

Tomasz Klimanek

Urząd Statystyczny w Poznaniu
e-mail: t.klimanek@stat.gov.pl

ESTYMACJA POŚREDNIA NIEPEŁNOSPRAWNOŚCI – BADANIE SYMULACYJNE¹

INDIRECT ESTIMATION OF DISABILITY – SIMULATION STUDY

DOI: 10.15611/pn.2017.469.09
JEL Classification: C81

Streszczenie: Niepełnosprawność postrzegana jest jako jeden z najważniejszych problemów, przed jakimi stoi dziś społeczeństwo. Różne grupy odbiorców danych dotyczących niepełnosprawności mają zróżnicowane potrzeby informacyjne, które nie zawsze mogą być zaspokojone. W artykule przedstawione zostaną wyniki badania symulacyjnego, którego celem jest oszacowanie, z wykorzystaniem statystyki małych obszarów (SMO) – w oparciu o estymatory opracowane w ramach projektu EURAREA, odsetka osób niepełnosprawnych z uwzględnieniem problemu nieplanowanych domen (tzw. lokalnych rynków pracy) na podstawie danych z Narodowego Spisu Ludności i Mieszkań 2002. W badaniu symulacyjnym wykorzystane zostaną dane dla województwa wielkopolskiego, a przyjęty schemat losowania próby został skonstruowany tak, aby był jak najbardziej zbliżony do tego, który był zastosowany w Narodowym Spisie Ludności i Mieszkań w 2011 r. W artykule podjęta zostanie także próba oceny zastosowanych estymatorów.

Słowa kluczowe: niepełnosprawność, estymatory projektu EURAREA, statystyka małych obszarów, lokalne rynki pracy.

Summary: The phenomenon of disability is regarded as one of the most serious social problems facing contemporary society. Different groups of data users express the need for data on disability, also at the level of so called unplanned domains. The author will present the results of simulation study, which is aimed at estimating the percentage of disabled persons for illustrative delimited local labour markets in 2002. Individual data from the 2002 census for Wielkopolska region will be analysed using a simulation approach. The sampling desing for the simulation study was adopted as to resemble as much as possible the one applied in the 2011 Census. In the paper the author will also try to evaluate the performance of applied estimators.

Keywords: disability, EURAREA project estimators, small area statistics, local labour markets.

¹ Praca zrealizowana w ramach projektu, który został sfinansowany ze środków Narodowego Centrum Nauki, przyznanych na podstawie decyzji numer DEC-2013/11/B/HS4/01472.

1. Wstęp

Zgodnie z ostatnim raportem Światowej Organizacji Zdrowia (WHO) zjawisko niepełnosprawności dotyka ponad 1 miliarda ludzi na całym świecie, co stanowi ok. 15% populacji całego globu [WHO 2014]. Powoduje to, że zjawisko niepełnosprawności staje się jednym najpoważniejszych problemów społecznych, z jakimi zmagają się współczesna cywilizacja. Liczba osób z niepełnosprawnością systematycznie rośnie, co jest efektem między innymi procesu starzenia się ludności. Realizacja zadań związanych z włączeniem tej grupy społecznej wymaga jednak szczegółowej i aktualnej informacji. Na niższych szczeblach podziału terytorialnego kraju zapotrzebowanie na informacje dotyczące niepełnosprawności niejednokrotnie przewyższa zakres i szczegółowość danych udostępnianych przez statystykę publiczną. Znajduje to wyraz w krytycznych opiniach formułowanych przez środowiska naukowe zwłaszcza w stosunku do ostatniego Narodowego Spisu Ludności i Mieszkań 2011 [Slany 2014].

