył, Tomasz Klimanek, Jacek Kowalewski , An attempt to use the gravity model in the analysis of commuters.....	143
Marta Dziechciarz-Duda, Anna Król, Klaudia Przybysz , Subsistence minimum versus factors influencing tendency to benefit from social care. Classification of households	152
Hanna Dudek , Subjective equivalence scales – analysis based on data about satisfaction with incomes.....	162
Joanicjusz Nazarko, Ewa Chodakowska, Marta Jarocka , Segmentation of universities using cluster analysis versus technological competitors determined by the DEA method – a comparative study	172
Ewa Chodakowska , Selected methods of classification in schools' rating.....	181
Bartosz Soliński , Renewable energy sector in the European Union – classification in the light of change management strategy	191
Krzysztof Szwarz , Classification of Wielkopolska voivodeship due to the demographic situation	201

Elżbieta Gołata, Grażyna Dehnel , Administrative registers in business analysis.....	211
Katarzyna Chudy, Marek Sobolewski, Kinga Stępień , Application of taxonomic methods in forecasting the profitability ratios of listed banks in Poland.....	221
Katarzyna Dębowska , Modeling bankruptcy of firms by using discrimination and regression methods.....	230
Alina Bojan , Identification of variables which influence attractiveness of given investments with the usage of multivariate analysis.....	239
Justyna Brzezińska , Log-linear analysis in the study of mortality in EU.....	246
Aneta Rybicka, Bartłomiej Jefmański, Marcin Pelka , Latent class analysis in student satisfaction surveys.....	254
Bartłomiej Jefmański , The respondent's opinions measurement in the R program with an application of fuzzy sets theory.....	264
Julita Stańczuk , A comparison of the results of multistate classification of economic objects using discriminant analysis and artificial neural networks.....	274
Jerzy Krawczuk , Effectiveness of classification methods in S&P500 stock index direction changes forecasting.....	282
Anna Czapkiewicz, Beata Basiura , The simulation study of the utility of the Copula-GARCH models for clustering financial time series.....	290
Radosław Pietrzyk , Timing and selectivity in mutual funds performance measurement.....	305
Aleksandra Witkowska, Marek Witkowski , Use of the Panzar-Rosse method to assess of the competition level in the cooperative banks sector.....	314
Marcin Pelka , Ensemble learning with the application of <i>boosting</i> in symbolic data analysis.....	322
Justyna Wilk , Comparative study of symbolic data classification software.....	332
Tomasz Bartłomowicz, Justyna Wilk , Application of symbolic data analysis methods for domain database searching.....	341
Kamila Migdał-Najman , A proposal of hybrid clustering method based on self-learning networks.....	351
Dorota Rozmus , Comparison of accuracy of spectral clustering and cluster ensembles stability based on bagging idea.....	360
Krzysztof Najman , A dynamic grouping based on self-learning GNG networks.....	369
Małgorzata Misztal , Influence of data imputation methods on the results of object classification using classification trees in the case of small data sets – simulation assessment.....	379
Mariusz Kubus , The application of pre-conditioning of explanatory variable for feature selection.....	386
Barbara Batóg, Jacek Batóg , Application of discriminant analysis to the identification of factors determining the rate of return on the capital market.....	395

