

Grażyna Dehnel

Akademia Ekonomiczna w Poznaniu

WPŁYW KLASYFIKACJI PODMIOTÓW GOSPODARCZYCH NA EFEKTYWNOŚĆ MODELOWANIA PODSTAWOWYCH KATEGORII EKONOMICZNYCH MAŁYCH PRZEDSIĘBIORSTW

1. Wstęp

Mikroprzedsiębiorstwa, w których liczba pracujących nie przekracza 9 osób, stanowią ponad 95% wszystkich przedsiębiorstw działających na terenie naszego kraju.

Od roku 1999 GUS prowadzi coroczne badanie reprezentacyjne SP3 dotyczące tego sektora przedsiębiorstw. W publikacji GUS (por. [GUS 2002]) wyniki badania prezentowane są jedynie w dwóch podziałach: ze względu na rodzaj prowadzonej działalności (sekcje PKD) oraz w podziale na województwa. Należy jednak podkreślić, że województw dotyczą zaledwie trzy tablice. Zawierają one informacje o liczbie podmiotów gospodarczych, liczbie pracujących i zatrudnionych, przeciętnym miesięcznym wynagrodzeniu brutto oraz o przychodach i kosztach.

Tradycyjne metody estymacji stosowane w badaniu działalności gospodarczej mikroprzedsiębiorstw pozwalają na dokonanie uogólnień jedynie w takich przekrojach.

Ze względu na fakt, że zapotrzebowanie na dane na niższym poziomie agregacji dla tzw. małych domen jest coraz większe, prowadzone są eksperymentalne badania, które mają na celu dostarczenie szacunków w przekroju województw i sekcji PKD.

Pionierską próbę estymacji informacji dotyczących mikroprzedsiębiorstw na tym poziomie podjęto w Katedrze Statystyki Akademii Ekonomicznej w Poznaniu we współpracy z poznańskim Urzędem Statystycznym. Zejście na tak niski poziom było możliwe dzięki zastosowaniu estymacji pośredniej oraz połączeniu informacji z trzech różnych źródeł: badania SP3, Bazy Jednostek Statystycznych (BJS), rejestru podatkowego Ministerstwa Finansów.

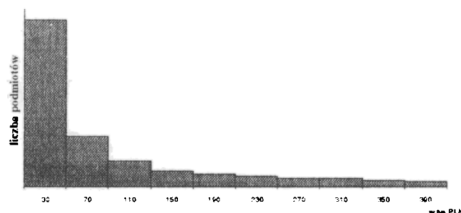
Podstawowe problemy, które napotkano i które w znacznej mierze obciążały otrzymane szacunki, to:

- skrajnie asymetryczne rozkłady badanych zmiennych,
- duża liczba jednostek, dla których badana zmienna przyjmuje wartości zerowe,
- zmienne dodatkowe przybierające wartości skrajne.

Rozkłady sumy wypłacanych wynagrodzeń i przychodów zaprezentowano na wykresach (por. rys. 1, 2).



Rys. 1. Rozkład podmiotów gospodarczych według sumy wypłaconych wynagrodzeń w roku
Źródło: SP3, 2001.



Rys. 2. Rozkład podmiotów gospodarczych według przychodów
Źródło: SP 3, 2001.

Wymienione własności rozkładów zmiennych z zakresu statystyki gospodarczej zostały już wielokrotnie opisane w literaturze (por. [Chambers 1996; Karlberg 2000; Chambers i in. 2001]). Nieustannie poszukuje się nowych rozwiązań prowadzących do otrzymania wiarygodnych szacunków w przekroju małych domen. Jedno z nich, zaproponowane przez Raya Chambersa (por. [Chambers i in. 2001]), polega na wprowadzeniu do estymacji typu GREG dodatkowej zmiennej – z – oraz na przypisaniu szczególnej roli wagom nadawanym obserwacjom.

W niniejszym artykule podjęto próbę empirycznej weryfikacji możliwości wykorzystania modelu Chambersa do szacowania informacji o działalności gospodarczej małych przedsiębiorstw w przekroju województw i sekcji PKD. Jego celem będzie ocena wpływu klasyfikacji podmiotów gospodarczych na efektywności estymacji oraz określenie zasadności zastosowania modelu Chambersa w statystyce gospodarczej.

