

*Wojciech Gamrot**

O WYKORZYSTANIU METODY WAŻENIA DANYCH DO ESTYMACJI KOWARIANCJI PRZY BRAKACH ODPOWIEDZI

Streszczenie. Metoda ważenia danych w przypadku wystąpienia braków odpowiedzi opiera się na przypisaniu zaobserwowanym wartościom badanej cechy odpowiednich wag kompensujących niekompletność danych w próbie. W niniejszym opracowaniu podjęto próbę zastosowania tej metody do estymacji złożonego parametru populacji jakim jest kowariancja dwóch cech. Zaproponowano estymatory kowariancji wykorzystujące ważenie danych. Ich własności zbadano z wykorzystaniem symulacji komputerowej.

Słowa kluczowe: Ważenie danych, estymacja, kowariancja, brak odpowiedzi.

I. WPROWADZENIE

Niekompletność danych w próbie statystycznej jest problemem z jakim spotyka się organizator każdego badania statystycznego prowadzonego na dużą skalę. Jego wystąpienie pociąga za sobą obciążenie oszacowań parametrów populacji oraz utratę precyzji estymacji wynikającą ze zmniejszenia liczebności próby. Wiele wysiłku poświęca się ograniczeniu tych niepożądanych efektów. Szerokie zastosowanie znajdują tu metody estymacji wykorzystujące związki pomiędzy wartościami cech pomocniczych i skłonnością poszczególnych jednostek populacji do uczestnictwa w badaniu statystycznym. Znaczącą grupę wśród nich stanowią tak zwane metody ważenia danych (ang. weighting adjustment methods), polegające na tworzeniu oszacowań parametrów populacji będących kombinacjami liniowymi wartości cechy poprzez przypisanie obserwacjom cechy zarejestrowanym w niekompletnej próbie odpowiednich wag równoważących zarówno fakt próbkowania populacji, jak też niekompletność obserwacji w niej. Wagi te tworzy się często jako odwrotności znanych prawdopodobieństw inkluzji i szacowanych na podstawie dostępnej informacji o wartościach cech pomocniczych indywidualnych prawdopodobieństw odpowiedzi. Stosunkowo niewiele uwagi poświęcono dotychczas w literaturze przedmiotu możliwościom wykorzystania tego podejścia do estymacji bardziej złożonych parametrów populacji, nie będących liniowymi kombinacjami obserwacji cechy, takich jak

* Dr, Katedra Statystyki, Uniwersytet Ekonomiczny w Katowicach.

wariancja, kowariancja, czy też współczynnik zmienności. W niniejszym opracowaniu podjęto próbę wypełnienia tej luki, poprzez konstrukcję opartych o ważenie danych estymatorów kowariancji oraz przez zbadanie ich własności poprzez przeprowadzenie symulacji komputerowej.

II. ESTYMACJA WARTOŚCI GLOBALNEJ

Rozważmy skończoną populację U o rozmiarze N i pewną cechę X przyjmującą dla poszczególnych jednostek populacji ustalone wartości $x_1, \dots, x_N$. Celem badania jest estymacja wartości globalnej $t(X) = \sum_{i \in U} x_i$ cechy X . Niech s ozna-

cza próbę losowaną z populacji U z wykorzystaniem pewnego planu losowania $p(s)$ charakteryzującego się prawdopodobieństwami inkluzji rzędu pierwszego odpowiednio równymi $\pi_i = \sum_{s \ni i} p(s)$ dla $i \in U$ oraz prawdopodobieństwami in-

kluzji drugiego rzędu odpowiednio równymi $\pi_{ij} = \sum_{s \ni i, j} p(s)$ dla $i \neq j \in U$. Przyj-

mijmy, że wskutek wystąpienia braków odpowiedzi wartości badanych cech obserwowane są jedynie w pewnym podzbiorze s_1 próby s . Zbiór jednostek populacji dla których nie uzyskano obserwacji cechy oznaczmy symbolem s_2 . Podobnie jak Cassel i inni (1983) przyjmujemy, że brak odpowiedzi jest zjawiskiem losowym i każdemu możliwemu podzbirowi s_1 próby s odpowiada pewne prawdopodobieństwo jego realizacji równe $q(s_1 | s)$. Warunkowy rozkład prawdopodobieństwa $q(s_1 | s)$ nazwiemy rozkładem odpowiedzi (ang. *response distribution*). Określa on prawdopodobieństwa uzyskania odpowiedzi od każdej i -tej jednostki populacji wynoszące $\rho_{i|s} = \sum_{s_1 \ni i} q(s_1 | s)$ dla $i \in U$ oraz prawdopo-

dobieństwa równoczesnego uzyskania odpowiedzi od i -tej i j -tej jednostki populacji wynoszące $\rho_{ij|s} = \sum_{s_1 \ni i, j} q(s_1 | s)$ dla $i \neq j \in U$. W przypadku pełnej odpowiedzi,