Mając powyższe na względzie, w niniejszym artykule podjęto próbę wykorzystania technik estymacji pośredniej w szacowaniu odsetka osób niepełnosprawnych wyłącznie biologicznie dla domen, które zostały zdefiniowane jako przekrój tzw. lokalnych rynków pracy i wieku ekonomicznego ludności w województwie wielkopolskim. Należy podkreślić, że są to przekroje, które nie były uwzględniane w schemacie losowania próby (tzw. *unplanned domains*). Wykorzystanie w badaniu symulacyjnym danych dotyczących 2002 roku z jednej strony wynikało z dostępności do danych jednostkowych, a z drugiej strony, ze względu na znajomość wartości w populacji, umożliwiało ocenę tzw. obciążenia empirycznego wykorzystanych estymatorów. Przyjęty schemat losowania próby został skonstruowany tak, aby był jak najbardziej zbliżony do tego zastosowanego w Narodowym Spisie Ludności i Mieszkań w 2011 roku.

2. Obszary rynków pracy

W 2007 roku podczas przeglądu wdrażania regulacji dotyczących NUTS (Klasyfikacji Jednostek Terytorialnych do Celów Statystycznych – *Classification of Territorial Units for Statistics*), nazywanej także *Nomenclature of Territorial Units for Statistics*), krajowe urzędy statystyczne zaproponowały przetestowanie alternatywnej klasyfikacji podziału administracyjnego kraju. Pierwszym krokiem miało być wypracowanie wspólnego (w ramach całej Unii Europejskiej) podejścia w zakresie obszarów rynków pracy. Przyjęto, że obszar rynku pracy ma być geograficznie określonym i ekonomicznie zintegrowanym obszarem funkcjonalnym, w ramach którego jego mieszkaniec może znaleźć miejsce pracy (zakładając pewną określoną odległość dojazdu do pracy) lub może zmienić zatrudnienie bez zmiany miejsca zamieszkania.

Od 2015 roku, ze względu na doświadczenia w zakresie opracowania dojazdów do pracy, Ośrodek Statystyki Miast Urzędu Statystycznego w Poznaniu uczestniczy w granicy Eurostatu w zakresie delimitacji obszarów rynków pracy w oparciu o algorytm wypracowany w Wielkiej Brytanii TTWA (*Travel To Work Areas*) (https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en). Algorytm ten opiera się na następującej nierówności:

$$\frac{\min_{SC}}{\text{tar}_{SC}} \leq \left(1 - \left(1 - \frac{\min_{SC}}{\text{tar}_{SC}} \right) * \text{MAX} \left(\frac{\text{tar}_{SZ} - SZ}{\text{tar}_{SZ} - \min_{SZ}}, 0 \right) \right) * \left(\frac{\min(SZ, \text{tar}_{SC})}{\text{tar}_{SC}} \right), \quad (1)$$

gdzie: SC – wskaźnik samodzielności lokalnego rynku pracy (tzw. *self-containment*), SZ – liczba pracujących mieszkańców danego obszaru, \min_{SC} – minimalna wartość wskaźnika samodzielności lokalnego rynku pracy, tar_{SC} – docelowa wartość wskaźnika samodzielności lokalnego rynku pracy, \min_{SZ} – minimalna liczba pracujących mieszkańców, tar_{SZ} – docelowa liczba pracujących mieszkańców.

Wskaźnik samodzielności lokalnego rynku pracy to mniejsza spośród dwóch wartości: a) stosunku liczby osób zarówno mieszkających, jak i pracujących na danym obszarze do liczby wszystkich pracujących mieszkańców tego obszaru (strona podażowa); b) stosunku liczby osób zarówno mieszkających, jak i pracujących na danym obszarze do liczby wszystkich osób pracujących na tym obszarze (strona popytowa).

Delimitacja obszarów rynków pracy opiera się na następujących etapach:

1. Każda gmina uważana za potencjalny obszar rynku pracy (p-LMA).
2. Sprawdzenie dla każdego p-LMA wartości wynikającej z nierówności (1).
3. Znalezienie p-LMA o najmniejszej wartości prawej strony nierówności (1) (p-LMA_X).
4. Dołączenie do p-LMA_X najsilniej z nim powiązanego potencjalnego obszaru rynku pracy (miara pp dana równaniem (2), przyjmująca największą wartość) i powtórzenie kroków 2 i 3.

