

Jan Paradysz, Karolina Paradysz

Uniwersytet Ekonomiczny w Poznaniu

BENCHMARKING W STATYSTYCE MAŁYCH OBSZARÓW

Streszczenie: Badania reprezentacyjne dostarczają wiarygodnych szacunków nie tylko dla całej zbiorowości, lecz także dla wielu subpopulacji, np. regionów czy powiatów. Wielkość prób w małych domenach, szczególnie w małych przestrzennie obszarach, rzadko pozwala na szacunki bezpośrednie. Zatem estymacja pośrednia odgrywa istotną rolę w zaspokajaniu stale rosnących potrzeb na wiarygodną statystykę, nawet gdy dostępne są tylko bardzo małe próby. Do dokonania estymacji dla małych obszarów niezbędne jest „pożyczanie mocy” celem zwiększenia precyzji szacunku. W zależności od modelu estymacji oraz zmiennych wspomagających dla każdego małego obszaru uzyskujemy wiele szacunków. W związku z tym powstają problemy z wyborem tego właściwego. Benchmarking pozwala na właściwy wybór szacunku. Zaproponowano trzy rodzaje kryteriów: poziomu, porządku i dystansu.

Słowa kluczowe: Estymacja dla małych obszarów, benchmarking, empiryczna estymacja bayesowska, hierarchiczna estymacja bayesowska, raking.

Wielu autorów zajmujących się badaniami reprezentacyjnymi, a zwłaszcza estymacją na podstawie relatywnie małych liczebnie prób, podkreśla dużą użyteczność praktyczną metodologii wykorzystujących informację spoza danego konkretnego badania. Coraz częściej dostrzega się potrzebę pełnego wykorzystania źródeł informacji z wcześniejszych badań podobnego rodzaju oraz wszelkiej innej wiedzy dotyczącej przedmiotu estymacji. Tym przedmiotem estymacji najczęściej bywają małe obszary, czyli jednostki przestrzenne, którym tradycyjna metoda reprezentacyjna nie mogła niczego zaoferować ze względu na zbyt małe liczebności prób¹. Pod wpływem dużego zapotrzebowania na informację rozwijają się nowe techniki estymacji, pozwalające pokonać bariery związane z bezwzględnie małymi próbami. Statystykę małych obszarów można by więc nazwać nową filozofią badań sta-

¹ W estymacji dla małych obszarów dopuszcza się również próby zerowe (por. np. [Pfeffermann, Sverchkov 2007, s. 2; D’Alò i in. 2006]). W klasycznej metodzie reprezentacyjnej unika się takich przypadków, a badacze reprezentujący ten kierunek badań wymagają pewnej minimalnej wielkości próby – także w estymacji dla małych obszarów (por. [Bracha i in. 2003, s. 8 i 27]).

tystycznych, gdyż rozluźniając rygory stawiane przez klasyczną metodę reprezentacyjną, bardzo mocno akcentuje problem pełnego wykorzystania wszystkich możliwych informacji. Można powiedzieć, że statystyka małych obszarów podejmuje wyzwania badawcze w tych warunkach, gdzie tradycyjne metody estymacji okazały się niewydolne lub nieekonomiczne.

