

Grażyna Dehnel

Akademia Ekonomiczna w Poznaniu

OCENA CHARAKTERYSTYK EKONOMICZNYCH MAŁYCH PODMIOTÓW Z WYKORZYSTANIEM ESTYMACJI BAYESOWSKIEJ

1. Cel opracowania

Od blisko 2 lat CSR we współpracy z Urzędem Skarbowym w Poznaniu prowadzi badania eksperymentalne nad możliwością wykorzystania danych ze źródeł administracyjnych w szacowaniu informacji o działalności gospodarczej małych przedsiębiorstw.

W tym czasie zmieniano koncepcję badania, stosowano różne metody estymacji, otrzymywano mniej lub bardziej zadowalające oceny i charakterystyki precyzji. Przez cały ten okres niezmiennie pozostawało przekonanie o konieczności wykorzystania rejestrów dla poprawy precyzji. Przedstawione poniżej wyniki analizy stanowią zatem kontynuację prowadzonych wcześniej prac mających charakter eksperymentalny (por. [3; 5]).

Analiza dotyczy roku 2001. Wykorzystano w niej udostępnione przez GUS wyniki badania reprezentacyjnego SP3, którym objęta została 5-procentowa próba podmiotów gospodarczych o liczbie pracujących do 9 osób (por. [4]). W badaniu tym zastosowano schemat losowania warstwowego. Warstwami były zbiorowości utworzone na podstawie dwóch kryteriów: lokalizacji podmiotu na poziomie województwa i rodzaju działalności prowadzonej przez niego według PKD. Niezależnie od tego podziału dodatkowo zostały wydzielone warstwy ze względu na podstawową formę prawną podmiotu oraz rozmiar przedsiębiorstwa.

Jako dodatkowe źródła informacji posłużyły: rejestr podatkowy Ministerstwa Finansów oraz Baza Jednostek Statystycznych GUS (zwane dalej odpowiednio rejestrem podatkowym i BJS)¹.

¹ Charakterystyka wyróżnionych źródeł informacji zawarta jest w artykule [3].

Poruszone w artykule zagadnienia stanowią realizację 3 zadań. Określono je w następujący sposób:

- próba empirycznej weryfikacji możliwości zastosowania metod empirycznej estymacji bayesowskiej do szacowania informacji o działalności gospodarczej małych przedsiębiorstw w przekroju województw i sekcji PKD,
- ocena efektywności empirycznej estymacji bayesowskiej,
- określenie zasadności zastosowania estymacji bayesowskiej w porównaniu z estymacją bezpośrednią i syntetyczną regresyjną w przekroju województw i sekcji PKD.

Szacunek przeprowadzono dla czterech podstawowych wielkości charakteryzujących działalność gospodarczą: liczby zatrudnionych na podstawie umowy o pracę, przeciętnego miesięcznego wynagrodzenia, średnich przychodów na 1 zakład i średnich kosztów na 1 zakład.

Ze względu na dużą niejednorodność zbiorowości, analizy dokonano nie na poziomie jednostkowym, a na poziomie wyróżnionych domen, tzn. województw, sekcji oraz województw i sekcji łącznie. Jako zmienne objaśniające przyjęto wartości średnie cech dodatkowych dla domen.

Estymacji dokonano dla następujących przekrojów:

- województw, tj. 16 domen oznaczonych symbolem Id2_W,
- rodzaju działalności gospodarczej w rozumieniu sekcji PKD, tj. 11² domen oznaczonych symbolem Id2_S,
- województw i sekcji PKD łącznie, tj. 176 domen oznaczonych symbolem Id_3.

Z uwagi na ograniczone ramy niniejszego opracowania przedstawione zostaną szacunki jedynie wybranych zmiennych. Są to średnia liczba zatrudnionych na 1 zakład oraz przeciętna wielkość przychodu na 1 zakład. Wyniki zaprezentowano w dwóch przekrojach Id_2, tj. dla województw (Id2_W) i dla sekcji (Id2_S) oraz w dwóch przekrojach Id_3, tj. dla sekcji przetwórstwo przemysłowe (D) oraz handel i naprawy (G). Pozostałe szacunki przedstawione zostały w raporcie sporządzonym dla GUS i są do wglądu w Centrum Statystyki Regionalnej.