Katarzyna Wójcik, Janusz Tuchowski , Comparative analysis of text documents similarity measures based on frequency matrix and based on domain knowledge.....	405
Iwona Staniec , Factor analysis in the identification of areas that determine the improvement of management systems in Polish organizations.....	415
Marek Lubicz, Maciej Zięba, Adam Rzechonek, Konrad Pawełczyk, Jerzy Kołodziej, Jerzy Błaszczyk , Comparative analysis of selected data mining approaches to the classification of medical data with missing values (covariates).....	425
Iwona Foryś , The log-linear analysis using to select the factors determining the attractiveness of the price of flats on the secondary market on the example of local housing market.....	435
Ewa Genge , Trimming approach to the mixtures of normal distributions.....	443
Jerzy Korzeniewski , Efficiency assessment of Ichino method and mean value method of selecting variables in cluster analysis.....	450
Andrzej Dudek , SMS – proposal of new clustering algorithm.....	459
Artur Mikulec , Evaluation methods for the grouping result in cluster analysis.....	468
Małgorzata Machowska-Szewczyk , Fuzzy clustering algorithm for objects described by symbolic or fuzzy variables.....	478
Artur Zaborski , PROFIT analysis and its using in the research of preferences.....	487
Karolina Bartos , Cluster analysis of selected countries due to the structure of their citizens' consumer expenditures – the use of Kohonen networks.....	495
Barbara Batóg, Magdalena Mojsiewicz, Katarzyna Wawrzyniak , Classification of households according to the impulses of concluding the insurance contract by means of qualitative variable models.....	504
Izabela Kurzawa , The application of LA/AIDS model to examine price elasticities of demand of households in the urban-rural relationship.....	512
Aleksandra Luczak, Feliks Wysocki , Linear ordering methods of objects described by a set of metric and ordinal characteristics.....	522
Agnieszka Sompolska-Rzechuła , The comparison of the classical and positional taxonomic analysis of the quality of life differentiation in Zachodniopomorskie voivodeship.....	531
Joanna Banaś, Małgorzata Machowska-Szewczyk , Evaluation of intensity of mailboxes using with the ordered probit model.....	540
Iwona Bąk , Segmentation of pensioners and annuitants households in terms of expenditures on recreation and culture.....	551
Aneta Becker , Application of ANP method to organize Polish voivodships in terms of dynamics of the use of ICT in 2008-2010.....	561
Katarzyna Dębowska , The classification of sectors' financial situation using the methods of multivariate statistical analysis.....	570

Anna Domagała , Proposal of a new method for variable selection in DEA models (combined forward stepwise selection method).....	579
Henryk Gierszal, Karina Pawlina, Maria Urbańska , Statistical analysis in demand research of ICT services in mobile networks.....	589
Hanna Gruchociak , Construction of regression estimator for two-level data	600
Tomasz Klimanek, Marcin Szymkowiak , Application of spatial models in indirect estimation of some labor market characteristics	609
Jarosław Lira , Forecasting of hog livestock production profitability in Poland	618
Christian Lis , The utilization of taxonomic methods in the appraisal of competitiveness of south Baltic ports	627
Beata Bieszk-Stolorz, Iwona Markowicz , The application of the multinomial logit model in evaluating employment odds for the unemployed job seekers	636
Lucyna Przezbórska-Skobiej, Jarosław Lira , Agritourism space of Poland and its valuation.....	645
Paweł Ulman , Model of expenses distribution and demand functions.....	654
Maria Urbańska, Tadeusz Mizera, Henryk Gierszal , Methods of statistical analysis in research of molluscs	663

Sylwia Filas-Przybył

Urząd Statystyczny w Poznaniu, Ośrodek Statystyki Miast

Tomasz Klimanek

Uniwersytet Ekonomiczny w Poznaniu

Jacek Kowalewski

Urząd Statystyczny w Poznaniu

ANALIZA DOJAZDÓW DO PRACY ZA POMOCĄ MODELU GRAWITACJI

Streszczenie: W artykule podjęto próbę wykorzystania modelu grawitacji do opisu ciężenia określonych jednostek podziału przestrzennego kraju do ośrodków miejskich, odgrywających istotną rolę w analizie lokalnych i regionalnych rynków pracy. Dotychczas możliwość aplikacji modeli grawitacji w warunkach polskich była ograniczona dostępnością danych, a także niewielkimi możliwościami technicznymi związanymi z przetwarzaniem informacji przestrzennej. Dopiero prace podjęte w 2008 r. przez Ośrodek Statystyki Miast Urzędu Statystycznego w Poznaniu wypełniły lukę informacyjną w zakresie badań dojazdów do pracy, która istniała w Polsce od 1988 r.

Słowa kluczowe: dojazdy do pracy, geoinformacja, GIS, model grawitacji, przepływy ludności.