Jednakże, jak napisał D. Pfeffermann [2002], przy estymacji dla małych obszarów mamy do czynienia z dwoma rodzajami problemów. Po pierwsze, jak oszacować wiarygodne charakterystyki dla małych obszarów lub domen w oparciu o bardzo małe próby². Po drugie, jak oszacować standardowe błędy szacunku dla tych charakterystyk. Tej filozofii badań wyszło naprzeciw jedno z największych przedsięwzięć w estymacji dla małych obszarów, jakim był projekt EURAREA³. W tym projekcie międzynarodowym realizowanym na użytek Eurostatu, w którym uczestniczyli także pracownicy Katedry Statystyki Akademii Ekonomicznej w Poznaniu, po wyselekcjonowaniu siedmiu estymatorów z zakresu SMO podjęto próbę ich wykorzystania w badaniu rynku pracy i warunków bytu na podstawie budżetów gospodarstw domowych w 5 krajach Unii Europejskiej (Wielka Brytania – koordynator badania, Finlandia, Hiszpania, Szwecja, Włochy) i Polski. Punktem wyjścia były narodowe spisy powszechne, przeprowadzone w poszczególnych krajach na przełomie lat osiemdziesiątych i dziewięćdziesiątych XX w. Jedyny wyjątek stanowiła Polska, w której NSP 1988 był zbyt odległy od momentu pierwszych badań siły roboczej (1992 r.) i dotyczył warunków życia jeszcze w systemie gospodarki socjalistycznej. Wobec tego dla Polski podstawą do konstrukcji bazy danych łączącej badania reprezentacyjne ze spisem był rok 1995, w którym przeprowadzono mikrospis. Na podstawie mikrospisu, badania budżetów gospodarstw domowych oraz Banku Danych Regionalnych utworzono coś w rodzaju populacji generalnej o nazwie Poldata, z której losowano próbki zgodnie z metodą Monte Carlo (por. [Gołata 2004, s. 302-303]). Wiarygodność wyników estymacji według poszczególnych estymatorów była oceniana empirycznie, podobnie jak jakość ich precyzji. Można powiedzieć, że każdy mały obszar był oceniany w zasadzie oddzielnie, bez uwzględnia ich wzajemnych relacji oraz zmian w czasie⁴.

Dopiero w 2008 roku, niezależnie od prac Ghosha, Datty i Pfeffermanna, sformułowaliśmy kryteria oceny jakości szacunków dla małych obszarów, uwzględniające relacje między nimi (por. [Paradysz 2008, 2008a]). Jak się nam wydaje, jest to właściwy krok w kierunku konstrukcji estymatorów benchmarkowych. Wyróżniliśmy

² W późniejszych pracach D. Pfeffermann koncentruje się bardziej na ocenie jakości uzyskanych wyników i wiele miejsca poświęca benchmarkingowi w statystyce małych obszarów (por. [Pfeffermann, Tiller 2003; 2006; Pfeffermann i in. 2005]).

³ Odnośnie do celów i wyników projektu EURAREA patrz: [Heady, Ralphs 2005; Gołata 2004; Kordos 2004].

⁴ W ówczesnych warunkach polskich postulat dynamicznej oceny procesu estymacji był trudny do spełnienia, gdyż dysponowaliśmy jedynie danymi z jednego spisu ludności.

3 grupy kryteriów oceny dobroci estymacji dla małych obszarów (domen): 1) formalne wewnętrzne, 2) formalne zewnętrzne, 3) merytorycznej ocena wyników estymacji pośredniej.

Kryterium formalnej wewnętrznej zgodności wymaga, żeby na poszczególnych stopniach podziału przestrzennego, była zachowana struktura według wartości szacowanej cechy. To znaczy, suma wartości szacowanej cechy dla małych obszarów (a), na przykład dla gmin, powinna dać analogiczne parametry dla powiatów, podregionów, województw i dla całego kraju. W przypadku braku tej zgodności formalnej w literaturze z zakresu benchmarkingu w statystyce małych obszarów stosowano wyrównywanie (raking) lub metodę korygowania wskaźnika (*ratio adjustment method*). Pierwsza z nich polegała na podsumowaniu wszystkich wartości dla małych obszarów, $\hat{Y} = \sum \hat{y}_a$, i podzieleniu tej sumy przez wartość dla dużego Y . Wynik dzielenia to I_y . Gdyby iloraz I_y był większy od jednościi, to przez niego należało podzielić szacunki pierwotne dla małych obszarów – \hat{y}_a . Jeśli natomiast $I_y < 1$, to $\hat{y}_a \cdot I_y$. Druga z tych metod odnosi się do estymatorów występujących w postaci wskaźnika natężenia bądź struktury i polega na przemnożeniu każdego \hat{y}_a przez współczynnik korygujący, adekwatnego do wskaźnika dla dużego obszaru.

