

Elżbieta Gołata, Jan Paradysz

Akademia Ekonomiczna w Poznaniu

DEMOGRAFICZNE KORZENIE I PERSPEKTYWY STATYSTYKI MAŁYCH OBSZARÓW

1. Wstęp

Podręczniki i publikacje przeglądowe omawiające zagadnienia statystyki małych obszarów SMO podkreślają jej demograficzną genezę. Autorzy tych opracowań ograniczają się w zasadzie do wskazania na szacunki liczby ludności w okresach międzyspisowych. W niniejszym opracowaniu przedstawiono przyczyny warunkujące rozwój metodologii estymacji pośredniej na gruncie badań ludnościowych, obszernie omówiono zakres dotychczasowych aplikacji oraz wskazano perspektywy dalszych zastosowań. Zwrócono uwagę, iż w dobie szczupłości zasobów finansowych, również w ramach innych dziedzin nauki, wykorzystanie idei integracji danych z różnych źródeł może przyczynić się do poprawy jakości wnioskowania statystycznego.

2. Uwarunkowania rozwoju metod estymacji pośredniej w badaniach demograficznych

W opinii wielu badaczy estymacja dla małych obszarów ma dłuższą historię aniżeli sama nazwa nowej dziedziny badań. Jak zauważają M. Gosh i J.N.K. Rao, pierwsze prace w zakresie estymacji dla małych domen dotyczyły szacunku liczby ludności w okresach międzyspisowych dla małych jednostek terytorialnych [26]. Podkreślają oni dorobek takich demografów, jak R.C. Schmidt, S. Shapiro i J. Schachter, W.H. Grabill i Lee-Jay Cho (por. [54; 55; 27]). Powstających w owym czasie prac nie określano jeszcze mianem statystyki małych obszarów (*small area statistics*). Jednymi z pierwszych, którzy dostrzegli powstanie nowej dziedziny badań, byli L. Kish i N.J. Purcell (por. [48]). Czynniki, które zadecydowały o tym, że demografowie zostali pionierami technik estymacji pośredniej to:

- 1) jakość źródeł danych o ludności
- a) brak informacji, w szczególności o strukturach ludności,

- b) niekompletna statystyka ludności;
- 2) działanie czynników perturbacyjnych;
- 3) różnorodność źródeł
- a) narodowe Spisy Powszechne,
- b) bieżąca rejestracja ruchu naturalnego,
- c) badania specjalne;
- 4) uporządkowana i spójna metodologia badań demograficznych.

Panuje powszechna opinia, że statystka ludności jest bardzo rzetelna i godna zaufania, a zakres informacji wyczerpujący. Tymczasem jakość danych demograficznych jest bardzo zróżnicowana zarówno w przekroju terytorialnym, jak i w zależności od rodzaju zdarzeń. Pomiedzy krajami występuje silne zróżnicowanie statystyki ludności uzależnione od tradycji i poziomu badań. W krajach uznawanych za rozwinięte znajdujemy również wiele ograniczeń „statystycznych” uniemożliwiających przeprowadzenie bezpośrednich analiz.

Do estymacji pośredniej uciekamy się w demografii w przypadku braku rejestracji zdarzeń demograficznych lub braku odpowiednich struktur demograficznych (np. według płci, wieku, miejsca zamieszkania). Z pierwszym przypadkiem mamy do czynienia w wielu krajach afrykańskich. Tak było również w Polsce lat pięćdziesiątych, gdy po wielu perturbacjach dziejowych nie zdołano jeszcze zaprowadzić bieżącej, powszechnej, kompletnej i odpowiednio ukierunkowanej rejestracji ruchu ludności. Graniczną datą, od której dysponujemy w Polsce kompletną rejestracją urodzeń ogółem, zgonów, małżeństw i rozwodów, jest rok 1950, a np. dla urodzeń małżeńskich dopiero rok 1975. W przypadku zawierania i rozpadu małżeństw w zasadzie do dzisiaj nie dysponujemy odpowiednio dobrymi informacjami. W tych wszystkich przypadkach pewną nadzieję mogą stworzyć badania specjalne przeprowadzone metodą reprezentacyjną oraz różne techniki estymacji pośredniej.

Podobnie utrudnione jest bieżące korzystanie z informacji o urodzeniach według wieku i wykształcenia matek, gdyż nie są znane struktury kobiet według wieku i wykształcenia. Można mieć przy tym poważne wątpliwości co do możliwości wykonania tego rodzaju analiz nawet w okresach spisowych. Wątpliwości biorą się stąd, że w badaniach specjalnych, do których należą także spisy ludności, częstym przypadkiem jest zawyżanie poziomu wykształcenia [43]. Również w tych sytuacjach lepszym rozwiązaniem może okazać się estymacja pośrednia [13].

Wiele trudności pojawia się przy pomiarze migracji ludności, przede wszystkim dotyczy to migracji zagranicznych, ale także wewnętrznych. Niekompletność i niedoskonałość tej statystyki w sposób zarówno bezpośredni jak i pośredni wpływa na struktury ludności oraz szacowane na ich podstawie mierniki natężenia procesów demograficznych. Selektywność migracji ze względu na wiek, płeć czy miejsce zamieszkania powoduje, iż rozkład błędów jest nierównomierny. W szczególności można poruszyć problem badania wpływu nie zarejestrowanej emigracji zagranicznej Polaków na stan i strukturę ludności w kraju. Na skalę problemów wskazują wyniki badań B. Sakson [51] czy wcześniejszych opracowań M. Kędel-

skiego [32] i M. Okólskiego [41]. Pokazują one braki oficjalnej statystyki migracji międzynarodowych, a także skalę błędów oszacowania stanu ludności czy współczynników urodzeń i zgonów. Po przystąpieniu Polski do Unii Europejskiej nastąpiła kolejna fala bardzo licznej emigracji młodych ludzi do krajów zachodnich, które otworzyły swoje rynki pracy. W efekcie struktury ludności określone na podstawie wyników NSP'2002 prawdopodobnie uległy znacznej deformacji. Problemem jest również pomiar migracji wewnętrznych związanych m.in. z coraz gwałtowniejszym w ostatnich latach rozwojem budownictwa na terenach podmiejskich otaczających duże aglomeracje.

Migracje ludności są uwzględniane w analizach demograficznych jako jeden z tzw. czynników perturbacyjnych, tj. takich, które powodują zakłócenia. W klasycznej analizie demograficznej oprócz ruchów wędrowkowych do czynników takich zalicza się zgony. Uwzględnienie zjawisk perturbacyjnych jest niezmiernie ważne szczególnie w przypadku przeprowadzania analizy na podstawie informacji retrospektywnej. Powodują one wykluczenie z obserwacji pewnej części badanej populacji. Migracje mogą ponadto wpływać na odroczenia i rekompensaty zdarzeń, których realizacja zależy od woli jednostki. W tej sytuacji demografia wypracowała system współczynników, które pozwalają na oszacowanie czystej postaci zjawiska. Oparte są one na założeniach sformułowanych przez L. Henry'ego [30] w postaci warunków, które określił mianem niezależności i ciągłości¹. Przykłady takiego podejścia są przedstawiane w podręcznikach demografii [33; 24] oraz stosowane w przeprowadzanych analizach [43].

Czynnikiem perturbacyjnym o nieco innym charakterze jest mała liczebność populacji. Z sytuacją taką spotykają się demografowie przeprowadzający badania w ujęciu regionalnym. Mała liczebność populacji zamieszkującej badane terytorium powoduje duże zróżnicowanie liczby zdarzeń demograficznych, takich jak np. urodzenia czy zgony. Dodatkowo zróżnicowanie to potęgowane jest przez wpływ struktury ludności według wieku, płci czy stanu cywilnego. Mała liczebność próby powodująca dużą wariancję estymatorów bezpośrednich jest też jednym z głównych czynników uzasadniających potrzebę stosowania metod estymacji pośredniej.