Sprawdzanie powiązań między obszarami A i B odbywa się poprzez analizę następującego równania²:

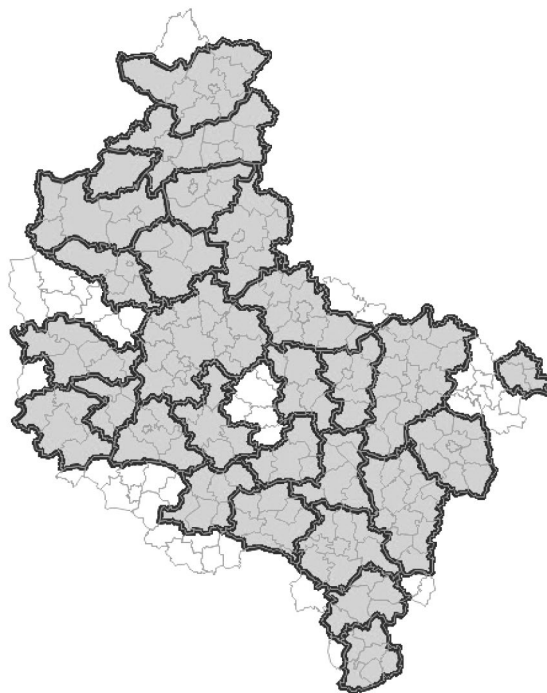
$$pp = \frac{\text{dojazdy}(\text{obszar}_A \rightarrow \text{obszar}_B)^2}{\text{mieszk}_{\text{obszarA}} * \text{pracuj}_{\text{obszarB}}} + \frac{\text{dojazdy}(\text{obszar}_B \rightarrow \text{obszar}_A)^2}{\text{mieszk}_{\text{obszarB}} * \text{pracuj}_{\text{obszarA}}}. \quad (2)$$

5. Gminy, których nie da się przyporządkować w trakcie kroków 2-4, zostają na końcu przypisane „ręcznie” do utworzonych LMA.

Spośród szeregu testowanych parametrów, wykorzystywanych w algorytmach delimitacji obszarów rynków pracy dla województwa wielkopolskiego, na podstawie danych z NSP2011 na potrzeby niniejszego artykułu wybrano następujący ze-

² Gdzie pp oznacza powiązany potencjalny obszar rynku pracy, mieszk – liczbę mieszkańców danego obszaru, a pracuj – liczbę pracujących w danym obszarze.

staw: $\text{minSZ} = 2000$, $\text{tarSZ} = 15\ 000$, $\text{minSC} = 0,55$, $\text{tarSC} = 0,75^3$, a geograficzną prezentację wybranej delimitacji przedstawiono na rys. 1.



Rys. 1. Obszary rynków pracy dla województwa wielkopolskiego na podstawie danych z NSP2011⁴

Źródło: opracowanie własne na podstawie danych Ośrodka Statystyki Miast Urzędu Stat. w Poznaniu.

3. Charakterystyka badania symulacyjnego

Spośród danych jednostkowych pochodzących ze Spisu Ludności i Mieszkań 2002 roku dla województwa wielkopolskiego pobrano próbę zgodnie ze schematem losowania jak najbardziej zbliżonym do tego, który był zastosowany w spisie powszechnym 2011 roku. Liczebność próby w badaniu symulacyjnym wyniosła 240 391 mieszkań, w których zamieszkiwały 724 773 osoby w wieku 15 lat i więcej. Badaną

³ W czasie przygotowywania niniejszego artykułu prace nad wyznaczeniem optymalnej delimitacji obszarów rynków pracy dla województwa wielkopolskiego wciąż trwały, co oznacza, że zaprezentowany zestaw parametrów należy traktować jedynie jako jeden z wielu możliwych.

