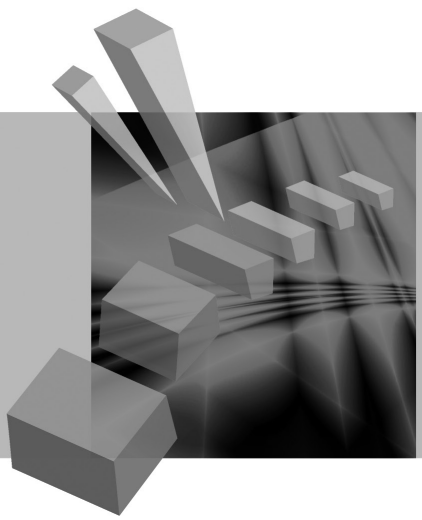
PRACE NAUKOWE
Uniwersytetu Ekonomicznego we Wrocławiu
RESEARCH PAPERS
of Wrocław University of Economics

279

Taksonomia 21

Klasyfikacja i analiza danych – teoria i zastosowania



Redaktorzy naukowi
Krzysztof Jajuga
Marek Walesiak

Wydawnictwo Uniwersytetu Ekonomicznego we Wrocławiu
Wrocław 2013

Redaktor Wydawnictwa: Aleksandra Śliwka
Redaktor techniczny: Barbara Łopusiewicz
Korektor: Barbara Cibis
Łamanie: Małgorzata Czupryńska
Projekt okładki: Beata Dębska

Publikacja jest dostępna w Internecie na stronach:
www.ibuk.pl, www.ebscohost.com,
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

Informacje o naborze artykułów i zasadach recenzowania znajdują się
na stronie internetowej Wydawnictwa
www.wydawnictwo.ue.wroc.pl

Tytuł dofinansowany ze środków Narodowego Banku Polskiego
oraz ze środków Sekcji Klasyfikacji i Analizy danych PTS

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

Wersja pierwotna: publikacja drukowana
Druk: Drukarnia TOTEM

Spis treści

Summaries

PRACE NAUKOWE UNIWERSYTETU EKONOMICZNEGO WE WROCŁAWIU
RESEARCH PAPERS OF WROCŁAW UNIVERSITY OF ECONOMICS nr 279 • 2013

Taksonomia 21. Klasyfikacja i analiza danych – teoria i zastosowania ISSN 1899-3192

Karolina Paradysz
Centrum Statystyki Regionalnej w Poznaniu

BENCHMARKOWA ANALIZA ESTYMACJI DLA MAŁYCH OBSZARÓW NA LOKALNYCH RYNKACH PRACY

Streszczenie: Statystyka małych obszarów (SMO) znajduje zastosowanie w warunkach niedostatecznej liczebności próby. Na początku XXI wieku w Głównym Urzędzie Statystycznym postanowiono sprawdzić, czy metodologia SMO mogłaby być wykorzystana w Polsce. Zadanie to powierzono zespołowi pod kierunkiem C. Brachy. Na podstawie BAEL w latach 1995-2002 dokonano estymacji z Badania Aktywności Ekonomicznej Ludności na poziomie powiatów dla lat 1995-2002, wykorzystując dodatkowo wyniki NSP 2002. Natomiast na podstawie danych BAEL z 2003 r. zweryfikowano możliwość wykorzystania złożonych metod estymacji do dezagregacji danych na poziomie powiatów. W obu tych opracowaniach dokonano formalnej oceny jakości szacunków, wykorzystując w tym celu parametry stochastycznej struktury estymatorów klasy SMO (klasycznych, syntetycznych, złożonych). Przedmiotem niniejszego opracowania jest analiza krytycznej oceny wyników dokonanych przez zespół metodologiczny w GUS. W artykule podejmiemy próbę dalszej weryfikacji metodologii SMO z punktu widzenia kryteriów zaproponowanych przez J. Paradysza [2008].

Słowa kluczowe: statystyka małych obszarów, benchmarking, rynek pracy, Badanie Aktywności Ekonomicznej Ludności.

