

PRACE NAUKOWE

Uniwersytetu Ekonomicznego we Wrocławiu

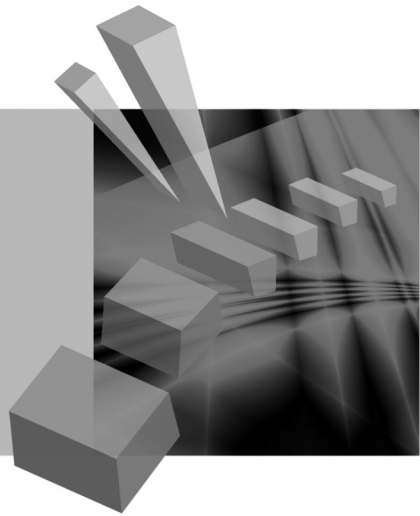
RESEARCH PAPERS

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,
Mirosław 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
Sylwia 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 Szwarc , 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

Jan Paradysz, Karolina Paradysz

Uniwersytet Ekonomiczny w Poznaniu

OBSZARY BEZROBOCIA W POLSCE – PROBLEM BENCHMARKOWY

Streszczenie: W statystyce małych obszarów istnieje wiele estymatorów pośrednich, co do których powstaje problem kryteriów wyboru oraz oceny wyników estymacji. Jedno z bardziej perspektywicznych rozwiązań stwarza statystyka lokalnego rynku pracy. Badanie aktywności ekonomicznej ludności jest systematyczne, powtarzalne, obejmuje wszystkie gospodarstwa domowe. Można je zintegrować ze statystyką ludności oraz z rejestrami administracyjnymi, dzięki czemu dysponujemy względnie stałą bazą zmiennych wspomagających. Szacując wszystkie elementy rynku pracy i sumując realizację odpowiednich estymatorów, mogliśmy porównać je z liczbą ludności i ocenić ich dokładność. W wyniku tego porównania trochę lepszy okazał się estymator EB. Estymatory pośrednie charakteryzowały się też mniejszymi standardowymi błędami szacunku.

Słowa kluczowe: estymacja dla małych obszarów, lokalny rynek pracy, benchmarking statystyczny.

Estymacja dla małych obszarów jest sztuką mniej lub bardziej arbitralnych wyborów, wśród których znalezienie pewnych obiektywnych kryteriów oceny trafności wyników byłoby znacznym postępem. Zaproponowane niedawno przez J. Paradysza [2008] kryteria dobroci estymacji dla małych obszarów nie uwzględniały konkretnych wzorców (benchmarków), do których można byłoby porównać dane wyniki. Sformułowany tam postulat zbudowania syntetycznego wskaźnika rozwoju zjawiska, będącego przedmiotem estymacji, wydaje się interesującą, ale dotychczas niesprawdzoną propozycją rozwiązania tego problemu¹. W niniejszym opracowaniu podejmiemy do oceny dobroci estymacji dla małych obszarów bardziej konkretnie i zaproponujemy częściowe rozwiązanie problemu, jakie nam stwarza rynek pracy. Mianowicie, po oszacowaniu wszystkich kategorii ludności na rynku pracy – bezrobotnych, pracujących i biernych – ich sumę skonfrontujemy z liczbą ludności według faktycznego miejsca zamieszkania. Badanie aktywności ekonomicznej ludności

¹ Inna z metod polegająca na porównaniu sumy realizacji danego estymatora dla małych obszarów, czyli J. Paradysza [2008, wzór 1] kryterium poziomym, była przedmiotem naszego poprzedniego opracowania. Chodziło tam o tzw. raking (por. [Paradysz, Paradysz 2011]), gdzie wykorzystano realizację estymatorów z książki E. Gołaty [2004].