Bayesowski estymator wartości globalnej ma postać daną wzorem (1) (por. [Kordos, Paradysz 2000, s. 691]):

$$\hat{y}_a(EB) = \frac{s^2(\hat{y}_a)}{s^2(\hat{y}_a) + s^2(r_a)} \hat{y}_a(r) + \frac{s^2(r_a)}{s^2(\hat{y}_a) + s^2(r_a)} \hat{y}_a. \quad (1)$$

Wariancja estymatora (1) dana jest wzorem:

$$s^2[\hat{y}_a(EB)] = \frac{s^2(\hat{y}_a) \cdot s^2(r_a)}{s^2(\hat{y}_a) + s^2(r_a)}, \quad (2)$$

- gdzie: $\hat{y}_a(EB)$ – estymator bayesowski wartości globalnej w małym obszarze a , który wykorzystuje informacje zarówno z modelu regresji, jak i z próby,
- $\hat{y}_a(r)$ – estymator regresyjny wartości globalnej w małym obszarze a , który wykorzystuje tylko informacje z modelu regresji,
- \hat{y}_a – estymator bezpośredni wartości globalnej w małym obszarze a , który wykorzystuje tylko informacje z próby,
- $s^2(\hat{y}_a)$ – wariancja estymatora bezpośredniego wartości globalnej w małym obszarze a , który wykorzystuje tylko informacje z próby,
- $s^2(r_a)$ – wariancja estymatora regresyjnego wartości globalnej w małym obszarze a , który wykorzystuje tylko informacje z modelu regresji,

$s^2[\hat{y}_a(EB)]$ – wariancja estymatora bayesowskiego wartości globalnej w małym obszarze a , który wykorzystuje informacje zarówno z modelu regresji, jak i z próby.

Tabela 1. Raking liczb bezrobotnych i pracujących w powiatach woj. wielkopolskiego na podstawie estymatora bayesowskiego, 1995 r.

Powiaty	Estymator bayesowski		Estymator bezpośredni		Raking	
	bezrobotni	pracujący	bezrobotni	pracujący	bezrobotni	pracujący
Ogółem	177 964	1 244 419	133 909	973 186	133 909	973 186
Chodzieski	3 597	20 670	3 125	15 700	2 707	16 165
Czarnkowsko-trzcianecki	7 158	42 748	6 655	34 430	5 386	33 431
Gnieźnieński	9 610	61 904	9 933	47 063	7 231	48 411
Gostyński	4 502	28 360	3 957	29 980	3 388	22 179
Grodziski	2 383	19 538	2 636	17 828	1 793	15 280
Jarociński	5 061	32 654	2 601	21 234	3 808	25 537
Kaliski	12 012	77 758	3 175	46 063	9 038	60 810
Kępiański	3 170	24 049	2 459	24 010	2 385	18 807
Kolski	7 475	41 438	5 544	39 384	5 625	32 406
Koniński	15 236	88 618	6 228	50 928	11 464	69 303
Kościański	3 164	31 038	3 034	30 158	2 381	24 273
Krotoszyński	5 183	31 052	4 160	30 226	3 900	24 284
Leszczyński	5 485	42 901	1 466	17 661	4 127	33 550
Międzychodzki	2 407	17 951	1 877	14 446	1 811	14 038
Nowotomyski	2 167	23 667	2 019	26 095	1 631	18 509
Obornicki	2 831	24 423	2 458	18 631	2 130	19 100
Ostrowski	11 643	65 806	10 615	65 061	8 761	51 463
Ostrzeszowski	4 355	36 699	2 983	22 240	3 277	28 700
Piński	8 885	54 584	7 420	38 525	6 686	42 687
Pleszewski	5 334	29 535	3 384	23 737	4 014	23 098
Poznański	12 186	163 572	5 648	89 532	9 169	127 920
Rawicki	3 204	24 249	3 464	26 584	2 411	18 964
Słupecki	5 635	27 252	3 696	20 280	4 240	21 312
Szamotulski	3 681	34 662	3 937	34 237	2 770	27 107
Średzki	2 628	20 228	4 868	39 332	1 977	15 819
Śremski	2 708	22 760	2 632	19 064	2 038	17 799
Turecki	7 117	41 136	4 740	32 916	5 355	32 170
Wągrowiecki	5 487	29 857	5 308	23 583	4 129	23 349
Wolsztyński	2 718	23 443	2 712	22 641	2 045	18 333
Wrzesiński	4 157	29 139	4 200	28 892	3 128	22 788
Złotowski	6 785	32 728	6 975	22 725	5 105	25 595

Źródło: na podstawie: [Gołata 2004, s. 265-266] oraz mikrospis 1995.

