Grażyna Dehnel

SZACOWANIE WARIANCJI ESTYMATORA W STATYSTYCE MAŁYCH OBSZARÓW

W związku z coraz większym zapotrzebowaniem na ocenę parametrów dla bardzo szczegółowego podziału populacji na podpopulacje pojawia się wiele opracowań nowych metod estymacji parametrów, dotyczących małego obszaru bądź pewnej dziedziny studiów. Zaczynają one uzupełniać tradycyjne techniki, które okazują się bardziej czasochłonne, kosztowne i mniej dokładne. Pozwalają też uniknąć nawet przy bardzo małych liczebnościach próby wielkich błędów szacunku, które występują w wypadku stosowania klasycznych metod.

W literaturze przedmiotu istnieje dość znaczna liczba estymatorów szacujących parametry rozkładu badanej cechy. Nie ma jednak jednoznacznego określenia, który z nich, przy posiadanych informacjach i w danych warunkach, jest najodpowiedniejszy.W przypadku estymatorów proponowanych w statystyce małych obszarów nieznana jest a priori ich precyzja szacunku. Należą one do tej klasy estymatorów nieprostych, dla których z reguły nie daje się wyprowadzić wzoru na ich wariancję. Sensowne zatem wydaje się szacowanie rozmiarów precyzji estymatorów pośrednich na podstawie symulacji komputerowej. W wyniku kolejnych badań symulacyjnych poznajemy ich pewne własności, które mogą być pomocne przy wyborze estymatorów w ich praktycznym zastosowaniu.

Niniejsze opracowanie poświęcono badaniu własności dwunastu estymatorów statystyki małych obszarów w większości uwzględniających wartości cechy dodatkowej. Opierając się na pracach G. H. Choudry'ego i J. N. K. Rao (1992), A. C. Skinner'a (1991), a szczególnie na opracowaniu Cz. Brachy (1994) przeprowadzono badanie symulacyjne dla populacji rzeczywistej, gdzie próba losowana była według schematu dwustopniowego. Analizy dokonano

na podstawie danych pochodzących z Badania Dzietności prowadzonego w ramach NSP'88.

Dla potrzeb symulacji komputerowej zbiorowość 7691 gospodarstw domowych potraktowano jako populację generalną, z której wylosowano podpróbki o stałej liczebności 200 elementów. Rozpatrywane były trzy cechy:

– powierzchnia użytkowa mieszkania (y),
– liczba izb w mieszkaniu (x),
– płeć osoby reprezentującej gospodarstwo domowe (d).

Szacowano wartość pierwszej z wyżej wymienionych cech – przeciętną użytkową powierzchnię mieszkania. Druga – liczba izb – traktowana była jako cecha dodatkowa. Za przyjęciem takiego schematu przemawia silna dodatnia zależność między cechami ($r_{xy} = 0{,}8$). Cecha trzecia stanowiła wyróżnik przy podziale populacji generalnej na dziedziny.

W ramach populacji wydzielono trzy warstwy :

– gospodarstwa domowe w Poznaniu,
– gospodarstwa domowe w innych miastach,
– gospodarstwa domowe znajdujące się na wsi.

Badanie zostało ograniczone do pierwszej warstwy, liczącej 5120 jednostek. W jej ramach wyróżniono 40 grup po 128 jednostek każda. Niezależnie od tego podziału, w oparciu o trzecią cechę, a więc płeć osoby reprezentującej dane gospodarstwo, dokonano podziału populacji na dwie dziedziny.

