

Tomasz Klimanek

Akademia Ekonomiczna w Poznaniu

ESTYMACJA POŚREDNIA JAKO METODA SZACOWANIA INFORMACJI O DZIAŁALNOŚCI GOSPODARCZEJ DLA MAŁYCH PRZEDSIĘBIORSTW W PRZEKROJU DOMEN

1. Wstęp

Polska statystyka gospodarcza stała się w latach dziewięćdziesiątych przed poważnym wyzwaniem związanym z gwałtownie zmieniającą się strukturą przedsiębiorstw, powstawaniem dużych obszarów szarej strefy oraz popytem na informację na możliwie najniższym poziomie agregacji przestrzennej i sektorowej. Do tych wyzwań doszły nowe możliwości, jakie dają alternatywne źródła zasilania informacyjnego, możliwości przetwarzania danych oraz nowoczesne techniki estymacji.

W Polsce Główny Urząd Statystyczny prowadzi kilka badań działalności przedsiębiorstw dostosowanych do ich wielkości pod względem liczby zatrudnionych. Największe trudności techniczne oraz metodologiczne dotyczą mikroprzedsiębiorstw zatrudniających do 9 osób. Ten rodzaj przedsiębiorstw jest objęty badaniem SP3¹. Ze względu na dość duże błędy szacunku badanie to zapewnia estymację jedynie na poziomie województw i sekcji PKD. Dlatego w Głównym Urzędzie Statystycznym powstała myśl, aby w celu dokonania szacunków w przekroju województw i sekcji PKD jednocześnie wykorzystać dorobek metodologiczny projektu EURAREA oraz innych prac badawczych w zakresie statystyki małych obszarów. Niniejszy artykuł prezentuje rezultaty takiego eksperymentalnego podejścia z wykorzystaniem estymatora syntetycznego regresyjnego².

¹ Szerzej o źródłach administracyjnych wykorzystywanych w szacowaniu informacji o działalności gospodarczej małych przedsiębiorstw i badaniu SP3, por. [6].

² Szerzej o wynikach uzyskanych dla estymatora złożonego opartego na liniowej kombinacji estymatora bezpośredniego i syntetycznego uwzględniającego zoptymalizowane wagi por. [5].

2. Cele i etapy badania

Główny celem badania była próba empirycznej weryfikacji możliwości zastosowania metod estymacji pośredniej do szacowania informacji o działalności gospodarczej małych przedsiębiorstw w przekroju województw i sekcji PKD jednocześnie. Niniejszy artykuł prezentuje wyniki uzyskane za pomocą estymatora syntetycznego regresyjnego. Autor postawił sobie także dwa inne cele. Są to:

a) ocena precyzji estymacji syntetycznej regresyjnej w przekroju województw i sekcji PKD,

b) próba określenia zgodności oszacowań pośrednich i bezpośrednich.

W pierwszym etapie badania podjęta została decyzja co do wyboru zmiennych objaśnianych, zmiennych objaśniających oraz co do zastosowanej strategii budowy syntetycznego estymatora regresyjnego. Estymację syntetyczną regresyjną przeprowadzono w dwóch ujęciach:

a) przyjmując za zmienne objaśniające wartości średnie dla domen cech z dodatkowych źródeł informacji; a mianowicie BJS i rejestru Ministerstwa Finansów;

b) przyjmując za zmienne objaśniające wartości bezpośrednich estymatorów średnich określonych dla domen na podstawie badania SP3.

Ze względu na ramy niniejszego opracowania zaprezentowane zostaną wyniki jedynie pierwszego z wyróżnionych ujęć. Dla każdego z analizowanych ujęć przekrojowych (odpowiednio dla województw, sekcji PKD oraz województw i sekcji PKD łącznie) oszacowano model regresji, w którym zmiennymi objaśnianymi było 5 podstawowych charakterystyk działalności gospodarczej małych przedsiębiorstw:

- przeciętna liczba zatrudnionych w przeliczeniu na podmiot gospodarczy,
- suma rocznych wynagrodzeń w przeliczeniu na podmiot gospodarczy,
- przeciętne miesięczne wynagrodzenie w przeliczeniu na jednego zatrudnionego,
- przeciętny przychód w przeliczeniu na podmiot gospodarczy,
- przeciętny koszt w przeliczeniu na podmiot gospodarczy.