2. Empiryczna estymacja bayesowska

W przeprowadzonym badaniu zastosowano estymację bayesowską, wzorując się na opracowaniach opublikowanych przez C. Brachę [1] oraz C. Brachę, B. Lednickiego i R. Wieczorkowskiego [2, s. 32-36].

Estymator bayesowski należy do grupy estymatorów złożonych, których wartość jest średnią ważoną z dwóch estymatorów: bezpośredniego i syntetycznego. Taka kombinacja ma na celu zbilansowanie obciążenia estymatora syntetycznego z niestabilnością estymatora bezpośredniego. Udział każdego z estymatorów określony jest wielkością wagi.

² Ze względu na małą liczbę podmiotów, w analizie nie uwzględniono sekcji A, B, C, E. Ujęto je w grupie „pozostałe”.

$$\hat{Y}_{d(COM)} = \gamma_d \hat{Y}_{d(DIR)} + (1 - \gamma_d) \hat{Y}_{d(SYN,REG)}, \quad (1)$$

gdzie: γ_d , $0 \leq \gamma_d \leq 1$ jest wagą z przedziału $(0; 1)$ minimalizującą $MSE(\hat{Y}_{d(COM)})$,

$d = 1 \dots D$ – domena,

$\hat{Y}_{d(DIR)}$ – estymator bezpośredni szacowanego parametru z badania SP3,

$\hat{Y}_{d(SYN,REG)}$ – syntetyczny estymator regresyjny wartości średniej szacowanej zmiennej w domenie d .

$$\hat{Y}_{d(SYN,REG)} = \bar{\mathbf{x}}'_d \hat{\boldsymbol{\beta}},$$

gdzie: $\hat{\boldsymbol{\beta}}$ – wektor współczynników regresji drugiego rodzaju, otrzymany metodą najmniejszych kwadratów,

$\bar{\mathbf{x}}_d$ – wartość średnia cechy objaśniającej dla domeny d wyznaczona na podstawie dodatkowego źródła informacji (rejestrów podatkowych bądź BJS).

Dla każdej z wyróżnionych w badaniu domen obliczono kombinację liniową dwóch składników: wartości estymatora klasycznego i wartości teoretycznej uzyskanej z modelu, ważąc udział każdego z nich. W przypadku estymacji bayesowskiej idea wyznaczania wag sprowadza się do minimalizowania średniego błędu kwadratowego estymatora złożonego $\hat{Y}_{d(COM)}$, przy założeniu, że $cov(\hat{Y}_{d(DIR)}, \hat{Y}_{d(SYN,REG)}) = 0$.

Wagę γ_d wyznaczono według wzoru (por. [8, s. 179; 6]):

$$\gamma_d(opt) = \frac{V(\hat{Y}_{d(SYN,REG)})}{MSE(\hat{Y}_{d(DIR)}) + V(\hat{Y}_{d(SYN,REG)})}, \quad (2)$$

gdzie: $MSE(\hat{Y}_{d(DIR)})$ – błąd średniokwadratowy estymatora bezpośredniego $\hat{Y}_{d(DIR)}$,

$V(\hat{Y}_{d(SYN,REG)})$ – wariancja estymatora syntetycznego regresyjnego.

Oszacowanie wariancji estymatora złożonego przeprowadzono na podstawie wariancji komponentów:

$$V(\hat{Y}_{d(COM)}) = \frac{V(\hat{Y}_{d(DIR)}) \cdot V(\hat{Y}_{d(SYN,REG)})}{V(\hat{Y}_{d(DIR)}) + V(\hat{Y}_{d(SYN,REG)})}. \quad (3)$$

Podstawiając w miejsce nieznanymi wariancji ich estymatory, otrzymujemy:

$$\hat{V}(\hat{Y}_{d(COM)}) = \frac{\hat{V}(\hat{Y}_{d(DIR)}) \cdot \hat{V}(\hat{Y}_{d(SYN,REG)})}{\hat{V}(\hat{Y}_{d(DIR)}) + \hat{V}(\hat{Y}_{d(SYN,REG)})}. \quad (4)$$

W ocenie jakości estymacji pośredniej wykorzystano względny błąd szacunku

$$REE = \frac{\sqrt{M\hat{S}E}}{\hat{Y}_d}. \quad (5)$$

Wyznaczono również wartość zaproponowanego przez L. Kisha współczynnika *deff* jako względnej miary efektywności. Jest on zdefiniowany następującym wzorem:

$$deff(\hat{Y}_{COM}) = \frac{M\hat{S}E(Y_{COM})}{\hat{V}(Y_{DIR})}. \quad (6)$$

Współczynnik ten pozwala na porównanie precyzji zastosowanej metody estymacji z efektywnością estymatora bezpośredniego. Wartość współczynnika *deff* mniejsza od jedności oznacza, że zastosowana estymacja bayesowska jest bardziej efektywna od estymacji bezpośredniej.