J. Kordos [1999] wprowadził do użycia statystykę k_a^2 jako iloraz wariancji: estymatorów bezpośredniego $[s(\hat{y}_a)]$ i regresyjnego $[s^2(r_a)]$. Dzięki statystyce Kordo-

sa łatwo można się zorientować, co do zysku na efektywności estymatora bayesowskiego nad estymatorem bezpośrednim. Przekształcenie Kordosa umożliwia także przedstawienie formuł (1) i (2) na estymator bayesowski i jego wariancję w nowej, wygodniejszej formie. Estymator wartości globalnej można napisać w następującej formie:

$$\hat{y}_a(EB) = \frac{k_a^2}{1+k_a^2} \hat{y}_a(r) + \frac{1}{1+k_a^2} \hat{y}_a. \quad (3)$$

Natomiast wariancję estymatora bayesowskiego jako:

$$s^2[\hat{y}_a(EB)] = \frac{1}{1+k_a^2} s^2(\hat{y}_a). \quad (4)$$

Stosując estymator bayesowski, osiąga się tym większy zysk na efektywności, im wyższe jest k_a^2 . Dzięki formule (4) łatwo można zauważyć, że duże wartości ilorazu Kordosa sprawiają, iż tak zwany empiryczny estymator bayesowski jest bardziej efektywny od estymatora bezpośredniego. Jeśli k_a^2 jest równe 1, to empiryczny estymator bayesowski jest dwukrotnie bardziej efektywny od estymatora bezpośredniego.

Na podstawie wzorów na estymator 3 i 4 E. Gołata oszacowała liczby bezrobotnych i pracujących w oparciu o dane z mikrospisu 1995. Jako zmienną pomocniczą wykorzystano wielkość domeny, czyli liczby ludności w badanych powiatach. Pominięto przy tym powiaty grodzkie. Wartości „ogółem” wynoszące w tab. 1 odpowiednio 177 964 bezrobotnych, 1 244 419 pracujących nie zostały obliczone na podstawie empirycznego estymatora bayesowskiego, lecz stanowią sumę powiatów ziemskich. Estymator bezpośredni dla całego województwa wielkopolskiego z uwzględnieniem miast grodzkich wykazał 171 240 bezrobotnych i 1 293 631 pracujących, z czego w powiatach ziemskich było odpowiednio 133 909 i 973 186 osób. Zatem suma wartości estymatora empiryczno-bayesowskiego dla powiatów ziemskich była znacznie wyższa niż bezpośredniego. W przypadku bezrobotnych o 32,9%, a pracujących o 27,9%. Traktując estymator bezpośredni dla dużego obszaru jako bardziej wiarygodny, należałoby stwierdzić, że estymator empiryczno-bayesowski mocno zawyżył poziom obu wartości globalnych. Oceniając uzyskane wyniki w punktu widzenia kryteriów J. Paradysza [2008], można zauważyć znaczne przekroczenie kryterium poziomu. Przestrzenny układ wartości globalnych w poszczególnych powiatach ziemskich wydaje się wskazywać, że uporządkowanie według estymatora empiryczno-bayesowskiego jest właściwsze. Zatem można zastosować metodę rakingu (patrz dwie ostatnie kolumny w tab. 1).

Opisany już raking i metoda „korekcja szacunków” są bardziej uspokojeniem sumienia niezbyt wnikliwych użytkowników produkowanych przez nas statystyk niż rzeczywistym rozwiązaniem problemu. Po pierwsze, automatyczna zmiana we-