1. Wstęp

Metody statystyki przestrzennej w ciągu ostatnich lat są coraz silniej obecne w polskiej literaturze naukowej z zakresu analizy danych. Chociaż w warstwie rozważań teoretycznych publikacje krajowe nie odbiegały od prac autorów zachodnioeuropejskich, to jednak przez długie lata raził ewidentny brak artykułów i publikacji zwartych, prezentujących najnowsze odkrycia z rozbudowaną warstwą zastosowań. Wydaje się, że główną przyczyną takiego stanu rzeczy było przede wszystkim zapóźnienie technologiczne (brak dostępu do specjalistycznego oprogramowania służącego przetwarzaniu map numerycznych oraz mało wydajny sprzęt komputerowy), ale także brak danych statystycznych zorientowanych przestrzennie, tzn. mających komponent odpowiadający za lokalizację (współrzędne geograficzne lub inny rodzaj odwzorowania układu kartezjańskiego X0Y).

Delimitacja przestrzeni wymaga działania na podstawie wypracowanej jednoznacznej metodologii uwzględniającej określone kryteria. Prace badawcze umożliwiające określenie związków międzyregionalnych i wyznaczenie granic oddziaływania różnych ośrodków na otaczające je obszary w swej metodyce uwzględniają jako jedno z głównych kryteriów delimitacyjnych dojazd do pracy. W polskiej praktyce statystycznej brakowało dotąd aktualnych danych o tym zjawisku. Dostrzeżenie tej luki, która ujawniła się zwłaszcza w związku z projektem Urban Audit, skłoniło do podjęcia próby określenia możliwości oszacowania natężenia i kierunków dojazdów do pracy przy wykorzystaniu tzw. źródeł administracyjnych. Prace w tym zakresie podjęto i zrealizowano w Ośrodku Statystyki Miast Urzędu Statystycznego w Poznaniu [Kruszka 2009].

Wyniki badania przeprowadzonego przez Ośrodek Statystyki Miast mogą stanowić doskonałą bazę do prowadzenia dalszych analiz przestrzennych. W niniejszym artykule autorzy podejmują, pierwszą w polskich warunkach, próbę wykorzystania danych z badania dojazdów do pracy związanych z zatrudnieniem do konstrukcji modelu grawitacji – jednego z podstawowych modeli wykorzystywanych w analizach oddziaływania jednostek (gmin, powiatów itd.) w przestrzeni. Kolejnym etapem badań będzie próba stworzenia tzw. mapy dostępności powiatów w analizie lokalnych rynków pracy.

2. Przepływy ludności związane z zatrudnieniem w 2006 r. – charakterystyka badania

Na podstawie zapisu w „Programie badań statystycznych statystyki publicznej” Ministerstwo Finansów udostępniło na potrzeby badania informacje znajdujące się na formularzach podatkowych dotyczących roku 2006 (PIT-11/8B i PIT-40), które wpłynęły do urzędów skarbowych na terenie całego kraju, a sporządzane zostały wyłącznie przez pracodawcę. Odpowiednie zbiory informacji zostały utworzone przez urzędy skarbowe według aplikacji przygotowanej przez Izbę Skarbową w Poznaniu, a Ministerstwo Finansów, z zachowaniem wszelkich procedur ochrony danych, przekazało je do statystyki publicznej. W strukturze zbioru dla każdego rekordu uwzględnione zostały wszystkie niezbędne informacje pozwalające na realizację badania zgodnie z metodologią opracowaną i przetestowaną wcześniej w postępowaniu pilotażowym „Przepływy ludności związane z zatrudnieniem w województwie wielkopolskim w 2004 r.”, które również wykonano w Urzędzie Statystycznym w Poznaniu.

Biorąc pod uwagę cel badania i wyżej określone administracyjne źródło danych, przyjęto, że zbiorowość dojeżdżających do pracy będzie obejmowała pracowników najemnych, którzy mieszkają w gminie innej niż ta, gdzie znajduje się ich miejsce pracy, i jednocześnie mają zwiększone koszty uzyskania przychodu z tytułu dojazdów do pracy. Generowanie tak zdefiniowanej zbiorowości wymagało etapowości prac zmierzających do jak najlepszej identyfikacji tworzących ją jednostek.