Próba losowana była dwustopniowo. Na obu stopniach zastosowano losowanie według schematu prostego bez zwracania. Jednostkę pierwszego stopnia stanowił obwód spisowy, natomiast drugiego stopnia – gospodarstwo domowe. W ramach jednej próby losowano 10 grup, a następnie po 20 jednostek z każdej z nich. Tak więc liczebność próby była równa $20 \times 10 = 200$ jednostek. Wylosowano 1000 prób i w ramach próby obliczono dla każdej z dwóch dziedzin wartości 12 estymatorów. Służą one do estymacji wartości globalnej cechy. Znając jednak liczebności poszczególnych grup można na ich podstawie oszacować wartość średnią cechy. W przypadku badania wielkości mieszkań większy sens ma średnia arytmetyczna niż wartość globalna. Tak więc w oparciu o estymatory szacowano przeciętną powierzchnię mieszkania.

Przy opisie wszystkich estymatorów posłużono się stałymi oznaczeniami. S_{hd} stanowi tę część wylosowanej próby, która należy do dziedziny d i warstwy h, m_{hd} to wylosowana liczba grup należących do dziedziny d i warstwy h (dla $h = 1, 2$ $d = 1, 2$). $I = 1, 2, ..., m$ oznacza numery kolejnych grup biorących udział w próbie, natomiast $j = 1, 2, ..., k$ to gospodarstwa domowe wylosowane w ramach grupy. Liczebność próby jest zatem równa $m \times k = n$. Wartość cechy $Y(X)$ dla gospodarstwa domowego wylosowanego w j-tym ciągnieniu z grupy za i-tym razem z h-tej warstwy oznaczono przez $y_{hij}(x_{hij})$ dla $j = 1, 2, ..., n$, $i = 1, 2, ..., m$, $h = 1, 2$.

1. Estymator Horvitza-Thompsona (HT)

Estymator HT należy do grupy tak zwanych estymatorów bezpośrednich. Warto go zaprezentować w celu porównania jego rozkładu i efektywności z pozostałymi estymatorami. Estymator HT wykorzystuje jedynie wartości dotyczące badanej cechy, bez żadnych dodatkowych informacji.

$$\hat{Y}_{hd}^{(1)} = \frac{N_h}{n_h} \sum_{i \in s_{hd}} \sum_{j=1}^{k} y_{ij} . \tag{1}$$

2. Estymatory POS

Nazwa tego typu estymatorów jest skrótem angielskiego terminu – poststratified. Do ich zastosowania konieczna jest znajomość liczebności populacji w poszczególnych dziedzinach – N_{hd}. Podział na warstwy następuje dopiero po wylosowaniu próby. W tej grupie przedstawione są trzy rodzaje estymatorów:

a. Estymator zwykły POS – wykorzystujący jedynie informacje o badanej cesze Y i liczebności populacji dla danej dziedziny N_{hd}:

$$\hat{Y}_{hd}^{(2)} = \frac{N_{hd}}{\hat{N}_{hd}} \hat{Y}_{hd}^{(1)} = N_{hd} \hat{\bar{Y}}_{hd} , \tag{2}$$

gdzie

$$\hat{\bar{Y}}_{hd} = \hat{Y}_{hd}^{(1)} / \hat{N}_{hd} \tag{3}$$

$$\hat{N}_{hd} = \frac{N_h}{n_h} m_{hd} k . \tag{4}$$

Trzy kolejne estymatory wykorzystują oprócz informacji o badanej cesze i liczebności populacji dla danej dziedzin N_{hd}, także wartości cechy dodatkowej X.

b. Estymator ilorazowy POS

$$\hat{Y}_{hd}^{(3)} = \frac{\hat{Y}_{hd}^{(1)}}{\hat{X}_{hd}} X_{hd} = r_{hd} X_{hd} , \tag{5}$$

gdzie

$$r_{hd} = \frac{\hat{Y}_{hd}^{(1)}}{\hat{X}_{hd}^{(1)}}, \tag{6}$$

$\hat{X}_{hd}^{(1)}$ estymator (1) wartości globalnej cechy X.