3. Wyniki estymacji podstawowych informacji o działalności gospodarczej małych przedsiębiorstw

Poniżej zamieszczono wyniki szacunku wybranych informacji o działalności gospodarczej małych przedsiębiorstw w przekroju województw, sekcji PKD oraz sekcji i województw zgodnie z opisaną w poprzednim punkcie metodą. Tam, gdzie było to możliwe, dla porównania dołączono szacunki publikowane przez GUS. Ponadto w tablicach dotyczących przekroju według województw lub sekcji zamieszczono informację o liczebności próby.

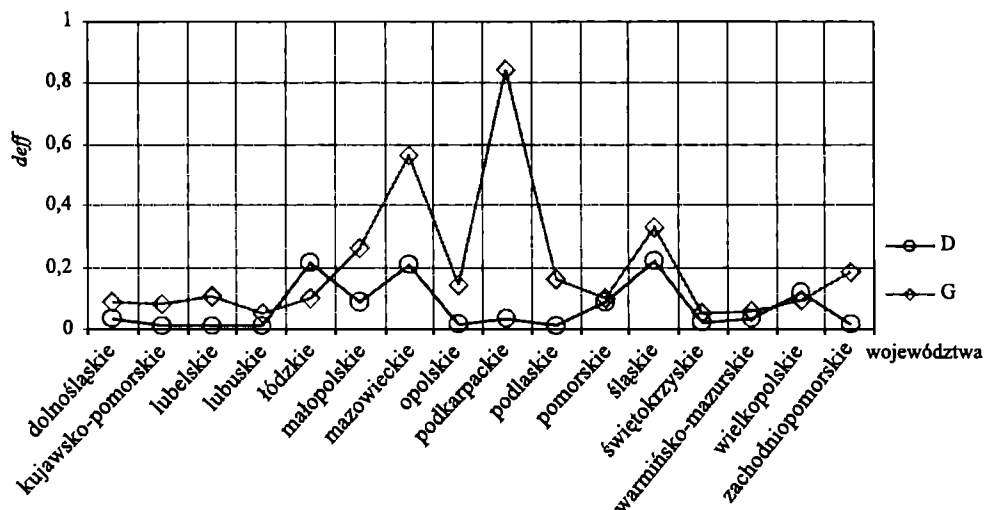
Wartości estymatora bayesowskiego charakteryzują się mniejszym zróżnicowaniem niż estymacja bezpośrednia. Estymator syntetyczny zaś, zgodnie z ideą jego konstrukcji, uśrednia szacowaną wielkość na poziomie przeciętnym dla domeny na wyższym poziomie agregacji. Z kolei estymator bayesowski przybliża szacowaną wartość do oceny otrzymanej metodą klasyczną.

Oceniając precyzję, można zauważyć, że względne błędy szacunku w przypadku estymacji bezpośredniej są zdecydowanie bardziej zróżnicowane w porównaniu z estymacją bayesowską.