a więc wtedy gdy $\rho_{i|s} = \rho_{ij|s} = 1$ dla $i, j \in U$ nieobciążonym estymatorem wartości globalnej w populacji jest statystyka zaproponowana przez Horvitz i Thompsona (1952). Niestety, w przeciwnym wypadku estymator ten jest obciążony a jego obciążenie nie maleje wraz ze wzrostem liczebności próby (nie jest nawet asymptotycznie nieobciążony) co wykazali Nargundkar i Joshi (1975). W tej sytuacji jako oszacowanie wartości globalnej cechy X w populacji rozważa się statystykę (por. Lessler i Kalsbeek 1992):

$$\hat{t}_\rho(X) = \sum_{i \in S_1} \frac{x_i}{\pi_i \hat{\rho}_i} \quad (1)$$

gdzie $\hat{\rho}_i$ oznacza oszacowanie indywidualnego prawdopodobieństwa odpowiedzi ρ_i . Można wykazać, że gdyby oszacowania $\hat{\rho}_i$ były bezbłędne, to powyższa statystyka byłaby nieobciążonym estymatorem $t(X)$. W praktyce niestety indywidualne prawdopodobieństwa odpowiedzi nie są znane i zachodzi potrzeba ich estymacji.

III. ESTYMACJA PRAWDOPODOBIENSTW ODPOWIEDZI

W najprostszym przypadku dla oszacowania prawdopodobieństw odpowiedzi można się posłużyć modelem, stanowiącym, iż udzielenie lub nie udzielenie odpowiedzi przez dowolną i -tą i j -tą spośród jednostek wylosowanych do próby to niezależne zdarzenia losowe o jednakowym prawdopodobieństwie niezależnym od próby s co można zapisać formułami:

$$\forall_{i \in U} \rho_{i|s} = \rho; \quad (2)$$

$$\forall_{i,j \in U} \rho_{ij|s} = \rho_{i|s} \cdot \rho_{j|s} = \rho^2 \quad (3)$$

Estymator największej wiarygodności parametru ρ przyjmuje postać:

$$\hat{\rho}_{UNI} = \frac{1}{N} \sum_{i \in S_1} \frac{1}{\pi_i} \quad (4)$$

W konsekwencji otrzymujemy estymator wartości globalnej $t(X)$ w postaci:

$$\hat{t}_{UNI}(X) = \sum_{i \in S_1} \frac{x_i}{\pi_i \hat{\rho}_{UNI}} \quad (5)$$

Bethlehem (1988) wykazał, że przybliżone obciążenie tego estymatora wyznaczone poprzez rozwinięcie go w szereg Taylora z dokładnością do wyrazów pierwszego rzędu jest proporcjonalne do kowariancji pomiędzy prawdopodobieństwami odpowiedzi i wartościami cechy badanej a ponadto wyznaczył jego przybliżoną wariancję.

Inny sposób konstrukcji oszacowań indywidualnych prawdopodobieństw odpowiedzi opiera się na wykorzystaniu zależności pomiędzy wartościami cech

pomocniczych obserwowanymi w całej próbie s , oraz prawdopodobieństwami odpowiedzi. Mogą one okazać się przydatne do wyznaczenia dokładniejszych oszacowań indywidualnych prawdopodobieństw odpowiedzi. W tym celu przydatny będzie model regresji logistycznej (por np. Cassel i inni 1983), opisywany formułą:

$$\rho_i(\beta_0, \beta_1) = (1 + \exp(\beta_0 + \beta_1 z_i))^{-1} \quad (6)$$

gdzie $z_1, \dots, z_N$ oznaczają wartości pewnej cechy pomocniczej obserwowanej w całej próbie s , natomiast β_0 i β_1 to nieznane parametry. Oceny $\hat{\beta}_0$ i $\hat{\beta}_1$ tych parametrów można uzyskać maksymalizując funkcję wiarygodności (por. np. Thiel 1979):

$$L(\beta_0, \beta_1) = \prod_{i \in s_1} \rho_i(\beta_0, \beta_1) \cdot \prod_{i \in s_2} (1 - \rho_i(\beta_0, \beta_1)) \quad (7)$$

Rozwiązanie tak postawionego problemu można uzyskać z wykorzystaniem metod iteracyjnych. W niniejszej pracy zastosowano do tego celu metodę największego spadku Rosena (1969). Wykorzystując uzyskane w ten sposób oszacowania $\hat{\beta}_0$ i $\hat{\beta}_1$ można skonstruować estymator indywidualnego prawdopodobieństwa odpowiedzi ρ_i w postaci:

$$\hat{\rho}_{iLOG} = (1 + \exp(\hat{\beta}_0 + \hat{\beta}_1 z_i))^{-1} \quad (8)$$

Stąd estymator wartości globalnej $t(X)$ przyjmuje postać:

$$\hat{t}_{LOG}(X) = \sum_{i \in s_1} \frac{x_i}{\pi_i \hat{\rho}_{iLOG}} \quad (9)$$