1. Wstęp

Estymacja dla małych obszarów jest działem statystyki, który zajmuje się metodami wykorzystywania informacji statystycznych uzyskanych dla całej populacji do wnioskowania o badanych cechach w wyróżnionych podpopulacjach (podpopulacje te noszą nazwę właśnie małych obszarów, dziedzin lub domen), por. [Domański, Pruska 2001, s. 36]. Za początek studiów w zakresie statystyki małych obszarów w Polsce można przyjąć rok 1992, kiedy ówczesny wiceprezes GUS oraz przewodniczący PTS prof. Jan Kordos zorganizował międzynarodową konferencję o zasięgu światowym, por. [Kalton, Kordos, Platek 2003][1]. Jednakże prowadzone od 20 lat

[1] Równie duże znaczenie dla rozwoju estymacji dla małych obszarów miała konferencja w Rydze w 1999 r., której współorganizatorem był także J. Kordos, por. [International... 1999]. Aktywny udział polskiej reprezentacji w konferencji ryskiej miał wpływ na zaproszenie Polski do konsorcjum

badania i opracowania miały dotychczas charakter metodologiczny bądź sprawozdawczy będący weryfikacją metod proponowanych w literaturze światowej[2].

Dopiero jednak próba podjęta przez zespół pod kierunkiem Brachy w GUS nosi znamiona kompleksowej weryfikacji statystyki małych obszarów dla praktyki GUS. Zespół ten opublikował 2 prace (por. [Bracha, Lednicki, Wieczorkowski 2004; Bracha 2003]) dotyczące rynku pracy na podstawie Badania Aktywności Ekonomicznej Ludności (BAEL). Wyjątkowość tych prac polega na zastosowaniu metod estymacji pośredniej w BAEL-u przez zespół, który dokonywał losowania próby i najlepiej znał związaną z tym pragmatykę[3]. Miał dostęp do danych jednostkowych na wszystkich poziomach podziału administracyjnego kraju oraz do danych wspomagających. Niestety, żadne z tych dwóch opracowań zespołu Brachy nie zostało poddane wnikliwej analizie, na jaką zasługuje. O ile nam wiadomo, nie ukazała się żadna recenzja ani – co bardziej istotne – analiza jakości wyników estymacji[4].

Ze względu na obszerność opracowań związanych z oceną wykonanych prac w GUS ograniczamy się tylko do późniejszego opracowania, które dotyczy tylko 2003 r. Bracha i in. [2004] wykorzystali 5 estymatorów klasy SMO dla pracujących (aktywnych zawodowo), biernych i bezrobotnych. Celem pracy jest próba oceny jakości wyników (uzyskanych w toku estymacji wielkości charakterystycznych dla rynku pracy) z punktu widzenia kryteriów zaproponowanych niegdyś przez J. Paradysza [2008]. Przedstawimy rozmiary różnic, jakie występują na różnych poziomach agregacji. Spróbujemy odpowiedzieć na pytanie, czy oceniane estymatory dają wartości absurdalne[5]. W ocenianej pracy po wyznaczeniu wartości estymatorów bayesowskich dla powiatów ogółem dokonano dodatkowej korekty estymatorów podzielonych na składniki: miasto/wieś oraz płeć w celu spełnienia warunku sumowania się szacunków z poziomów bardziej szczegółowych na poziom wyższy. Zastosowano metodę opisaną w pracy [You, Rao, Dick 2004], tzw. benchmarking estymatorów[6].

EURAREA, które pod kierunkiem P. Heady'ego testowało możliwości wykorzystania statystyki małych obszarów w krajach Unii Europejskiej.

[2] Ograniczając się tylko do większych opracowań monograficznych, Dehnel [2003; 2010] weryfikowała przydatność metody estymacji pośredniej w statystyce gospodarczej, Gołata [2004] oceniała siedem estymatorów przyjętych jako standard w projekcie EURAREA dla estymacji na lokalnym rynku pracy, natomiast T. Żądło [2008] zajął się numerycznymi aplikacjami statystyki małych obszarów w języku R.

[3] Opisy metodologii badań w Głównym Urzędzie Statystycznym zamieszczane w publikacjach wynikowych są zbyt ogólne jak na potrzeby estymacji pośredniej.

[4] Jak się zdaje, niniejsza próba oceny jest pierwszą i zaledwie dotyka problemu benchmarkingu w estymacji dla małych obszarów.