(BAEL) pozwala nam na estymację poszczególnych elementów rynku pracy: bezrobotnych, pracujących, biernych i aktywnych zawodowo. Nie stanowi dla nas problemu, że ostatnia z tych kategorii ludności jest sumą bezrobotnych i pracujących. Wręcz przeciwnie, pozwala nam to na podejście dwuwariantowe w naszej propozycji: a) oszacowanie każdej z trzech kategorii osobno, b) oszacowanie tylko biernych i aktywnych. Pierwszy z wariantów ma większe znaczenie merytoryczne, gdyż różni te kategorie ludności, co do których informacja ma szczególne znaczenie społeczne i ekonomiczne. Ludność aktywna zawodowo, jako suma bezrobotnych i pracujących, budzi zapewne mniejsze zainteresowanie odbiorców, ale łatwiej poddaje się modelowaniu ekonometrycznemu.

Celem prezentacji metody wykorzystujemy BAEL z 2005 r.² w odniesieniu do powiatów liczących powyżej 100 tysięcy mieszkańców „faktycznie zamieszkałych”³. Według tej kategorii ludności przeprowadzane są wszystkie badania aktywności ekonomicznej ludności. Ze względu na znaczny nakład pracy, niejednoznaczność ujęcia⁴ oraz metodologiczny charakter pracy nie połączyliśmy powiatów ziemskich z grodzkimi, co dalej może mieć pewien wpływ na jakość wyników poznawczych. Z analizy wyłączyliśmy trzy największe miasta Polski, które przed podziałem administracyjnym z 1999 r. były w zasadzie miastami na prawach województw i mają swoją specyfikę. Wstępna analiza współzależności między elementami składowymi rynku pracy a potencjalnymi zmiennymi objaśniającymi pokazała, że Warszawa, Łódź i Kraków znacznie odbiegają od reszty powiatów w Polsce.

Będąc do naszej dyspozycji dane z BAEL 2005 były w postaci obszernych tabeli statystycznych obejmujących poszczególne elementy rynku pracy według powiatów, płci oraz wieku. W niniejszej pracy ograniczymy się tylko do ludności ogółem w wieku zdolności do pracy bez rozróżniania płci i wieku. W tabeli 1 przedstawiamy wybranych 14 powiatów, z czego 7 charakteryzuje się najniższym odsetkiem bezrobotnych, a w 7 pozostałych jest on najwyższy.

Do oszacowania struktury rynku pracy w poszczególnych powiatach wykorzystamy 3 estymatory, które należą do najczęściej wykorzystywanych w estymacji pośredniej⁵: Horwitza-Thompsona (HT_p), regresyjny ($\hat{y}_d(p)$) i empiryczno-bayesowski ($\hat{y}_{ap}(EB)$). HT_{ap} jest estymatorem bezpośrednim liczby jednostek wy-

² Wybór roku analizy był podyktowany naszą dostępnością do danych, ale ma także uzasadnienie merytoryczne. Jest to pierwszy rok po wstąpieniu Polski do Unii Europejskiej i szerszego dostępu dla Polaków do rynku pracy za granicą.

³ Ludność faktyczna obejmuje liczbę mieszkańców stałych w danej miejscowości powiększona o saldo zameldowanych i wymeldowanych na pobyt czasowy ponad 3 miesiące.

⁴ O ile połączenie niedużych powiatów grodzkich, jak na przykład miasto Krosno, z ziemskim wydaje się celowe, o tyle połączenie miasta Poznania z poznańskim może być już dyskusyjne. Jeszcze więcej wątpliwości mogą nasuwać powiaty ziemskie, które są tylko częścią strefy przylegającej do powiatu grodzkiego, na przykład wrocławski i m. Wrocław.

⁵ Poszczególne modele estymacji pośredniej, które tutaj wykorzystujemy, zaliczają się do jedno-poziomowego systemu analizy na poziomie przestrzennym (*area level models*), por. [Rao 2003, s. 115 i nast.].

Tabela 1. Liczba ludności faktycznie zamieszkałej, w wieku produkcyjnym, oraz jej struktura według sytuacji na rynku pracy w powiatach powyżej 100 tys. mieszkańców w 2005 r.