Założono, że tzw. momentem krytycznym badania będzie 31.12.2006 r. Zatem zbiór informacji został ograniczony do rekordów (formularzy), które w polu zamykającym okres rozliczeniowy zawierały tę datę. Na takim zbiorze prowadzono następnie prace sprawdzające zgodność słownych zapisów cech adresowych płatnika oraz podatnika z zastosowanymi symbolami terytorialnymi z rejestru TERYT. Porównania dokonano na podstawie wykazu jednostek podziału terytorialnego kraju według stanu na 1.01.2007 r. Podstawowymi jednostkami terytorialnymi, pomiędzy którymi następował przepływ ludności związany z zatrudnieniem, w omawianym badaniu były gminy. W przypadku gmin miejsko-wiejskich dla potrzeb podatkowych wystarczający jest symbol terytorialny obejmujący całą gminę. W badaniu przepływów ludności istotna jest jednak także relacja „miasto-wieś”. Stąd, wykorzystując wszystkie cechy adresowe występujące na formularzach podatkowych, na podstawie zawartości rejestru TERYT (wykaz miejscowości, wykaz ulic) dokonano podziału na część miejską i wiejską we wszystkich gminach miejsko-wiejskich. Zgodnie z przyjętymi założeniami zbiorowość dojeżdżających do pracy miała obejmować tylko pracowników najemnych. Jednakże informacje zawarte na formularzach PIT 11/8B i PIT-40 dotyczyły populacji szerszej zakresowo pod względem źródeł dochodu. Zapisy ustawowe powodowały również, że ten sam podatnik mógł wystąpić w bazie danych kilkakrotnie z tymi samymi lub różnymi źródłami dochodu. Dlatego w kolejnym etapie prac – z przygotowanego wcześniej zbioru rekordów – programowo wygenerowana została baza danych z unikatowymi numerami NIP tych podatników, dla których źródłem przychodu w 2006 r. była praca najemna. Trafiły do niej te rekordy, które w dziale dotyczącym przychodów miały wypełniony wiersz uwzględniający wynagrodzenia ze stosunku: pracy, służbowego, spółdzielczego i z pracy nakładczej, a także zasiłki pieniężne z ubezpieczenia społecznego wypłacane przez zakład pracy. Równocześnie wyeliminowane zostały powtarzające się numery NIP podatnika z wypełnioną pozycją dotyczącą wynagrodzeń, do czego wykorzystano kryterium wysokości dochodu.

W opisany powyżej sposób powstał zbiór podatników liczący w skali kraju ponad 9,5 mln osób. Stanowił on bazę danych, z której wyodrębniono zbiorowość osób dojeżdżających do pracy. Zgodnie z przyjętą definicją osoba dojeżdżająca do pracy miała spełniać następujące warunki: gmina jej zamieszkania musiała być różna od gminy miejsca pracy i jednocześnie zwiększone były koszty uzyskania przychodu z tytułu dojazdu do pracy. Zbiór osób spełniających wyżej zapisane założenia liczył w skali kraju ponad 2,3 mln osób. W końcowym etapie opracowania danych przystąpiono do weryfikacji otrzymanych wcześniej wyników, głównie pod kątem prawidłowej lokalizacji miejsca pracy, wykorzystując do tego celu informacje zawarte w BJS (Baza Jednostek Statystycznych). W rezultacie otrzymano macierz przepływu pracowników najemnych, którzy dojeżdżają do pracy poza gminę swojego zamieszkania.

3. Model grawitacji w analizie dojazdów do pracy

Model grawitacji jako koncepcja naukowa związany jest z XVII-wiecznymi badaniami I. Newtona i R. Hooke'a, których spór o autorstwo tej koncepcji nawet dziś trudno jest rozstrzygnąć¹. Klasyczna postać modelu grawitacji jest następująca:

$$F = G \frac{m_1 m_2}{r^2}, \quad (1)$$

gdzie: F – grawitacja (siła wzajemnego oddziaływania dwóch ciał),
 G – stała grawitacji,
 m_1, m_2 – masy ciał odpowiednio: pierwszego i drugiego,
 r – odległość między ciałami.

Zastosowania teorii grawitacji w ekonomii pojawiły się dopiero w połowie XX wieku i związane są z pracami A. Voorheesa, który zastosował model grawitacji w analizie problemów transportowych, oraz J. Tinbergen, który podjął próbę wykorzystania modelu grawitacji do analizy wymiany międzynarodowej [Anderson 2010].