c. Estymator regresyjny POS-1

$$\hat{Y}_{hd}^{(4)} = N_{hd}[\hat{\bar{Y}}_{hd} + b_{hd}(\bar{X}_{hd} - \hat{\bar{X}}_{hd})] = \hat{Y}_{hd}^{(2)} + b_{hd}(X_{hd} - \hat{X}_{hd}^{(2)}), \tag{7}$$

gdzie

$$b_{hd} = \frac{\sum_{i \in s_{hd}} \sum_{j=1}^{k} (x_{hij} - \bar{x}_{hd})(y_{hij} - \bar{y}_{hd})}{\sum_{i \in s_{hd}} \sum_{j=1}^{k} (x_{hij} - \bar{x}_{hd})^2} \tag{8}$$

$$\bar{X} = \frac{X_{hd}}{N_{hd}}, \tag{9}$$

$X^{(2)}$ – estymator (2) wartości globalnej cechy X.

d. Regresyjny estymator POS-2

$$\hat{Y}_{hd}^{(5)} = N_{hd}\left[\bar{y}_{hd} + \frac{\hat{y}_h}{\hat{x}_h}(\bar{X}_{hd} - \bar{x}_{hd})\right]. \tag{10}$$

3. Estymatory syntetyczne

Oparte są na założeniu, że struktura populacji jest identyczna (niewiele różniąca się) jak struktura w poszczególnych dziedzinach. Takie założenie pozwala na podział oszacowanej wartości globalnej dla całej populacji na części odpowiadające poszczególnym dziedzinom. Podziału tego dokonuje się proporcjonalnie do liczby jednostek z danej dziedziny bądź do wartości globalnej cechy dodatkowej dla dziedziny. Do tej grupy zaliczają się następujące estymatory:

a. Estymator zwykły SYN wykorzystujący informacje o badanej cesze w całej warstwie oraz informacje o wielkości dziedziny d w stosunku do warstwy h:

$$\hat{Y}_{hd}^{(6)} = N_{hd} \frac{\hat{Y}_h}{N_h} \tag{11}$$

$$\hat{Y}_h = \frac{N_h}{n_h} \sum_i \sum_j y_{ij}, \tag{12}$$

$\hat{Y}_h$ – estymator (1) wartości globalnej Y_h.

W dwóch kolejnych estymatorach przy szacowaniu wartości cechy Y będą brane pod uwagę dane dotyczące cechy X.

b. Estymator ilorazowy SYN

$$\hat{Y}_{hd}^{(7)} = \frac{\hat{Y}_h}{\hat{X}_h} X_{hd} = r_h X_{hd}, \tag{13}$$

gdzie

$$r_h = \frac{Y_h}{\hat{X}_h}, \tag{14}$$

$\hat{X}_{hd}$ estymator wartości globalnej cechy X obliczony na podstawie wzoru (11).

c. Estymator regresyjny SYN

$$\hat{Y}_{hd}^{(8)} = N_{hd}[\hat{\bar{Y}}_h + b_h(\bar{X}_h - \hat{\bar{X}}_h)], \tag{15}$$

gdzie

$$b_h = \frac{\sum_{i \in s_{hd}} \sum_{j=1}^{k} (x_{hij} - \bar{x}_{hi})(y_{hij} - \bar{y}_{hi})}{\sum_{i \in s_{hd}} \sum_{j=1}^{k} (x_{hij} - \bar{x}_{hi})^2} \tag{16}$$

$$\hat{\bar{Y}}_h = \frac{\hat{Y}_h}{N_h}, \tag{17}$$

$\hat{\bar{Y}}_h(\hat{\bar{X}}_h)$ – nieobciążony estymator średniej $\bar{Y}_h(\bar{X}_h)$.