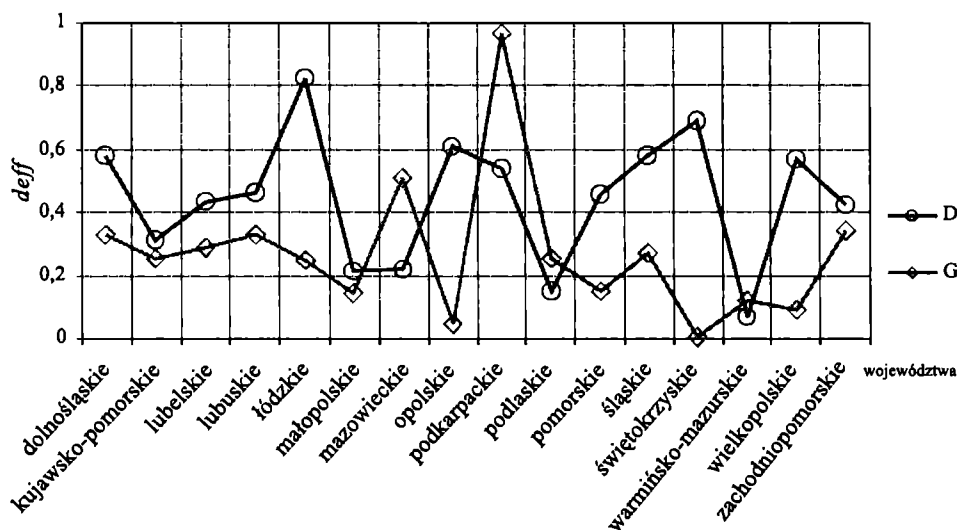
Największy zysk na efektywności pojawia się w przypadku najmniejszych wyróżnionych domen, tj. w przekroju Id_3. Należy jednak podkreślić, że jest on bardziej istotny w przypadku liczby zatrudnionych niż w przypadku przychodu.

Porównując wartości względnych błędów szacunku w przekroju Id_2, czyli sekcji lub województw, można wskazać, że większa redukcja błędu ma miejsce dla województw niż dla sekcji. Może to wynikać z tego, że dla sekcji wartości estymatora syntetycznego są bardziej obciążone. Zróżnicowanie pomiędzy sekcjami jest większe niż pomiędzy województwami. Zgodnie z założeniami estymacji syntetycznej wpływa to na znaczne obciążenie szacunków, a w konsekwencji zmniejsza precyzję estymacji złożonej.

Jak pokazują wyniki badania największa redukcja wartości błędu w przypadku estymacji złożonej ma miejsce wówczas, gdy dla estymacji bezpośredniej przyjmuje on duże wartości. Jeśli zaś *REE* jest stosunkowo mały dla estymacji bezpośredniej, to po zastosowaniu estymatora bayesowskiego zmniejsza się on nieznacznie.



Rys. 1. Wartość *deff* – dla przeciętnej liczby osób zatrudnionych w przeliczeniu na jeden zakład w województwach w 2001 r. w przekroju sekcji przetwórstwo przemysłowe (D) i handel (G)
Źródło: opracowanie własne.



Rys. 2. Wartość *deff* – dla przychodów przeliczonych na 1 zakład w województwach w 2001 r. w przekroju sekcji przetwórstwo przemysłowe (D) i handel (G)
Źródło: opracowanie własne.

Tabela 1. Estymacja przeciętnej liczby zatrudnionych na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw

Województwo	Estymator			REE		
	<i>DIR</i>	<i>SYN, REG</i>	<i>COM</i>	<i>DIR</i>	<i>SYN, REG</i>	<i>COM</i>
Dolnośląskie	0,62	0,67	0,67	0,059	0,023	0,021
Kujawsko-pomorskie	0,61	0,64	0,63	0,088	0,023	0,022
Lubelskie	0,57	0,58	0,58	0,087	0,043	0,039
Lubuskie	0,69	0,6	0,61	0,098	0,043	0,040
Łódzkie	0,61	0,69	0,67	0,057	0,037	0,031
Małopolskie	0,63	0,65	0,64	0,048	0,035	0,028
Mazowieckie	0,76	0,76	0,76	0,040	0,060	0,033
Opolskie	0,75	0,69	0,69	0,087	0,030	0,029
Podkarpackie	0,51	0,51	0,51	0,067	0,089	0,053
Podlaskie	0,55	0,60	0,59	0,104	0,050	0,045
Pomorskie	0,75	0,74	0,75	0,062	0,040	0,034
Śląskie	0,67	0,68	0,67	0,049	0,047	0,034
Świętokrzyskie	0,59	0,60	0,6	0,120	0,049	0,045
Warmińsko-mazurskie	0,69	0,68	0,68	0,075	0,045	0,039
Wielkopolskie	0,76	0,67	0,68	0,063	0,025	0,023
Zachodniopomorskie	0,62	0,63	0,63	0,073	0,031	0,028

Tabela 2. Estymacja przeciętnej liczby zatrudnionych na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla wybranych sekcji PKD