W pewnych sytuacjach wykorzystanie metod parametrycznych do szacowania indywidualnych prawdopodobieństw odpowiedzi napotyka trudności. Dzieje się tak wtedy gdy brak jest przesłanek mogących uzasadnić wybór postaci modelu i gdy zachodzi niebezpieczeństwo nietrafnej jego specyfikacji. Do szacowania prawdopodobieństw odpowiedzi można wtedy wykorzystać metody nieparametryczne, a w szczególności estymatory jądrowe (por. np. Kulczycki (2005)), nie wymagające jawnego określania analitycznej postaci zależności pomiędzy cechami pomocniczymi i prawdopodobieństwami odpowiedzi. W niniejszym artykule wykorzystano rozważany wcześniej przez Giommiego (1987), estymator jądrowy:

$$\hat{\rho}_{iKRN} = \sum_{j \in S_1} \exp\left(\frac{-(z_i - z_j)^2}{2h^2}\right) / \sum_{j \in S} \exp\left(\frac{-(z_i - z_j)^2}{2h^2}\right) \quad (10)$$

gdzie stała h określana jest zwykle mianem parametru wygładzania lub szerokości pasma (ang. bandwidth). Określa ona z jednej strony czułość metody na lokalne anomalie rozkładu prawdopodobieństwa odpowiedzi, a z drugiej strony zróżnicowanie uzyskiwanych ocen tego prawdopodobieństwa. Zwiększenie wartości h oznacza zatem zwiększenie wariancji oszacowań względem obciążenia, a zmniejszenie skutkuje zmniejszeniem obciążenia kosztem wzrostu wariancji. Na potrzeby niniejszego artykułu przyjęto arbitralnie, że $h=1$. W rezultacie otrzymujemy kolejny estymator wartości globalnej $t(X)$ w postaci:

$$\hat{t}_{KRN}(X) = \sum_{i \in S_1} \frac{x_i}{\pi_i \hat{\rho}_{iKRN}} \quad (11)$$

IV. ESTYMACJA KOWARIANCJI

Rozważmy dwie cechy X i Y przyjmujące w skończonej populacji U odpowiednio wartości $x_1, \dots, x_N$ oraz $y_1, \dots, y_N$. Kowariancję cech X i Y można zapisać za pomocą formuły:

$$C(X, Y) = \frac{1}{N-1} \sum_{i \in U} \left(x_i - \frac{1}{N} \sum_{i \in U} x_i \right) \left(y_i - \frac{1}{N} \sum_{i \in U} y_i \right) \quad (12)$$

Dla potrzeb estymacji parametru $C(X, Y)$ w sytuacji gdy w badaniu statystycznym występują braki odpowiedzi zapiszemy powyższe wyrażenie w równoważnej postaci:

$$C(X, Y) = \frac{1}{N-1} (N^{-1}t(XY) - N^{-2}t(X)t(Y)) \quad (13)$$

gdzie $t(X) = \sum_{i \in U} x_i$, $t(Y) = \sum_{i \in U} y_i$ oraz $t(XY) = \sum_{i \in U} x_i y_i$. W przypadku pełnej odpowiedzi jednym z powszechnie znanych sposobów estymacji parametru $C(X, Y)$ jest zastąpienie nieznanymi wartościami globalnymi $t(X)$, $t(Y)$, $t(XY)$ odpowiednimi estymatorami Horvitz-Thompsona (por. np. Särndal et al 1992). W przypadku braku odpowiedzi estymator Horvitz-Thompsona nie jest jednak

nawet asymptotycznie nieobciążony, wskutek czego również estymator kowariancji będzie się charakteryzował obciążeniem nie malejącym do zera wraz ze wzrostem liczebności próby. Dlatego też, w miejsce statystyk Horvitz-Thompsona wykorzystamy omawiane wyżej estymatory wykorzystujące ważne dane. Rozważmy zatem następujące trzy estymatory kowariancji:

$$\hat{C}_{\text{UNI}}(X, Y) = \frac{1}{N-1} (N^{-1} \hat{t}_{\text{UNI}}(XY) - N^{-2} \hat{t}_{\text{UNI}}(X) \hat{t}_{\text{UNI}}(Y)) \quad (14)$$

$$\hat{C}_{\text{LOG}}(X, Y) = \frac{1}{N-1} (N^{-1} \hat{t}_{\text{LOG}}(XY) - N^{-2} \hat{t}_{\text{LOG}}(X) \hat{t}_{\text{LOG}}(Y)) \quad (15)$$

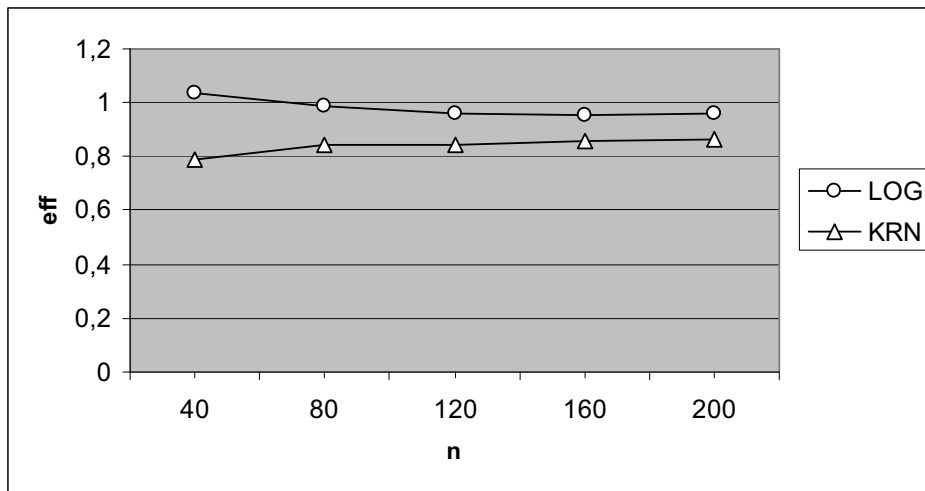
$$\hat{C}_{\text{KRN}}(X, Y) = \frac{1}{N-1} (N^{-1} \hat{t}_{\text{KRN}}(XY) - N^{-2} \hat{t}_{\text{KRN}}(X) \hat{t}_{\text{KRN}}(Y)) \quad (16)$$