4. Uogólnione estymatory regresyjne (MRE)

Składają się z dwóch części. Pierwszą stanowi odpowiedni estymator syntetyczny. Druga to składnik korygujący, wykorzystujący reszty między rzeczywistymi wartościami cechy Y jednostek wylosowanych do próby a wartościami cechy Y, szacowanymi w oparciu o estymator ilorazowy oraz regresyjny.

a. Estymator ilorazowy MRE, czyli skorygowany o reszty ilorazowy estymator syntetyczny

$$\hat{Y}_{hd}^{(9)} = \hat{Y}_{hd}^{(7)} + N_{hd}\, e_{qdh} / \hat{N}_{hd}, \tag{18}$$

gdzie

$$e_{qdh} = \frac{N_h}{n_h} \sum_{i \in s_{hd}} \sum_{j=1}^{k} e_{qhij} \tag{19}$$

$$e_{qhij} = y_{hij} - r_h, \tag{20}$$

dla $j = 1, 2, \ldots, k$; $i \in s_{hd}$.

b. Estymator regresyjny MRE – estymator regresyjny syntetyczny skorygowany o reszty

$$\hat{Y}_{hd}^{(10)} = \hat{Y}_{hd}^{(8)} + N_{hd}\, e_{lrdh} / \hat{N}_{hd}, \tag{21}$$

gdzie

$$e_{lrdh} = \frac{N_h}{n_h} \sum_{i \in s_{hd}} \sum_{j=1}^{k} e_{lrhij} \tag{22}$$

$$e_{lrhij} = y_{hij} - y_h - b_h(\overline{X}_h - \overline{x}_h), \tag{23}$$

dla $j = 1, 2, ..., k$; $i \in s_{hd}$.

5. Estymatory złożone

Zależne są od wielkości próby (Sample-Size Dependent Estimators) stanowią ostatnią z pięciu grup estymatorów. Każdy z nich oparty jest na dwóch estymatorach omawianych już wcześniej. Przy czym wpływ poszczególnych estymatorów na szacowaną wartość cechy zależy od tego, jak duża jest liczebność próby w dziedzinie w stosunku do całej próby.

a. W **estymatorze złożonym SSD** udział estymatorów składowych określony jest np. przez stosunek $w_{hd} = \frac{n_{hd}}{n_h}$ i $W_{hd} = \frac{N_{hd}}{N_h}$.

$$\hat{Y}_{hd}^{(11)} = \alpha_{hd}\,\hat{Y}_{hd}^{(2)} + (1 - \alpha_{hd})\,\hat{Y}_{hd}^{(7)}, \tag{24}$$

gdzie

$$\alpha_{hd} = \begin{cases} 1, & \text{jeśli} \quad w_{hd} \geq \delta\, W_{hd} \\ 1/\delta \left(\dfrac{w_{hd}}{W_{hd}} \right) & \text{jeśli} \quad w_{hd} < W_{hd}, \end{cases} \tag{25}$$

δ – parametr wybierany subiektywnie, ma wpływ na wielkość udziału estymatora syntetycznego w szacowaniu wartości estymatora (11). Wraz z jego wzrostem rośnie udział tego estymatora składowego. Najczęściej parametr przybiera wartość równą 1 (G. H. Choudry, J. N. K. Rao, 1992).

b. Estymator złożony SSD* wykorzystuje estymatory zbudowane w oparciu o cechę dodatkową X_{hd}

$$\hat{Y}_{hd}^{(12)} = \alpha_{hd}^{*}\,\hat{Y}_{hd}^{(5)} + (1 - \alpha_{hd}^{*})\,\hat{Y}_{hd}^{(7)}, \tag{26}$$

gdzie

$$\alpha^{*}_{hd} = \begin{cases} 1, & \text{jeśli} \quad w_{hd} \geq W_{hd} \\ \left(\frac{w_{hd}}{W_{hd}}\right)^{h-1}, & \text{jeśli} \quad w_{hd} < W_{hd}, \end{cases} \tag{27}$$

parametr h dobiera się subiektywnie określając tym samym wpływ estymatora syntetycznego na szacowanie wartości cechy Y (C. E. Särndal i M. A. Hidiroglou – 1989 – sugerują , aby h było równe 2).