Wartości zmiennych objaśnianych zostały określone jako oceny estymatorów bezpośrednich odpowiednich wartości średnich na poziomie domen. Natomiast jako zmienne objaśniające uwzględniono wielkości średnie cech dodatkowych na poziomie domen z alternatywnych źródeł informacji. Były to następujące cechy z rejestrów podatkowych oraz BJS:

- przychód (rejestry podatkowe),
- koszt (rejestry podatkowe),
- strata (rejestry podatkowe),
- dochód (rejestry podatkowe),
- liczba pracujących (BJS).

Po oszacowaniu parametrów modelu obliczane były teoretyczne wartości cech objaśnianych dla poszczególnych przekrojów.

3. Estymacja syntetyczna regresyjna

Estymator syntetyczny regresyjny został zdefiniowany w sposób poniższy:

$$\hat{\bar{Y}}_{d(SYN,REG)} = \bar{X}_d \hat{\beta}, \quad (1)$$

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

$\bar{X}_d$ – wartość średnia cechy objaśniającej dla domeny d wyznaczona na podstawie dodatkowego źródła informacji.

Wartości zmiennych z dodatkowego źródła informacji traktujemy jako odpowiadające wielkościom z populacji generalnej, choć oczywiście jesteśmy świadomi niedostatków wykorzystywanych rejestrów administracyjnych. Zakładamy jednak, iż wartości q zmiennych pomocniczych są dostępne w przekroju wszystkich analizowanych domen. Przez Dom oznaczmy zbiór indeksów d , dla których $d = 1, \dots, Dom$. Wartości cech dodatkowych dla Dom zapiszemy w postaci wektorów, zakładając, że $q < Dom$.

Przyjmujemy więc, że:

$$\hat{\bar{Y}}_{d(SYN,REG)} = \bar{\mathbf{x}}_d^T \mathbf{b} + u_d, \quad (2)$$

gdzie $\mathbf{b} = \left[b_q \right]_{q \times 1}$

jest nieznanym wektorem współczynników regresji oraz zmienne losowe u_d są stochastycznie niezależne i mają identyczne rozkłady $u_d \sim N(0, \sigma_d^2)$.

Powyższy model można zapisać w równoważnej postaci macierzowej:

$$\hat{\bar{Y}}_{(SYN,REG)} = \mathbf{X} \mathbf{b} + \mathbf{u}, \quad (3)$$

gdzie: $\hat{\bar{Y}}_{(SYN,REG)} = \left[\hat{\bar{Y}}_{d(SYN,REG)} \right]_{d \times 1}$; $\mathbf{X} = \left[\bar{x}_{dq} \right]_{d \times q}$; $\mathbf{u} = \left[u_d \right]_{d \times 1}$; $\mathbf{u} \sim N_d(\mathbf{0}, \sigma_u^2 \mathbf{I}_{(d)})$,

przy czym $\mathbf{0}$ jest Dom -wymiarowym wektorem zerowym, a $\mathbf{I}_{(d)}$ jest macierzą jednostkową stopnia Dom . Estymatorem uzyskanym KMNK (klasyczną metodą najmniejszych kwadratów) wektora $\mathbf{b}$ jest statystyka:

$$\hat{\mathbf{b}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\bar{Y}}, \quad (4)$$

gdzie: $\mathbf{X}$ jest macierzą wartości średnich cech dodatkowych dla domen (zgodnie z założeniem jest to macierz wartości ustalonych), $\hat{\bar{Y}}$ zaś jest wektorem wartości średnich badanej cechy dla tych samych domen. Współrzędnymi wektora $\hat{\bar{Y}}$ są estymatory wartości średniej badanej cechy otrzymane z badania reprezentacyjnego (SP3).