⁴ Do dalszych analiz wykorzystywane zostały tylko te obszary rynku pracy, na które w całości składały się gminy znajdujące się w obrębie województwa wielkopolskiego. Te obszary zostały zaznaczone na rysunku kolorem jasnozielonym. Pozostałe gminy województwa wielkopolskiego tworzyły obszary rynków pracy wspólnie z gminami z innych województw.

cechą był odsetek osób niepełnosprawnych wyłącznie biologicznie w populacji osób w wieku 15 lat i więcej, przy czym osobę niepełnosprawną wyłącznie biologicznie definiuje się jako osobę, która nie posiada orzeczenia o niepełnosprawności wydane-go przez uprawniony organ, lecz odczuwa ograniczenie sprawności w wykonywaniu czynności podstawowych dla swojego wieku (zabawa, nauka, praca, samoobsługa)⁵. Szacunków dokonywano dla domen określonych jako przekrój obszarów rynków pracy oraz wieku ekonomicznego ludności (osoby w wieku przedprodukcyjnym, produkcyjnym i poprodukcyjnym). Obszary rynków pracy dla 2002 roku zostały wyznaczone zgodnie z algorytmem przedstawionym powyżej i wypracowanym na podstawie danych dla 2011 roku⁶.

W badaniu symulacyjnym zastosowano estymatory projektu EURAREA [ONS 2004]⁷. Przeprowadzono także pogłębioną analizę korelacji, której efektem był wybór zmiennych pomocniczych do modeli wykorzystanych w poniższych estymatorach (fakt bycia osobą niepełnosprawną wyłącznie prawnie, płeć, miejsce zamieszkania (w podziale na miasto i wieś) oraz przynależność do grup wieku: 60-64, 70-74, 90-94).

Estymator bezpośredni Horwitza-Thompsona (HT):

$$\hat{Y}^{(1)} = \hat{Y}_d^{Direct} = \frac{1}{\hat{N}_d} \sum_{i \in u_d} w_{id} y_{id}, \quad (3)$$

$$\hat{N}_d = \sum_{i \in u_d} w_{id}; \quad w_{id} = 1 / \pi_{id},$$

gdzie: i – jednostka, d – domena, π_{id} – prawdopodobieństwo inkluzji (wylosowania jednostki i z domeny d).

Estymator Greg – uogólniony estymator regresyjny:

$$\hat{Y}^{(2)} = \hat{Y}_d^{Greg} = \frac{1}{\hat{N}_d} \sum_{i \in u_d} w_{id} y_{id} + \left(\bar{\mathbf{X}}_d - \frac{1}{\hat{N}_d} \sum_{i \in u_d} w_{id} \mathbf{x}_{id} \right)^T \hat{\boldsymbol{\beta}}, \quad (4)$$

gdzie: $\bar{\mathbf{X}}_d = (\bar{X}_{d,1}, \dots, \bar{X}_{d,p})^T$ jest wektorem p średnich zmiennych pomocniczych, $\mathbf{x}_{id} = (x_{id,1}, \dots, x_{id,p})^T$ jest wektorem jednostkowych obserwacji p zmiennych pomocniczych, a $\hat{\boldsymbol{\beta}} = \left(\sum_{i \in s_d} w_{id} \mathbf{x}_{id} \mathbf{x}_{id}^T \right)^{-1} \sum_{i \in s_d} w_{id} \mathbf{x}_{id} y_{id}$ jest wektorem oszacowań współczynników regresji liniowej za pomocą ważonej MNK.

⁵ Szacunki dla tej kategorii niepełnosprawności postulowano w artykule [Dehnel, Klimanek].

⁶ Takie uproszczenie wynikało z braku dostępności do odpowiednich danych.

⁷ Szczegółowy opis zastosowanych estymatorów zamieszczony jest w dokumentacji projektowej dostępnej pod adresem internetowym wskazanym w bibliografii.