dług określonego współczynnika korygującego wartości szacunków powoduje, że zmieniły się również oceny ich dyspersji, które także należałoby skorygować. Po drugie, przyczyną nietrafnego oszacowania poziomu jako sumy dla wszystkich małych obszarów mogą być różnice w błędach szacunku, nawet gdy mają one charakter losowy. Na przykład jednym powiatom trafniej oszacowaliśmy odsetek bezrobotnych niż innym. Ponadto mogą tutaj wystąpić błędy nielosowe, na przykład różne rozmiary szarej strefy w poszczególnych małych obszarach. Stąd bardzo istotne jest uwzględnienie dwóch dalszych kryteriów kolejności i dystansu. Kryterium kolejności polega na tym, żeby tak uporządkować małe obszary pod względem badanej zmiennej, jak jest w rzeczywistości. Kryterium dystansu nakłada dodatkowy warunek zachowania odpowiedniej odległości między obszarami. Estymację spełniającą wszystkie trzy kryteria można by nazwać doskonałą. Jej przeciwieństwem byłoby uporządkowanie w odwrotnej kolejności, czyli w ten sposób, że najgorszy z obszarów byłby pierwszy, przedostatni – drugi, a najlepszy na samym końcu. Jako miary dobroci szacunku dla małych obszarów można byłoby wykorzystać współczynniki korelacji liniowej. Problemem jest tu istnienie odpowiedniego wzorca porządku i odległości. W cytowanej już pracy J. Paradysz [2008] sugerował wykorzystanie syntetycznego wskaźnika rozwoju regionalnego silnie skorelowanego z badaną zmienną. Alternatywnymi sposobami oceny dobroci estymacji dla małych obszarów byłyby symulacje na wzór tych, które zastosowano w projekcie EURAREA lub odwołanie się do opinii ekspertów i lokalnych autorytetów z prośbą o ocenę wykonanych szacunków (por. [Paradysz 2008a]).

W ostatnich latach w wielu opracowaniach nałożono na estymatory klasy SMO z góry pewne ograniczenia, które im zapewniają pożądane własności, podobne do kryteriów zaproponowanych przez J. Paradysza [2008]. Punktem wyjścia jest hierarchiczny estymator bayesowski parametru Θ , który dla a -tego obszaru oznaczamy jako $\hat{\Theta}_a^{HB}$, a jego wariancję $\hat{V}(\hat{\Theta}_a^{HB})$. Natomiast $\hat{\Theta}_a^{BHB}$ oznacza benchmarkowy estymator hierarchiczno-bayesowski. $\hat{\Theta}_a^{BHB}$ jest funkcją estymatorów hierarchiczno-bayesowskich. Benchmarkowy estymator bayesowski będzie lepszy od estymatorów klasycznych jedynie wtedy, gdy wartość oczekiwana parametrów w rozkładzie *a priori* będzie bliska prawdziwej wartości parametru. Stąd też P. Dick, Y. You, i J.N.K. Rao [2002] zaproponowali benchmarkowy estymator bayesowski o postaci:

$$\hat{\Theta}_a^{BHB} = f(\hat{\Theta}_1^{HB}, \hat{\Theta}_2^{HB}, \dots, \hat{\Theta}_m^{HB}) \quad (5)$$

dla pewnej funkcji $f(\cdot)$, która zaspokaja następującą własność benchmarkową:

$$\sum_{a=1}^m \hat{\Theta}_a^{BHB} = \sum_{a=1}^m y_a. \quad (6)$$

Prawa strona równania (6) powinna pochodzić z alternatywnego i wiarygodnego źródła, na przykład jako wartość oszacowana na podstawie wskaźnika syntetycznego rozwoju danego małego obszaru a . Jako miara precyzji szacunku może posłużyć aposterioryczny średni błąd kwadratowy (*posteriori mean squared error* – *PMSE*):

$$PSME(\hat{\Theta}_a^{BHB}) = (\hat{\Theta}_a^{BHB} - \hat{\Theta}_a^{HB})^2 + V(\Theta_a | y), \quad (7)$$

gdzie prawa strona równania jest sumą korekty obciążenia oraz wariancji aposteriorycznej $V(\Theta_a | y)$ dla a -tego małego obszaru.

Alternatywne metody estymacji benchmarkowej w statystyce małych obszarów (domen) zaproponowali D. Pfeffermann i R. Tiller [2003, 2005, 2006] oraz G.S. Datta, M. Ghosh, R. Steorts i J. Maples [2009]. Szczególne znaczenie mają tutaj prace o charakterze ciągłym, kiedy dłuższa seria danych w czasie może służyć jako korekta najnowszych badań. Wdzięcznym obszarem zastosowań jest rynek pracy i badania ekonomiczne o pewnej regularności czasowej.