Zaprezentowana w niniejszym artykule propozycja zastosowania modelu grawitacji do analizy dojazdów do pracy jest pierwszą tego typu próbą w warunkach polskich. Zmodyfikowana postać modelu grawitacji jest następująca:

$$T_{ij} = k \frac{P_i^\lambda P_j^\beta}{d_{ij}^\alpha}, \quad (2)$$

gdzie: i – lokalizacja i -ta,
 j – lokalizacja j -ta,
 T_{ij} – dojazdy do pracy między lokalizacjami i oraz j (powiatem i oraz powiatem j) – liczba pracowników najemnych przemieszczających się pomiędzy powiatem i i powiatem j ,
 k – stała w modelu grawitacji,
 P_i – zmienna określająca rolę/masę lokalizacji i (lokalizacja pochodzenia) – liczba pracowników najemnych wyjeżdżających z powiatu i ,
 P_j – zmienna określająca rolę/masę lokalizacji j (lokalizacja celu) – liczba miejsc pracy dla pracowników najemnych w powiecie j ,
 d_{ij} – odległość między lokalizacjami i oraz j ,
 λ – parametr określający potencjał do generowania przepływów (emisyjność),
 β – parametr określający potencjał do przyciągania przepływów (atrakcyjność),
 α – parametr określający opór przestrzeni w generowaniu przepływów między lokalizacjami i oraz j .

¹ Niewątpliwą zasługą I. Newtona jest systematyczny wykład teorii grawitacji przedstawiony w dziele *Philosophiae Naturalis Principia Mathematica* z 1687 r.

Szacowanie parametrów modelu klasyczną metodą najmniejszych kwadratów (KMNK) przeprowadzane jest po jego zlogarytmowaniu do postaci liniowej:

$$\ln(T_{ij}) = \ln k + \lambda \ln P_i + \beta \ln P_j + \alpha \ln d_{ij}. \quad (3)$$

Pewnego wyjaśnienia wymaga sposób wyznaczania odległości między lokalizacjami i oraz j . Autorzy zdają sobie sprawę, że najwłaściwszym podejściem w analizie byłoby uwzględnienie kosztu lub czasu pokonania przestrzeni między tymi dwiema lokalizacjami. Niestety, jedynymi informacjami, które można wykorzystać do określenia odległości, była fizyczna odległość liczona w kilometrach w linii prostej pomiędzy centroidami obiektów geometrycznych, które tworzą figury powiatów na mapie numerycznej.

Pierwsze uzyskane modele charakteryzowały się już dosyć wysokim stopniem dopasowania ($R^2 = 0,67$), jednakże analiza reszt z modelu wykazała, że w przypadku par powiatów, takich jak: Poznań – powiat poznański ziemski, Kalisz – powiat kaliski, Leszno – powiat leszczyński oraz Konin – powiat koniński występowało olbrzymie przeszacowanie dojazdów do pracy w stosunku do danych rzeczywistych². Wynika to przede wszystkim z faktu, że odległości między centroidami w przypadku tych par są bardzo małe lub praktycznie równe zero (rys. 1). Dlatego w przypadku



Rys. 1. Centroidy powiatu poznańskiego ziemskiego i miasta Poznania

Źródło: opracowanie własne z pomocą programu SAS/GIS.

² Według uzyskanego modelu liczba osób dojeżdżających z powiatu poznańskiego do Poznania wynosiła 15 425 539, a z Poznania do powiatu poznańskiego 6 885 590.

wymienionych wyżej par powiatów autorzy postanowili połączyć każdą z par w jeden większy twór, który obejmuje w każdym z przypadków miasto na prawach powiatu oraz związany z danym miastem powiat ziemski.

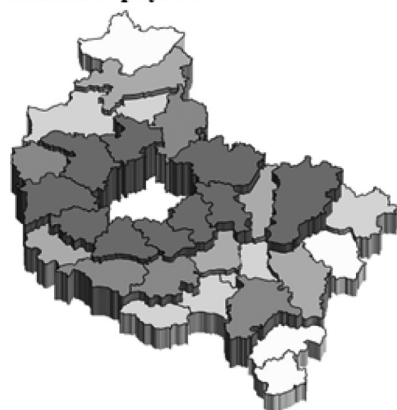
Stopień dopasowania modelu mierzony współczynnikiem determinacji wyniósł obecnie 71%, wszystkie parametry modelu okazały się istotne statystycznie (tab. 1), przy czym warto zwrócić uwagę, że parametr będący wykładnikiem potęgi przy odległości wynosi prawie 3. Świadczy to o tym, że w warunkach polskich odległość wciąż pozostaje bardzo dużą barierą przy przemieszczaniu się pracowników najemnych³.