gdzie statystyki $\hat{t}_{\text{UNI}}(X)$, $\hat{t}_{\text{UNI}}(Y)$ i $\hat{t}_{\text{UNI}}(XY)$ są odpowiednio estymatorami wartości globalnych $t(X)$, $t(Y)$ i $t(XY)$ wyznaczanymi zgodnie z wzorem (5), statystyki $\hat{t}_{\text{LOG}}(X)$, $\hat{t}_{\text{LOG}}(Y)$ i $\hat{t}_{\text{LOG}}(XY)$ są odpowiednio ich estymatorami wyznaczanymi z wykorzystaniem wzoru (9), natomiast statystyki $\hat{t}_{\text{KRN}}(X)$, $\hat{t}_{\text{KRN}}(Y)$ i $\hat{t}_{\text{KRN}}(XY)$ są odpowiednio ich estymatorami wyznaczanymi z wykorzystaniem wzoru (11).

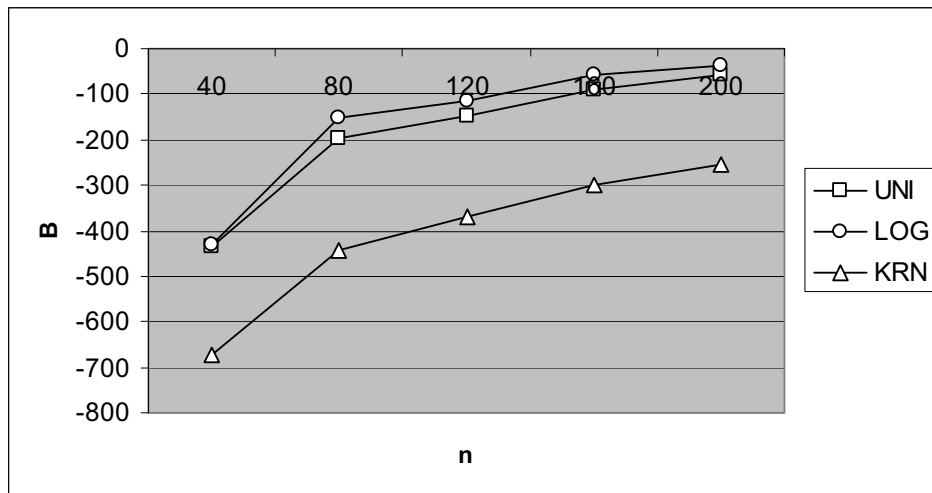
VI. WYNIKI EKSPERYMENTÓW SYMULACYJNYCH

Celem eksperymentów symulacyjnych było zbadanie własności stochastycznych zaproponowanych strategii estymacji. W badaniach wykorzystano dane uzyskane podczas spisu rolnego przeprowadzonego w roku 1996 w wybranych gminach powiatu Dąbrowa Tarnowska (gminy: Bolesław, Gręboszów i Radgoszcz). Dotyczyły one 2420 gospodarstw, które w symulacjach reprezentowały badaną populację. Wykorzystano trzy obserwowane zmienne: sprzedaż ogółem gospodarstwa (X), pogłowie bydła (Y) i powierzchnię gospodarstwa (Z). Szacowano kowariancję cech X i Y w populacji z wykorzystaniem cechy Z jako pomocniczej dla celów estymacji prawdopodobieństw odpowiedzi. Przyjęto, że prawdopodobieństwa te określone są formułą (6) przy czym wartości parametrów β_0 i β_1 przyjęto w każdym eksperymencie arbitralnie. Przeprowadzono trzy eksperymenty symulacyjne. W każdym z nich przeprowadzono 10000-krotną symulację losowania dla liczebności próby $n=40, 80, \dots, 200$.