Demografowie świadomi konieczności zapobiegania oddziaływaniu czynników perturbacyjnych stosowali różnorodne rozwiązania w celu zniwelowania skutków dużej zmienności liczby zdarzeń uniemożliwiającej przedstawienie wiarygodnych analiz czy prognoz ludności. Jednym z najprostszych sposobów jest obliczanie średniej wielkości współczynników dla kilku lat kalendarzowych (np. 2 czy 3 lat). Przykładowo E. Gołata [25] w analizie terytorialnego zróżnicowania czynników płodności kobiet w Polsce przeprowadzonej w przekroju makro i mikro, tj. województw oraz miast i gmin wiejskich między województwami i wewnątrz

¹ Warunek niezależności zakłada, że tej części generacji, która wyszła z pola obserwacji, można przypisać taki sam przebieg zjawiska w czystej postaci, jak tej części populacji, która podlegała obserwacji do końca. W wypadku zjawisk niepowtarzalnych dodatkowo zakłada się, że analizowane zjawisko nie wpływa na zjawisko zakłócające, np. urodzenie dziecka nie wpływa na umieralność kobiet w wieku zdolności rozrodczej.

województw, uwzględniła współczynniki średnie z 2 lat kalendarzowych. Przy dużym rozdrobnieniu terytorialnym podejście takie pozwoliło na pewną eliminację wahań przypadkowych.

Demografowie, świadomi konsekwencji przeprowadzania badania dla mało licznej populacji, czerpali z dodatkowej wiedzy, czyli „pożyczali moc”. Określenie to dobrze ujmuje istotę statystyki małych obszarów. W demografii podejście takie polegało m.in. na uwzględnianiu wiedzy o kształtowaniu się procesów ludnościowych w innych populacjach rozwijających się w zbliżonych warunkach w przeszłości, wykorzystywaniu różnych dostępnych źródeł informacji, analizowaniu rozwiązań modelowych. Źródła informacji o ludności należą do najstarszych, co sugeruje dobrą, wypracowaną przez lata doświadczeń jakość i dopracowaną metodologię. Warto podkreślić także dużą różnorodność źródeł danych demograficznych. Wymienić tutaj można przede wszystkim powszechne spisy ludności, bieżącą ewidencję ruchu naturalnego i wędrownego ludności, a także wiele badań specjalnych. Niejako więc w sposób naturalny demografowie wypracowali nawyk porównywania danych, dociekania przyczyn różnic, sprawdzania definicji, pojęć, sposobu zbierania informacji. Za wkład demografii w rozwój metod estymacji pośredniej można więc uznać badanie dostępnej infrastruktury statystycznej i jej wykorzystanie w celu poprawy jakości szacunku. Przede wszystkim należy podkreślić wkład w coraz powszechniejsze wykorzystanie rejestrów administracyjnych.

Demografia jest dziedziną badań, w której zastosowanie technik pośrednich wydaje się „naturalne”. Odpowiednio długo prowadzone bazy danych ludnościowych, odpowiadające systemowi wskaźników demograficznych, oraz precyzja ujęcia kohortowego i transversalnego umożliwiły rozwój skomplikowanych modeli i metod analizy.

3. Wkład demografii w rozwój statystyki małych obszarów

Demograficzne „korzenie” estymacji pośredniej są podkreślane przez większość statystyków zajmujących się tą problematyką. Wskazać tutaj można przede wszystkim jedyną w tym zakresie monografię J.N.K. Rao [49], w której wyodrębniono rozdział zatytułowany *Tradycyjne metody demograficzne* oraz wcześniejsze prace przeglądowe M. Gosh, J.N.K. Rao i J.N.K. Rao (por. [26; 50]). Rao podkreśla zastosowania metod estymacji pośredniej w szacowaniu liczby ludności dla małych jednostek terytorialnych oraz różnego rodzaju korekty sporządzane w okresach pospisowych. Omawiając podejście stosowane w demografii, wskazuje on na klasyfikację zaproponowaną przez Purcella i Kisha [48]: symptomatyczne techniki bilansowe (*Symptomatic Accounting Techniques* – SAT), procedury regresji symptomatycznej (*Regression Symptomatic Procedures*) oraz metody podwójnego systemu szacunku (*Dual-System Estimation*). Wśród pierwszych opracowań w zakresie metod SAT, J.N.K. Rao odwołuje się przede wszystkim do prac Bogue’a, Bogue’a i Duncana [6; 7] oraz opracowań metodologicznych Biura

Spisów USA [61]. W dziedzinie procedur regresji symptomatycznej J.N.K. Rao podaje przykłady opracowań Ericksena oraz Ericksena i Kadane [18-20]. Szczegółowy opis metod stosowanych w demografii zawierają również prace Purcella i Kisha oraz Zideka [48; 62]). Systematykę demograficznych metod estymacji pośredniej można przestawić za J.N.K. Rao następująco:

1) symptomatyczne techniki bilansowe (*Symptomatic Accounting Techniques – SAT*)

- a) metoda współczynników ruchu naturalnego (*Vital Rates Method – VR*),
- b) metoda złożona (*Composite Method*),
- c) metoda składnikowa (*Component Method – CM*),
- d) metoda gospodarstw domowych (*Housing Unit Method – HU*);

2) procedury regresji symptomatycznej (*Regression Symptomatic Procedures*)

a) metoda regresji (korelacji) ilorazowej i różnicowej (*Ratio Correlation and Difference Correlation Methods*),

b) metoda regresji próbkowej (*Sample Regression Method*);

3) metody podwójnego systemu szacunku (*Dual-system of Total Population*)

a) metoda podwójnego szacunku (*Dual-system Method*),

b) badania kontrolne (*Post-enumeration Surveys*).

Demografowie są również świadomi szerokiego stosowania metodologii estymacji pośredniej w badaniach ludnościowych. Wyrazem tego jest m.in. klasyfikacja metod szacunków pośrednich zawarta w ostatnio opublikowanym podręczniku metodologicznym dla ekspertów ONZ (*Manual X: Indirect techniques for demographic estimation*, UN 2005). Przedstawiona tam klasyfikacja jest następująca:

1) estymacja pośrednia całkowita,

2) estymacja pośrednia częściowa,

3) estymacja pośrednia na podstawie modeli demograficznych.

W kolejnych dwóch punktach omówione zostaną demograficzne podstawy statystyki małych obszarów według ujęcia statystyków (klasyfikacja według Rao) jak i od strony demografów (klasyfikacja ekspertów ONZ).

3.1. Tradycyjne metody demograficzne według J.N.K. Rao

Cechą charakterystyczną technik określanych mianem SAT jest wykorzystanie bieżących danych rejestrów administracyjnych w połączeniu z odpowiednimi informacjami ostatniego spisu powszechnego. Zróżnicowane informacje z rejestrów administracyjnych wykorzystywane w USA zawierały takie zmienne „symptomatyczne”, jak: liczba urodzeń i zgonów, liczba nowo powstałych budynków, listy zapisu dzieci do szkoły, których zróżnicowanie jest silnie uwarunkowane zmianami liczby ludności.

Ideę szacunku liczby ludności wykorzystującą symptomatyczne techniki bilansowe można zaprezentować na przykładzie **metody współczynników ruchu naturalnego**. Stosuje się tutaj tylko informacje o urodzeniach i zgonach dla danego roku kalendarzowego, przy czym zmienne te są wykorzystywane jako informacje

o charakterze symptomatycznym, a nie jako komponenty zmian liczby ludności. Ponadto zakłada się dostępność informacji z oficjalnego źródła o bieżącej liczbie ludności w większym obszarze zawierającym interesujący nas mały obszar.