Sekcja	Estymator			REE		
	<i>DIR</i>	<i>SYN, REG</i>	<i>COM</i>	<i>DIR</i>	<i>SYN, REG</i>	<i>COM</i>
Przetwórstwo przemysłowe	1,12	0,99	1,07	0,034	0,048	0,028
Budownictwo	0,72	0,86	0,79	0,057	0,049	0,037
Handel i naprawy	0,75	0,76	0,75	0,026	0,089	0,025
Hotele i restauracje	0,9	0,91	0,91	0,088	0,069	0,054
Transport, gospodarka magazynowa i łączność	0,38	0,47	0,4	0,074	0,113	0,062
Pośrednictwo finansowe	0,18	0,14	0,18	0,086	0,448	0,085
Obsługa nieruchomości i firm, nauka	0,51	0,44	0,5	0,054	0,138	0,051
Edukacja	0,21	0,29	0,23	0,123	0,168	0,100
Ochrona zdrowia i opieka społeczna	0,28	0,25	0,27	0,121	0,200	0,104
Pozostała działalność usługowa komunalna, społeczna i indywidualna	0,48	0,43	0,46	0,089	0,112	0,069

Tabela 3. Estymacja przeciętnej liczby zatrudnionych na 1 zakład na podstawie wyników badań SP3 z 2001 r. z wykorzystaniem informacji z BJS i Ministerstwa Finansów – przekrój województw i sekcji PKD – Przetwórstwo przemysłowe

Województwo	Przetwórstwo przemysłowe (sekcja D)					
	Estymator			REE		
	DIR	SYN, REG	COM	DIR	SYN, REG	COM
Dolnośląskie	1,12	0,99	0,99	0,140	0,034	0,033
Kujawsko-pomorskie	1,21	0,90	0,91	0,243	0,033	0,033
Lubelskie	1,47	0,94	0,95	0,205	0,033	0,033
Lubuskie	1,59	1,02	1,03	0,241	0,037	0,037
Łódzkie	1,11	1,15	1,14	0,078	0,041	0,036
Małopolskie	0,96	0,98	0,98	0,106	0,037	0,035
Mazowieckie	1,13	1,03	1,06	0,078	0,047	0,040
Opolskie	1,45	1,12	1,13	0,193	0,038	0,038
Podkarpackie	1,14	0,94	0,95	0,143	0,032	0,032
Podlaskie	1,14	0,95	0,96	0,230	0,034	0,034
Pomorskie	0,89	0,99	0,98	0,114	0,033	0,032
Śląskie	1,07	1,14	1,12	0,079	0,039	0,035
Świętokrzyskie	1,29	1,04	1,05	0,190	0,040	0,039
Warmińsko-mazurskie	0,99	0,97	0,97	0,163	0,033	0,032
Wielkopolskie	1,13	1,08	1,09	0,102	0,037	0,035
Zachodniopomorskie	1,25	0,99	0,99	0,179	0,033	0,033

Tabela 4. Estymacja przeciętnej liczby zatrudnionych na 1 zakład na podstawie wyników badań SP3 z 2001 r. z wykorzystaniem informacji z BJS i Ministerstwa Finansów – przekrój województw i sekcji PKD – Handel

Województwo	Handel (sekcja G)					
	Estymator			REE		
	DIR	SYN, REG	COM	DIR	SYN, REG	COM
Dolnośląskie	0,75	0,65	0,66	0,088	0,041	0,037
Kujawsko-pomorskie	0,59	0,60	0,60	0,116	0,044	0,041
Lubelskie	0,58	0,57	0,57	0,106	0,040	0,038
Lubuskie	0,78	0,58	0,59	0,112	0,043	0,040
Łódzkie	0,66	0,62	0,63	0,094	0,038	0,035
Małopolskie	0,80	0,64	0,67	0,073	0,045	0,038
Mazowieckie	0,82	0,64	0,74	0,065	0,084	0,052
Opolskie	0,76	0,68	0,69	0,104	0,040	0,038
Podkarpackie	0,60	1,00	0,63	0,091	0,201	0,084
Podlaskie	0,61	0,62	0,62	0,123	0,045	0,042
Pomorskie	0,90	0,72	0,74	0,106	0,042	0,039
Śląskie	0,77	0,74	0,75	0,068	0,038	0,033
Świętokrzyskie	0,61	0,60	0,60	0,138	0,042	0,040
Warmińsko-mazurskie	0,81	0,60	0,62	0,102	0,039	0,036
Wielkopolskie	0,94	0,68	0,70	0,102	0,043	0,040
Zachodniopomorskie	0,63	0,68	0,67	0,111	0,043	0,040