Sposób uzyskania wariancji estymatora syntetycznego regresyjnego przedstawimy, wzorując się na opracowaniu [3, s. 34 i nast.]. Wartość oczekiwaną predyktora badanego parametru dla domeny d skonstruowanego na podstawie rejestrów Ministerstwa Finansów zapisać można jako:

$$\begin{aligned} E(\hat{\bar{Y}}_{d(SYN,REG)}) &= E(\mathbf{x}_d^T \hat{\mathbf{b}}) = E(\mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \hat{\mathbf{Y}}) = E(\mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X} \mathbf{b} + \mathbf{u})) = \\ &= E(\mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T (\mathbf{X} \mathbf{b} + \mathbf{u})) = E(\mathbf{x}_d^T \mathbf{b} + \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}) = \mathbf{x}_d^T \mathbf{b}. \end{aligned} \quad (5)$$

Różnica między oceną estymatora a wartością oczekiwaną jest więc równa:

$$\hat{\bar{Y}}_{d(SYN,REG)} - E(\hat{\bar{Y}}_d) = \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u}, \quad (6)$$

a wariancja jest równa:

$$V(\hat{\bar{Y}}_{d(SYN,REG)}) = V(\mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{u} \mathbf{u}^T \mathbf{X} (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d) = \sigma_u^2 \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d. \quad (7)$$

Estymator wariancji definiujemy wzorem:

$$S^2(\hat{u}) = \frac{1}{Dom - q} \sum_{d=1}^{Dom} \hat{u}_d^2, \quad (8)$$

gdzie: $\hat{u}_d = \hat{\bar{Y}}_{d(EXP)} - \hat{\bar{Y}}_{d(SYN,REG)}$; $d = 1, \dots, Dom$ są resztami modelu. Ostatecznie estymator wariancji $D^2(\tilde{\theta}_d)$ określony jest wzorem:

$$V(\hat{\bar{Y}}_{d(SYN,REG)}) = S^2(\hat{u}) \mathbf{x}_d^T (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{x}_d. \quad (9)$$

4. Wyniki estymacji

Tabela 1 zawiera syntetyczny opis oszacowanych modeli regresji dla każdej ze zmiennych objaśnianych³.

Analiza tab. 1 potwierdza, że najlepsze, w sensie najwyższej wartości współczynnika determinacji, są modele skonstruowane na poziomie sekcji PKD (liczba zatrudnionych, przychody, koszty). Jedynie w przypadku sumy wynagrodzeń wyższą wartością statystyki R^2 charakteryzuje się model skonstruowany na poziomie województw. W przekrojach województwo i sekcja PKD łącznie jakość modeli można uznać za zadowalającą. Wartości skorygowanego współczynnika R^2 wahają się od 40 do 60%.

³ Wartości estymatorów w przypadku zmiennej przeciętne miesięczne wynagrodzenie w przeliczeniu na jednego zatrudnionego uzyskano, dzieląc wartość estymatora dla zmiennej suma wynagrodzeń przez 12 (liczba miesięcy) oraz przez liczbę zatrudnionych.

Tabela 1. Syntetyczna charakterystyka modeli opisujących informacje o działalności gospodarczej małych przedsiębiorstw

Y	DOMENA	n	S(u)	R ²	R ² _skorygowane
SUMWYN	WOJ	16	904,30	0,912	0,868
SUMWYN	ID_SEKC	11	2 155,05	0,841	0,682
SUMWYN	woj_id_sekc	176	3 477,87	0,521	0,510
LZAT	WOJ	16	0,06	0,637	0,456
LZAT	ID_SEKC	11	0,11	0,935	0,870
LZAT	woj_id_sekc	176	0,24	0,609	0,600
PRZYCH	WOJ	16	49 026,70	0,458	0,187
PRZYCH	ID_SEKC	11	74 132,38	0,887	0,774
PRZYCH	woj_id_sekc	176	195 975,83	0,465	0,453
KOSZT	WOJ	16	49 090,58	0,429	0,144
KOSZT	ID_SEKC	11	72 140,24	0,886	0,772
KOSZT	woj_id_sekc	176	194 754,04	0,450	0,437

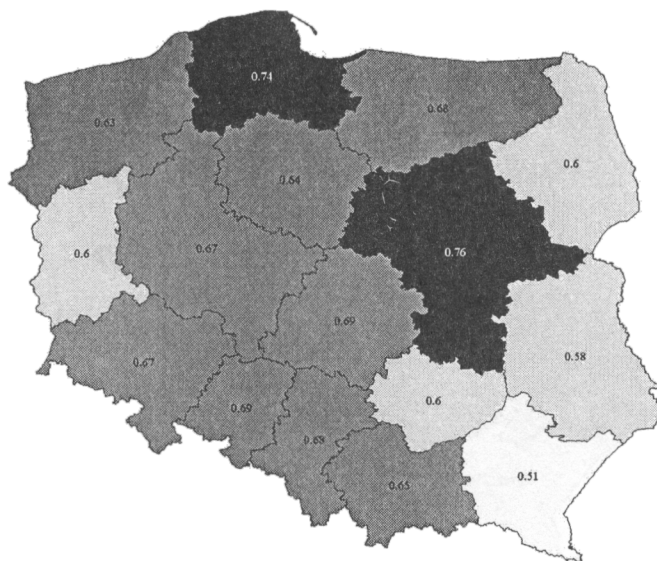
Źródło: obliczenia własne.