Przy znajomości wielkości b_t, d_t – urodzeń i zgonów w roku bieżącym t dla małego obszaru, wielkości stopy urodzeń i zgonów są szacowane według następującego wzoru [46]:

$$\hat{r}_{1t} = r_{10} \cdot \frac{\hat{R}_{1t}}{R_{10}} = r_{10} \cdot \hat{\phi}_1 \text{ oraz } \hat{r}_{2t} = r_{20} \cdot \frac{\hat{R}_{2t}}{R_{20}} = r_{20} \cdot \hat{\phi}_2. \quad (1)$$

gdzie: $\hat{r}_{1t}, \hat{r}_{2t}$ – oszacowania współczynników urodzeń i zgonów dla małego obszaru w roku bieżącym t , r_{10}, r_{20} – surowe współczynniki urodzeń i zgonów (odpowiednio 1 i 2) w małym obszarze w roku spisowym 0, $\hat{R}_{1t}, \hat{R}_{2t}, R_{10}, R_{20}$ – surowe współczynniki zgonów i urodzeń w dużym obszarze w roku bieżącym t (oszacowania) oraz w roku spisowym 0:

$$\hat{R}_{1t} = \frac{B_t}{\hat{P}_t}, \quad \hat{R}_{2t} = \frac{D_t}{\hat{P}_t}, \quad R_{10} = \frac{B_0}{P_0}, \quad R_{20} = \frac{D_0}{P_0}.$$

Metoda VR polega na szacowaniu współczynników urodzeń i zgonów dla małego obszaru w roku bieżącym t , przyjmując pewną stopę zmian, o której zakłada się, że jest taka sama, jak w dużym obszarze. To właśnie dzięki wykorzystaniu informacji o liczbie ludności dla dużego obszaru z dodatkowego

źródła można określić wartości współczynników $\hat{R}_{1t}$ oraz $\hat{R}_{2t}$ i relację: $\hat{\phi}_1 = \frac{\hat{R}_{1t}}{R_{10}}$

oraz $\hat{\phi}_2 = \frac{\hat{R}_{2t}}{R_{20}}$. Wówczas oszacowanie bieżącej liczby ludności w małym obszarze

możliwe jest jako:

$$\hat{p}_t = \frac{1}{2} \left(\frac{b_t}{\hat{r}_{1t}} + \frac{d_t}{\hat{r}_{2t}} \right). \quad (2)$$

Poprawność szacunków otrzymanych z zastosowaniem metody VR zależy od prawdziwości przyjętych założeń dotyczących współczynników ϕ_1, ϕ_2 służących do aktualizacji stóp urodzeń i zgonów dla małego obszaru oraz szacunku liczby ludności dla dużego obszaru zawierającego ów mały obszar. Założenie takie jest oczywiście co najmniej dyskusyjne.

Metoda złożona jest udoskonaleniem metody współczynników ruchu naturalnego VR. Wykorzystuje się tutaj metodę VR w celu niezależnego wyznaczenia liczby ludności według grup płci i wieku, a następnie sumuje się owe „składniki”, aby otrzymać „złożony szacunek” bieżącej liczby ludności. Metoda ta wymaga

znajomości liczby urodzeń i zgonów według wyróżnionych grup (wieku) oraz liczby ludności w wyróżnionych grupach w większym obszarze. Można tutaj wykorzystywać alternatywne źródła informacji. Przykładowo w USA ludność w wieku szkolnym 5-14 lat jest szacowana na podstawie list zapisów do szkół oraz współczynników zapisów do szkół.

W przeciwieństwie do poprzednich, **metoda składnikowa** dostarcza szacunku liczby ludności, uwzględniając dodatkowo oszacowanie migracji netto. Dysponując informacją spisową o liczbie ludności p_0 w małym obszarze oraz przyjmując, że $b_{0,t}$, $d_{0,t}$ oraz $m_{0,t}$ oznaczają odpowiednio liczbę urodzeń, zgonów i migracji netto² w domenie w przedziale czasu $[0, t]$, można skonstruować następujące równanie bilansowe bieżącej liczby ludności w małym obszarze:

$$\hat{p}_t = p_0 + b_{0,t} - d_{0,t} + m_{0,t}. \quad (3)$$

Rejestracja urodzeń i zgonów jest w większości rozwiniętych państw świata powszechna i w miarę kompletna. Jednak wiarygodne informacje dotyczące migracji netto zazwyczaj nie są dostępne. W USA migracje netto są dzielone na migracje cywilne i „militarne”, przy czym informacje o migracjach „militarnych” są dostępne w rejestrach administracyjnych. Do szacunku migracji cywilnych stosuje się metodę składnikową II (CM-II)³ wykorzystującą listy zapisów do szkół, podczas gdy metoda Rejestrów Administracyjnych (*Administrative Records AR*) opiera się na danych podatkowych. W Kanadzie informacje o emigracjach oraz migracjach wewnętrznych szacowane są również na podstawie danych podatkowych dla poszczególnych małych obszarów.

Pospisowe szacunki ludności według płci i wieku oraz stanu cywilnego są przeprowadzane w Kanadzie z zastosowaniem kohortowej metody składnikowej, która jest analogiczna do metody składnikowej, ale wprowadza odpowiednie modyfikacje wynikające z dostosowania do naturalnego ujęcia wzdłużnego.

W **metodzie gospodarstw domowych**⁴ oraz miejsc zbiorowego zamieszkania (*housing units & group quarters*) bieżąca liczba ludności jest przedstawiona jako:

$$p_t = (h_t)(pph_t) + gq_t, \quad (4)$$

gdzie: h_t – liczba mieszkań zamieszkałych w momencie t , pph_t – średnia liczba osób w mieszkaniu w momencie t , gq_t – liczba osób w gospodarstwach zbiorowych w momencie t .

² Migracja netto $m_{0,t}$ to suma imigracji $i_{0,t}$ oraz migracji wewnętrznych netto $n_{0,t}$ minus liczba emigracji $e_{0,t}$.

³ Jest to późniejsza wersja metody CM-I.

⁴ Metoda gospodarstw domowych stosowana jest w USA do oszacowania ludności w przekroju *county* (odpowiednik gmin). Jest to metoda atrakcyjna, ale problemem jest uzyskanie niezbędnych informacji.

Wszystkie wielkości uwzględnione w powyższym równaniu bilansowym muszą być oszacowane. W szczególności wielkość h_t jest szacowana na podstawie informacji o zmianach liczby mieszkań określonych przez liczbę pozwoleń na budowę wydanych w okresie $[0, t]$ skorygowaną o czas potrzebny do zakończenia budowy⁵. Ponadto uwzględnia się liczbę mieszkań zlikwidowanych w czasie jaki minął od ostatniego spisu. Wartości p_{ph_t} oraz gq_t szacuje się na podstawie p_{ph_0} ostatniego spisu, bądź poprzez ekstrapolację wielkości z dwóch ostatnich spisów.

W opinii Markera (por. [38]) większość ze wskazanych wyżej metod określić można mianem wielowymiarowej regresji liniowej. Metody regresji symptomatycznej również wykorzystują procedury wielowymiarowej regresji liniowej ze zmiennymi symptomatycznymi jako zmiennymi niezależnymi w równaniu regresji. Wskazać tutaj można **metodę regresji ilorazowej i różnicowej**. Przyjmijmy że dysponujemy informacjami o liczbie ludności oraz wartościach zmiennych symptomatycznych z dwóch kolejnych spisów powszechnych, oznaczonymi jako: p_{ia} (ludność w i -tej domenie w czasie a), s_{ija} (wartość j -tej zmiennej symptomatycznej $j (= 1, \dots, p)$ w i -tej domenie $i (= 1, \dots, m)$ w czasie a ($a (= 0, 1, t)$) oraz informacją o proporcji liczby ludności i zmiennej symptomatycznej w obszarze a do pewnego dużego obszaru zawierającego mały obszar. Względna zmianę⁶ wielkości zmiennej zależnej U_i w okresach między spisami dla obszaru i -tego można przedstawić w formie zależności od zmian względnych wartości zmiennych symptomatycznych z_{ij} dla j -tej zmiennej w obszarze i -tym. Relację tę można wyrazić w postaci regresji wielowymiarowej:

$$U_i = \gamma_0 + \gamma_1 z_{i1} + \dots + \gamma_p z_{ip} + u_i, \quad (5)$$

gdzie: u_i – składnik resztowy z założenia nieskorelowany o wartości oczekiwanej równej zero i wariancji określonej jako σ_u^2 .

Oceny parametrów równania (5) otrzymuje się z zastosowaniem Klasycznej Metody Najmniejszych Kwadratów KMNK. Wówczas pospisowa zmiana wielkości Y otrzymywana jest w formie predykcji na podstawie równania

⁵ W literaturze proponowane były różne modyfikacje i udoskonalenia metody gospodarstw domowych. Przykładowo h_t może być uzyskiwane jako: $h_t = h_0 + rec_t - rec_0$, gdzie rec_t jest liczbą aktywnych (używanych) urządzeń elektrycznych w momencie x . Inną wersją opartą na takim samym źródle byłoby $h_t = (h_0 / rec_0) rec_t$.