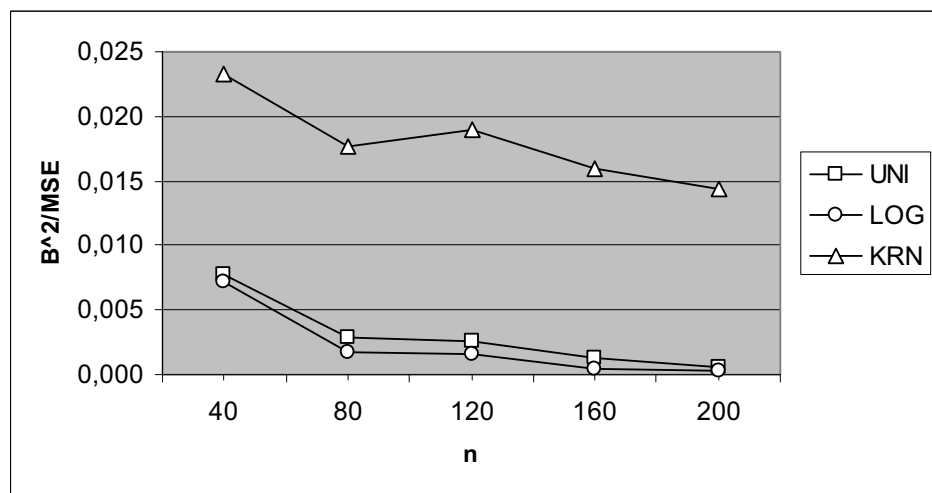
W **pierwszym eksperymencie** zbadano własności proponowanych estymatorów w sytuacji gdy $\beta_0=\beta_1=0$. Do porównania dokładności wykorzystano względny wskaźnik efektywności, obliczany jako stosunek średniego błędu kwadratowego danego estymatora, do średniego błędu kwadratowego estymatora $\hat{C}_{\text{UNI}}(X, Y)$. Względne wskaźniki efektywności estymatorów, ich obciążenia oraz udziały obciążeń w MSE przedstawiono odpowiednio na rys. 1., 2. i 3. Estymatory $\hat{C}_{\text{UNI}}(X, Y)$ i $\hat{C}_{\text{LOG}}(X, Y)$ cechują się zbliżoną dokładnością. Estymator $\hat{C}_{\text{KRN}}(X, Y)$ jest dokładniejszy od dwóch pozostałych. Obciążenie estymatorów jest ujemne. Jego wartość bezwzględna maleje wraz ze wzrostem rozmiaru próby. Obciążenie estymatorów $\hat{C}_{\text{UNI}}(X, Y)$ oraz $\hat{C}_{\text{LOG}}(X, Y)$ jest kilkukrotnie mniejsze niż estymatora $\hat{C}_{\text{KRN}}(X, Y)$, zwłaszcza dla dużych prób. Udział obciążenia w MSE nie przekracza 1% dla $\hat{C}_{\text{UNI}}(X, Y)$ oraz $\hat{C}_{\text{LOG}}(X, Y)$ oraz 2,5% dla $\hat{C}_{\text{KRN}}(X, Y)$ a więc jest znikomy.



Rys. 1. Zależność względnego wskaźnika efektywności (eff) estymatorów kowariancji od liczebności n próby s dla $\beta_0=0$, $\beta_1=0$



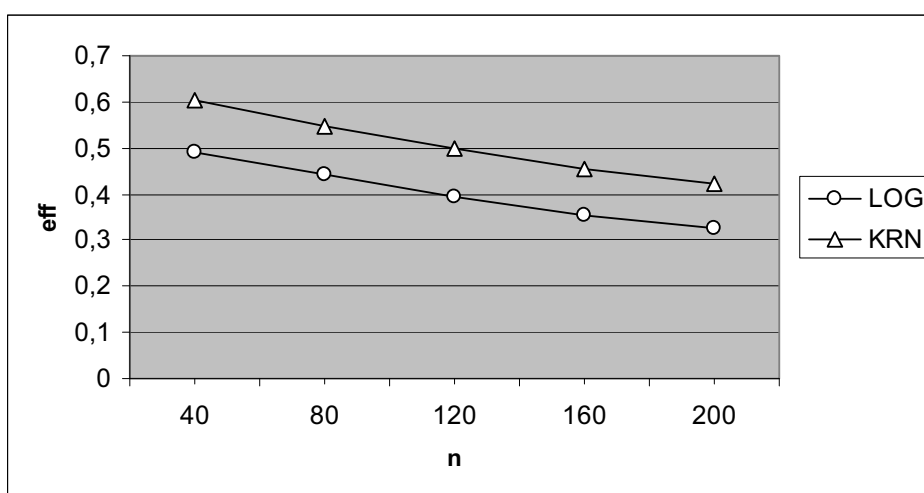
Rys 2. Zależność obciążenia (B) estymatorów kowariancji od liczebności n próby s dla $\beta_0=0$, oraz $\beta_1=0$