Tabela 5. Estymacja przychodów na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw

Województwo	Liczebność próby	Szacunek GUS	Estymator			REE		
			DIR	SYN, REG	COM	DIR	SYN, REG	COM
Dolnośląskie	3 190	240,7	240,7	258,7	250,8	0,06	0,05	0,04
Kujawsko-pomorskie	1 546	197,3	197,3	252,9	232,7	0,08	0,05	0,04
Lubelskie	2 244	278,2	278,2	261,1	264,3	0,16	0,08	0,07
Lubuskie	1 398	217,8	217,8	235,5	225,3	0,09	0,09	0,06
Łódzkie	2 975	221,3	221,3	268,4	237,1	0,07	0,08	0,05
Małopolskie	4 126	282,5	282,5	288,0	285,8	0,08	0,07	0,05
Mazowieckie	6 926	373,5	373,5	357,0	369,7	0,06	0,11	0,05
Opolskie	1 559	339,0	339,0	283,1	287,0	0,19	0,06	0,06
Podkarpackie	2 755	253,8	253,8	247,9	251,8	0,11	0,16	0,09
Podlaskie	1 526	224,3	224,3	257,8	235,6	0,08	0,10	0,06
Pomorskie	3 071	282,0	282,0	272,6	277,8	0,08	0,09	0,06
Śląskie	5 234	294,4	294,4	338,9	308,3	0,06	0,08	0,05
Świętokrzyskie	1 449	376,1	376,1	286,2	289,0	0,37	0,09	0,08
Warmińsko-mazurskie	1 468	257,4	257,4	207,8	228,9	0,12	0,12	0,09
Wielkopolskie	3 346	295,0	295,0	292,9	293,3	0,10	0,05	0,04
Zachodniopomorskie	1 994	221,4	221,4	246,0	232,3	0,06	0,07	0,05

Tabela 6. Estymacja przychodów na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla wybranych sekcji PKD

Sekcja	Liczebność próby	Szacunek GUS	Estymator			REE		
			DIR	SYN, REG	COM	DIR	SYN, REG	COM
Przetwórstwo przemysłowe	8 674	236,50	248,1	190,3	238,0	0,06	0,17	0,06
Budownictwo	3 443	169,20	174,7	168,4	174,0	0,06	0,17	0,06
Handel i naprawy	15 435	454,30	442,5	486,1	447,7	0,04	0,10	0,04
Hotele i restauracje	803	116,70	125,3	96,7	123,7	0,08	0,44	0,08
Transport, gospodarka magazynowa i łączność	2 600	172,60	168,8	221,5	174,2	0,07	0,16	0,07
Pośrednictwo finansowe	3 222	543,00	556,8	482,4	497,9	0,15	0,09	0,08
Obsługa nieruchomości i firm, nauka	4 439	159,40	163,8	91,4	161,1	0,05	0,45	0,05
Edukacja	955	78,40	80,4	113,3	82,6	0,11	0,29	0,10
Ochrona zdrowia i opieka społeczna	2 183	66,00	66,0	97,8	66,9	0,08	0,35	0,08
Pozostała działalność usługowa komunalna, społeczna i indywidualna	2 071	88,60	87,8	90,8	88,1	0,12	0,36	0,11

Tabela 7. Estymacja przychodów na 1 zakład na podstawie wyników badań SP3 z 2001 r.
z wykorzystaniem informacji z BJS i Ministerstwa Finansów – przekrój województw i sekcji PKD –
Przetwórstwo przemysłowe