⁶ W metodzie regresji stosunkowej do pomiaru zmian wykorzystuje się następujące wskaźniki wielkości stosunkowych: $U_i = (p_{i1} : P_1) / (p_{i0} : P_0)$, $z_{ij} = (s_{ij1} : S_{j1}) / (s_{ij0} : S_{j0})$. Natomiast metoda regresji różnicowej do pomiaru zmian wykorzystuje następujące różnice odpowiednich wielkości stosunkowych: $U_i = p_{i1} : P_1 - p_{i0} : P_0$ oraz $z_{ij} = s_{ij1} : S_{j1} - s_{ij0} : S_{j0}$.

$$Y_i^* = \hat{\gamma}_0 + \hat{\gamma}_1 x_{i1} + \dots + \hat{\gamma}_p x_{ip} \quad (6)$$

z wykorzystaniem znanych informacji o zmianach w postaci stosunków lub różnic⁷.

Wszystkie problemy, jakie pojawiają się przy szacowaniu regresji, są rozwiązywane tradycyjnymi metodami stosowanymi w ekonometrii. Przykładowo w Statistics Canada w walce z heteroskedastycznością czy aby uporać się z problemem współliniowości (*multicollinearity*) stosuje się metodę regresji ważonej oraz metodę regresji grzbietowej (*ridge regression*). W Kanadzie stosowano metodę regresji do otrzymania wstępnych szacunków ludności dla małych obszarów, jednak ostateczne szacunki były oparte na metodzie składnikowej.

Omówione metody regresji są oparte na współczynnikach regresji wyznaczonych na podstawie danych z dwóch ostatnich spisów. Oczywiście szacując liczbę ludności w okresie wykraczającym poza przedział czasu między dwoma spisami, trzeba liczyć się z możliwą zmianą obserwowanej relacji, co w konsekwencji prowadzi do obciążeń predykcji Y_i^* . Oznacza to, że bieżąca relacja: $Y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + v_i$ może zawierać współczynniki regresji β_j istotnie różne od γ_j z równania (5). **Metoda regresji próbkowej** zaproponowana przez Ericksena [18] pozwala wyeliminować powyższy problem dzięki konstrukcji modelu odzwierciedlającego aktualną relację na podstawie danych badania reprezentacyjnego. Załóżmy, że badanie częściowe dostarcza szacunków $\hat{Y}_i$ dla k spośród m małych obszarów. Równanie regresji dopasowane do danych z próby losowej dostępnej dla k małych obszarów można zastosować w celu oszacowania wartości Y_i dla wszystkich obszarów (zarówno tych wylosowanych jak i nie wylosowanych do próby), wykorzystując znaną informację o zmianach zmiennych symptomatycznych:

$$\tilde{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \dots + \hat{\beta}_p x_{ip}. \quad (7)$$

Bieżące szacunki p_{it} są otrzymane podobnie jak poprzednio – podstawiając $\tilde{Y}_i$ w miejsce Y_i^* . Ericksen porównał średni błąd względny regresji próbkowej oraz w metodzie regresji stosunkowej na podstawie danych spisowych z 1970 r. Rezultaty te wskazały, że redukcja wartości błędu średniego jest raczej niewielka. Jednakże w przypadku dużych błędów (powyżej 10%) redukcja MSE⁸ w porównaniu z metodą regresji ilorazowej była znaczna.

Pod koniec lat siedemdziesiątych G. Feeney i R. Chambers [9] zaproponowali technikę estymacji SPREE – „zachowującą struktury”. Wykorzystuje się w niej

⁷ W metodzie regresji stosunkowej otrzymujemy: $Y_i = (p_{it} : P_t) / (p_{i1} : P_1)$ oraz $x_{ij} = (s_{ijt} : S_{jt}) / (s_{ji1} : S_{j1})$ natomiast w metodzie regresji różnicowej: $Y_i = p_{it} : P_t - p_{i1} : P_1$ oraz $x_{ij} = s_{ijt} : S_{jt} - s_{ji1} : S_{j1}$.

⁸ MSE oznacza w estymacji *Mean Square Error*, czyli średni błąd kwadratowy.

proporcjonalne dopasowanie wielkości w tablicach wielowymiarowych do rozkładów brzegowych. Jako wielkości brzegowe przyjmuje się oceny estymatorów bezpośrednich (tradycyjnych). Natomiast do oszacowania wielkości warunkowych wykorzystuje się technikę estymacji syntetycznej.

Powszechnie uznaje się, że estymacja pośrednia powinna się opierać na modelu ujmującym relację między szacowaną zmienną w poszczególnych domenach a dodatkowymi, uzupełniającymi informacjami pochodzącymi z alternatywnych źródeł. W sytuacji niewielkiej liczebności próby z badanej domeny wykorzystanie dodatkowych informacji traktuje się jako sposób na poprawę estymacji. R.E. Fay i R.A. Herriot [22], dokonując szacunku dochodu na osobę w przekroju lokalnych jednostek samorządowych⁹, spostrzegli duże różnice w jakości dopasowania. Wśród 39 tys. jednostek samorządowych (*civil divisions*), dla których dokonali szacunku, ok. 15 tys. było zamieszkiwanych przez mniej niż 500 osób. Dla jednostki liczącej ok. 500 mieszkańców względny błąd szacunku wynosił 13%, podczas gdy dla jednostki zamieszkiwanej przez ok. 100 osób wzrastał do ponad 30%. Autorzy ci zaproponowali zastosowanie szacunków uwzględniających procedurę Jamesa-Steina (1961), łączącą estymację na podstawie próby z estymacją opartą na dodatkowych źródłach. Wśród innych najbardziej znanych zastosowań modelu mieszanego można wskazać opracowania Ericksena i Kadane [19; 20], dotyczące międzypisowych szacunków ludności.

3.2. Klasyfikacja metod estymacji pośredniej w badaniach demograficznych według ekspertów ONZ

Powyższe przykłady wyraźnie wskazują, jakoby metodologia stanowiąca podwaliny estymacji pośredniej rozwinęła się w związku z szacunkiem liczby ludności dla małych jednostek terytorialnych w okresach pospisowych¹⁰. Zwróćmy uwagę, że klasyfikacja demograficznych metod SMO zaproponowana przez Rao [49] sprowadza się do szacowania liczby ludności. Można sądzić, że wynika to z podstawowego charakteru tej zmiennej jako wielkości niezbędnej w większości zastosowań procedur estymacyjnych. Ponadto jest to często wykorzystywana zmienna pomocnicza w innych (poza demograficznymi) zastosowaniach estymacji pośredniej. W klasyfikacji tej uwzględniono specyficzne metody szacowania liczby ludności, ale też do tej podstawowej wielkości się ograniczono. Naszym zdaniem, demograficzne podwaliny statystyki małych obszarów są znacznie głębsze. Odwołując się do opinii ekspertów ONZ, można rozróżnić 3 rodzaje estymacji pośredniej w badaniach demograficznych (por. [60], t. 3, rozdz. 11): całkowitą, częściową oraz estymację na podstawie modeli demograficznych. Do estymacji

⁹ Szacunek ten został przeprowadzony na zlecenie Departamentu Skarbu USA, na podstawie mikrospisu (ok. 20-procentowa próba losowa) z 1970 r. oraz rejestrów administracyjnych [22].

¹⁰ Zdaniem J. Paradyśa [44] do kategorii estymacji pośredniej można zaliczyć również szacunek liczby ludności na podstawie różnych źródeł podatkowych (podymne, świętopietrze) [36, s. 233-248].

pośredniej uciekamy się w przypadku braku rejestracji zdarzeń demograficznych lub braku odpowiednich struktur demograficznych.

Całkowita estymacja pośrednia w przypadku np. analizy płodności opiera się na 4 rodzajach danych oraz odpowiadających im procedurach:

- 1) informacjach o dzieciach „kiedykolwiek” urodzonych (*children-ever-born* CEB),
- 2) informacjach o ludności według płci i wieku,
- 3) szacunkach współczynników płodności na podstawie modeli regresji,
- 4) empirycznych zależnościach między miarami płodności umożliwiającymi przechodzenie od jednych ujęć do drugih.