Po przeprowadzeniu badania symulacyjnego dla każdego z dwunastu estymatorów obliczono przybliżone wartości : wariancji estymatora, względnego obciążenia estymatora oraz względnego błędu szacunku.
Wielkość:

$$\hat{Y}_{hd} = \frac{1}{1000} \sum_{p=1}^{1000} \hat{Y}_{hd} \tag{28}$$

z uwagi na dużą liczbę prób można potraktować jako przybliżoną wartość oczekiwaną estymatora

$$\hat{\bar{Y}}_{hd} = E(\hat{Y}_{hd}). \tag{29}$$

Wariancja natomiast:

$$S^2(\hat{Y}_{hd}) = \frac{1}{999} \sum_{p=1}^{1000} (\hat{Y} - \hat{\bar{Y}})^2 \tag{30}$$

może być uważana za przybliżoną wartość wariancji estymatora

$$S^2(\hat{Y}_{hd}) \approx D^2(\hat{Y}_{hd}). \tag{31}$$

Na podstawie wzoru (32) określono (w %) względne obciążenie estymatora:

$$RBE = \frac{\left|\hat{Y}_{hd} - Y_{hd}\right|}{Y_{hd}}. \tag{32}$$

Względne błędy szacunku (w %) obliczone zostały zgodnie ze wzorem:

$$REE = \frac{1}{1000} \sum_{p=1}^{1000} \left(\frac{\left|\hat{Y}_{hd} - Y_{hd}\right|}{Y_{hd}} \right). \tag{33}$$

Wartości wyżej wymienionych trzech parametrów przedstawione w tabeli 1 pozwalają na wstępną ocenę wyników przeprowadzonego badania.

Tabela 1. Porównanie parametrów estymatorów

	HT	POS	POS.I	POS.R	POS.RE	SYN	SYN.I	SYN.R	MRE	MRE	SSD	SSD*
$S^2(x)$	314,1	8,8	5,3	3,5	5,4	3,9	1,5	1,4	5,9	9,6	4,4	3
RBE (%)	0,15	0,36	0,23	0,2	0,26	0,32	0,72	0,32	0,06	0,07	0,09	0,04
REE (%)	22	3,7	2,4	2,3	2,3	2,5	1,6	1,6	1,3	0,3	0,6	0,9

$S^2(x)$ – wariancja estymatora,
RBE – względne obciążenie estymatora (w %),
REE – względne błędy szacunku (w %).

Rozpatrując wariancję estymatorów można zauważyć, że dla pierwszego z nich jest ona nieporównywalnie duża. W kolejności zmniejszania się wartości wariancji mamy grupę estymatorów typu POS, uogólnione estymatory regresyjne, następnie estymatory złożone oraz grupę estymatorów typu SYN – gdzie wartości wariancji są najmniejsze. Zarówno wewnątrz grupy estymatorów typu POS, jak i typu SYN największą wariancję mają estymatory zwykłe. Jej wartość jest dwukrotnie większa od wartości wariancji estymatorów ilorazowych i regresyjnych z tej samej grupy.

Oceniając ogólnie estymatory pod względem ich wariancji najlepiej prezentują się estymatory : syntetyczny ilorazowy i syntetyczny regresyjny.

O jakości estymatora świadczy między innymi również jego obciążenie. Pod tym względem zdecydowanie najlepsze są estymatory MRE oraz grupa estymatorów złożonych. Większe obciążenie posiadają estymatory typu POS, wśród których estymator regresyjny ma najmniejsze obciążenie.

Estymatory typu SYN teoretycznie powinny mieć mniejsze obciążenie od estymatorów typu POS. Zbudowane są bowiem w oparciu o całą próbę, a nie jedynie na podstawie jednostek wchodzących w skład danej dziedziny. Taka sytuacja miałaby miejsce, gdyby struktury cechy badanej oraz dodatkowej w każdej dziedzinie, jak i w całej warstwie były identyczne (mało różniące się między sobą). W praktyce jednak takie sytuacje nie zdarzają się i dlatego estymatory typu SYN są obciążone. W miarę wzrostu różnicy między strukturą dla dziedziny i strukturą dla warstwy wzrasta obciążenie estymatorów. Korzystny wpływ na wartość obciążenia ma duża zależność między cechą badaną i dodatkową. Wartości obciążenia estymatorów typu SYN są w tym badaniu stosunkowo dość znaczne. Najlepszy z nich jest, są podobnie jak w grupie estymatorów POS, estymator regresyjny.