Realizacja tak określonych zadań pozwoli na weryfikację hipotezy zakładającej, że na efektywność szacunku podstawowych kategorii ekonomicznych małych przedsiębiorstw wpływa m.in. klasyfikacja podmiotów gospodarczych.

2. Estymacja typu GREG

Tradycyjny estymator typu GREG (*Generalised REGression estimator*) jest estymatorem bezpośrednim, uwzględniającym – za pomocą modelu regresji – różnice

między wartościami zmiennych dodatkowych w populacji i w próbie w danej małej domenie.

$$\hat{Y}_{reg} = \sum_{k \in N} \hat{y}_k + \sum_{k \in n} a_k e_k, \quad (1)$$

gdzie: $\hat{y}_k = \mathbf{x}'_k \hat{\beta}$, e_k – reszty, a_k – wagi.

Jednym z założeń modelu regresji liniowej jest to, że składniki losowe mają jednakową wariancję (homoscedastyczność). W przypadku naruszenia tego założenia (heteroscedastyczności) oceny parametrów regresji są nieefektywne. Zastosowanie odpowiedniej transformacji odwrotnościowej może jednak w znacznym stopniu poprawić sytuację.

Model Chambersa, który odnosi się do estymatorów typu GREG, proponuje transformację polegającą na włączeniu dodatkowej zmiennej pomocniczej z .

$$\hat{Y}_{reg} = \sum_{k \in N} \hat{y}_k + \sum_{k \in n} w_k e_k, \quad (2)$$

gdzie: $\hat{y}_k = \mathbf{x}'_k \hat{\beta}$, e_k – reszty, w_k – wagi,

$$\hat{\beta} = \left(\sum_{k \in n} a_k \mathbf{x}_k \mathbf{x}'_k / z_k^\gamma \right)^{-1} \left(\sum_{k \in n} a_k \mathbf{x}_k y_k / z_k^\gamma \right), \quad (3)$$

z_k^γ – zmienna pomocnicza określająca heteroscedastyczność w regresji y względem x ,

γ – parametr określający poziom heteroscedastyczności.

Wzór oznaczony numerem (2) można przedstawić również w postaci:

$$\hat{Y}_{reg} = \sum_{k \in n} a_k g_k y_k, \quad (4)$$

gdzie g_k jest wagą k -tej jednostki zdefiniowaną jako:

$$g_k = 1 + \left(X - \hat{X}_{HT} \right)' \left(\sum_{k \in n} a_k \mathbf{x}_k \mathbf{x}'_k / z_k^\gamma \right)^{-1} \left(\mathbf{x}_k / z_k^\gamma \right), \quad (5)$$

$\hat{X}_{HT}$ – estymator bezpośredni Horvitz-Thompsona wyznaczony dla zmiennej x .

Wagi g_k wyznaczone dla estymatora typu GREG pozostają w silnym związku z miarą odległości $DFBETA$, która pozwala na zidentyfikowanie wartości skrajnych.

$$g_k = 1 + \frac{1}{w} \left(\frac{1 - h_k}{e_k} \right) \left(X - \hat{X}_{HT} \right)' DFBETA_k, \quad (6)$$

gdzie: $h_k = \left(\mathbf{x}_k / z_k^\gamma \right)' \left(\sum_{k \in n} \mathbf{x}_k \mathbf{x}'_k / z_k^\gamma \right)^{-1} \left(\mathbf{x}_k / z_k^\gamma \right)$.

Im wartość parametru $\gamma > 0$, tym silniejsza jest tendencja do rozproszenia punktów od linii wyznaczonej przez funkcję. Z praktyki badań statystycznych wiadomo,

że w przypadku wielu populacji wartość parametru γ zawiera się w przedziale $1 \leq \gamma \leq 2$ (por. [Särndal, Swensson, Wretman 1992, s. 255]).