Przez dzieci „kiedykolwiek” urodzone rozumie się informacje o liczbie żywo urodzonych dzieci przez daną generację kobiet bez precyzowania czasu ich przyjścia na świat lub ich wieku (por. np. dane GUS z 1971 r.). Mając tak szczegółową informację, wzmocnioną czasami danymi o liczebności generacji, można oszacować cząstkowe współczynniki płodności ogólnej, a niekiedy także płodności małżeńskiej. W literaturze przedmiotu zaproponowano wiele metod, które taką estymację umożliwiają. Należą do nich m.in. techniki:

- 1) E. Arriagi dla płodności ogólnej,
- 2) L. Henry’ego dla płodności małżeńskiej,
- 3) G. Feeneya.

Technika E. Arriagi opiera się na informacjach o liczbie „kiedykolwiek” urodzonych dzieci w 2 okresach, np. w czasie 2 spisów [60], oraz na założeniu stałości procesów demograficznych, tj. jednokierunkowych ich zmian między obydwoma spisami. Na zupełnie innych zasadach oparta jest technika L. Henry’ego [30], chociaż prowadzi do podobnych rezultatów, tj. oszacowania współczynników dzietności teoretycznej.

Wyjściowym źródłem informacji w **metodzie L. Henry’ego** są ogólne informacje o liczbie urodzonych dzieci według kolejności urodzenia oraz ewentualnie statystyka „bieżących” odstępów między urodzeniami. W przypadku braku statystyki bieżących odstępów między urodzeniami L. Henry (por. [30]) zaproponował tzw. typowe odstępów intergenetyczne. Na podstawie stosunkowo rzadkich na początku lat pięćdziesiątych danych o odstępach intergenetycznych L. Henry [30, s. 126-128] opracował tablice typowych rozkładów pomiędzy urodzeniami¹¹. W tym celu wykorzystał wyniki badań specjalnych przeprowadzonych we Francji w 1907 r. oraz bieżącą statystykę odstępów intergenetycznych w byłej Czechosłowacji¹². Z tego powodu, poszczególne typowe rozkłady odstępów zostały nazwane: francuski, czeski, morawski, słowacki i rusko-podkarpacki. Zdaniem L. Henry’ego, 3 pierwsze typy odpowiadają bardziej rodzinom maltuzjańskim (świadomie sterującym procesem rozrodu), a 2 pozostałe charakteryzują populacje niemaltuzjańskie.

¹¹ Dla polskiego czytelnika są one dostępne także w polskim przekładzie podręcznika R. Pressata (por. [46] s. 520-521).

¹² Czechosłowacja miała najstarszą (od 1925 r.) bieżącą rejestrację odstępów intergenetycznych.

Naszym zdaniem, takie postawienie problemu może się niekiedy łączyć z jego znacznym uproszczeniem, gdyż rodziny mogą stosować bardzo różną strategię swego rozwoju i antykoncepcji: opóźnienia urodzeń, zatrzymania bądź rozstawienia w czasie. Jednakże, jeśli liczby urodzeń według kolejności nie ulegają z roku na rok zbyt dużym wahaniom, to wybór rozkładu typowego nie ma większego znaczenia. Można się posłużyć którymkolwiek rozkładem, nawet dość odległym od rzeczywistego. Jak wykazano to na przykładzie byłej Czechosłowacji [42, s. 175-194], metoda pośrednia na podstawie odstępów typowych daje dostatecznie dobre wyniki, które niewiele różnią się od rezultatów metody bezpośredniej.

Przykładem innej techniki szacowania cząstkowych współczynników płodności jest **Metoda własnych dzieci** (*Own Children Method*). Na podstawie danych spisu powszechnego umożliwia ona szacunek cząstkowych współczynników płodności dla okresu 10-15 lat poprzedzających moment spisu. Dane spisowe o liczbie dzieci klasyfikuje się według rocznych grup wieku i wieku matki. Wiek matki może być sprecyzowany tylko dla dzieci z tego samego gospodarstwa domowego, w którym spisana została matka, stąd nazwa metody, tj. dzieci własne dla kobiety należącej do danego gospodarstwa domowego. Algorytm tej metody sprowadza się do 3 następujących etapów:

- 1) estymacji liczników cząstkowych współczynników płodności, tj. liczby urodzeń w latach poprzedzających spis według wieku matki,
- 2) estymacji mianowników cząstkowych współczynników płodności, tj. liczby kobiet według wieku w latach poprzedzających spis,
- 3) estymacji współczynników płodności jako ilorazów wielkości oszacowanych w punkcie pierwszym i drugim. Przy czym mianowniki oblicza się jako średnią liczbę kobiet dla poszczególnych lat poprzedzających spis.

Ten sposób postępowania umożliwia oszacowanie współczynników płodności z uwzględnieniem tylko tych dzieci i tych kobiet, które dożyły momentu spisu. Podejście to wymaga więc korekty ze względu na umieralność zarówno kobiet, jak i dzieci.

Początki Metody własnych dzieci sięgają tabulacji sporządzonych na podstawie spisów przeprowadzonych w USA w latach 1910 i 1940 [23]. Jednakże jej rozwój zawdzięczamy opracowaniom W.H. Grabilla oraz Lee-Jay Cho z lat sześćdziesiątych [27; 28]. Pierwsze zastosowania metody dotyczą danych spisu koreańskiego z 1966 r.

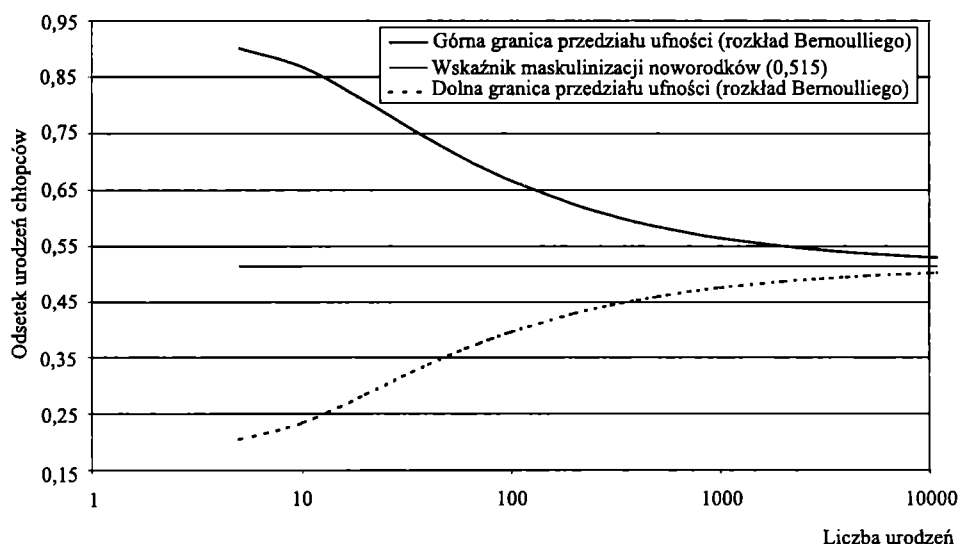
Omawiając metody estymacji pośredniej stosowane w demografii, należy docenić także dorobek polskich demografów i statystyków. Pierwszeństwo trzeba oddać pracy J. Kordosa [36], który oszacował **strukturę ludności według dochodów**, uwzględniając różne źródła informacji. Kordos skonstruował modele łączące informacje ze spisu ludności przeprowadzonego w 1960 r., wyniki badania budżetów gospodarstw domowych oraz bieżącą statystykę wynagrodzeń. Kontynuację tych badań, z wykorzystaniem empirycznej estymacji bayesowskiej, stanowią prace w ramach projektu EURAREA [21] oraz przedstawione w niniejszym tomie opracowanie T. Klimanka [35].

Szacunek ogólnej liczby poczęć oraz struktury rozwiązań ciąż w Polsce został przedstawiony przez J. Jóźwiak i J. Paradysza [31]. W estymacji pośredniej wykorzystali oni informacje z Czech i Słowacji o współczynniku dzietności teoretycznej, liczbie poronień oraz aborcji w przeliczeniu na kobietę w wieku 15-49 lat.

Do oszacowania migracji zagranicznych M. Kędelski [32]) zastosował metody tzw. statystyki lustrzanej. Polega ona na szacunku na podstawie 2 niezależnych źródeł informacji, w tym przypadku statystyki polskiej i niemieckiej.

Jednym z najnowszych przykładów zastosowań estymacji pośredniej w polskiej demografii jest **estymacja względnego ryzyka zgonu według przyczyn** przedstawiona przez J. Bijaka [5]. Przedstawił on analizę porównawczą metody tradycyjnej oraz podejścia bayesowskiego do szacunku względnego ryzyka zgonów na przykładzie Czech, Holandii i Polski w latach 1994-1996.