Analiza danych zawartych w tab. 2 wskazuje, że najwyższą średnią liczbą zatrudnionych w małych przedsiębiorstwach charakteryzują się firmy w województwach: mazowieckim, pomorskim i wielkopolskim. Z kolei najniższe wartości estymator zarówno bezpośredni (EXP), jak i syntetyczny (SYN, REG) osiąga

Tabela 2. Estymacja przeciętnej liczby zatrudnionych na 1 zakład na podstawie badania SP3 z 2001 roku z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw

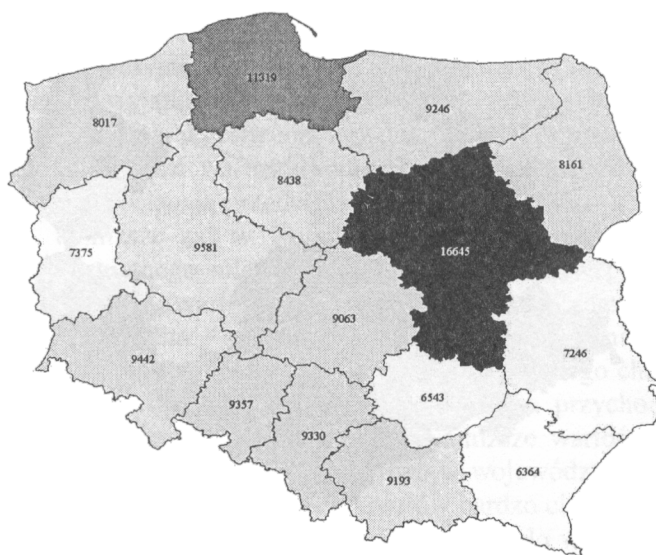
Województwo	Estymator		REE	
	EXP	SYN, REG	EXP	SYN, REG
Dolnośląskie	0,62	0,67	0,059	0,023
Kujawsko-pomorskie	0,61	0,64	0,088	0,023
Lubelskie	0,57	0,58	0,087	0,043
Lubuskie	0,69	0,60	0,098	0,043
Łódzkie	0,61	0,69	0,057	0,037
Małopolskie	0,63	0,65	0,048	0,035
Mazowieckie	0,76	0,76	0,040	0,060
Opolskie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

Źródło: obliczenia własne.



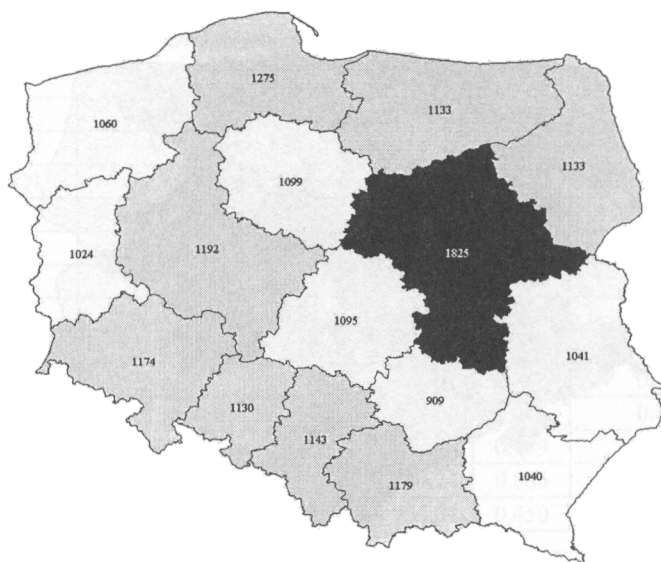
Rys. 1. Przestrzenny rozkład wartości estymatora syntetycznego regresyjnego przeciętnej liczby zatrudnionych na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw

Źródło: opracowanie własne.