W przeprowadzonym badaniu przeanalizowano następujące estymatory:

1. $\hat{Y}_{\text{exp}}$ – estymator bezpośredni.
2. $\hat{Y}_{\text{reg}}^0 - \gamma = 0 \Rightarrow z_k^0$ estymator regresyjny oparty na liniowym modelu regresji zakładającym homoscedastyczność.
3. $\hat{Y}_{\text{reg}}^1 - \gamma = 1 \Rightarrow z_k^1$ jw. założenie o heteroscedastyczności.
4. $\hat{Y}_{\text{reg}}^{1.5} - \gamma = 1,5 \Rightarrow z_k^{1.5}$ jw. założenie o heteroscedastyczności.
5. $\hat{Y}_{\text{reg}}^2 - \gamma = 2 \Rightarrow z_k^2$ jw. założenie o heteroscedastyczności.

3. Ocena estymacji

Do wyznaczenia ocen precyzji badanych estymatorów zastosowano jedną ze znanych metod przybliżonych, opartą na podróbkach – metodę bootstrapową. Wykonano 500 iteracji, na podstawie których wyznaczono wartość wariancji badanego parametru

$$\text{Var} = \frac{1}{500-1} \sum_{b=1}^{500} (\hat{Y}_i - \hat{Y})^2. \quad (7)$$

Dla każdej iteracji wyznaczono zmodyfikowane wagi $a_{k(b)} = a_k \frac{n}{n-1} m_{k(b)}$, gdzie:

- $m_{k(b)}$ oznacza, ile razy jednostka i -ta została wybrana do podróbki b ,
- $a_{k(b)}$ – waga dla jednostki k w podpróbce b ; a_k – oryginalna waga jednostki k .

Oceny precyzji danego estymatora dokonano na podstawie dwóch parametrów:

1) współczynnika zmienności estymatora: $CV = \frac{\sqrt{\text{Var}}}{\hat{Y}},$

2) współczynnika wariancji: $VR = \frac{\text{Var}\hat{Y}}{\text{Var}\hat{Y}_{\text{exp}}}.$

4. Schemat badania

W badaniu wykorzystano 3 źródła informacji.

1. **Badanie reprezentacyjne SP3**, które zawierało informacje o zmiennych badanych y . Próba wylosowana do badania SP3 w roku 2001 liczyła ponad 114 tys.

jednostek (4%). Jednak ostatecznie informacje pozyskano jedynie od 44 807 podmiotów gospodarczych.

2. **Bazę Jednostek Statystycznych (BJS)**, zawierającą informacje z Krajowego Rejestru Urzędowego Podmiotów Gospodarki Narodowej REGON. Ogólna liczba małych podmiotów w systemie REGON w 2001 r. wynosiła 2 855 497.

3. **Zbiory danych z systemu podatkowego Ministerstwa Finansów** (rejestr podatkowy) – stanowiło je 907 580 zeznań podatkowych od osób fizycznych i prawnych.

Zarówno BJS, jak i rejestr podatkowy wykorzystano jako źródła cech dodatkowych: x i z , których zadaniem jest poszerzenie informacji uzyskanych z badania SP3.

Próba połączenia trzech zbiorów danych udała się jedynie w ok. 60%. Wykorzystanie tych informacji na poziomie jednostek indywidualnych wymagało więc zmniejszenia próby z SP3 z 44 807 do ok. 27 000 jednostek. Główną przyczyną tak znacznej redukcji próby było przede wszystkim to, że zbiory uzyskane z Ministerstwa Finansów nie zawierały informacji o wszystkich podmiotach gospodarczych. Nie obejmowały one 25% podmiotów prowadzących księgę przychodów i rozchodów oraz 12% podmiotów rozliczających się na podstawie karty podatkowej.

Należy podkreślić, że nieuwzględnienie określonej grupy podmiotów może powodować poważne obciążenia wyników przeprowadzonej analizy. Ze względu jednak na fakt, że badanie miało charakter eksperymentalny, a zmniejszona próba nie traciła własności, którymi charakteryzowała się cała próba SP3, do szacunków wykorzystano próbę zredukowaną.

Estymacji dokonano dla czterech zmiennych (y): sumy wypłaconych w ciągu roku wynagrodzeń, liczby osób zatrudnionych, przychodów i kosztów.