Od lat do określania tzw. scenariuszy w **projekcjach demograficznych** wykorzystywano podobieństwo rozwoju procesów demograficznych w różnych krajach i regionach o zbliżonym poziomie społeczno-ekonomicznym [14-16].



Rys. 1. Krzywe ufności dla wskaźnika maskulinizacji ($p = 0,05$) zgodnie z rozkładem Bernoulliego w zależności od liczby urodzeń, współczynnik ufności 0,9445, skala półlogarytmiczna

Źródło: opracowanie własne na podstawie [45].

J. Paradysz [45] z kolei zwraca uwagę na deformację proporcji urodzeń według płci przy konstrukcji prognoz dla małych jednostek terytorialnych, co ilustruje rys. 1 przedstawiający granice przedziałów ufności dla współczynnika maskulinizacji w zależności od liczby urodzeń. Zgodnie z rozkładem dwumianowym, dla liczby urodzeń rzędu 100 dzieci, „teoretyczny” ich rozkład według płci powinien być następujący: 51-52 chłopców i 48-49 dziewczynek. Jak pokazuje statystyka uro-

dzeń w gminach, w praktyce często zdarza się 40 chłopców i 60 dziewczynek albo odwrotnie. Również przy liczbie urodzeń wzrastającej do 200 czy 300 precyzja szacunku określona rozpiętością przedziału ufności jest ciągle niewielka (0,445; 0,605). W konsekwencji szacunek liczby urodzeń według płci jest obciążony dużymi błędami losowymi.

Wskazując czynniki decydujące o praktycznej wartości prognozy demograficznej, oprócz właściwej struktury ludności według płci, wieku i regionu, trafności założeń dotyczących ruchu naturalnego i migracyjnego, J. Paradysz [45] wymienia wielkość subpopulacji, dla których sporządzona jest prognoza. Dla mniejszych zbiorowości uzyskamy wyższe błędy losowe prognozy. Wielkość błędów losowych zależy także od wariancji cechy. Gdy szacujemy frakcję, dla wartości $p = 0,5$ otrzymamy maksymalną wariancję wynoszącą $\text{Var}(p) = 0,25$, ale już dla prawdopodobieństwa $p = 0,1$ lub $0,9$ wariancja wyniesie $\text{Var}(p) = 0,09$. Jest to wartość ponad dwa i pół razy mniejsza niż w pierwszym przypadku. A zatem, *ceteris paribus*, także wariancja estymatora będzie dwukrotnie mniejsza.

Na zakończenie tych rozważań podjęliśmy próbę zaproponowania wielokryteriowej klasyfikacji metod estymacji pośredniej stosowanych w badaniach demograficznych. Uwzględnia ona klasyfikację demograficznych metod estymacji pośredniej w ujęciu J.N.K. Rao, a także opinię ekspertów ONZ w dziedzinie demografii oraz nasze doświadczenia własne. Chcieliśmy w ten sposób podkreślić szerokie zastosowania estymacji pośredniej w analizach ludnościowych. Nie bez znaczenia jest także wkład demografów w rozwój metodologii statystyki małych obszarów, jak podkreślana już estymacja regresyjna syntetyczna, technika SPREE, której prekursorami są R. Chambers i G. Feeney [9] czy rozwój metod estymacji opartej na modelu zapoczątkowanej w pracach R.E. Faya i R.A. Herriota oraz E.P. Ericksena i J.B. Kadane [9; 18-20]. Proponowana przez nas klasyfikacja metod estymacji pośredniej w demografii uwzględnia następujące kryteria.

1. Rodzaj wykorzystanych źródeł informacji:

- a) dane administracyjne dla analizowanego obszaru,
- b) informacje o tych samych parametrach, ale dla innych, podobnych obszarów,
- c) informacje dla tej samej charakterystyki w badanym obszarze, otrzymane przy okazji innych badań przeprowadzanych w innym czasie (spisy powszechne, mikrospisy).

2. Metody estymacji pośredniej:

- a) syntetyczna (ilorazowa i regresyjna),
- b) złożona,
- c) bayesowska (empiryczna i hierarchiczna),
- d) *Structure Preserving Estimation SPREE*.

3. Stopień zaangażowania specjalistycznej metodologii badań demograficznych:

- a) całkowita,
- b) częściowa,

- c) na podstawie modeli demograficznych.
- 4. Rodzaj analizowanej generacji:
 - a) rzeczywista,
 - b) hipotetyczna.
- 5. Rodzaj szacowanych charakterystyk i zdarzeń demograficznych:
 - a) stan i struktura ludności,
 - b) umieralność i trwanie życia,
 - c) płodność,
 - d) reprodukcja ludności,
 - e) analiza migracji,
 - f) prognozowanie demograficzne,
 - g) rynek pracy,
 - h) ubóstwo i poziom życia,
 - i) dochody ludności,
 - j) inne.

4. Perspektywy rozwoju estymacji pośredniej w demografii regionalnej i lokalnej

Podkreślając wkład demografii w rozwój estymacji pośredniej, można podjąć próbę syntezy. Na tym etapie badań nie będzie ona wyczerpująca i ograniczy się do wskazania perspektyw rozwoju statystyki małych obszarów przez pryzmat jej kształtowania się i rozwoju na gruncie badań demograficznych. Wyróżniono tutaj 3 zasadnicze grupy zagadnień, tj. rozwój badań metodologicznych, eksplorację dostępnej infrastruktury statystycznej oraz integrację baz danych, w tym przede wszystkim dążenie do integracji źródeł pełnych z wynikami badań reprezentacyjnych. W odniesieniach tych odwołamy się do polskiej infrastruktury statystycznej:

A. Rozwój metodologii estymacji pośredniej

1. Równania bilansowe w szacowaniu liczby ludności, metoda złożona – potrzeba kalibracji, rozwój technik typu SPREE.
2. Idea technik symptomatycznych – szacowanie zmian populacji zamieszkującej badane terytorium zgodnie ze zmianami obserwowanymi w odniesieniu do czynników silnie uzależnionych od zmian liczby ludności (rozwój budownictwa, szkolnictwa itp.).
3. Rozwój estymacji syntetycznej – wykorzystanie relacji między zmiennymi szacowanymi a symptomatycznymi (pomocniczymi).
4. Estymacja ilorazowa.
5. Estymacja regresyjna (modyfikacja estymatorów regresyjnych przez wprowadzenie regresji próbkowej i ujęcia modelowego).
6. Uwzględnianie kontekstowych baz danych – rozwój modelowania wielopoziomowego.

7. Prognozowanie dla małych obszarów.

B. Eksploracja dostępnej infrastruktury statystycznej

1. Badania demograficzne w warunkach informacji wyczerpującej

- a) bieżąca statystyka ludności (tradycyjne źródła danych: spis powszechny, ewidencja ruchu naturalnego ludności),
- b) rejestry administracyjne¹³ (np. PESEL),
- c) nowe, niekonwencjonalne źródła zasilania informacyjnego:
 - rejestry podatkowe w statystyce dojazdów do pracy,
 - System Informacji Oświatowej (SIO), np. listy zapisów uczniów do szkół,
 - Systemy informacji geodezyjnej jako podstawa integracji nowych wieloobiektowych baz danych (np. rejestr pozwoleń na budowę, rejestr budynków zlikwidowanych, lokalne plany zagospodarowania przestrzennego),
- d) wieloobiektowe bazy danych na podstawie poszczególnych rejestrów (np. rejestru bezrobotnych oraz pomocy społecznej: PULS + POMOST = SYRIUSZ).

2. Reprezentacyjne badania demograficzne

- a) rozwój metodologii z zakresu statystyki małych domen i metodologii estymacji pośredniej w modelowaniu demograficznym (np. metoda „własnych dzieci”, por. [10; 11],
- b) coraz powszechniejsze zastosowanie metody „własnych dzieci” w demografii historycznej [8], w krajach o gorszej statystyce ludności [1-4], ale także w krajach rozwiniętych w odniesieniu do takich determinant dzietności kobiet, jak migracje [53; 40], czy religia [52].

c) estymacja wybranych parametrów w ankiecie rodzinnej (badanie postaw matrymonialnych i prokreacyjnych w duchu Indianapolis Study) za pomocą statystyki małych obszarów,

d) wykorzystanie SMO w prognozowaniu regionalnym i lokalnym [45].