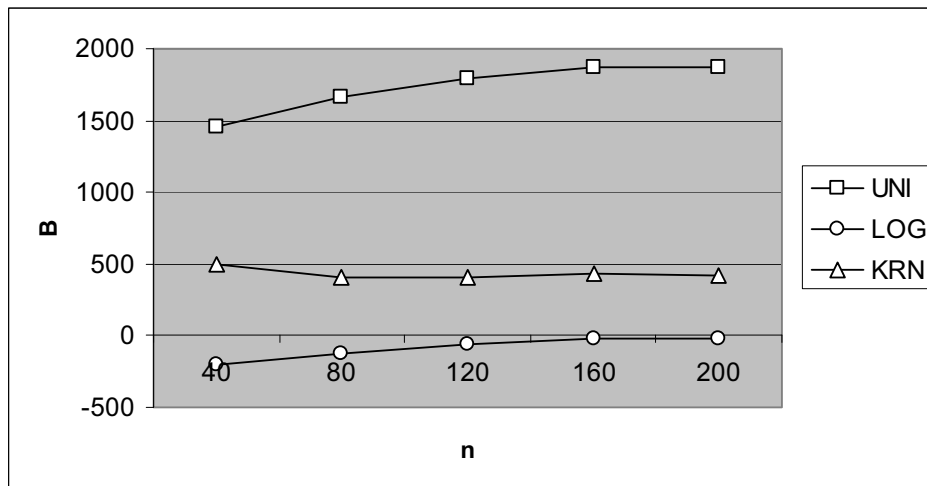
Rys 3. Zależność udziału (B^2/MSE) obciążenia w średnim błędzie kwadratowym estymatorów kowariancji od liczebności n próby s dla $\beta_0=0$, $\beta_1=0$

W **drugim eksperymencie** zbadano własności proponowanych estymatorów kowariancji w sytuacji gdy $\beta_0=0$ oraz $\beta_1=-1$. Względne wskaźniki efektywności estymatorów, ich obciążenia oraz udział obciążeń w MSE przedstawiono na rys. 4. 5. i 6. Estymatory $\hat{C}_{LOG}(X, Y)$ oraz $\hat{C}_{KRN}(X, Y)$ są dokładniejsze od esty-

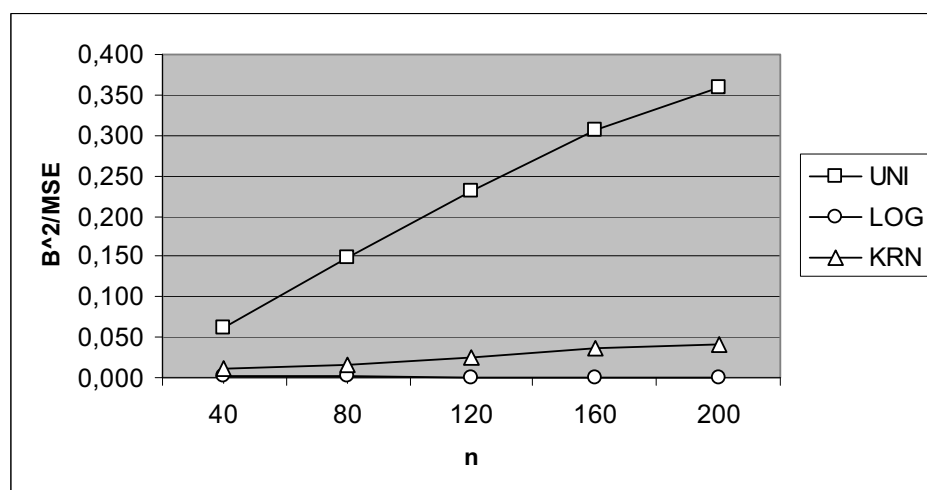
matora $\hat{C}_{\text{UNI}}(X, Y)$ w sensie średniego błędu kwadratowego a ich przewaga wzrasta wraz z rozmiarem próby. Obciążenie estymatora $\hat{C}_{\text{LOG}}(X, Y)$ pozostaje ujemne dla wszystkich badanych rozmiarów próby i co do wartości bezwzględnej maleje wraz z n . Estymator $\hat{C}_{\text{KRN}}(X, Y)$ charakteryzuje się wyraźnie większym co do modułu, dodatnim obciążeniem które wraz ze wzrostem n zmienia się minimalnie i raczej nieregularnie – jego zmiany trudno interpretować jako wyraźną tendencję. Największe co do wartości bezwzględnej, dodatnie obciążenie zarejestrowano dla estymatora $\hat{C}_{\text{UNI}}(X, Y)$. Wrasta ono wraz ze zwiększeniem n . Udział obciążenia w MSE dla estymatora $\hat{C}_{\text{LOG}}(X, Y)$ nie przekracza 1% i maleje wraz ze wzrostem n . Dla estymatora $\hat{C}_{\text{KRN}}(X, Y)$ rośnie wraz ze wzrostem n , utrzymując się w przedziale $\langle 1\%, 5\% \rangle$. Dla estymatora $\hat{C}_{\text{UNI}}(X, Y)$ szybko rośnie przekraczając 35% dla $n=200$.