Województwo	Przetwórstwo przemysłowe (sekcja D)					
	Estymator			REE		
	DIR	SYN, REG	COM	DIR	SYN, REG	COM
Dolnośląskie	203,5	174,1	191,2	0,118	0,163	0,096
Kujawsko-pomorskie	220,2	190,1	199,5	0,167	0,131	0,103
Lubelskie	179,2	152,3	164,1	0,166	0,172	0,120
Lubuskie	259,6	140,8	196,0	0,131	0,225	0,118
Łódzkie	163,8	119,9	156,2	0,110	0,329	0,105
Małopolskie	249,9	166,8	184,7	0,229	0,180	0,144
Mazowieckie	389,2	229,3	264,8	0,195	0,177	0,135
Opolskie	205,7	157,3	186,9	0,138	0,227	0,119
Podkarpackie	163,9	138,3	152,1	0,143	0,183	0,113
Podlaskie	263,0	149,4	166,8	0,244	0,183	0,151
Pomorskie	231,5	185,7	206,9	0,127	0,147	0,097
Śląskie	268,0	135,8	212,3	0,117	0,271	0,113
Świętokrzyskie	168,7	139,5	159,8	0,138	0,251	0,121
Warmińsko-mazurskie	310,4	164,3	174,8	0,309	0,162	0,147
Wielkopolskie	217,7	206,6	212,9	0,132	0,160	0,102
Zachodniopomorskie	226,0	168,1	192,6	0,141	0,162	0,108

Tabela 8. Estymacja przychodów na 1 zakład na podstawie wyników badań SP3 z 2001 r.
z wykorzystaniem informacji z BJS i Ministerstwa Finansów – przekrój województw i sekcji PKD – Handel

Województwo	Handel (sekcja G)					
	Estymator			REE		
	DIR	SYN, REG	COM	DIR	SYN, REG	COM
Dolnośląskie	399,0	363,1	374,9	0,079	0,061	0,048
Kujawsko-pomorskie	312,8	366,4	352,7	0,119	0,059	0,053
Lubelskie	324,1	317,5	319,4	0,093	0,061	0,051
Lubuskie	313,1	317,4	316,0	0,094	0,066	0,054
Łódzkie	351,6	320,2	328,0	0,096	0,061	0,052
Małopolskie	487,4	386,3	401,1	0,118	0,062	0,055
Mazowieckie	531,5	615,4	572,5	0,083	0,074	0,055
Opolskie	490,7	324,3	331,5	0,217	0,070	0,067
Podkarpackie	363,3	363,8	363,3	0,088	0,459	0,086
Podlaskie	365,0	367,0	366,5	0,109	0,063	0,055
Pomorskie	482,6	375,4	391,7	0,124	0,067	0,059
Śląskie	467,8	357,2	387,5	0,083	0,067	0,053
Świętokrzyskie	761,2	285,2	287,0	0,449	0,074	0,074
Warmińsko-mazurskie	372,6	322,0	328,3	0,138	0,060	0,055
Wielkopolskie	510,5	380,2	392,4	0,149	0,064	0,059
Zachodniopomorskie	348,8	352,4	351,2	0,096	0,069	0,056

Ocena precyzji szacunków oparta na wartościach współczynnika *deff* wskazuje na to, że we wszystkich rozpatrywanych przekrojach zastosowanie estymacji bayesowskiej wpłynęło na zwiększenie efektywności (por. rys. 1 i 2). We wszystkich przypadkach współczynnik *deff* przyjął wartość mniejszą od jedności. Warto zauważyć, że większy zysk odnotowano w przypadku liczby osób zatrudnionych. Dla przychodu wartości współczynnika są większe, ale w żadnym przypadku nie przekraczają jedności.

Podsumowując, można stwierdzić, że zastosowanie estymacji bayesowskiej niezależnie od rozpatrywanego przekroju poprawia precyzję szacunku.

4. Wnioski

Przeprowadzona analiza pozwala na sformułowanie następujących wniosków:

1. Wyniki szacunków z zastosowaniem podejścia bazującego na modelu regresji wydają się mniej obciążone w przekroju regionalnym aniżeli według rodzaju działalności gospodarczej.

2. Każdy przekrój powinien podlegać odrębnemu modelowaniu i uwzględniać pogłębioną analizę jakości modelu. Stanowi to istotne utrudnienie przy szerszym zastosowaniu.

3. Otrzymane rezultaty nie są w pełni zadowalające, jednak oszacowania są „rozsądne” i zachęcają do dalszych badań.

4. Zastosowanie estymacji złożonej bilansującej obciążenie podejścia syntetycznego oraz niestabilność estymacji bezpośredniej poprawia precyzję szacunku. Zysk jest tym większy, im mniejsza jest liczebność próby w domenie oraz im bardziej szczegółowy jest podział na domeny.