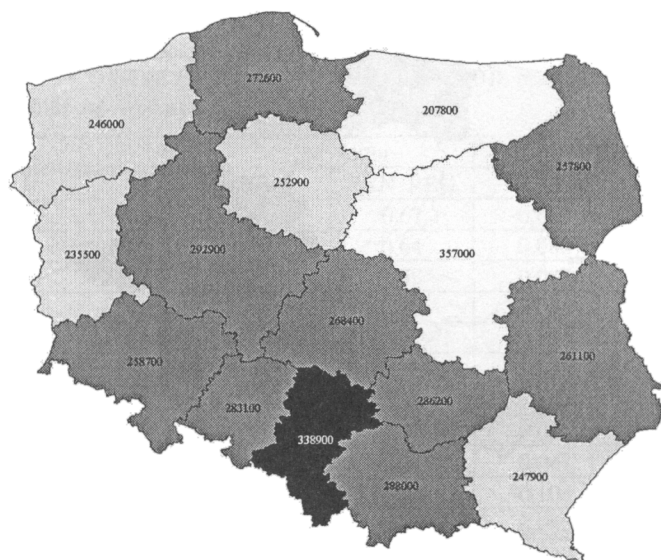


Rys. 2. Przestrzenny rozkład wartości estymatora syntetycznego regresyjnego sumy rocznych wynagrodzeń w przeliczeniu na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw

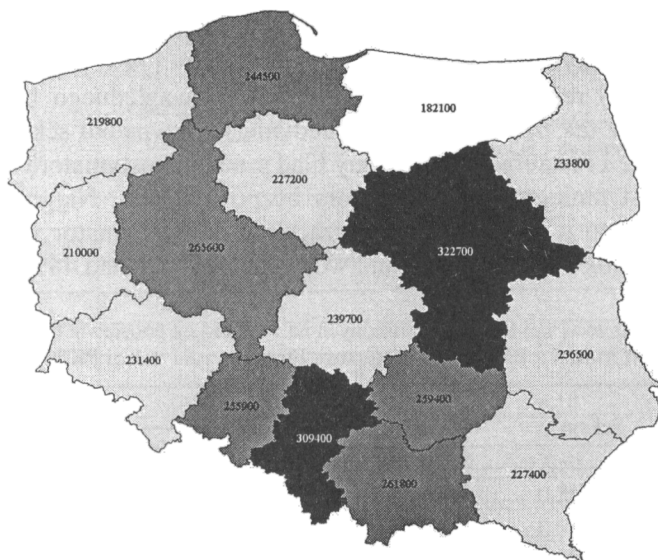
Źródło: opracowanie własne.



Rys. 3. Przestrenny rozkład wartości estymatora syntetycznego regresyjnego przeciętnego miesięcznego wynagrodzenia w przeliczeniu na jednego zatrudnionego na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw
Źródło: opracowanie własne.



Rys. 4. Przestrenny rozkład wartości estymatora syntetycznego regresyjnego przeciętnej wysokości przychodów na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw
Źródło: opracowanie własne.



Rys. 5. Przestrzenny rozkład wartości estymatora syntetycznego regresyjnego przeciętnej wysokości kosztów na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla województw

Źródło: opracowanie własne.

w województwie podkarpackim. Względne błędy szacunku (REE) w przypadku estymatora syntetycznego regresyjnego w większości przypadków są niższe niż dla estymatora bezpośredniego z wyjątkiem województw: mazowieckiego i podkarpackiego.

Rysunki 1-5 przedstawiają przestrzenny rozkład wartości estymatora syntetycznego regresyjnego. Najwyższe wartości sumy rocznych wynagrodzeń w przeliczeniu na jeden podmiot osiągają przedsiębiorstwa w województwach mazowieckim i pomorskim, najniższe zaś w województwach: podkarpackim i łódzkim. W odniesieniu do przeciętnego miesięcznego wynagrodzenia w przeliczeniu na 1 zatrudnionego najwyższą wartość estymator syntetyczny regresyjny osiąga w województwie mazowieckim – wynosi ona 1825 zł. Najniższymi przeciętnymi miesięcznymi wynagrodzeniami w przeliczeniu na 1 zatrudnionego charakteryzują się przedsiębiorstwa w województwie łódzkim. Najwyższe przychody i koszty mają przedsiębiorstwa w województwie śląskim. Najniższe wartości estymatora notują dla tych dwóch zmiennych przedsiębiorstwa w województwie warmińsko-mazurskim. W odniesieniu do przychodów i kosztów bardzo ciekawa jest pozycja województwa mazowieckiego. Z jednej strony należy ono do województw o najniższych przychodach w przeliczeniu na 1 podmiot, natomiast z drugiej – ma wysokie koszty w przeliczeniu na 1 podmiot, pod tym względem jest ono drugie po województwie śląskim.