O precyzji estymatora świadczy także względny błąd szacunku. Wyniki badania pokazują, że estymator HT jest najmniej efektywny z przedstawio-

nych. Kolejne miejsce zajmuje zwykły estymator typu POS. Lepszy od nich jest trzeci z estymatorów wykorzystujących cechę dodatkową – zwykły estymator typu SYN.

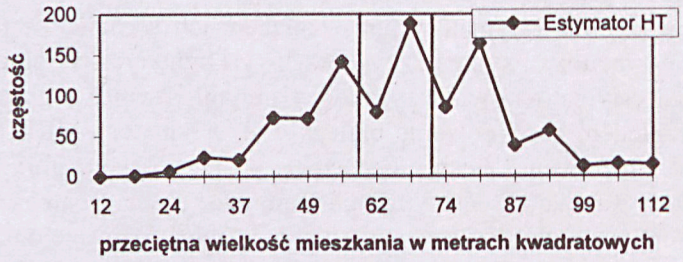
Wśród tych estymatorów, które cechę dodatkową wykorzystują, ogólnie rzecz biorąc, estymatory typu SYN są efektywniejsze od odpowiednich estymatorów typu POS.

Zwiększenie efektywności wiąże się między innymi z tym, że większa jest liczebność próby dla estymatora typu SYN. Zbudowane są one na podstawie całej próby, w przeciwieństwie do estymatora typu POS oszacowanego na podstawie jednostek należących do rozpatrywanej dziedziny.

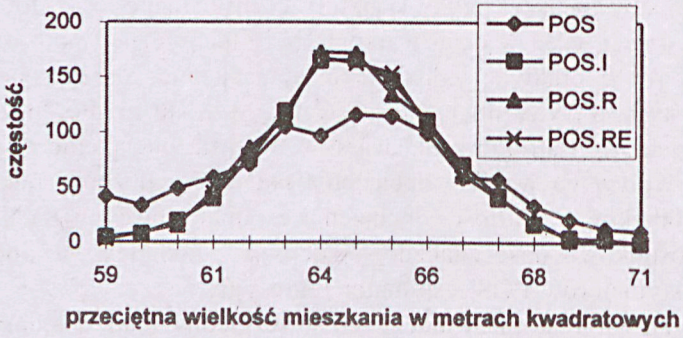
Kolejna grupa, a więc uogólnione estymatory regresyjne, charakteryzują się jeszcze mniejszym błędem szacunku od swoich odpowiedników w poprzednich dwóch grupach. Podobnie korzystnie przedstawia się sytuacja w grupie piątej, a więc wśród estymatorów złożonych.

W oparciu o wartości estymatorów we wszystkich próbach wyznaczono ich rozkłady (rys.1-5). Badanie istotności różnic na poziomie istotności α=0,001 nie dało podstaw do odrzucenia hipotezy o normalności rozkładu dla grupy estymatorów syntetycznych oraz dla estymatora złożonego SSD*.

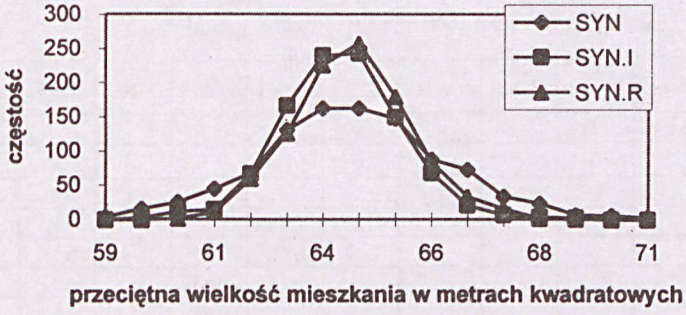
Na rys. 4 dodatkowo oznaczono pionowymi liniami obszar zmienności szacowanej wartości cechy dla wszystkich pozostałych estymatorów.