Literatura

- D'Alò M., Di Consiglio L., Falorsi S., Solari F., *Small area estimation of the Italian poverty rate*, „Statistics in Transition” 2006, vol. 7, no. 4, s. 771-784.
- Bracha Cz., Lednicki B., Wieczorkowski R., *Estymacja danych z badania aktywności ekonomicznej ludności na poziomie powiatów dla lat 1995-2002*, Główny Urząd Statystyczny, Warszawa 2003.
- Datta G.S., Ghosh M., Steorts R., Maples J., *Bayesian Benchmarking with Applications to Small Area Estimation*, Statistical Research Division U.S. Census Bureau, Washington 2009, Research report series (Statistics #2009-01).
- Dick P., Yong You, Rao J.N.K., *Benchmarking Hierarchical Bayes Small Area Estimators with Application in Census Undercoverage Estimation*, SSC Annual Meeting, Proceedings of the Survey Methods Section, 2002.
- Gołata E., *Estymacja pośrednia bezrobocia na lokalnym rynku pracy*, Prace Habilitacyjne nr 11, Wyd. Akademii Ekonomicznej w Poznaniu, Poznań 2004.
- Heady P., Ralphs M., *EURAREA: An overview of the project and its findings*, „Statistics in Transition” 2005, vol. 7, no. 3, s. 557-570.
- Kordos J., *Problemy estymacji danych dla małych obszarów*, „Wiadomości Statystyczne” 1999, nr 1, s. 85-101.
- Kordos J., *Niektóre aspekty jakości w statystyce małych i średnich obszarów. Tradycja i obecne zadania statystyki w Polsce*, A. Zeliaś (red.), Wydawnictwo Akademii Ekonomicznej w Krakowie, Kraków 2004.
- Kordos J., Paradysz J., *Some experiments in small area estimation in Poland*, „Statistics in Transition” 2000, vol. 4, no. 4, s. 679-697.
- Paradysz J., *Kryteria dobroci estymacji dla małych obszarów*. Konferencja naukowa z okazji jubileuszu 90-lecia Głównego Urzędu Statystycznego: „Statystyka społeczna dokonania – szanse – perspektywy”, Kraków, 28-30.01.2008.

- Paradysz J., , *Kryteria samorządowe oceny estymacji pośredniej dla małych obszarów*, [w:] J. Olaszewski, M. Słodowa-Helpa (red.), *Koncepcje oraz czynniki rozwoju regionalnego i lokalnego w warunkach funkcjonowania Polski w strukturach zintegrowanej Europy*, Wydawnictwo Wyższej Szkoły Komunikacji i Zarządzania w Poznaniu, Poznań 2008a, s. 103-114.
- Pfeffermann D., *Small area estimation: New developments and directions*, „International Statistical Review” 2002, no. 1, s. 125-143.
- Pfeffermann D., Tiller R., *State-Space Modeling with Correlated Measurements with Application to Small Area Estimation Under Benchmark Constraints*, 2003, www.oecd.org/dataoecd/0/55/2640279.pdf.
- Pfeffermann D., Tiller R., Brown S., *Small Area Estimation with Stochastic Benchmark Constraints: Theory and Practical Application in US Labor Statistics*, Workshop on Frontiers in Benchmarking Techniques and Their Application to Official Statistics Luxembourg, 7-8 April 2005.
- Pfeffermann D., Tiller R., *Small area estimation with state space models subject to benchmark constraints*, „Journal of the American Statistical Association” 2006, vol. 101, no. 476, s. 1387-1397.
- Pfeffermann D., Sverchkov M., *Small Area Estimation Under Informative Probability Sampling of Areas and Within The Selected Areas*, Southampton Statistical Sciences Research Institute Methodology Working Paper M07/06, University of Southampton, 2007.
- Rao J.N.K., *Small Area Estimation*, Wiley, New York 2003.

BENCHMARKING IN THE STATISTICS OF SMALL AREAS

Summary: Sample surveys are used to produce credible estimates not only for the total population but also for a lot of subpopulations for example the regions or counties. However, sample size in small domains, particularly small geographical areas, is rarely big enough to provide reliable direct estimates for specific small areas. Small area estimation plays the leading role in the growing demand for reliable statistics even when only very small samples are available for these areas. In making estimates for small areas, it is necessary to “borrow strength” from related areas to form indirect estimators that increase the precision. For every small area we can produce a lot of estimates according to different models and auxiliary variables. We have problems of choosing right estimates. Benchmarking in small area statistics allows to choose right estimates for every region. We propose three kinds of criteria for right indirect estimates evaluation: level, order and distance