Tabela 1. Parametry oszacowanego modelu grawitacji

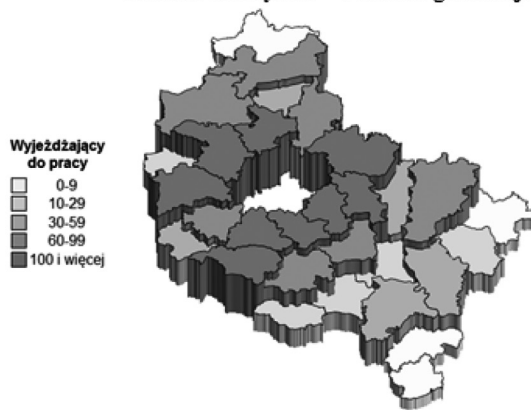
Zmienna	Oszacowanie parametru	Błąd standardowy	Statystyka <i>t</i> -Studenta	Prawdopodobieństwo testowe
Wyraz wolny	-5,84943	1,02152	-5,73	<,0001
LN_X1	0,69755	0,06735	10,36	<,0001
LN_X2	1,34890	0,06266	21,53	<,0001
LN_D	-2,94240	0,08992	-32,72	<,0001

Źródło: obliczenia własne z pomocą programu SAS.

wartości empiryczne



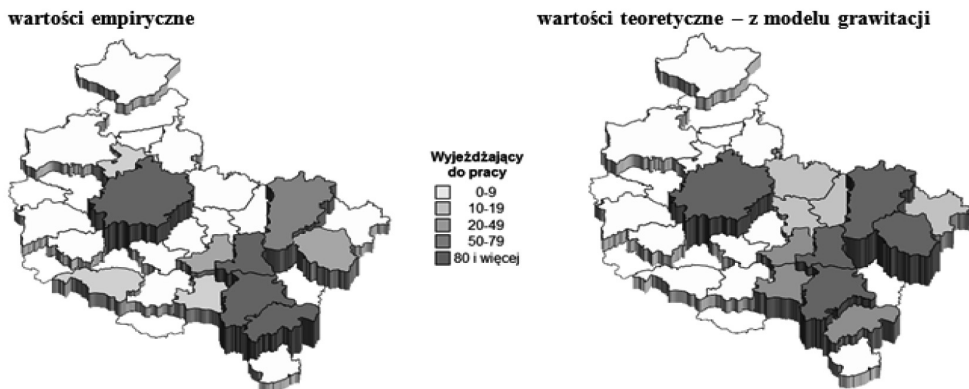
wartości teoretyczne – z modelu grawitacji



Rys. 2. Stopień dopasowania modelu grawitacji do danych empirycznych dla powiatu Poznań (powiat grodzki wraz z powiatem ziemskim)

Źródło: opracowanie własne z pomocą programu SAS/GIS.

³ W krajach zachodnioeuropejskich „opór przestrzeni” jest dużo niższy. W pracy opisującej badania szwedzkie [Eliasson, Strom 2008, s. 17-18] parametr będący wykładnikiem potęgi przy odległości wynosi zaledwie 0,1.



Rys. 3. Stopień dopasowania modelu grawitacji do danych empirycznych dla powiatu kaliskiego (powiat grodzki wraz z powiatem ziemskim)

Źródło: opracowanie własne z pomocą programu SAS/GIS.

Analiza reszt nie wykazała znacznych różnic między wartościami empirycznymi a wartościami teoretycznymi wynikającymi z modelu (rys. 2 i 3)⁴.

4. Wnioski

Badanie przeprowadzone przez Ośrodek Statystyki Miast Urzędu Statystycznego w Poznaniu – „Przepływy ludności związane z zatrudnieniem w 2006 r.” – oraz jego wyniki udostępnione na stronie internetowej urzędu w postaci macierzy przepływów [Internet 1] umożliwiają prowadzenie wielu prac analitycznych z zakresu geostatystyki. Celem niniejszego artykułu było zaprezentowanie próby zbudowania modelu opisującego kształtowanie się dojazdów do pracy za pomocą newtonowskiej koncepcji grawitacji.