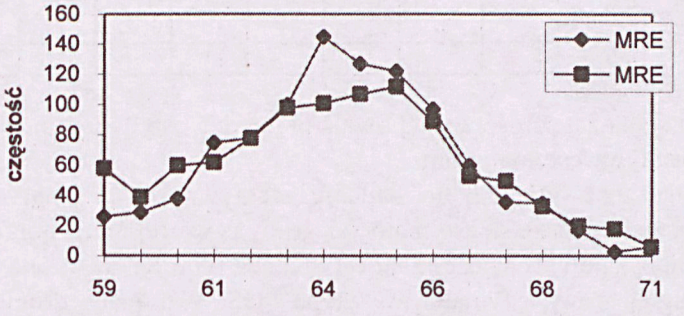
Rys. 1. Estymator Horvitza-Thompsona



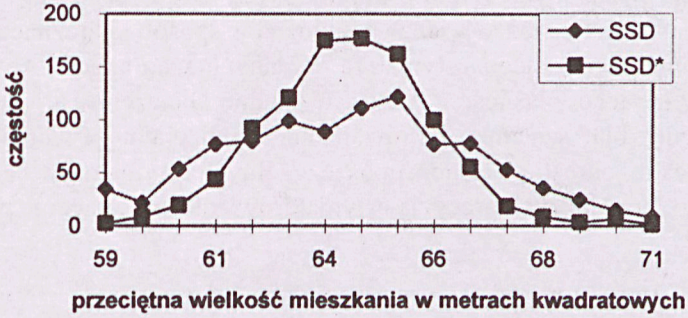
Rys. 2. Estymatory typu POS



Rys. 3. Estymatory typu SYN



Rys. 4. Estymatory typu MRE



Rys. 5. Estymatory typu SSD

Przeprowadzono również badanie zależności między wartościami poszczególnych estymatorów (tabela 2).

Silną dodatnią zależność wykazał estymator złożony SSD w powiązaniu z estymatorem ilorazowym POS oraz z regresyjnym MRE. Natomiast estymator SSD* ujawnił wyraźny związek z estymatorami typu POS (reg.1; reg.2). Mocną korelacją charakteryzowały się również następujące pary estymatorów : syntetyczne (regresyjny i ilorazowy), typu POS (reg.1 i reg.2), ilorazowy POS i regresyjny MRE.

Tabela 2. Zestawienie zależności między estymatorami (r_{xy})

	POS	POS.I	POS.R	POS.RE	SYN	SYN.I	SYN.R	MRE	MRE	SSD	SSD*
HT	0,09	0,12	0,05	0,08	0,09	0,02	0,01	0,1	0,08	0,08	0,07
POS		0,5	0,59	0,71	0,68	0,46	0,37	0,87	**0,99**	**0,98**	0,68
POS.I			0,7	0,72	0,36	0,51	0,49	0,57	0,48	0,55	0,71
POS.R				**0,97**	0,39	0,68	0,68	0,73	0,56	0,67	**0,95**
POS.RE					0,48	0,69	0,67	0,8	0,68	0,77	**0,98**
SYN						0,68	0,55	0,83	0,7	0,75	0,58
SYN.I							**0,98**	0,57	0,44	0,63	0,83
SYN.R								0,46	0,33	0,54	0,81
MRE									0,87	0,89	0,79
MRE.1										**0,97**	0,65
SSD											0,78

Bardzo słaba zależność występowała pomiędzy estymatorem HT a wszystkimi pozostałymi estymatorami.