W przekroju sekcji najwyższe wartości estymator bezpośredni przeciętnej liczby zatrudnionych na jeden zakład osiąga w przypadku przetwórstwa przemys-

słowego oraz sekcji „hotele i restauracje”. Podobnie jest w przypadku estymatora syntetycznego regresyjnego. W przekroju województw i sekcji przewaga estymatora syntetycznego regresyjnego w odniesieniu do względnego błędu szacunku (REE) nie jest już tak bardzo widoczna. Jedynie w przypadku sekcji „budownictwo” oraz „hotele i restauracje” względny błąd szacunku estymatora syntetycznego regresyjnego jest mniejszy niż estymatora bezpośredniego. Najgorsze wyniki w sensie największego względnego błędu szacunku daje estymator syntetyczny, regresyjny w przypadku sekcji „pośrednictwo finansowe” (ponad 44%).

Tabela 3. Estymacja przeciętnej liczby zatrudnionych na 1 zakład na podstawie badania SP3 z 2001 r. z wykorzystaniem informacji z BJS oraz Ministerstwa Finansów dla sekcji PKD

Seksja	Estymator		REE	
	EXP	SYN, REG	EXP	SYN, REG
Przetwórstwo przemysłowe	1,12	0,99	0,034	0,048
Budownictwo	0,72	0,86	0,057	0,049
Handel i naprawy	0,75	0,76	0,026	0,089
Hotele i restauracje	0,90	0,91	0,088	0,069
Transport, gospodarka magazynowa i łączność	0,38	0,47	0,074	0,113
Pośrednictwo finansowe	0,18	0,14	0,086	0,448
Obsługa nieruchomości i firm, nauka	0,51	0,44	0,054	0,138
Edukacja	0,21	0,29	0,123	0,168
Ochrona zdrowia i opieka społeczna	0,28	0,25	0,121	0,200
Pozostała działalność usługowa komunalna, społeczna i indywidualna	0,48	0,43	0,089	0,112

Źródło: obliczenia własne.

Dla syntetycznej oceny estymacji w przekroju województwo i sekcja PKD jednocześnie wykorzystano miarę oceniającą zysk na efektywności. Jest ona zdefiniowana następująco:

$$deff = \frac{V(\hat{Y}_{d(SYN,REG)})}{V(\hat{Y}_{d(DIRECT)})} \quad (10)$$

Odnosi ona wariancję estymatora pośredniego do wariancji estymatora bezpośredniego. Jeśli $deff < 1$, estymator pośredni można uznać za bardziej efektywny niż estymator bezpośredni. Należy jednak pamiętać, że miara ta ocenia jedynie wariancje estymatorów, abstrahując od obciążenia estymatorów.

Zaprezentowane w tab. 4 wyniki dowodzą, że stosowanie estymatora syntetycznego regresyjnego daje mniejszą wariancję estymatora w porównaniu z estymatorem bezpośrednim. Zysk na efektywności jest dość duży. Współczynnik $deff$ kształtuje się od 1 do 47%. Wariancja estymatora pośredniego jest 100-krotnie mniejsza od wariancji estymatora bezpośredniego w przypadku sekcji „przetwó-

Tabela 4. Zysk na efektywności (*deff*) dla sumy wynagrodzeń w przekroju województw i wybranych sekcji PKD