Jako zmienne pomocnicze typu x i z wykorzystano: liczbę osób pracujących (BJS), przychody (rejestry podatkowe) i koszty (rejestry podatkowe), przy czym dana zmienna pomocnicza mogła być wykorzystana jednocześnie jako x i z . Przeanalizowano wiele podejść różniących się składem zmiennych. W doborze zmiennych dodatkowych kierowano się przede wszystkim stopniem skorelowania informacji z badania SP3 oraz rejestru podatkowego i BJS-u. Własność (znana z teorii metody reprezentacyjnej) mówiąca o tym, że silna zależność między cechami znacznie poprawia precyzję estymacji, potwierdziły wyniki przeprowadzonego wcześniej badania symulacyjnego.

Tabela 1. Liczba jednostek w populacji generalnej (PG) i w próbie biorących udział w badaniu w podziale na domeny

Sekcje PKD			Populacja generalna	Próba
Przemysł	D	4	5 994	333
Budownictwo	F	6	4 348	97
Handel	G	7	26 244	587
Hotele	H	8	2 738	31
SUMA			39 324	1 048

Źródło: wyniki badania SP3.

Estymacji dokonano w przekroju województw i rodzaju prowadzonej działalności gospodarczej (sekcja PKD). Wyróżniono 160 domen (16 województw x 10 sekcji PKD), dla każdej szacowano wartości czterech zmiennych, rozważając przy tym różne podejścia. Prezentację wyników w artykule zawężono do jednej zmiennej – sumy wypłat wynagrodzeń – w czterech sekcjach województwa dolnośląskiego (por. tab. 1).

W tabeli 2 zamieszczono wartości współczynnika zmienności estymatora (CV) i współczynnika wariancji (VR) dla wszystkich analizowanych estymatorów w podziale na sekcje PKD w województwie dolnośląskim. Pogrubioną czcionką zaznaczono przykłady sytuacji, w których zastosowanie estymacji typu GREG poprawiło precyzję. Natomiast podkreślone wartości odnoszą się do przypadków, w których estymacja bezpośrednia charakteryzowała się większą precyzją niż estymacja pośrednia.

Tabela 2. Precyzja estymacji sumy wypłacanych wynagrodzeń w podziale na sekcje PKD w województwie dolnośląskim

x	z	$CV(\hat{Y}_{exp})$	$CV(\hat{Y}_{reg}^0)$	$CV(\hat{Y}_{reg}^1)$	$CV(\hat{Y}_{reg}^{1,5})$	$CV(\hat{Y}_{reg}^2)$	$VR(\hat{Y}_{reg}^0)$	$VR(\hat{Y}_{reg}^1)$	$VR(\hat{Y}_{reg}^{1,5})$	$VR(\hat{Y}_{reg}^2)$
Przemysł										
p	p	0,153	0,113	0,105	0,098	0,100	1,0149	1,1161	1,0292	0,9297
p	lp	0,161	0,118	0,123	0,128	0,135	1,0157	1,0428	1,0706	1,1128
lp	lp	0,148	0,141	0,138	0,136	0,137	1,2136	1,1920	1,1796	1,1860
Budownictwo										
p	p	<u>0,236</u>	<u>0,249</u>	<u>0,339</u>	<u>0,642</u>	<u>2,660</u>	1,0240	1,1810	1,8157	4,0401
p	lp	0,242	0,266	0,261	0,260	0,260	1,0668	1,0556	1,0518	1,0510
lp	lp	0,241	0,407	0,369	0,353	0,337	1,2552	1,1888	1,1586	1,1265
Handel										
p	p	<u>0,104</u>	<u>0,086</u>	<u>1,729</u>	<u>-1,377</u>	<u>-1,045</u>	0,8724	6,9087	33,7324	162,167
p	lp	0,099	0,093	0,088	0,087	0,086	0,9890	0,9328	0,9139	0,9011
lp	lp	0,098	0,107	0,099	0,097	0,095	1,1196	1,0365	1,0081	0,9891
Hotele										
p	p	0,363	0,360	0,361	0,376	0,384	1,3951	1,5575	1,6794	1,7546
p	lp	0,386	0,417	0,395	0,404	0,415	1,5182	1,5561	1,7099	1,9128
lp	lp	<u>0,356</u>	<u>1,657</u>	<u>1,132</u>	<u>1,011</u>	<u>0,940</u>	2,4971	2,2941	2,3549	2,4782

p – przychód.

lp – liczba osób pracujących.