Powiat (a)	Liczba ludności N_a	Liczebność próby N_a	Odsetek (p)			
			pracują- cych (1)	bezrobotnych (2)	biernych (3)	brak da- nych (4)
Powiaty o najniższym odsetku bezrobotnych w próbie						
Piaseczyński	93 346	97	60,8	2,1	36,1	1,0
Warszawski zachodni	65 199	187	51,9	2,7	44,9	0,5
Zamojski	64 016	355	66,5	2,8	30,1	0,6
Nowotarski	109 782	232	45,3	3,4	48,7	2,6
M. Gdynia	165 973	752	48,5	4,3	46,8	0,4
Mielecki	82 323	328	47,6	4,3	47,0	1,2
Garwoliński	63 392	525	53,5	4,4	42,1	0,0
Powiaty o najwyższym odsetku bezrobotnych w próbie						
Dzierżoniowski	67 529	463	34,3	15,6	49,5	0,6
Lubiński	71 874	254	46,1	15,7	37,8	0,4
M. Radom	147 387	657	31,8	16,0	51,9	0,3
Kłódzki	108 044	618	25,9	16,7	57,1	0,3
Radomski	88 519	366	39,3	16,7	44,0	0,0
Ostrowiecki	74 217	402	37,6	16,9	45,3	0,2
M. Włocławek	79 941	512	37,7	17,8	44,3	0,2

Uwaga: powiaty uszeregowano rosnąco według odsetka bezrobotnych w próbie.

Źródło: opracowanie własne na podstawie BAEL 2005.

różnionych (pracujących, bezrobotnych bądź biernych) w małym obszarze (domenie) a , który wykorzystuje tylko informacje z próby:

$$HT_{a,p} = F_a \cdot \sum_{i=1}^{n_a} y_{api}, \quad (1)$$

gdzie: a – mały obszar (powiat),
 p – sytuacja (stan) na rynku pracy: 1 – pracujący, 2 – bezrobotny, 3 – bierny, 4 – brak danych,

$F_a = N_a/n_a$ – empiryczny wskaźnik uogólnienia dla a -tego małego obszaru,
 N_a – liczba badanych elementów w populacji generalnej w małym obszarze a ,
 n_a – liczba badanych elementów w próbie w małym obszarze a .

Fracja osób w małym obszarze w stanie p :

$$P_{a,p} = \frac{HT_{a,p}}{HT_a}. \quad (2)$$

Szacunkiem wariancji estymatora bezpośredniego frakcji osób wyróżnionych w stanie p jest wyrażenie

$$S^2(P_{a,p}) = P_{a,p} \cdot (1 - P_{a,p}). \quad (3)$$

Przy założeniu losowania nieograniczonego proseggo zależnego⁶ ocenę wariancji estymatora bezpośredniego liczby osób wyróżnionych w stanie p przedstawia wzór, por. [Zasepa 1972, s. 108]:

$$S^2(HT_{a,p}) = \frac{N_a^2(N_a - n_a)}{N_a - 1} \cdot \frac{S^2(P_{a,p})}{n_a}. \quad (4)$$

Jeśli N jest relatywnie duże, a n stosunkowo małe, to można uprościć sobie wzór (4) do wyrażenia (5):

$$S^2(HT_{a,p}) = N_a^2 \cdot \frac{S^2(P_{a,p})}{n_a}. \quad (5)$$

Estymator regresyjny szacowany jest na podstawie modelu regresji, w którym zmienną objaśnianą jest liczba ludności w odpowiednim stanie, a zmiennymi objaśniającymi (wspomagającymi) wskaźniki z banku danych lokalnych GUS. Z kilkudziesięciu zmiennych objaśniających wybrano do poszczególnych modeli następujące: X_1 – liczba ludności w miastach „ogółem”, X_2 – liczba kobiet „ogółem”, X_3 – bezrobotni zarejestrowani pozostający bez pracy dłużej niż 1 rok, ogółem, X_4 – liczba małżeństw, X_5 – liczba małżeństw cywilnych, X_6 – liczba urodzeń żywych, X_7 – liczba pracujących w głównym miejscu pracy, ogółem, X_8 – liczba osób w wieku 70-74 lata mężczyzn i kobiet razem.