C. Łączenie źródeł z badań pełnych i reprezentacyjnych

1. PULS jako źródło danych pomocniczych w estymacji parametrów dla małych domen w BAEL-u.

2. POMOST jako źródło danych pomocniczych w estymacji parametrów dla małych domen w badaniu budżetów gospodarstw domowych.

3. Rejestry podatkowe jako źródło danych pomocniczych w estymacji parametrów dla małych domen w SP3.

4. Wykorzystanie wyników pełnego spisu ludności jako źródła danych pomocniczych w estymacji parametrów dla małych domen w BAEL-u, badaniu budżetów gospodarstw domowych, badaniu dzietności kobiet itp.

Powyższe zestawienie wskazuje, że kontynuacja badań i analiz demograficznych prowadzonych na potrzeby władz centralnych i samorządowych oraz biznesu generuje dalszy rozwój statystyki małych obszarów. Mianem fundamentalnych

¹³ Opierając się na opracowaniu Gosha i Rao [26], można wskazać na metodę szacowania migracji netto na podstawie rejestrów indywidualnych, którą od wykorzystanych źródeł informacji nazwano metodą rejestrów administracyjnych *Administrative Records* (AR) [57].

można określić zasługi demografii w rozwoju metodologii estymacji pośredniej. Z jej „inspiracji” rozwijało się wiele metod estymacji, wyróżnić tu można estymację syntetyczną regresyjną czy techniki SPREE. Szczególną grupę technik estymacji pośredniej stanowią metody specyficzne, powstające w wyniku pogłębianej analizy procesów demograficznych, ich modelowania oraz źródeł danych o ludności. Technik takich jest coraz więcej. Omówioną w tym opracowaniu metodę własnych dzieci należy traktować jako przykładową ilustrację. We wstępie do kolejnej wersji podręcznika metodologicznego [59] eksperci ONZ podkreślają, iż choć dołożyli wszelkich starań, ogromny rozwój technik estymacji pośredniej spowodował, że spełnienie oczekiwań odbiorców w zakresie aktualności i kompletności prezentowanych metod, przekraczało ich możliwości.

Oprócz zagadnień metodologicznych, bardzo duże znaczenie dla rozwoju statystyki małych obszarów ma eksploracja istniejącej infrastruktury statystycznej, jej poznanie i integracja. Niezmiernie ważne wydaje się łączenie baz danych, w tym ze źródeł pełnych i reprezentacyjnych oraz rozwój kontekstowych baz danych. Realizacji tych zadań sprzyja rozwój informatyki. Wszystkie 3 wskazane nurty badań: metodologiczny, eksploracji dostępnej infrastruktury statystycznej oraz integracji baz danych, nawzajem się uzupełniają i warunkują. Wiążą się one z organizacją systemu statystyki publicznej podporządkowanej wymogom gospodarki rynkowej i rozwojowi samorządności regionalnej i lokalnej.

Literatura

- [1] Abbasi-Shavazi M.J., *Recent Changes and the Future of Fertility in Iran. Population and Development in the United Nations System. Completing the Fertility Transition*, <http://www.un.org/esa/population/publications/completingfertility/2RevisedABBASIpaper> 2005.
- [2] Abbasi-Shavazi M.J. *An Assessment of the Own-children Method of Estimating Fertility by Birthplace In Australia*, „Journal of Population Research” 1997, 14 (2), 167-186.
- [3] Arnaldo C., *Regional Fertility Trends in Mozambique*, „Journal of Population Research” 2004, 11 (1).
- [4] Bhat P.N.M., *India's Changing Dates with Replacement Fertility: A Review of Recent Fertility Trends and Future Prospects*, Population and development in the United Nations system (2005. Index of /esa/population/publications. Completing the fertility transition), <http://www.un.org/esa/population/publications/completingfertility/2RevisedBHATpaper>, PDF, s. 376-391.
- [5] Bijak J., *Estymacja względnego ryzyka zgonu według przyczyn: porównanie metody tradycyjnej oraz empirycznego podejścia bayesowskiego na przykładzie Czech, Holandii i Polski w latach 1994-1996*, „Studia Demograficzne” 2005, 147, Warszawa 3-29.
- [6] Bogue, D.J., *A Technique for Making Extensive Postcensal Estimates*, „Journal of the American Statistical Association” 1950, 45 149-163.
- [7] Bouge D.J., Duncan B.D., *A Composite Method of Estimating Post Censal Population of Small Areas by Age, Sex and Colour*, Vital Statistics-Special Report 47, No. 6, National Office of Vital Statistics, Washington, DC 1959.
- [8] Breschi M., De Santis G., *The Own-Children Method of Fertility Estimation in Historical Demography*, Forum, Udine 1997.

- [9] Chambers R., Feeney G., *Log-linear Models for Small Area Estimation*, Australian Bureau of Statistics, unpublished paper (1977).
- [10] Childs G., *Demographic Analysis of Small Populations Using the Own-children Method*, Washington University, Washington 2004.
- [11] Childs G., Goldstein M.C., Ben Jiao, Beall C.M., *Tibetan Fertility Transitions in China and South Asia*, „Population and Development Review” 2005, nr 2, 337-349.
- [12] Cho L.-J., Retherford R.D., Minja K.C., *The Own-children Method of Fertility Estimation*, East-West Center, Honolulu 1986.
- [13] Cicali D., De Santis G., *Estimating Cohort Fertility Levels and Differentials by Educational Level with the Own-children Method and Pseudo-panels: Italy, Cohorts 1945-55*, Università degli studi di Firenze, Dipartimento Statistico, Firenze 2006; <http://www.ds.unifi.it/ricerca/interessi/demografia/bassa-fecondita/messina/pubblicazioni/Cicali01.pdf>
- [14] Cieślak M., *Prognostowanie analogowe*, [w:] *Nieklasyczne metody prognozowania*, red. M. Cieślak, PWN, Warszawa 1983.
- [15] Cieślak M., *Liczba i struktury ludności*, [w:] *Procesy demograficzne w byłych województwach dolnośląskich w latach 1975-1997*, red. M. Cieślak, AE, Wrocław 1999.
- [16] Cieślak M., *Prognostowanie analogowe*, [w:] *Prognostowanie gospodarcze. Metody i zastosowanie*, Wydawnictwo Naukowe PWN, Warszawa 2001.
- [17] Coale A.J., *Factors Associated with the Development of Low Fertility: An Historic Summary Proceedings of the World Conference*, New York 1967.
- [18] Ericksen E.P., *A Regression Method for Estimating Populations of Local Areas*, „Journal of the American Statistical Association” 1974, 69, 867-875.
- [19] Ericksen E.P., Kadane J.B., *Estimating the Population in a Census Year (with discussion)*, „Journal of the American Statistical Association” 1985, 80, 98-131.
- [20] Ericksen E.P., Kadane J.B., *Sensitivity Analysis of Local Estimates of Undercount in the 1980 U.S. Census*, [w:] *Small Area Statistics. An International Symposium*, eds. R. Platek, J.N.K. Rao, C.E. Särndal, M.P. Singh, John Wiley & Sons, New York 1987, s. 23-45.
- [21] *EURAREA Project Reference Volume*, raport końcowy przygotowany w ramach Europejskiego Projektu Badawczego IST 2000-26290 EURAREA – *Enhancing Small Area Estimation Techniques to Meet European Needs*, Office for National Statistics, UK, London 2004, <http://www.statistics.gov.uk/eurarea/download.asp>.
- [22] Fay R.E., Herriot R.A., *Estimates for Income in Small Places: An Application of James-Stein Procedures to Census Data*, „Journal of the American Statistical Association” 1979, 74, 269-277.
- [23] Feeney G., *The Own-children Method of Estimating Age-specific Fertility Rates: Lecture Notes Prepared for a Demographic Analysis and Data Evaluation Workshop*, Department of Statistics, Kuala Lumpur, Malaysia, 10-28 March, The East-West Center, Honolulu 1975.
- [24] Frątczak E., Kotowska I.E., Kurkiewicz J., Kuropka I., Paradysz J., *Procesy demograficzne i metody ich analizy*, red. J. Kurkiewicz, AE, Kraków 2006.
- [25] Gołata E., *Studia nad terytorialnym zróżnicowaniem płodności kobiet w Polsce*, Monografie i Opracowania nr 322, SGPiS, Warszawa 1990.
- [26] Ghosh M., Rao J.N.K., *Small Area Estimation: An Appraisal*, „Statistical Science” 1994, 9(1).
- [27] Grabill W.H., Lee-Jay Cho, *Methodology for the Measurement of Current Fertility from Population Data on Young Children*, „Demography” 1965, 1, 50-73.
- [28] Grabill W.H., Lee-Jay Cho, Bouge D.J., *Differential Current Fertility in the United States*, Community and Family Study Center, University of Chicago, 1970.
- [29] Heady P., Hennel S., *Small Area Estimation and the Ecological Effect – Modifying Standard Theory for Practical Situations*, Office for National Statistics, London 2002, raport badawczy przygotowany w ramach Europejskiego Projektu Badawczego IST 2000-26290 EURAREA, *Enhancing Small Area Estimation Techniques to Meet European Needs*.
- [30] Henry L., *Fécondité des mariages. Nouvelle méthode de mesure*, „Travaux et Documents” nr 16, Presses Universitaires de France, Paris 1953.