Źródło: obliczenia własne.

Otrzymane wyniki wskazują na znaczne zróżnicowanie precyzji estymacji. Nie można jednoznacznie wskazać, który z zaprezentowanych estymatorów charakteryzuje się największą precyzją. W każdym przypadku wymagany jest dobór odpowiedniego modelu i zmiennych. Szczególną uwagę zwracają jednak wyniki w sekcji *Handel*, w podejściu, w którym zarówno zmienną x , jak i z jest przychód. Przyjęte modele w przypadku estymatorów $\hat{Y}_{reg}^{1,5}$ i $\hat{Y}_{reg}^2$ doprowadziły do ujemnych osza-

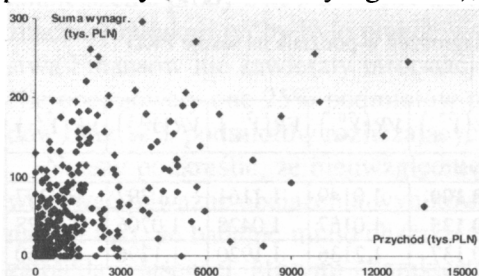
cowań. Biorąc pod uwagę fakt, że zmienną badaną jest suma wypłacanych wynagrodzeń, taki wynik estymacji wymaga głębszej analizy.

Tabela 3. Współczynniki korelacji liniowej pomiędzy wartościami zmiennej badanej i zmiennymi dodatkowymi na podstawie badania SP3, BJS i rejestru podatkowego z 2001 r.

PRZEMYSŁ	Liczba prac.	Przychód	BUDOWNICTWO	Liczba prac.	Przychód
suma wynagrodzeń	0,538	0,479	suma wynagrodzeń	0,587	0,301
HANDEL	liczba prac.	przychód	HOTELE	liczba prac.	przychód
suma wynagrodzeń	0,579	0,517	suma wynagrodzeń	0,536	0,402

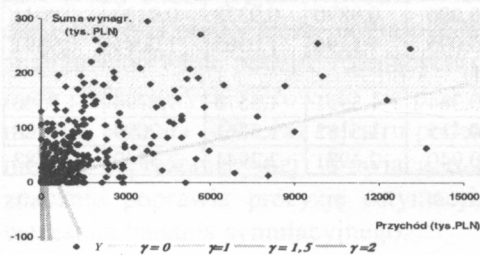
Źródło: obliczenia własne.

Diagram rozrzutu prezentuje strukturę domeny *Handel* x – przychód, y – suma wynagrodzeń (por. rys. 3). Widoczne są na nim obserwacje odstające zarówno w przestrzeni y (duża suma wynagrodzeń), jak i w przestrzeni x (duży przychód).



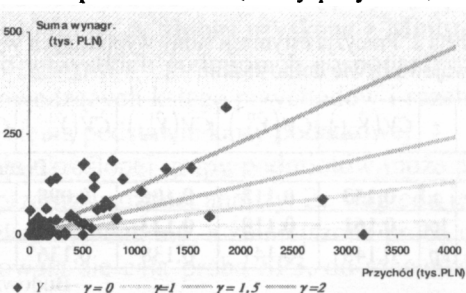
Rys. 3. Diagram rozrzutu przedstawiający zależność pomiędzy sumą wynagrodzeń a przychodem na podstawie badania SP3 – sekcja *Handel*

Źródło: opracowanie własne.



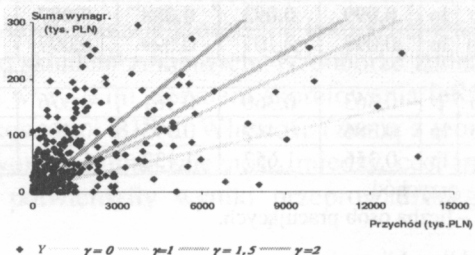
Rys. 5. Linie regresji wyznaczone na podstawie badanych estymatorów $\hat{Y}_{GREG}^{\gamma}$ – sekcja *Handel*

Źródło: opracowanie własne.