Estymator regresyjny w ogólnej postaci dla konkretnego małego obszaru przedstawiamy według wzoru:

$$y_{ap} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k + \varepsilon, \quad (6)$$

wariancja estymatora regresyjnego a -tym obszarze:

$$S^2(\hat{y}_{ap}) = S^2(\varepsilon) \cdot \left(\frac{1}{n} + \frac{(x_{1a} - \bar{x}_1)^2}{\sum_i (x_{1i} - \bar{x}_1)^2} + \frac{(x_{2a} - \bar{x}_2)^2}{\sum_i (x_{2i} - \bar{x}_2)^2} + \dots + \frac{(x_{ka} - \bar{x}_k)^2}{\sum_i (x_{ki} - \bar{x}_k)^2} \right), \quad (7)$$

⁶ Potrzeba takiego założenia wynika z braku informacji o współczynnikach uogólnienia na poziomie powiatów.

gdzie: β_0 – wyraz wolny;
 $\beta_1, \beta_2, \dots, \beta_k$ – współczynniki regresji;
 ε – składnik losowy,
 $i = 1, 2, \dots, n$ – jest numerem w uporządkowaniu n małych obszarów,
 $S^2(\varepsilon)$ – wariancja składnika losowego.

Wśród modeli regresji dla pracujących najlepszy był model (8) z liczbą kobiet „ogółem” w powiecie jako zmienną wspomagającą, który objaśnił 94,2% wariancji wspólnej⁷:

$$y_{11} = -4366,78 + 0,5905 \cdot x_2 + \varepsilon \quad (8)$$

[1318,6] [0,0138] [6841,7]

$$R^2 = 0,9426; \quad R^2_{\text{skor}} = 0,9421; \quad V(\varepsilon) = 15,1\%; \quad ss = 112.$$

Porównanie standardowych błędów szacunku parametrów równania (8) z ich ocenami świadczy o ich statystycznej istotności. Współczynnik zmienności losowej ($V(\varepsilon) = 15,1\%$) nie jest zbyt wysoki i dla dalszej prezentacji problemu estymacji dla małych obszarów uznaliśmy go za wystarczająco dobry. Zupełnie do zaakceptowania jest także drugi z kolei modeli (9). Wielce istotne parametry modelu, wysoki współczynnik determinacji i relatywnie niski współczynnik zmienności resztowej ($R^2 = 0,916$; $V(\varepsilon) = 18,37\%$) z inną, alternatywną zmienną objaśnianą pozwala nam na niezależne oszacowanie liczby pracujących w poszczególnych powiatach.

$$y_{12} = -8362,7 + 61,3828 \cdot x_4 + \varepsilon \quad (9)$$

[1713,69] [1,7564] [8275,6]

$$R^2 = 0,9160; \quad R^2_{\text{skor}} = 0,9150; \quad V(\varepsilon) = 18,37\%; \quad ss = 112.$$

Wobec powyższego obydwa modele (8) i (9) stanowią też punkt wyjścia do konstrukcji estymatora empiryczno-bayesowskiego, który także wystąpi w dwóch wariantach, por. [Kordos, Paradysz 1999]:

$$\hat{y}_{ap}(EB_1) = \frac{S^2(HT_{a,p})}{S^2(HT_{a,p}) + S^2(\hat{y}_{p1})} \hat{y}_{ap1} + \frac{S^2(\hat{y}_{p1})}{S^2(HT_{a,p}) + S^2(\hat{y}_{p1})} HT_{a,p} \quad (10)$$

oraz

$$\hat{y}_{ap}(EB_2) = \frac{S^2(HT_{a,p})}{S^2(HT_{a,p}) + S^2(\hat{y}_{p2})} \hat{y}_{ap2} + \frac{S^2(\hat{y}_{p2})}{S^2(HT_{a,p}) + S^2(\hat{y}_{p2})} HT_{a,p}. \quad (11)$$

⁷ W równaniach regresji (4), (5) i dalszych, które zawierają składnik losowy ε , opuszczamy „dąsek” nad zmienną objaśnianą y , natomiast estymator regresyjny, jego realizacje oraz odpowiadające im wariancje oznaczamy przez \hat{y} .