[5] Przed tym autorzy opracowania pod kierunkiem Brachy zabezpieczyli się, wprowadzając pojęcie „minimalnej wielkości obserwacji" 50 jednostek losowania pierwszego stopnia. Nowoczesne metody z zakresu statystyki małych obszarów pozwalają jednak na szacowanie w warunkach mniejszych prób.

[6] Kwestia benchmarkingu została poruszona w referacie umieszczonym w [Paradysz, Paradysz 2011]. W literaturze światowej pionierami w tej dziedzinie są Rao, Ghosh, Pfefferman czy Datt, por. *Bayesian Benchmarking with Applications to Small Area Estimation,* G.S. Datta, M. Ghosh, R. Steorts and J. Maples, University of Georgia, University of Florida and US Bureau of the Census.

2. Założenia kryterium poziomu [Paradysz 2008]

Wychodząc od postulatów oceny jakości estymacji dla małych obszarów sformułowanych przez J. Paradysza [2008], spróbujemy ocenić wyniki Cz. Brachy i in. [2004] ze względu na kryterium poziomu. Według Paradysz [2008], „kryterium poziomu oznacza, że suma wartości dla wszystkich małych obszarów istotnie nie odbiega od prawdziwej nieznanej wartości parametru dla dużego obszaru". W klasycznej estymacji bezpośredniej ten postulat nazywa się nieobciążonością. Ocenę jakości dobroci estymacji dla małych obszarów w odniesieniu do bezrobocia przeprowadzamy na trzech poziomach podziału administracyjnego: NUTS 2 – województwa, NUTS 3 – podregiony, NUTS 4 – powiaty.

Postulat, który w pracy J. Paradysz został nazwany nieobciążonością, wyraża się wzorem:

$$\sum_{a} \hat{y}_a^p = \hat{Y}, \tag{1}$$

gdzie: $\hat{y}_a^p$ to wartość globalna cechy y w małym obszarze a na poziomie p; p to poziom danego podziału administracyjnego kraju, na przykład w przypadku klasyfikacji NUTS: $p = 1$ dla makroregionów, $p = 2$ dla województw, $p = 3$ – podregiony, $p = 4$ – powiaty.

Kryterium poziomu odnosimy do kategorii „bezrobotni". W odniesieniu do tej kategorii występującej w BAEL w pracy Brachy i in. [2004] została oszacowana liczba bezrobotnych z podziałem na płeć. We wspomnianym opracowaniu zastosowano estymatory:

– klasyczny na 3 poziomach NUTS – całego kraju, województw i podregionów;
– syntetyczny na 3 poziomach NUTS – województw, podregionów i powiatów;
– złożony na poziomie województw i podregionów.

3. Charakterystyki estymatorów oraz metoda oceny ich jakości, prezentacja wyników

Dostępność danych pozwala na analizę na poziomie NUTS 4 (czyli biorąc pod uwagę wszystkie powiaty i miasta na prawach powiatu) tylko 3 estymatorów – HB, EB i syntetycznego. Wartości estymatorów klasycznego i złożonego zostały wyznaczone dla powiatów i miast na prawach powiatu, w których wylosowano powyżej 50 jednostek. Każdy z nich jest oparty na innych założeniach i posiada swoje charakterystyczne własności. Estymator syntetyczny zakłada udział pracujących w poszczególnych województwa w 2003 r. taki sam jak w NSP 2002, co można przedstawić następującymi wzorami ze względu na poziom agregacji:

– dla województw:

$$x_w = t \cdot f_w, \tag{2}$$

gdzie f_w jest udziałem wartości danej zmiennej (z NSP 2002) w w-tym województwie w stosunku do całego kraju, natomiast t jest estymatorem klasycznym dla Polski;

– dla podregionów:

$$x_{ws} = t_w \cdot f_{ws}, \tag{3}$$

gdzie f_{ws} jest udziałem wartości danej zmiennej (z NSP 2002) w s-tym podregionie w stosunku do w-tego województwa, natomiast t_w jest estymatorem klasycznym dla w-tego województwa [Bracha, Lednicki, Wieczorkowski 2004, s. 19];

– dla powiatu:

$$x_{wp} = t_w \cdot f_{wp}, \tag{4}$$

gdzie f_{wp} jest udziałem wartości danej zmiennej (z NSP 2002) w p-tym powiecie w stosunku do w-tego województwa.