- [31] Józwiak J., Paradysz J., *Demograficzny wymiar aborcji*, „Studia Demograficzne” 1993, 1 (Warszawa), 31-42.
- [32] Kędelski M., *Fikcja demograficzna w Polsce i RFN (Ze studiów nad migracjami zagranicznymi)*, „Studia Demograficzne” 1990, 1, (Warszawa), 21-55.
- [33] Kędelski M., Paradysz J., *Demografia*, AE, Poznań 2006.
- [34] Klimanek T., *Wielopoziomowa analiza struktury agrarnej gminy w systemie GEO-INFO*, praca doktorska, Wydział Zarządzania Akademii Ekonomicznej w Poznaniu (2003).
- [35] Klimanek T., *Próba wykorzystania statystyki małych obszarów w szacowaniu dochodów gospodarstw domowych*, [w:] *Statystyka w praktyce społeczno-gospodarczej*, red. W. Ostasiewicz, Prace Naukowe Akademii Ekonomicznej nr 1163, AE, Wrocław 2006.
- [36] Kordos J., *Rozkład ludności pozarolniczej według wysokości dochodów na osobę w 1960 r.*, Biuletyn KPZK PAN 1963, nr 8.
- [37] Kotowska I.E., *Prognozowanie gospodarstw domowych. Problemy i metody*, Monografie i Opracowania nr 396, SGH, Warszawa 1994.
- [38] Marker D.A., *Organization of Small Area Estimators*, Proceedings of Survey Research Methods Section, 409-414, American Statistical Association, Washington, DC 1983.
- [39] Mencarini L., *An Analysis of Fertility and Infant Mortality in South Africa based on 1993 LSMS data*, paper presented for the Third African Population Conference „African Population in the 21st Century”, Durban, South Africa, 6-10 December 1999.
- [40] Ng E., Nault F., *Fertility among Recent Immigrant Women to Canada, 1991: An Examination of the Disruption Hypothesis*, „International Migration” 1997, 35 (4), 559-580.
- [41] Okólski M., *Statistics of Immigration in Poland. Conditions of Correctness, Assessment of the Current State, Proposals of New Solutions*, Working Papers ISS, „Prace Migracyjne”, Uniwersytet Warszawski, Warszawa 1997.
- [42] Paradysz J., *Reprodukcja ludności w Polsce. Studium metodologiczno-poznawcze*, Monografie i Opracowania nr 11(312), SGPiS, Warszawa 1990.
- [43] Paradysz J., *Dzielnosc kobiet w Polsce*, Narodowy Spis Powszechny 1988, Materiały i opracowania statystyczne, GUS, Warszawa 1992.
- [44] Paradysz J., *Small Area Statistics in Poland – First Experiences and Application Possibilities*, „Statistics in Transition” 1998, vol. 3(5).
- [45] Paradysz J., *Analiza i prognoza rozwoju demograficznego miasta Poznania w latach 1999-2030*, AE, Poznań 2005.
- [46] Pressat R., *Analiza demograficzna*, PWN, Warszawa 1966.
- [47] *Procesy demograficzne i metody ich analizy*, red. J. Kurkiewicz, AE, Kraków 2006 (maszynopis).
- [48] Purcell N.J., Kish L., *Postcensal Estimates for Local Areas (Domains)*, „International Statistical Review” 1980, 48, 3-18.
- [49] Rao J.N.K., *Small Area Estimation*, Wiley, Hoboken 2003.
- [50] Rao J.N.K., *Some Recent Advances in Model-Based Small Area Estimation*, „Survey Methodology” 1999, 25, Statistics Canada.
- [51] Sakson B., *Wpływ „niewidzialnych” migracji zagranicznych lat osiemdziesiątych na struktury demograficzne Polski*, Monografie i opracowania nr 481, SGH, Warszawa 2002.
- [52] Schellekens J., Eisenbach Z., *Religiosity and Marital Fertility: Israeli Arab Muslims, 1955-72*, University of Southampton (2005), School of Social Sciences, Seminars: <http://www.socsci.soton.ac.uk/socstats/Research/seminars/2005-11-15-Schellekens.pdf>.
- [53] Schmertmann C.P., Swicegood C.G., Bean F.D., *Immigration's Impact on US Fertility: An Exploratory Analysis of Recent Trends. Working Paper*. (2001) http://mailer.fsu.edu/~schmert/paa02/schmertmann_swicegood_bean_paa_abstract.pdf
- [54] Schmidt R.C., *Short-cut Methods for Estimating County Populations*, „Journal of the American Statistical Association” 1952, 47.
- [55] Shapiro S., Schachter J., *Birth Registration Completeness in the United States and Geographic Areas, 1950*, Vital Statistics Special Reports, Selected Studies (1964).

- [56] Smith M.D., *Preparation of Fertility Estimates Using the Own-child Method: From the Labour Force Survey, through the East-West Fertility Program, with Excel Output*, Background Paper (2003) no 3 July, <http://www.apsoc.ox.ac.uk/Oxpop/publications%20files/WP14.pdf>.
- [57] Starsinic D.E., *Development of Population Estimates for Revenue Sharing Areas*, Census Tract Papers, Ser. GE40, No. 10 U.S. Government Printing Office, Washington, DC 1974.
- [58] Timaeus I., *The Own-children Method of Fertility Estimation*, „Population Studies” 1988, 42(1), 143-144.
- [59] UN Manual X: *Indirect Techniques for Demographic Estimation*. Population and development in the United Nations system (2005), [podręcznik dostępny tylko w wersji elektronicznej] http://www.un.org/esa/population/publications/Manual_X/Manual_X.html.
- [60] UNPF *Readings in Population Research Methodology*. Published for United Nations Population Fund by Social Development Center, Chicago, Ill. 1993.
- [61] U.S. Bureau of the Census *Methods of Population Estimation: Part I, Illustrative Procedure of the Bureau's Component Method II. Current Population Reports*, Series P-25, No. 339. U.S. Government Printing Office, Washington DC 1966.
- [62] Zidek J.V., *A Review of Methods for Estimating the Populations of Local Areas*, Technical Report 82-4, University of British Columbia, Vancouver 1982.

DEMOGRAPHIC ROOTS AND PERSPECTIVES OF SMALL AREA STATISTICS

Summary

Small area estimation is becoming important in survey sampling due to a growing demand for reliable small area statistics from both public and private sectors. The paper refers to demographic background of small area estimation SAE, its determinants, applications and perspectives. Usually the demographic origin of SAE is restricted to methods for local estimation of population in post-censal years. That is why demographics contribution in the development of SAE techniques is presented according to the classification by J.N.K. Rao and the one used by UN experts in demographic research. The studies conducted by Polish demographers are described and own classification of SAE methods in demography, combining the two above mentioned as well as own experience is proposed. The present article is largely an appraisal of some of the methods. Among others we focus on E. Arriaga, G. Feeney and L. Henry methods for fertility estimation.

Underlining the impact of demography in SAE studies we tried to mention the most important directions of further development. First of all, estimation methodology, data mining and integration of electronic data bases are considered. Special attention is focused on integration of administration records and survey studies.