Podsumowując symulacyjne badanie efektywności estymatorów można stwierdzić, że najlepsze są estymatory z grupy czwartej MRE oraz estymatory złożone. Słabszą efektywność mają estymatory typu SYN, są one jednak lepsze od drugiej grupy estymatorów – typu POS. W ramach drugiej i trzeciej grupy najbardziej efektywny jest estymator regresyjny, a najmniej – zwykły estymator. Wyniki te zasadniczo pokrywają się z wynikami badań symulacyjnych przeprowadzonych między innymi przez Cz. Brachę (1994). Na tym etapie badań wydaje się, że jeszcze nie można w sposób jednoznaczny określić zastosowania konkretnego estymatora w danych warunkach. Aby to zrobić trzeba poznać jego własności (a więc względne obciążenie, a przede wszystkim względny błąd szacunku). Prowadzona symulacja może jednak, w pewien ogólny sposób, określić własności poszczególnych estymatorów bądź wskazać na wymogi, przy których precyzja estymatorów zostanie znacznie poprawiona.

Literatura

Bracha Cz. (1992), Próba zbadania własności estymatorów wartości globalnej dla dziedzin w losowaniu dwustopniowym, „Wiadomości Statystyczne", nr 37.11, s. 9-13.

Bracha Cz. (1994), Metodologiczne aspekty badania małych obszarów, „Studia i Materiały" nr 43, GUS, Warszawa.

Cassel C. M., Särndal C. E. i **Wretman J. H.** (1977), Foundations of Inference in Survey Sampling, JW&S, New York.

Choudhry G. H., Rao J. N. K. (1992), Evaluation of small area estimators and empirical study, referat na międzynarodową konferencję naukową, „Small Area Statistics and Survey Designs", GUS, Warszawa 30 września – 3 października 1992.

Elliot D. (1992), A project to examine the potential for small area estimates of employment and health data in the UK – Some preliminary thoughts, referat na międzynarodową konferencję naukową, „Small Area Statistics and Survey Designs", GUS, Warszawa 30 września – 3 października 1992.

Elvers E., Särndal C. E., Wretman J. H. i **Örnber G.** (1985), Regression analysis and ratio analysis for domains: a randomization-theory approach. Canadian Journal of Statistics 13.2, s. 185-199.

Platek R., Rao J. N. K., Särndal C. E. i **Singh M. P.** (1987), Small Area Statistics, An International Symposium, JW&S, New York.

Särndal C. E. (1981), Frameworks for inference in survey sampling with applications to small area estimations and dajustment for nonresponse, Bulletin of the ISI 49, s. 494-513.

Särndal C. E. (1981), Design-consistent versus model-dependent estimation for small domains, JASA 79, s. 624-631.

Särndal C. E. (1992), Design-based approaches in estimation for domains, referat na międzynarodową konferencję naukową, „Small Area Statistics and Survey Designs", GUS, Warszawa 30 września – 3 października 1992.

Särndal C .E. i **Hidiroglou M. A.** (1989), Small domain estimation: a conditional analysis, JASA 84, s. 266-275.

Skinner C. (1991), The use of synthetic estimation techniques to produce small area estimates, Department of Social Statistics Uniwersity of Southampton, maszynopis powielony.

Steczkowski J. (1995), Reprezentacyjne badania jakości wyrobów, Wydawnictwo PLATAN, Kraków.

Welfe A. (1995), Ekonometria, PWE, Warszawa.

Wywiał J. (1991), O planie losowania próby proporcjonalnym do średniej z próby cechy pomocniczej, „Wiadomości Statystyczne", s. 21-22.

Wywiał J. (1992), Statystyczna metoda reprezentacyjna w badaniach ekonomicznych (Optymalizacja badań próbkowych), Prace Naukowe Akademii Ekonomicznej w Katowicach.

Wywiał J. (1995), Wielowymiarowe aspekty metody reprezentacyjnej, Ossolineum, Wrocław.