5. Estymacja bayesowska charakteryzuje się mniejszym zróżnicowaniem i jest bardziej uśredniona w porównaniu z estymacją bezpośrednią

6. Estymacja bayesowska wymaga, aby dla każdego szacowanego parametru oraz przekroju odrębnie wyznaczać wartość wagi,

7. Ze względu na dużą niejednorodność rozkładów szacowanych zmiennych należy poszukiwać dalszych modyfikacji zastosowanych rozwiązań, jak np. warstwowanie podmiotów w zależności od liczby zatrudnionych: 0 zatrudnionych; 1-5 zatrudnionych oraz 6-9 zatrudnionych.

8. Przeprowadzone badanie stanowi pierwszą próbę szacunków podstawowych parametrów z zakresu statystyki gospodarczej małych przedsiębiorstw w przekroju województwo i sekcja, w którym wykorzystano rejestr podatkowy.

9. Poprawę jakości szacunków może zapewnić korzystanie z kompletnych dodatkowych źródeł informacji (zbiory podatkowe) oraz źródeł informacji dodatkowej z kilku kolejnych lat, co wzbogaciłoby zasób zmiennych wspomagających.

Literatura

- [1] Bracha C. (2003), *Estymacja danych z badania aktywności ekonomicznej ludności na poziomie ludności powiatów dla lat 1995-2002*, GUS, Warszawa.
- [2] Bracha C., Lednicki B., Wieczorkowski R. (2004), *Wykorzystanie złożonych metod estymacji do dezagregacji danych z badania aktywności ekonomicznej ludności w roku 2003*, GUS, Warszawa.
- [3] Gołata E. (2006), *Wykorzystanie danych ze źródeł administracyjnych w szacowaniu informacji o działalności gospodarczej małych przedsiębiorstw w przekroju sekcji PKD i województw*, [w:] *Statystyka w badaniach społecznych*, red. S. Ostasiewicz, J. Paradysz, Prace Naukowe Akademii Ekonomicznej nr 1142, AE, Wrocław.
- [4] GUS (2002), *Działalność gospodarcza przedsiębiorstw o liczbie pracujących do 9 osób w 2001 r.*, Warszawa.
- [5] Klimanek T. (2006), *Estymacja pośrednia jako metoda szacowania informacji o działalności gospodarczej dla małych przedsiębiorstw w przekroju domen*, [w:] *Statystyka w badaniach społecznych*, red. S. Ostasiewicz, J. Paradysz, Prace Naukowe Akademii Ekonomicznej nr 1142, AE, Wrocław.
- [6] Kordos J., Paradysz J. (2000), *Some Experiments in Small Area Estimation in Poland*, „Statistics in Transition” nr 4.4, s. 679-697.
- [7] Kish L. (1996), *Stulecie zmagania o badania reprezentacyjne*, „Wiadomości Statystyczne” nr 8, GUS, Warszawa.
- [8] Rao J.N.K. (2003), *Small Area Estimation*, Wiley – Interscience, John Wiley and Sons, INC., Hoboken, New Jersey.

EMPIRICAL BAYES ESTIMATION OF SMALL BUSINESS CHARACTERISTICS IN POLAND

Summary

This paper describes the use of administrative data sources and indirect estimation techniques to estimate basic economic information about small business in the joint cross-section of Polish Classification of Economic Activities (PKD) and voivodships.

The study objective, specified as accounting for and applying tax data for a more effective use of a survey of small businesses with up to 9 employees, was understood in a twofold manner. First of all, it was a verification of the hypothesis concerning the possibility of improving estimation precision in studies available to date. Secondly, it was intended as a possible extension of estimation scope by joint distribution by voivodship and economic activity (PKD division). The basic economic information, for the aim of this study, was limited to the employment and revenues.

The Horvitz-Thompson estimates in the joint cross-sections of PKD and voivodships are presented and compared with the results of indirect: regression synthetic and composite estimates. The estimation precision characterizing economic activity of small enterprises is presented and analyzed for different types of domains: PKD sections, regions and joint cross-section of regions and economic activity.

Results obtained in the study entitle to draw the following conclusion. The application of indirect estimation to small business data requires consideration of the heterogeneity of its distribution. Nevertheless the results of the study present the practical possibilities and benefits of adopting the techniques of small area estimation to small business data in Poland.