Rys. 4. Zależność względnego wskaźnika efektywności (eff) estymatorów kowariancji od liczebności n próby s dla $\beta_0=0$, $\beta_1=-1$



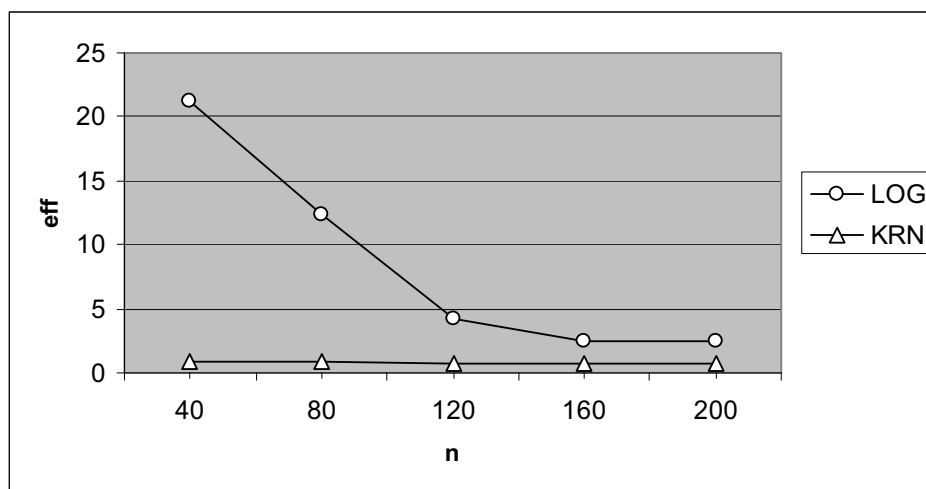
Rys 5. Zależność obciążenia (B) estymatorów kowariancji od liczebności n próby s dla $\beta_0=0$, oraz $\beta_1=-1$



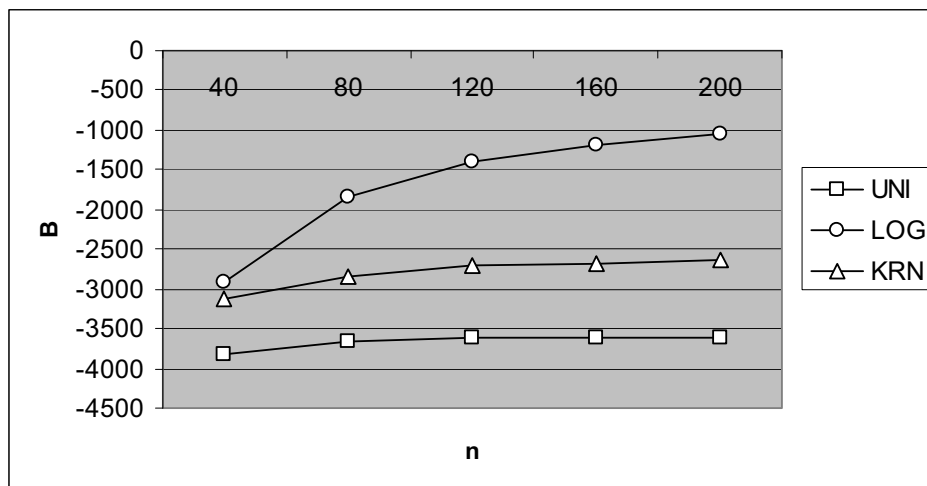
Rys 6. Zależność udziału (B^2/MSE) obciążenia w średnim błędzie kwadratowym estymatorów kowariancji od liczebności n próby s dla $\beta_0=0$, $\beta_1=-1$

W **trzecim eksperymencie** zbadano własności proponowanych estymatorów kowariancji w sytuacji gdy $\beta_0=0$ oraz $\beta_1=1$. Względne wskaźniki efektywności estymatorów, ich obciążenia oraz udziały obciążenia w MSE przedstawiono odpowiednio na rys. 7, 8 oraz 9. Względny wskaźnik efektywności estymatora $\hat{C}_{LOG}(X, Y)$ przyjmuje wartości znacznie większe od jedności co ozna-

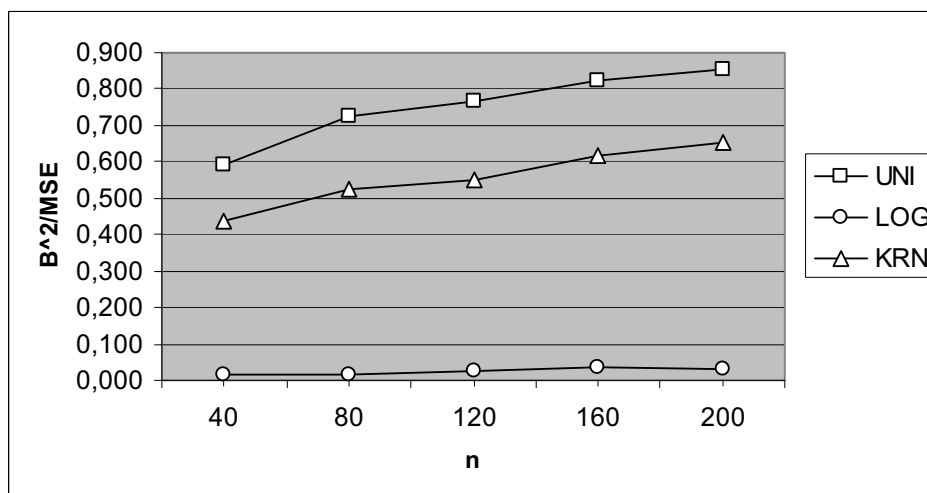
cza że estymator ten jest znacznie mniej dokładny od $\hat{C}_{\text{UNI}}(X, Y)$. Względny wskaźnik efektywności estymatora $\hat{C}_{\text{KRN}}(X, Y)$ utrzymuje się w przedziale $\langle 0.92; 0.69 \rangle$ co oznacza że estymator ten jest bardziej dokładny od estymatora $\hat{C}_{\text{UNI}}(X, Y)$, a przewaga w dokładności rośnie wraz ze wzrostem rozmiaru próby. Obciążenie wszystkich estymatorów jest w tym eksperymencie ujemne i maleje co do modułu wraz ze wzrostem n . Najmniejsze wartości bezwzględne a zarazem najszybszy spadek obciążenia wraz z n zaobserwowano dla estymatora $\hat{C}_{\text{LOG}}(X, Y)$. Największe wartości bezwzględne obciążenia i jego najwolniejszy spadek zaobserwowano dla estymatora $\hat{C}_{\text{UNI}}(X, Y)$. Tak więc, estymator o największym średnim błędzie kwadratowym charakteryzuje się najmniejszym obciążeniem. Udział obciążenia estymatora $\hat{C}_{\text{LOG}}(X, Y)$ w jego średnim błędzie kwadratowym jest pomijalnie mały, i nie przekracza 4%. Dla pozostałych estymatorów udział ten wynosi od czterdziestu do pięćdziesięciu procent i wzrasta wraz ze zwiększaniem rozmiaru próby.