Estymatory syntetyczne, z uwzględnieniem następujących modeli:

- liniowy model dwupoziomowy dla danych indywidualnych:

$$\hat{Y}^{(3)} = \hat{Y}_d^{Synth_A} = \bar{\mathbf{X}}_d^T \hat{\boldsymbol{\beta}}, \quad (5)$$

z wykorzystaniem modelu $y_{id} = x_{id}\boldsymbol{\beta} + u_d + e_{id}$, gdzie $u_d \sim \text{iid } N(0, \sigma_u^2)$, $e_{id} \sim \text{iid } N(0, \sigma_e^2)$, przy czym σ_u^2 i σ_e^2 oznaczają odpowiednio wariancję składnika resztowego na poziomie obszaru u_d i wariancję składnika resztowego na poziomie jednostki obserwacji e_{id} .

- liniowy model ze zmiennymi pomocniczymi na poziomie domen:

$$\hat{Y}^{(4)} = \hat{Y}_d^{Synth_B} = \bar{\mathbf{X}}_d^T \hat{\boldsymbol{\beta}} \quad (6)$$

z modelem $\bar{y}_d = \bar{\mathbf{X}}_d^T \boldsymbol{\beta} + u_d + \bar{e}_d$, gdzie $u_d + \bar{e}_d \sim \text{iid } N(0, \sigma_u^2 + \frac{\sigma_e^2}{n_d})$, przy czym n_d oznacza wielkość próby w domenie d .

Estymatory złożone EBLUP z uwzględnieniem następujących modeli:

- **EBLUP_A** – liniowy model dwupoziomowy dla danych indywidualnych,

$$\hat{Y}^{(5)} = \hat{Y}_d^{EBLUP_A} = w_d^{EBLUP_A} \hat{Y}_d^{Greg} + (1 - w_d^{EBLUP_A}) \hat{Y}_d^{Synth_A}, \quad (7)$$

gdzie: $w_d^{EBLUP_A} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + (\hat{\sigma}_e^2/n_d)}$ przy czym oszacowania $\hat{\sigma}_u^2$ i $\hat{\sigma}_e^2$ są uzyskiwane

iteracyjnie w procesie dopasowywania modelu dwupoziomowego.

- **EBLUP_B** – liniowy model ze zmiennymi pomocniczymi na poziomie domen

$$\hat{Y}^{(6)} = \hat{Y}_d^{EBLUP_B} = w_d^{EBLUP_B} \hat{Y}_d^{Direct} + (1 - w_d^{EBLUP_B}) \hat{Y}_d^{Synth_B}, \quad (8)$$

gdzie: $w_d^{EBLUP_B} = \frac{\hat{\sigma}_u^2}{\hat{\sigma}_u^2 + (\hat{\sigma}_e^2/n_d)}$, a oszacowania $\hat{\sigma}_u^2$ i $\hat{\sigma}_e^2$ są uzyskiwane iteracyj-

nie w procesie dopasowywania modelu dwupoziomowego.