Estymator złożony zaś jest liniową wypukłą kombinacją estymatorów klasycznego i syntetycznego. Dla podregionu określony on jest wzorem, por. [Bracha, Lednicki, Wieczorkowski 2004, s. 19]:

$$y_{ws} = v_{ws} t_{ws} + (1 - v_{ws}) x_{ws}, \tag{5}$$

gdzie v_{ws} jest wagą przypisaną estymatorowi klasycznemu. Estymator złożony dla województw i powiatów określony jest analogicznym wzorem.

W celu wykrycia różnic i określenia ich względnej siły i kierunku stosuje się wzory:

$$wsk_{IA} = \frac{\left(\sum_{i=1}^{t} R_{synt}{}^{k} - W_{synt}{}^{k} \right)}{\sum_{i=1}^{t} R_{synt}{}^{k}} \cdot 100, \tag{6}$$

$$wsk_{IB} = \frac{\left(\sum_{i=1}^{t} R_{złoż}{}^{k} - W_{złoż}{}^{k} \right)}{\sum_{i=1}^{t} R_{złoż}{}^{k}} \cdot 100, \tag{7}$$

gdzie odjemna występująca w liczniku to suma ocen estymatora syntetycznego bądź złożonego odnosząca się do k-tego podregionu, odjemnik zaś to wyestymowana ocena estymatora syntetycznego bądź złożonego dla k-tego podregionu. Indeksy dolne we wzorach oznaczają estymator (synt – syntetyczny, złoż – złożony). Suma od i do t oznacza sumę wartości estymatorów regionów R wchodzących w skład województwa W.

Po dokonaniu obliczeń wyniki zamieszczamy w tab. 1.

Tabela 1. Różnice względne między wartościami estymatora syntetycznego bezrobotnych we wszystkich województwach – porównanie sumy estymatorów na poziomie podregionów z estymatorem na poziomie województwa

Województwo	Syntetyczny			Złożony		
	ogółem	mężczyźni	kobiety	ogółem	mężczyźni	kobiety
Dolnośląskie	–7,84	–7,75	–7,94	–3,77	–3,73	–3,82
Kujawsko-pomorskie	3,58	–1,73	9,04	1,82	–0,86	4,73
Lubelskie	2,28	–0,72	5,86	1,15	–0,36	3,02
Lubuskie	4,62	5,09	4,06	2,37	2,61	2,07
Łódzkie	–1,09	3,52	–6,32	–0,54	1,79	–3,06
Małopolskie	–9,16	–13,37	–4,61	–4,38	–6,26	–2,25
Mazowieckie	–0,32	0,58	–1,38	–0,16	0,29	–0,69
Opolskie	0,00	0,00	0,00	0,00	0,00	0,00
Podkarpackie	1,08	0,25	1,99	0,54	0,13	1,01
Podlaskie	–5,31	–7,24	–3,19	–2,59	–3,49	–1,57
Pomorskie	2,33	0,47	4,29	1,18	0,24	2,19
Śląskie	–3,50	–2,32	–4,66	–1,72	–1,15	–2,27
Świętokrzyskie	0,00	0,00	0,00	0,00	0,00	0,00
Warmińsko-mazurskie	8,02	9,81	6,11	4,17	5,15	3,15
Wielkopolskie	2,43	4,78	0,04	1,23	2,45	0,02
Zachodniopomorskie	2,17	6,19	–2,39	1,10	3,20	–1,18

Źródło: opracowanie własne na podstawie [Bracha, Lednicki, Wieczorkowski 2004, Aneks].

Generalnie prawie w każdym z województw różnice względne (biorąc pod uwagę ich wartości bezwzględne) są niższe w przypadku estymatora złożonego. Województwo opolskie i świętokrzyskie są wyjątkowe pod tym względem, gdyż stanowią one zarazem podregion. W przypadku estymatora syntetycznego i złożonego wskaźnik obliczony według wzoru (6) charakteryzuje się stosunkowo niskimi wartościami (bezwzględnymi) w województwie mazowieckim (ze względu na kategorie „ogółem" i „mężczyźni") oraz wielkopolskim („kobiety"). Niska wartość bezwzględna tego wskaźnika świadczy o małych różnicach między wartościami estymatora syntetycznego w przypadku sumowania z poziomu podregionu do poziomu województwa.