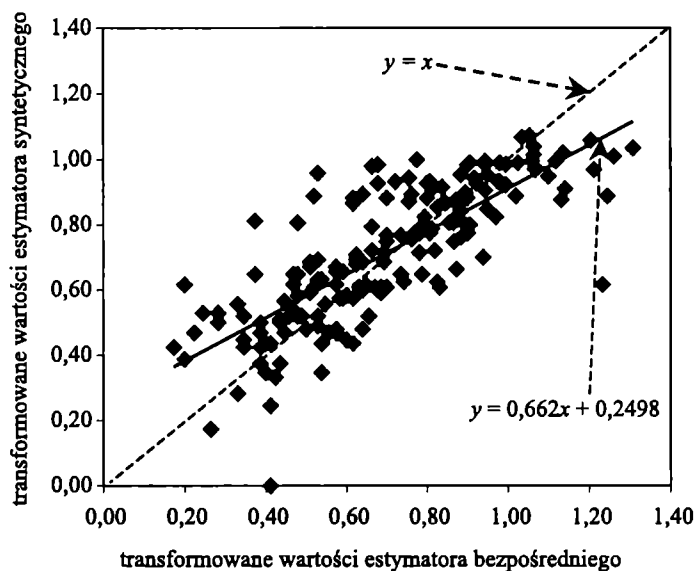
Nazwa województwa	Sekcja D – przetwórstwo przemysłowe	Sekcja F – budownictwo	Sekcja G – handel i naprawy	Sekcja K – obsługa nieruchomości i firm, nauka
Dolnośląskie	0,06	0,12	0,21	0,06
Kujawsko-pomorskie	0,01	0,01	0,14	0,07
Lubelskie	0,02	0,18	0,19	0,11
Lubuskie	0,02	0,02	0,12	0,09
Łódzkie	0,36	0,10	0,11	0,21
Małopolskie	0,15	0,22	0,24	0,04
Mazowieckie	0,16	0,20	0,18	0,16
Opolskie	0,03	0,04	0,15	0,07
Podkarpackie	0,04	0,24	18,50	0,05
Podlaskie	0,02	0,02	0,21	0,01
Pomorskie	0,10	0,03	0,07	0,03
Śląskie	0,20	0,10	0,38	0,08
Świętokrzyskie	0,06	0,47	0,17	0,04
Warmińsko-mazurskie	0,06	0,03	0,09	0,04
Wielkopolskie	0,18	0,05	0,08	0,09
Zachodniopomorskie	0,04	0,09	0,30	0,09

Źródło: opracowanie własne.

stwo przemysłowe” i „budownictwo” w województwie kujawsko-pomorskim oraz sekcji „obsługa nieruchomości i firm” oraz „nauka” w województwie podlaskim. Jedyne w przypadku sekcji „handel i naprawy” w województwie podkarpackim współczynnik *deff* przekroczył 18,5. Oznacza to ponad 18-krotnie większą wariancję estymatora pośredniego w stosunku do estymatora bezpośredniego.

5. Ocena obciążenia estymacji syntetycznej regresyjnej

Rozbieżności pomiędzy oszacowaniem bezpośrednim a ocenami estymatora syntetycznego regresyjnego przedstawiono w formie graficznej (por. rys. 6). Za podstawę porównań przyjęto estymator bezpośredni, zakładając, że jest on nie-obciążony. Ponieważ domeny, dla których przeprowadzono szacunek, są bardzo zróżnicowane pod względem wielkości, zgodnie z postulatem Chambersa dokonano transformacji ocen estymatorów funkcją potęgową o wykładniku równym 0,5. Sporządzono wykresy rozrzutu dla punktów o współrzędnych określonych przez wartości ocen estymatorów pośrednich dla poszczególnych cech oraz estymatora bezpośredniego. Następnie oszacowano parametry dopasowanej do tych punktów linii regresji $y = a_0 + a_1x$ i dokonano porównania z funkcją regresji o postaci $y = x$. Zastosowano w tym celu test istotności, by zweryfikować hipotezę, że różnica pomiędzy oceną współczynnika regresji a jednością jest statystycznie



Rys. 6. Ocena obciążenia estymacji syntetycznej regresyjnej na przykładzie estymatora wartości oczekiwanej liczby zatrudnionych

Źródło: opracowanie własne.

nieistotna dla przyjętego poziomu błędu pierwszego rodzaju (dla poziomu istotności $\alpha = 0,05$). Jeżeli oceny estymacji syntetycznej regresyjnej rozkładałyby się losowo względem nieobciążonych ocen bezpośrednich, to otrzymana na wykresie rozrzutu smuga byłaby zbliżona do dwusiecznej kąta prostego, a wartość współczynnika regresji zbliżona do jedności.