Wariancję estymatora empiryczno-bayesowskiego w obu przypadkach można wyrazić za pomocą wzorów (12) i (13):

$$s^2[\hat{y}_{ap}(EB_1)] = \frac{S^2(HT_{a,p}) \cdot S^2(\hat{y}_{p1})}{S^2(HT_{a,p}) + S^2(\hat{y}_{p1})} \quad (12)$$

oraz

$$s^2[\hat{y}_{ap}(EB_2)] = \frac{S^2(HT_{a,p}) \cdot S^2(\hat{y}_{p2})}{S^2(HT_{a,p}) + S^2(\hat{y}_{p2})}, \quad (13)$$

- gdzie: $HT_{a,p}$ – estymator bezpośredni wartości globalnej w małym obszarze (domenie) a , dla p -tej kategorii rynku pracy (pracujący $p = 1$, bezrobotni $p = 2$, bierni zawodowo $p = 3$); $P_{a,p}$ wykorzystuje tylko informacje z próby;
- \hat{y}_{ap1} – estymator regresyjny wartości globalnej w małym obszarze (domenie) a , który wykorzystuje tylko informacje z modelu regresji (jedyńka w indeksie dolnym oznacza numer pierwszy, „najlepszy” model regresji – litera p jak wyżej);
- $\hat{y}_{ap}(EB_1)$ – estymator bayesowski wartości globalnej w małym obszarze (domenie) a , wykorzystujący informacje zarówno z modelu pierwszego regresji dla kategorii rynku pracy p , jak i z próby – podobne oznaczenie stosuje się dla pozostałych estymatorów bayesowskich, przy czym zmianie ulegają indeksy dolne, litery w tych indeksach informują, o którą kategorię rynku pracy chodzi;
- $S^2(HT_{a,p})$ – wariancja estymatora bezpośredniego wartości globalnej w małym obszarze (domenie) a , który wykorzystuje tylko informacje z próby, czyli danej kategorii rynku pracy;
- $S^2(\hat{y}_{p1})$ – wariancja estymatora regresyjnego wartości globalnej w małym obszarze (domenie) a , który wykorzystuje tylko informacje z modelu regresji (jedyńka w indeksie dolnym oznacza numer pierwszy, „najlepszy” model regresji);
- $s^2[\hat{y}_{ap}(EB_1)]$ – to wariancja estymatora bayesowskiego wartości globalnej w małym obszarze (domenie) a , który wykorzystuje informacje zarówno z modelu regresji (pierwszego lub drugiego), jak i z próby, uwzględniając kategorie (pracujący, bierni, bezrobotni) danych rynku pracy.

Znacznie gorsze opracowania modeli regresji uzyskaliśmy w przypadku estymacji bezrobotnych. Za najlepsze uznano tutaj modele z liczbą ludności w miastach (x_1) oraz liczbą bezrobotnych zarejestrowanych pozostających bez pracy dłużej niż 1 rok – ogółem (x_3), jako zmiennymi wspomagającymi:

$$y_{21} = 954,922 + 0,02634 \cdot x_1 + 1,01872 \cdot x_3 + \varepsilon \quad (14)$$

[508,8823]
[0,0026]
[0,101832]
[2487,9]

$$R^2 = 0,8029; \quad R^2_{\text{skor}} = 0,7993; \quad V(\varepsilon) = 26,8\%; \quad ss = 111.$$

$$y_{22} = 775,13 + 1,587647 \cdot x_3 + \varepsilon \quad (15)$$

[702,3]
[0,1173]
[3435,8]

$$R^2 = 0,6205; \quad R^2_{\text{skor}} = 0,6172; \quad V(\varepsilon) = 37,0\%; \quad ss = 112.$$

Modele regresji dla szacowania liczby ludności biernej zawodowo – równania (16) i (17) – charakteryzowały się wysokimi współczynnikami determinacji liniowej, które znacznie przekraczały 90%. Za najlepsze uznano tutaj modele z takimi zmiennymi wspomagającymi, jak x_5 – liczba małżeństw cywilnych, x_6 – liczba urodzeń żywych, x_7 – liczba pracujących w głównym miejscu pracy – ogółem, x_8 – liczba osób w wieku 70-74 lata mężczyzn i kobiet razem:

$$y_{31} = 4580,62 + 55,194 \cdot x_5 + 14,26 \cdot x_6 + 0,208 \cdot x_7 + \varepsilon \quad (16)$$