Rys. 4. Linie regresji wyznaczone na podstawie badanych estymatorów $\hat{Y}_{GREG}^{\gamma}$ – sekcja *Budownictwo*

Źródło: opracowanie własne.



Rys. 6. Linie regresji wyznaczone na podstawie badanych estymatorów $\hat{Y}_{GREG}^{\gamma}$ po usunięciu obserwacji odstającej

Źródło: opracowanie własne.

Po naniesieniu na diagram rozrzutu linii regresji wyznaczonych na podstawie analizowanych estymatorów układ w przypadku trzech sekcji: *Przemysł*, *Budownictwo* i *Hotele*, był taki sam (por. rys. 4). Dla sekcji *Handel* wykres prezentował się inaczej (por. rys. 5). Jednakże w każdej z czterech analizowanych sekcji można

dostrzec tę samą prawidłowość. Regresja liniowa $\hat{Y}_{GREG}^0$, dla której $\gamma = 0$, w największym stopniu uwzględnia wartości odstające w przestrzeni x . Wraz ze wzrostem parametru γ w coraz większym stopniu brane pod uwagę są wartości odstające w przestrzeni y . Nadawana jest im większa wartość wagi g_k .

Rozkład wag dla sekcji *Handel* w dwóch podejściach przedstawiono w tab. 4 i 5. W przypadku modeli $\gamma = 0$ i $\gamma = 1$ wartości odstające w przestrzeni x otrzymały jedno z najwyższych wag w porównaniu z ich wartościami skrajnymi. Wraz ze wzrostem parametru γ wartości wag przenoszą ciężar estymacji z obserwacji skrajnych (x) na obserwacje skrajne (y). Parametr γ wpływa również na zakres zmienności wag. Im jego wartość jest większa, tym większa zmienność wag.

Tabela 4. Rozkład wartości wag w sekcji *Handel*, x – liczba osób pracujących, z – liczba osób pracujących

Handel x – l. osób prac. z – l. osób prac.	Min.	Max	Waga dla obserwacji odstającej (x)	Średnia	Odchylenie standardowe
$g_k 0$	-2,594	18,280	1,605	1,080	0,539
$g_k 1$	-2,289	14,631	2,040	1,068	0,465
$g_k 1,5$	-3,098	16,022	2,243	1,062	0,477
$g_k 2$	-4,015	24,801	2,421	1,057	0,533

Źródło: obliczenia własne.

Tabela 5. Rozkład wartości wag w sekcji *Handel*, x – przychód, z – przychód

Handel x – przychód z – przychód	Min.	Max	Waga dla obserwacji odstającej (x)	Średnia	Odchylenie standardowe
$g_k 0$	-2,54	59,55	0,98	1,05	0,65
$g_k 1$	-595,23	35,15	-233,78	0,96	5,38
$g_k 1,5$	-4040,69	6744,87	-1236,96	0,46	61,82
$g_k 2$	-30199,35	88315,93	-5821,17	-2,23	668,29

Źródło: obliczenia własne.

Tabela 6. Precyzja estymacji sumy wypłacanych wynagrodzeń w sekcji *Handel* w województwie dolnośląskim

$CV(\hat{Y}_{exp})$	$CV(\hat{Y}_{reg}^0)$	$CV(\hat{Y}_{reg}^1)$	$CV(\hat{Y}_{reg}^{1,5})$	$CV(\hat{Y}_{reg}^2)$	$VR(\hat{Y}_{reg}^0)$	$VR(\hat{Y}_{reg}^1)$	$VR(\hat{Y}_{reg}^{1,5})$	$VR(\hat{Y}_{reg}^2)$
0,104	0,086	1,729	-1,377	-1,045	0,8724	6,9087	33,7324	162,1675
Po zastąpieniu obserwacji odstającej wartością mediany								
0,097	0,089	0,088	0,093	0,126	0,9635	1,0119	1,0895	1,5613
Postratyfikacja								
0,1063	0,1104	0,1111	0,1186	0,1753	1,1393	1,3669	1,5692	2,5939

Źródło: obliczenia własne.