Podobnej analizy można dokonać, wykorzystując oceny estymatora syntetycznego oszacowanego dla bezrobotnych w opracowaniu Cz. Brachy dla poszczególnych powiatów. Wykorzystując również z tegoż opracowania oceny estymatora syntetycznego w kategorii aktywnych zawodowo dla poszczególnych powiatów, można sprawdzić, czy suma ocen poszczególnych estymatorów syntetycznych w powiatach równa jest ocenie estymatora syntetycznego (dla tejże kategorii) w danym województwie.

Poniżej znajduje się tabela wynikowa (tab. 2) i wzór, na podstawie którego dokonano obliczeń.

$$wsk_{II} = \frac{\left(\sum_{i=1}^{t} P_{synt}{}^{k} - W_{synt}{}^{k} \right)}{\sum_{i=1}^{t} P_{synt}{}^{k}} \times 100, \tag{8}$$

gdzie: t – oznacza liczbę powiatów,

k – oznacza numer województwa,

W_{synt} – oznacza ocenę parametru estymatora w danym podregionie t,

$\sum_{i=1}^{t} P_{synt}{}^{k}$ – oznacza sumę wartości estymatora syntetycznego w tych powiatach (od i do t), które wchodzą w skład województwa W.

Tabela 2. Różnice względne między wartościami estymatora syntetycznego bezrobotnych we wszystkich województwach – porównanie sumy estymatorów na poziomie powiatów z estymatorem na poziomie województwa

Województwo	Estymator syntetyczny		
	ogółem	mężczyźni	kobiety
Dolnośląskie	7,27	7,20	7,35
Kujawsko-pomorskie	–3,72	1,70	–9,93
Lubelskie	–2,33	0,72	–6,22
Lubuskie	–4,85	–5,37	–4,23
Łódzkie	1,08	–3,65	5,94
Małopolskie	8,39	11,79	4,40
Mazowieckie	0,32	–0,58	1,37
Opolskie	–3,37	–1,40	–5,55
Podkarpackie	–1,09	–0,25	–2,04
Podlaskie	5,04	6,76	3,09
Pomorskie	–2,39	–0,48	–4,48
Śląskie	3,38	2,27	4,45
Świętokrzyskie	–18,60	–16,25	–21,47
Warmińsko-mazurskie	–8,72	–10,87	–6,50
Wielkopolskie	–2,49	–5,02	–0,04
Zachodniopomorskie	–2,22	–6,60	2,33

Źródło: opracowanie własne na podstawie [Bracha, Lednicki, Wieczorkowski 2004, Aneks].

Największe różnice względne odnotowano w przypadku (zarówno w kategorii „ogółem", jak i w rozbiciu ze względu na płeć) dla województwa świętokrzyskiego, najniższe zaś (biorąc pod uwagę wartości bezwzględne tych różnic) w Mazowieckiem. Stosunkowo niskimi różnicami względnymi charakteryzuje się również Podkarpacie.

BAEL nie jest jedynym źródłem danych o aktywności ekonomicznej ludności. Jednym ze źródeł, które zawiera dane na podobnych szczeblach agregacji, jest Bank Danych Lokalnych, który jest prowadzony i rozwijany przez Główny Urząd Statystyczny. BDL to największy w Polsce uporządkowany i udostępniany w Internecie zbiór informacji o sytuacji społeczno-gospodarczej, demograficznej, społecznej oraz stanie środowiska, opisujący województwa, powiaty oraz gminy jako podmioty systemu organizacji społecznej i administracyjnej państwa, a także regiony i podregiony stanowiące elementy nomenklatury jednostek terytorialnych do celów statystycznych.

W tej części opracowania dokonujemy oceny pod względem kryterium poziomu, porównując dwa źródła danych, biorąc pod uwagę estymatory EB, HB i syntetyczny dotyczący liczby bezrobotnych pochodzące z BAEL oraz liczby bezrobotnych rejestrowanych bezrobotnych z BDL. Wyjaśnienia wymagają charakterystyki estymatorów EB oraz HB.