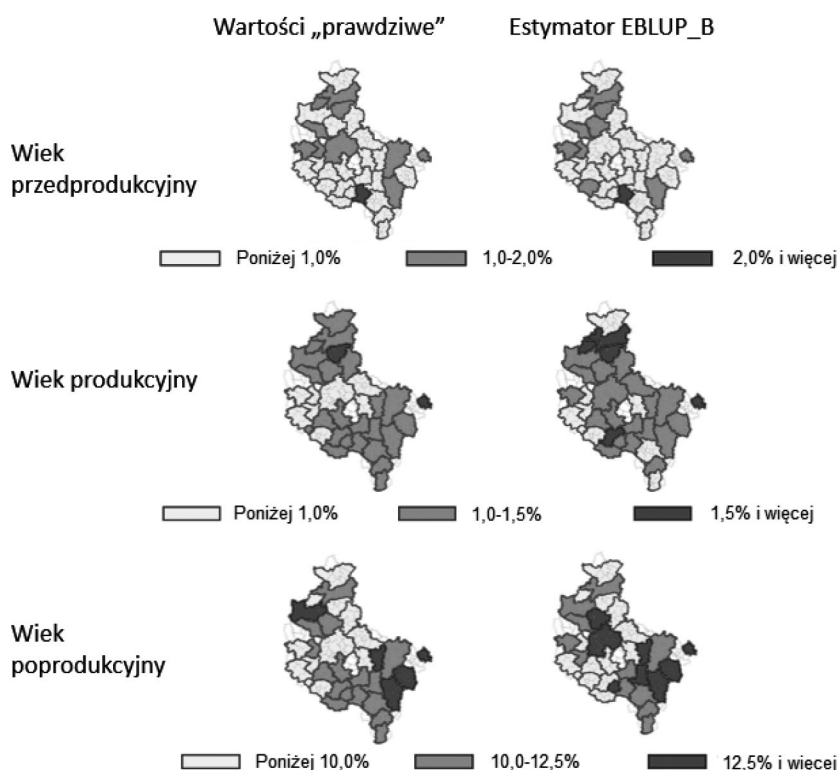
4. Wyniki badania

Natężenie zjawiska niepełnosprawności biologicznej nie jest jednakowe w analizowanych obszarach rynków pracy województwa wielkopolskiego bez względu na wiek ekonomiczny ludności. W przypadku wieku przedprodukcyjnego największy udział osób niepełnosprawnych wyłącznie biologicznie w ludności w wieku 15 lat i więcej występuje w obszarze rynku pracy Krotoszyna (południowa część województwa wielkopolskiego) i stanowi on ok. 2%. Najmniejszy udział niepełnosprawnych wyłącznie biologicznie w tej kategorii wieku ekonomicznego ludności wystę-

puje w obszarze rynku pracy Grodziska Wielkopolskiego (0,5%), znajdującym się w zachodniej części województwa.

Jeszcze niższymi udziałami osób niepełnosprawnych wyłącznie biologicznie w ludności w wieku 15 lat i więcej charakteryzuje się zbiorowość w wieku produkcyjnym. W tym przypadku najwyższe wskaźniki uzyskują obszary rynku pracy Chodzieży i Kłodawy (odpowiednio 1,8% i 1,6%). Najniższymi wskaźnikami charakteryzowały się obszary rynku pracy Grodziska Wielkopolskiego, Buku i Opalenicy (ok. 0,7%).

Wyraźnie wyższy był udział osób niepełnosprawnych wyłącznie biologicznie w wieku poprodukcyjnym i wahał się od 6,3% w obszarze rynku pracy Wolsztyna do 14,7% w obszarze rynku pracy Kłodawy. Wysokim udziałem osób niepełnosprawnych w tej grupie wieku ekonomicznego charakteryzowały się także obszary rynków pracy Kalisza, Turku i Słupcy w południowo-zachodniej Wielkopolsce oraz Czarnkowa, Trzcianki i Wielenia w północno-zachodniej części województwa (por. rys. 2).