Należy stwierdzić, że uzyskane wyniki są satysfakcjonujące, a uzyskany model jest zgodny nie tylko z teorią, ale także z praktyką badań prowadzonych w innych krajach zachodnioeuropejskich:

- stopień dopasowania modelu na poziomie 71% wskazuje, że spora część zmienności przepływów między powiatami województwa wielkopolskiego została wyjaśniona za pomocą skonstruowanego modelu. Należy także zwrócić uwagę, że modele prezentowane w cytowanych powyżej publikacjach zachodnioeuropejskich charakteryzują się współczynnikiem determinacji na poziomie 80%,

⁴ Z powodu ograniczeń nałożonych na autorów w artykule zamieszczono jedynie kartogramy z wynikami modelowania grawitacji dla powiatu Poznań oraz powiatu kaliskiego (w obydwu przypadkach chodzi o powiat grodzki wraz z powiatem ziemskim).

- przyjęcie „sztucznego” założenia w odniesieniu do czterech powiatów województwa wielkopolskiego (połączenie powiatów grodzkich i ziemskich) ograniczyło liczbę analizowanych jednostek i spowodowało większą agregację modelu, ale z drugiej strony spowodowało poprawę jakości modelu i wyeliminowanie dużych reszt wynikających z modelu,
- parametry modelu wskazują na to, że powiaty województwa wielkopolskiego cechują się większą atrakcyjnością niż emisyjnością – mają większą siłę przyciągania pracowników najemnych z innych powiatów niż tendencję do skłaniania pracowników najemnych z danego powiatu do poszukiwaniu pracy w innych powiatach,
- przestrzeń, w warunkach polskich, stanowi bardzo silną barierę dla mobilności pracowników najemnych przy przemieszczaniu się między powiatami.

Dalsze prace nad wykorzystaniem modelu grawitacji w analizie dojazdów do pracy związanych z zatrudnieniem będą się koncentrować na:

- 1) konstrukcji modelu grawitacji dla powiatów w innych województwach,
- 2) konstrukcji modelu grawitacji dla wszystkich powiatów w Polsce,
- 3) pracach metodologicznych nad rozwiązaniem problemu wyznaczania centroidów powiatów,
- 4) próbach wykorzystania innych niż odległość między centroidami zmiennych do oceny pomiaru oporów przestrzeni (czas, koszty),
- 5) pracach związanych z budową kartogramów przedstawiających dostępność lokalnych (powiaty) rynków pracy.

Literatura

- Anderson J.E., *The Gravity Model*, National Bureau of Economic Research, Cambridge 2010.
- Eliasson J., Strom M., *Commuting patterns in Sweden*, Jönköping International Business School, 2008.
- Harsman B., Quigley J.M., *Education, Job Requirements and Commuting an Analysis of Network Flows*, University of California Transportation Center, 1998.
- Kruszka K. (red.), *Przepływy ludności związane z zatrudnieniem w 2006 roku*, Urząd Statystyczny w Poznaniu, 2009.
- Signorino G., Pasetto R., Gatto E., Mucciardi M., La Rocca M., Mudu P., *Gravity models to classify commuting vs. resident workers. An application to the analysis of residential risk in a contaminated area*, „International Journal of Health Geographics” 2011, <http://www.ij-healthgeographics.com/content/pdf/1476-072X-10-11.pdf>.

Źródło internetowe

- [1] http://www.stat.gov.pl/poznan/69_567_PLK_HTML.htm.

AN ATTEMPT TO USE THE GRAVITY MODEL IN THE ANALYSIS OF COMMUTERS

Summary: The article is an attempt to use the gravity model to describe how some territorial units gravitate towards cities, which play a very important role in the analysis of local and regional labor markets. Until recently the applicability of gravity models in Poland had been rather difficult owing to the limited access to data about commuters and technical limitations connected with efficient spatial information processing. Only in 2008, thanks to work conducted by the Centre for Urban Statistics of the Statistical Office in Poznan, did it become possible to fill the information gap existing in the Polish statistics since 1988 with respect to data about commutes.

Keywords: commuters, geoinformation, GIS, gravity model, population flows.