Otrzymanie ujemnych szacunków dla sekcji *Handel* spowodowane było tym, że pojawiły się obserwacje odstające, w których wartość wypłacanych wynagrodzeń była bardzo duża (przypisano im w modelu dużą wagę), wartość przychodów zaś była nieporównywalnie mała. Do identyfikacji obserwacji odstających wykorzystano miarę odległości *DFBETA*.

Problemy wynikające z niejednorodności zbiorowości próbowano rozwiązać na dwa sposoby:

- zastępując obserwacje odstające wartością mediany,
- poprzez postratyfikację.

Na podstawie sumy wypłacanych wynagrodzeń dokonano klasyfikacji podmiotów gospodarczych. Wyróżniono dwie warstwy, a następnie przeprowadzono szacunek, niezależnie dla każdej z wyróżnionych warstw.

W przypadku pierwszej metody usunięcie jednej obserwacji i zastąpienie jej wartością mediany w znacznym stopniu wpłynęło na zmianę modelu (por. rys. 5 i 6). Podobna zmiana nastąpiła po zastosowaniu postratyfikacji. Każda z metod wpłynęła na znaczną poprawę precyzji szacunku (por. tab. 6).

5. Wnioski

1. Zastosowanie estymatorów typu GREG prowadzi do bardzo zróżnicowanych wyników.

2. Wprowadzenie dodatkowej zmiennej z pozwala na przeniesienie problemu estymacji z wartości skrajnych zmiennej pomocniczej na wartości skrajne zmiennej badanej.

3. Znaczną poprawę efektywności estymacji można osiągnąć przez: dobór odpowiedniego modelu, dobór odpowiednich zmiennych dodatkowych, postratyfikację na podstawie jednej ze zmiennych badanych, polepszenie jakości danych.

4. Model Chambersa znajduje szczególne zastosowanie w przypadku statystyki gospodarczej. Należy jednak podkreślić, że może być wykorzystany także w innych dziedzinach, w których problemem jest niejednorodność rozkładów zmiennych takich, jak np. demografia (por. [Gołata 2007]).

Literatura

- Chambers R.L. (1996), *Robust Case-Weighting for Multipurpose Establishment Surveys*, „Journal of Official Statistics” vol. 12, nr 1, s. 3-32.
- Chambers R.L., Falvey H., Hedlin D., Kokic P. (2001), *Does the Model Matter for GREG Estimation? A Business Survey Example*, „Journal of Official Statistics” vol. 17, nr 4, s. 527-544.

- Dehnel G., Gołata E. (2006), *Attempts to Estimate Basic Information for Small Business in Poland*, „Statistics in Transition” vol. 7, nr 4, March, s. 785-803.
- Gołata E. (2007), *Zastosowanie estymacji pośredniej w analizie zróżnicowania struktury gospodarstw domowych w Polsce według klasy miejsca zamieszkania*, *Taksonomia 14*, AE, Wrocław.
- GUS (2002), *Działalność gospodarcza przedsiębiorstw o liczbie pracujących do 9 osób w 2001 r.*, Warszawa.
- Karlberg F. (2000), *Survey Estimation for Highly Skewed Populations in the Presence of Zeroes*, „Journal of Official Statistics” vol. 16, nr 1, s. 229-241.
- Klimanek, T., Paradysz, J., *Adaptation of EURAREA Experience in Business Statistics in Poland*, „Statistics in Transition” vol. 7, nr 4, March, s. 841-852.
- Särndal C.E., Swensson B., Wretman J., (1992), *Model Assisted Survey Sampling*, Springer Verlag, New York.

EFFECT OF SMALL BUSINESS CLASSIFICATION ON THE ESTIMATION PRECISION OF BASIC ECONOMIC INFORMATION

Summary

According to Central Statistical Office there were about 1,6 million small (called *small business*) in Poland in 2001 – over 95% of all enterprises. Although the small economic entities constitute the basis of local economy there is no information about this group of enterprises available at regional and local level. One of the major problems involved in estimating information about economic activity across domains is the distribution of small companies variables. The study variable may be highly skewed, there may be a large proportion of zero responses and there may be several auxiliary variables that can be used to improve estimation but these may include some extreme values. This article presents Chamber's Model which examines the behaviour of GREG estimators taking into account data classification. It shows how an efficient Chambers Model was found for a business survey.