[1622,37]
[8,64]
[1,6397]
[0,048]
[6719,4]

$$R^2 = 0,9430; \quad R^2_{\text{skor}} = 0,9414; \quad V(\varepsilon) = 14,1\%; \quad ss = 110.$$

$$y_{32} = 1203 + 10,967 \cdot x_6 + 5,17 \cdot x_8 + \varepsilon \quad (17)$$

[1522,69]
[1,9914]
[0,4056]
[7058,02]

$$R^2 = 0,9365; \quad R^2_{\text{skor}} = 0,9353; \quad V(\varepsilon) = 14,9\%; \quad ss = 111.$$

Wyniki estymacji pozwalają nam na przedstawienie struktury rynku pracy według 5 powyższych estymatorów oraz precyzji ich szacunku. Próbkę tych wyników przedstawiamy w tab. 2.

Ludność faktycznie zamieszkała jako funkcja kryterium dobroci oszacowania pokazuje jakość estymacji pośredniej. Estymatory EB_1 , podobnie jak EB_2 , są trochę lepsze od bezpośredniego HT , którego obszar wahań różnic między ludnością ogółem oszacowaną na jego podstawie a faktyczną jest większy, od $-9,6\%$ do $9,4\%$. Jest to wynikiem względnie dużych prób, kiedy wewnątrz badanej populacji nie ma żadnych dalszych podziałów na wiek, płeć lub miejsce zamieszkania (miasto–wieś).

Rynek pracy jest uwarunkowany wieloma czynnikami po stronie zarówno podaży siły roboczej, jak i popytu na nią. Jeszcze większe problemy występują w pokryciu informacyjnym rynku pracy i jego determinant. Istnieje duże zróżnicowanie tych pojęć w czasie i w przestrzeni. Po pierwsze, za podobnej wielkości odsetkami bezrobotnych w dwóch wybranych regionach mogą się ukrywać rozmaitej jakości zasoby pracy, różny poziom rozwoju regionalnego, różne ich perspektywy gospodarcze i problemy społeczne. Na przykład wysoki wskaźnik poziomu bezrobocia poza sezonem w miejscowości atrakcyjnej pod względem turystycznym oznacza zupełnie coś innego od podobnej wielkości miernika w regionach pozbawionych możliwości rozwoju. Po drugie, problemem badania rynku pracy w Polsce, podobnie jak w innych krajach, jest rzetelność danych statystycznych i jednakowe rozumienie pojęć w odpowiednich przekrojach. Na przykład w publicystyce ekonomicznej podkreśla się

wpływ migracji zarobkowych na szarą strefę przez rejestrowanie się jako osoba bezrobotna i podejmowanie pracy za granicą. Bywają także osoby, które nie są zainteresowane jakimkolwiek zatrudnieniem lub nie mają szans na podjęcie pracy i wówczas nie zostaną ujęte w rejestrze bezrobotnych. Stąd potrzeba dwóch źródeł informacji: rejestru bezrobotnych oraz specjalnego badania siły roboczej, które w Polsce nosi nazwę badania aktywności ekonomicznej (BAEL).

Tabela 2. Ocena dobroci oszacowania rynku pracy na podstawie estymatorów HT oraz EB_1 przez porównanie z ludnością faktycznie zamieszkałą w dniu 30 czerwca 2005 r. Wszystkie liczby ludności (bezrobotnych, pracujących, biernych i razem) przedstawiono w tysiącach osób