Empirycznym estymatorem bayesowskim (w skrócie EB) wartości globalnej (liczby elementów z cechą wyróżnioną) dla *p*-tego powiatu będziemy nazywać statystykę (por. [Bracha, Lednicki, Wieczorkowski 2004, s. 33]):

$$y_p^{EB} = \alpha_p \hat{\theta}_p + (1-\alpha_p)\tilde{\theta}, \tag{9}$$

gdzie: α_p – to pewna stała spełniająca warunek $0 \le \alpha_p \le 1$, $\hat{\theta}_p$ to estymator rozpatrywanego parametru (θ_p) stosowany w badaniu reprezentacyjnym (w rozpatrywanym przypadku BAEL);

$\tilde{\theta}_p = x_p^T \hat{b}$ – to predyktor badanego parametru dla *p*-tego powiatu skonstruowany na podstawie danych z rejestrów pracy.

Z kolei hierarchiczne estymatory bayesowskie wymagają znajomości rozkładów **a priori** $f(\lambda)$ parametrów rozważanego modelu oraz rozkładów warunkowych $f(\mu, y|\lambda)$ parametrów małych obszarów μ (np. liczby bezrobotnych). Z reguły bezpośrednie obliczenia dla modeli stosowanych w praktyce wymagają stosowania skomplikowanych metod numerycznych, wykorzystuje się różne techniki symulacyjne, np. metodę Monte Carlo i łańcuchy Markowa [Bracha, Lednicki, Wieczorkowski 2004 , s. 45].

Porównania obliczone zostaną na podstawie współczynnika odchyleń danego estymatora od liczby bezrobotnych rejestrowanych. Obliczeń tych dokonano na podstawie wzoru:

$$Wsp = \frac{\sqrt{\frac{\sum (BDL - BAEL)^2}{n}}}{\frac{\sum BDL}{n}} \cdot 100. \qquad (10)$$

BDL dotyczy wielkości pobranych z Banku Danych Lokalnych, zaś BAEL dotyczy poszczególnych wartości estymatorów z Badania Aktywności Ekonomicznej Ludności.

W tabeli 3 przedstawiono wyniki obliczeń z wykorzystaniem powyższego wzoru.

Tabela 3. Współczynnik odchyleń wartości estymatorów z BAEL od liczby bezrobotnych rejestrowanych

Województwo	EB	HB	SYNT
Polska	37,17	37,46	40,95
Dolnośląskie	50,87	47,05	34,72
Kujawsko-pomorskie	28,82	27,11	22,35
Lubelskie	44,76	54,45	32,11
Lubuskie	24,32	36,77	12,53
Łódzkie	10,44	16,48	10,89
Małopolskie	47,88	42,35	72,99
Mazowieckie	53,84	53,39	75,15
Opolskie	21,30	22,41	16,10
Podkarpackie	27,83	33,72	22,87
Podlaskie	71,14	63,53	57,67
Pomorskie	22,24	30,78	28,59
Śląskie	25,64	21,10	30,08
Świętokrzyskie	42,30	42,67	32,70
Warmińsko-mazurskie	24,10	25,15	19,93
Wielkopolskie	30,80	33,73	34,23
Zachodniopomorskie	20,28	18,92	13,97

Źródło: opracowanie własne na podstawie [Bracha, Lednicki, Wieczorkowski 2004, Aneks].

Ze względu na wszystkie trzy estymatory najlepszym wynikiem ze względu na niską wartość parametru obliczonego według wzoru (10) charakteryzuje się województwo łódzkie (EB 10,44; HB 16,48; SYNT 10,89). Największe zaś wartości (co jest niekorzystnym zjawiskiem) odnotowuje się w zależności od estymatora w województwie: podlaskim 71,14 w przypadku EB; podlaskim 63,53 w przypadku HB; mazowieckim 75,15 w przypadku SYNT.

Najlepszym estymatorem ze względu na wartość dla Polski ogółem jest estymator EB, gdyż cechuje się najniższym współczynnikiem odchyleń, jednakże w przy-

padku podziału na województwa jest on najlepszy jedynie w 3 przypadkach (województwa wielkopolskie, pomorskie, łódzkie).