6. Wnioski

Należy podkreślić, że główny cel badania został osiągnięty. Przeprowadzone badania eksperymentalne potwierdziły istnienie możliwości zastosowania metod estymacji pośredniej do szacowania informacji o działalności gospodarczej małych przedsiębiorstw w przekroju województw i sekcji PKD jednocześnie. Wskazać trzeba jednak na konieczność poszukiwania nowych podejść. Nie we wszystkich przekrojach wyniki estymacji są zadowalające. Ponadto można wysnuć następujące wnioski:

1. Możliwość skorzystania z dodatkowych informacji pozwala na poprawę precyzji szacunku. Ma to odbicie w mniejszych względnych błędach szacunku i mniejszym od jedności współczynniku *deff*.

2. Poprawę jakości szacunków może zapewnić korzystanie z kompletnych dodatkowych źródeł informacji (zbiory podatkowe) oraz źródeł informacji dodatkowej z kilku kolejnych lat.

3. Należy poszukiwać takich transformacji zmiennych, które zapewniałyby bardziej symetryczne rozkłady.

4. Należy poszukiwać innych rozwiązań w zakresie stosowanych modeli. Pewną propozycją mogłyby tutaj być modele wielopoziomowe, w których oprócz zmiennych mierzonych na poziomie jednostek badania (przedsiębiorstw) wprowadzone zostały tzw. zmienne kontekstowe mierzone na poziomie domen.

Literatura

- [1] Bracha C. (1996), *Teoretyczne podstawy metody reprezentacyjnej*, Wydawnictwo Naukowe PWN, Warszawa.
- [2] Bracha C. (2003), *Estymacja danych z badania aktywności ekonomicznej ludności na poziomie ludności powiatów dla lat 1995-2002*, GUS, Warszawa.
- [3] Bracha C., Lednicki B., Wieczorkowski R. (2004), *Wykorzystanie złożonych metod estymacji do dezagregacji danych z badania aktywności ekonomicznej ludności w roku 2003*, GUS, Warszawa.
- [4] Brown G., Chambers R. Heady P., Heasman D. (2001), *Evaluation of Small Area Estimation Methods – an Application to Unemployment Estimates from the UK LFS*, Proceedings of Statistics Canada Symposium 2001, Achieving Data Quality in a Statistical Agency: a Methodological Perspective.
- [5] Dehnel G. (2006), *Ocena charakterystyk ekonomicznych małych podmiotów z wykorzystaniem estymacji bayesowskiej*, [w:] *Statystyka regionalna. Wielowymiarowa analiza statystyczna. Metoda reprezentacyjna w badaniach ekonomicznych*, AE, Wrocław.
- [6] Gołata E. (2006), *Wykorzystanie źródeł administracyjnych w szacowaniu informacji o działalności gospodarczej małych przedsiębiorstw w przekroju sekcji PKD i województw*, [w:] *Statystyka regionalna. Wielowymiarowa analiza statystyczna. Metoda reprezentacyjna w badaniach ekonomicznych*, AE, Wrocław.
- [7] Kordos J., Paradysz J. (2000), *Some Experiments in Small Area Estimation in Poland*, Statistics in Transition 4.4, s. 679-697.
- [8] Longford N. (2002), *Initial Theory Report: Small-area Estimation with Complex Sampling Design*, EURAREA, IST 2000-26290.

INDIRECT ESTIMATION OF SMALL BUSINESS CHARACTERISTICS

Summary

The Polish economic statistics faced a serious challenge in the nineties of the last century. That was because of a rapid change in the structure of enterprises, widening of the area the shadow economy and the increasing demand for the information on small territorial area and branch cross-section. Then quite new challenges emerged: the access to new alternative sources of data, improvement as far as data processing is concerned, and new estimation techniques. Based on the experiences of the EURAREA project, the paper presents an attempt to use small area statistics

methods to improve estimation precision with respect to basic information about economic activities of small business. The study was aimed at accounting for and applying tax data for a more effective use of a sample survey of small businesses with up to 9 employees. The domains were defined as voivodship and PKD category cross-section and the estimates for them were produced by the use of synthetic regression estimator. The final part of the paper focuses on the quality of estimation discussed from two perspectives: efficiency and bias.