Powiat	Bezrobotni		Pracujący		Bierni zawodowo		Razem na podstawie		Ludność faktyczna na 30 VI	% odchylenia od faktycznej	
	HT	EB_1	HT	EB_1	HT	EB_1	HT	EB_1		HT	EB_1
Powiaty o najniższej stopie bezrobocia											
Piaseczyński	1,9	1,9	56,8	51,4	33,7	36,8	92,4	90,2	93,3	-1,0	-3,4
Warszawski Zachodni	1,7	1,7	33,8	33,0	29,3	29,2	64,9	64,0	65,2	-0,5	-1,9
Zamojski	1,8	1,8	42,6	41,8	19,3	19,7	63,7	63,4	64,0	-0,6	-1,0
Nowotarski	3,8	3,8	49,7	49,8	53,5	52,3	106,9	106,0	109,8	-2,6	-3,4
Mielecki	3,5	3,5	39,2	38,8	38,7	38,7	81,3	81,1	82,3	-1,2	-1,5
Tarnowski	5,3	5,3	65,9	64,3	43,4	43,7	114,6	113,3	115,1	-0,4	-1,5
Bielski	1,6	1,6	18,9	19,1	14,1	14,3	34,5	35,0	34,6	-0,2	1,3
Powiaty o najwyższej stopie bezrobocia											
Starogardzki	11,2	11,1	31,2	31,3	33,7	34,0	76,0	76,4	76,4	-0,5	0,0
Nyski	14,2	14,0	35,8	36,1	44,4	44,1	94,3	94,1	94,5	-0,1	-0,4
Dzierżoniowski	10,5	10,2	23,2	23,4	33,4	33,3	67,1	66,9	67,5	-0,6	-0,9
M. Radom	23,6	23,4	46,9	49,4	76,5	74,7	146,9	147,4	147,4	-0,3	0,0
Radomski	14,8	14,6	34,8	35,2	38,9	38,7	88,5	88,6	88,5	0,0	0,1
Ostrowiecki	12,6	12,4	27,9	28,1	33,6	33,6	74,0	74,1	74,2	-0,2	-0,2
M. Włocławek	14,2	13,9	30,1	30,3	35,4	35,6	79,8	79,9	79,9	-0,2	-0,1
Minimum	1,6	1,6	17,5	17,8	14,1	14,3	34,5	35,0	34,6	-9,6	-8,6
Maksimum	38,6	38,2	204,0	200,4	181,6	185,1	424,1	423,7	426,3	9,4	7,6
Średnia	9,3	9,2	45,1	45,0	47,5	47,4	101,8	101,6	102,4	-0,5	-0,8

Źródło i uwagi jak w tab. 1.

Zastosowanie estymatorów pośrednich w niniejszej pracy przyniosło większy zysk o charakterze poznawczym, niż mogłoby się wydawać na podstawie wyników w tab. 2. Tym zyskiem jest pokazanie większej precyzji szacunku w przypadku estymatorów regresyjnych i empiryczno-bayesowskich, zob. tab. 3.

Na podstawie przeprowadzonych badań empirycznych (m.in. zamieszczonych w tab. 2 i 3) stwierdzamy wiele istotnych dla dalszych badań prawidłowości.

BAEL dostarcza w pełni wiarygodnych informacji do oszacowania struktury rynku pracy już na poziomie relatywnie małych jednostek przestrzennych, wielkości dużych powiatów, liczących powyżej 100 tysięcy mieszkańców.

W grupie dużych powiatów estymatory bezpośrednie charakteryzują znacznymi standardowymi błędami szacunku, które są odwrotnie proporcjonalne do wielkości

próby. Błędy średniokwadratowe (MSE) w stosunku do estymatora *HT* dla liczby bezrobotnych w takich powiatach, jak gnieźnieński, jarosławski i jasielski⁸, znacznie przekraczały 20%.