Rys. 7. Zależność względnego wskaźnika efektywności (eff) od n dla $\beta_0=0$, $\beta_1=1$



Rys 8. Zależność obciążenia (B) estymatorów od liczebności n próby s dla $\beta_0=0$, oraz $\beta_1=1$



Rys 9. Zależność udziału (B^2/MSE) obciążenia w średnim błędzie kwadratowym estymatorów od liczebności n próby s dla $\beta_0=0$, $\beta_1=1$

V. PODSUMOWANIE

Wyniki symulacji wskazują, że wykorzystanie dostępnych informacji o wartościach cech pomocniczych może w istotny sposób poprawić własności stochastyczne estymatorów kowariancji. Zarazem jednak trudno wśród badanych estymatorów wskazać taki, który byłby lepszy od pozostałych pod każdym względem. Z punktu widzenia obciążenia zdecydowanie najkorzystniej prezentuje się

estymator oparty o model logistyczny. Jednak jego dokładność mierzona błędem średniokwadratowym zależy wyraźnie od rozkładu odpowiedzi, i w sytuacji gdy skłonność respondentów do odpowiedzi jest negatywnie skorelowana z badanymi cechami jest ona bardzo niska. Gdy można oczekiwać występowania takich zależności godnym polecenia wydaje się estymator kowariancji skonstruowany z wykorzystaniem estymacji jądrowej. Uzyskane wyniki uzasadniają dalsze badanie tych strategii estymacji z wykorzystaniem metod analitycznych, oraz rozszerzenie dociekań na inne parametry populacji dające się przedstawić jako funkcje wartości globalnych.

BIBLIOGRAFIA

- Bethlehem J.G. (1988), Reduction of Nonresponse Bias Through Regression Estimation, *Journal of Official Statistics* Vol 4. No. 3. 251–160.
- Cassel C.M. Särndal C.E. Wretman J.H. (1983), Some Uses of Statistical Models in Connection with the Nonresponse Problem w: *Incomplete Data in Sample Surveys* W.G. Madow I.Olkin (red.) Academic Press New York.
- Giommi A. (1987), Nonparametric Methods for Estimating Individual Response Probabilities, *Survey Methodology*, Vol 13, No 2, 127–134.
- Horvitz D.G. Thompson D.J. (1952), A Generalization of Sampling Without Replacement from a Finite Universe, *Journal of the American Statistical Association* No 47, 663–685.
- Kulczycki P. (2005), *Estymatory jądrowe w analizie systemowej*. WNT Warszawa.
- Lessler J.T. Kalsbeek W.D. (1992), *Nonsampling Error in Surveys* Wiley, New York
- Nargundkar M.S. Joshi G.B. (1975), *Non-response in Sample Surveys* 40-th Session of International Statistical Institute – Warsaw 1975, Contributed Papers, 626–628.
- Rosen J.B. (1969), The gradient projection method for nonlinear programming, *Journal of Society for Industrial and Applied Mathematics*, Vol. 8(1), 181–217.
- Särndal C.E. Swensson B. Wretman J.H. (1992), *Model Assisted Survey Sampling* Springer–Verlag New York.
- Theil H. (1979), *Zasady ekonometrii*, PWN Warszawa.

Wojciech Gamrot

ON THE USE OF WEIGHTING ADJUSTMENTS TO ESTIMATE THE FINITE POPULATION COVARIANCE UNDER NONRESPONSE

Abstract

The weighting adjustment method associates some weight compensating for sample nonrespondents with each responding unit. These weights are usually constructed as reciprocals of individual response probabilities, estimated on the basis of available auxiliary information. In this paper an attempt is made to apply the weighting adjustment method to estimate the complex population parameter, namely the covariance between two population characteristics. A weighting adjustment estimator is proposed. Its properties are examined in a simulation study.