4. Wnioski

1. Podjęta próba oceny wyników pracy zespołu Cz. Brachy jest pierwszym i zbyt ogólnym spojrzeniem na stan tej pracy, aby można było autorytatywnie stwierdzić, że jej jakość jest dobra.

2. Oceniliśmy jedynie pod względem formalnym – poziom kryterium – estymatory liczby bezrobotnych bez uwzględnienia analizy merytorycznej (kryterium trzecie w klasyfikacji J. Paradysza).

3. Osobny problem stanowi porównanie stóp bezrobocia z BAEL z bezrobociem rejestrowanym, który tutaj został pominięty, a zostanie uwzględniony w dalszych badaniach.

4. Na podstawie przeprowadzonej analizy stwierdzono, że:

a) żaden z estymatorów nie daje wartości absurdalnych,

b) na poziomie województw lepszym estymatorem był estymator złożony niż syntetyczny,

c) w przypadku porównań dwóch źródeł danych odnośnie do liczby bezrobocia spośród 3 estymatorów (HB, EB, syntetyczny) najlepszy okazał się EB.

Literatura

Bracha Cz. (2003), *Estymacja danych z badania aktywności ekonomicznej ludności na poziomie powiatów dla lat 1995-2002*, GUS, Warszawa.

Bracha Cz., Lednicki B., Wieczorkowski R. (2004), *Wykorzystanie złożonych metod estymacji do dezagregacji danych z badania aktywności ekonomicznej ludności w roku 2003,* Z Prac Zakładu Badań Statystyczno-Ekonomicznych, zeszyt 300.

Domński Cz., Pruska K. (2001), *Metody statystyki małych obszarów*, Wyd. UŁ, Łódź 2001.

Gołata E. (2004), *Estymacja pośrednia aktywności ekonomicznej na potrzeby spisu opartego na rejestrach* , Pomiar Informacji w Gospodarce, Zeszyty Naukowe 149, Wyd. UE Poznań.

International Association of Survey Statisticians. Satellite Conference (1999), *Small Area Estimation – Conference Proceedings*, Riga, Latvia, August 1999.

Kalton G., Kordos J., Platek R. (1993), *Small Area Statistics and Survey Designs*, Vol. I: Invited Papers; Vol. II: Contributed Papers and Panel Discussion, Central Statistical Office, Warsaw.

Kostrzewa Z., Nowak L., Szałas D., Witkowski J., *Kierunki doskonalenia statystyki migracji zagranicznych ludności*, Wiadomości Statystyczne nr 5, maj 2010, Wyd. GUS.

Paradysz J. (2008), *Kryteria dobroci estymacji dla małych obszarów,* [w:] *Statystyka społeczna – dokonania, szanse, perspektywy*, red. K. Jakóbik, Biblioteka Wiadomości Statystycznych, tom 57, Główny Urząd Statystyczny, Warszawa 2008, s. 74-84.

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.

Śleszyński P. (2010), *Struktura przestrzeni i delimitacja obszarów społecznych w Warszawie*, Instytut Geografii i Przestrzennego Zagospodarowania PAN.

You Y., Rao J.N.K., P. Dick (2004), *Benchmarking Hierarchical Bayes Small Area Estimators in the Canadian Census Undercoverage Estimation*, Statistics in Transition 6(5), 631-640.

Żądło T. (2008), *Elementy statystyki małych obszarów z programem R*, Wydawnictwo AE Katowice, Katowice.

BENCHMARK ANALYSIS OF SMALL AREA ESTIMATION ON LOCAL LABOR MARKETS

Summary: Small area estimation is used in conditions when the sample size is too small to use a direct estimator. In the case of labor market research the most desirable information is believed to be this provided by poviat (NUTS 4). At the beginning of the twenty-first century the Central Statistical Office decided to use the SAE methodology in Poland. The result was published by Bracha et al. [2003; 2004]. On the base of Labour Force Survey (LFS, in Polish BAEL) the Bracha team with the SAE estimators (EB, HB, synthetic or composite) estimated the employed, the unemployed and the economically inactive taking poviats into consideration. We try to evaluate the Bracha team results from the point of view of the criteria proposed by J. Paradysz [2008].

Keywords: small area statistics, benchmarking, labor market, Labour Force Survey.