Tabela 3. Błędy średniokwadratowe (MSE) w % do danego estymatora

Powiat	Bezrobotni			Pracujący			Zawodowo bierni		
	<i>HT</i>	Reg1	<i>EB1</i>	<i>HT</i>	Reg1	<i>EB1</i>	<i>HT</i>	Reg1	<i>EB1</i>
Powiaty o najniższej stopie bezrobocia									
Piaseczyński	10,2	6,6	10,1	8,15	1,63	7,5	13,51	1,5	10,23
Warszawski zachodni	7,3	9,7	7,3	7,04	2,58	6,8	8,10	2,6	7,66
Zamojski	5,3	6,0	5,3	3,77	2,32	3,7	8,08	2,9	7,71
Nowotarski	6,6	4,7	6,5	7,22	1,28	6,4	6,74	1,4	6,07
Nowotarski	6,6	4,7	6,5	7,22	1,28	6,4	6,74	1,4	6,07
Mielecki	5,5	3,5	5,5	5,80	1,83	5,6	5,87	1,7	5,55
Tarnowski	4,6	4,7	4,5	3,94	1,22	3,8	5,87	1,6	5,45
Bielski	18,5	13,7	18,1	3,73	1,56	3,7	4,93	1,9	4,82
Powiaty o najwyższej stopie bezrobocia									
Starogardzki	4,8	3,1	4,7	5,80	2,04	5,6	5,42	1,8	5,18
Nyski	3,9	2,8	3,8	4,94	1,60	4,7	4,10	1,6	3,98
Dzierżoniowski	10,8	2,7	10,1	6,43	2,35	6,2	4,70	2,4	4,59
M. Radom	3,9	4,0	3,7	5,71	1,04	5,1	3,76	1,1	3,54
Radomski	5,2	4,5	5,0	6,49	1,67	6,1	5,90	2,0	5,61
Ostrowiecki	5,0	2,7	4,9	6,43	2,09	6,2	5,48	2,1	5,29
M. Włocławek	4,4	2,4	4,4	5,68	1,98	5,5	4,95	1,7	4,77
Minimum	2,5	2,4	2,5	2,8	0,9	2,2	2,9	1,1	2,3
Maksimum	31,1	13,7	25,2	12,3	2,6	11,1	13,5	2,9	10,2
Średnia	10,1	4,3	8,9	5,8	1,8	5,4	5,5	1,8	5,1

Źródło i uwagi jak w tab. 1.

1. Wybór najlepszego modelu regresji nie zawsze zapewnia nam najlepsze uporządkowanie małych obszarów zgodnie z przyjętym benchmarkiem. Konieczne jest poszukiwanie dalszych kryteriów oceny dobroci estymacji pośredniej, takich jak analiza rozkładów według wieku i płci.

2. Pogłębienie analizy przez uwzględnienie płci, wieku i typu miejsca zamieszkania pokaże większe znaczenie estymacji pośredniej nad bezpośrednimi.

⁸ Powiaty gnieźnieński, jarosławski i jasielski zostały pominięte w tab. 2 i 3, gdyż nie znalazły się ani wśród tych o najwyższym, ani o najniższym bezrobociu.

Literatura

- Gołata E., *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*, Wydawnictwo AE, Poznań 2004.
- Kordos J., Paradysz J., *Some experiments in small area estimation in Poland*, International Association of Survey Statisticians. Satellite Conference, Small Area Estimation, Conference Proceedings, Riga 1999.
- Paradysz J., *Kryteria dobroci estymacji dla małych obszarów*, [w:] *Statystyka społeczna – dokonania, szanse, perspektywy*, red. K. Jakóbk, Biblioteka Wiadomości Statystycznych, tom 57, Główny Urząd Statystyczny, Warszawa 2008.
- Paradysz J., Paradysz K., *Benchmarking w statystyce małych obszarów*, [w:] *Taksonomia 18, Klasyfikacja i analiza danych – teoria i zastosowanie*, red. K. Jajuga, M. Walesiak, Wydawnictwo UE, Wrocław 2011.
- Rao J.N.K., *Small Area Estimation*, Wiley – Interscience, John Wiley and Sons, INC., Hoboken, New Jersey 2003.
- Zasępa R., *Metoda reprezentacyjna*, Państwowe Wydawnictwo Ekonomiczne, Warszawa 1972.

AREAS OF UNEMPLOYMENT IN POLAND – BENCHMARK PROBLEM

Summary: In small area statistics there are many indirect estimators, for which there is a problem of selection criteria and evaluation of the estimation results. One of more promising solutions is created by local labour market statistics. The Polish Labour Force Survey (BAEL) is a systematic, repeatable and includes all households. It can be integrated with population statistics and administrative records, so we have a relatively solid base of auxiliary variables. Estimating all the elements of the labour market and summing the corresponding realizations of the estimators we compared them to population and assessed their accuracy. As a result of this comparison EB estimator was a little better. Indirect estimators were also characterized by a smaller mean square errors.

Keywords: small area estimation, local labour market, statistical benchmarking.