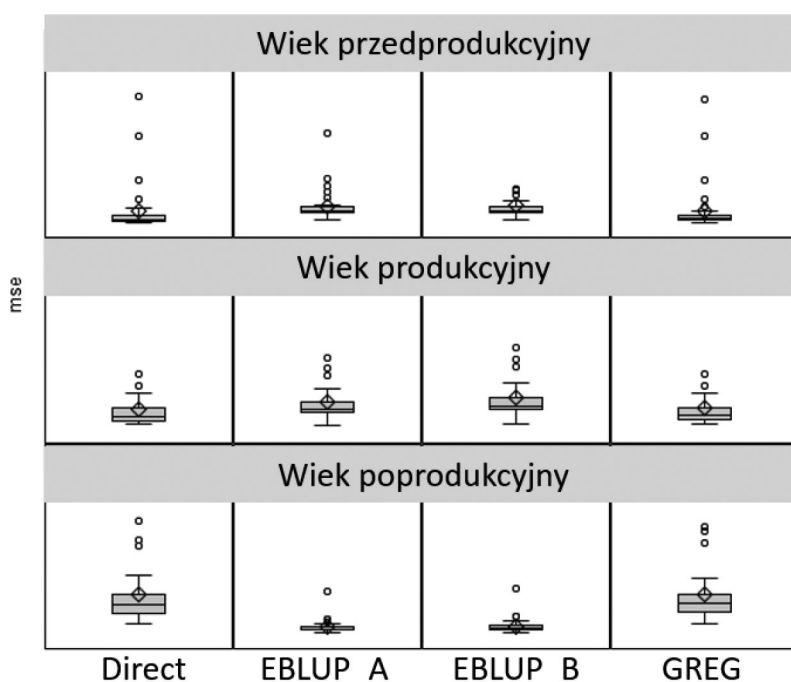


Rys. 2. Wybrane oszacowania odsetka osób niepełnosprawnych wyłącznie biologicznie dla zdefiniowanych domen w województwie wielkopolskim w 2002 roku

Źródło: opracowanie własne.

5. Ocena jakości szacunków

Chociaż wydaje się, że estymator EBLUP_B charakteryzuje się niewielkim obciążeniem empirycznym – rys. 2, to jednak analiza błędu średniokwadratowego analizowanych estymatorów (rys. 3) wskazuje, że jedynie w przypadku wieku poprodukcyjnego można mówić o wyraźnej poprawie jakości estymacji opartej na modelu w stosunku do estymacji bezpośredniej. W przypadku dwóch pozostałych kategorii wieku ekonomicznego estymacja z wykorzystaniem modeli nie daje poprawy jakości w znakomitej większości przypadków.



Rys. 3. Kształtowanie się błędu średniokwadratowego (MSE) wybranych estymatorów

Źródło: opracowanie własne.

6. Zakończenie

Przeprowadzone badanie symulacyjne pokazało, że chociaż EBLUP_B jest estymatorem charakteryzującym się niskim obciążeniem empirycznym w porównaniu z estymatorem bezpośrednim, to jednak nie wykazało istotnej poprawy precyzji estymatorów EBLUP_A i EBLUP_B, będących kombinacjami liniowymi odpowiednich estymatorów bezpośrednich (DIRECT, GREG) i syntetycznych (SYNTH_A

i SYNTH_B). Głównym powodem może być dość liczna próba w estymowanych przekrojach, która jest wystarczająca, aby z powodzeniem stosować wyłącznie estymator bezpośredni. W związku z tym planowane są dalsze prace z wykorzystaniem estymatorów pośrednich dla bardziej szczegółowych przekrojów (poziom powiatu lub niższy), charakteryzujących się niewielką liczebnością próby, gdzie oczekuje się uzyskania większego zysku na precyzji w porównaniu z estymatorami bezpośrednimi. W dalszych pracach rozważa się także ocenę wpływu zmian delimitacji obszarów rynków pracy na ocenę jakości estymacji.

Literatura

- Coombes M., Bond S., 2007, *Travel-to-Work Areas: the 2007 review*, Office for National Statistics, https://ec.europa.eu/eurostat/cros/content/labour-market-areas_en (14.03.2016).
- Dehnel G., Klimanek T. (w druku), *Niepełnosprawność w Narodowych Spisach Powszechnych 2002 i 2011 – porównanie zakresu informacyjnego*, Acta Universitatis Lodzianensis. Folia Oeconomica.
- ONS, 2004, EURAREA Project, Enhancing Small Area Estimation Techniques to Meet European Needs, Project documentation, <http://www.ons.gov.uk/ons/guide-method/method-quality/general-methodology/spatial-analysis-and-modelling/eurarea/index.html> (14.01.2012).
- Slany K., 2014, *Osoby niepełnosprawne w świetle Narodowego Spisu Powszechnego Ludności i Mieszkań z 2011 r. – wybrane aspekty*, Niepełnosprawność – zagadnienia, problemy, rozwiązania, nr II(11).
- WHO, 2014, Secretariat, WHO global disability action plan 2014–2021: better health for all people with disability, http://apps.who.int/iris/bitstream/10665/199544/1/9789241509619_eng.pdf (2.12.2015).